Kirsal, Yonal, 2013. Modelling and performance evaluation of wireless and mobile communication systems in heterogeneous environments. Available from Middlesex University's Research Repository.

Modelling and Performance Evaluation of Wireless and Mobile Communication Systems in Heterogeneous Environments

Yonal KIRSAL

School of Science and Technology

Middlesex University

A thesis submitted to Middlesex University in fulfilment of the requirements for degree of Doctor of Philosophy

April 2013

I would like to dedicate this thesis to my loving parents ...

Acknowledgements

The completion of the work for this thesis represents the opportunity to remember numerous individuals who have influenced my attitude towards my PhD studies. I would like to acknowledge my debt of gratitude to my supervisors, colleagues, friends and family.

First and foremost, I am deeply grateful to my supervisors, Professor Orhan Gemikonakli, Dr. Glenford Mapp and Dr. Enver Ever for their continuous guidance, kind support and great enthusiasm for high-quality research. Throughout my studies, I have benefited greatly from their invaluable advice on both, my research and career path. It has been a tremendous pleasure and inspiration to work with them. Definitely, this thesis would have been impossible without their help.

Special thanks to Dr. Altan Koctigit for the generous support with simulation programs and many useful discussions.

I would like to thank my sister Dr. Yoney Kirsal Ever, who has given me valuable advice and support during my study. She also shared good times with me and helped me through hard times.

I also would like to thank the staff of Middlesex University for their support during my studies. My sincere appreciation goes to Pia Wallington for ensuring the countless forms, conference registrations and paperwork were correctly completed and filed.

I am deeply indebted to my friends, in particular Nadir Timucin, Ramadan Kuslucali, Mehmet Simsekatan, Ersin Ozpinarli, Erman Bismis, Yunus Ozerdem, Urun Solyali, Ozan Aydingun and my office mates Aregbesola Akinola (Shola), Ammar Zayouna, Anhtuan Le, Arindam Gosh, Artemis Parvizi, Eser Gemikonakli, Krishna Doddapaneni, Mahbobeh Ghoreyshi, Nallini Selvaraj, Payam Rahmdel, Pawel Chwalinski, Ryan Zammit, Syed Murad and Javed Anjum Sheikh for their help with my research, and more for their great friendship and encouragement during my studies.

My deepest gratitude, love and affection belong to my parents, Hamide Kirsal and Ayhan Kirsal, to whom I owe all that I have ever accomplished. Their continued support, endless patience, and encouragement, in the times when it was most needed, have enabled me to get this far. To my parents, I owe everything I have achieved in my life.

Lastly and most importantly, I wish to thank my love Ayten Ozkirac for her understanding, continued love and support. Without her love, it is impossible for me to complete my PhD program.

Abstract

It is widely expected that next generation wireless communication systems will be heterogeneous, integrating a wide variety of wireless access networks. Of particular interest recently is the integration of cellular networks (GSM,GPRS,UMTS,EDGE and LTE) and wireless local area networks (WLANs) to provide complementary features in terms of coverage, capacity and mobility support. These different networks will work together using vertical handover techniques and hence understanding how well these mechanisms perform is a significant issue. In this thesis, these networks are modelled to yield performance results such as mean queue lengths and blocking probabilities over a range of different conditions. The results are then analysed using network constraints to yield operational graphs based on handover probabilities to different networks. Firstly, individual networks with horizontal handover are analysed using performability techniques. The thesis moves on to look at vertical handovers between cellular networks using pure performance models. Then the integration of cellular networks and WLAN is considered. While analysing these results it became clear that the common models that were being used were subjected to handover hysteresis resulting from feedback loops in the model. A new analytical model was developed which addressed this issue but was shown to be problematic in developing state probabilities for more complicated scenarios. Guard channels analysis, which is normally used to give priority to handover traffic in mobile networks, was employed as a practical solution to the observed handover hysteresis. Overall, using different analytical techniques as well as simulation, the results of this work form an important part in the design and development of future mobile systems.

Contents

List of Figures

List of Tables

Acronyms

ACK Acknowledgement

AP Access Point

BEB The Binary Exponential Backoff

BS Base Station

CDMA Code Division Multiple Access

CSMA/CA Carrier Sense Multiple Access with Collision Avoidance

CTMC Continuous Time Markov Chain

CTS Clear To Send

DCA Dynamic Channel Allocation

DCF Distributed Coordination Function

DES Discrete Event Simulation

EDGE Enhanced Data rates for GSM Evolution

FCA Fixed Channel Allocation

FCFS First Come First Served

FDMA Frequency Division Multiple Access

FIFO First In First Out

GPRS General Packet Radio Service

GSM Global System for Mobile Communication

ISM The industrial, scientific and medical

LTE Long-Term Evolution

MAC Media Access Control

MQL Mean Queue Length

MRM Markov Reward Model

MS Mobile Subscriber

MTBF Mean Time Between Failure

MTTR Mean Time To Repair

PCF Point Coordination Function

pdf probability density function

PNs Petri Nets

PoA Point of Attachment

P_B Blocking Probability

QBD Quasi Birth and Death

QoE Quality of Experience

QoS Quality of Service

QSMB Quasi Simultaneous Multiple Births

RAT Radio Access Technology

RTS Request To Send

SMD Simultaneous Multiple Deaths

SPN Stochastic Petri Net

TDMA Time Division Multiple Access

UMTS Universal Mobile Telecommunications System

WCDMA Wideband Code Division Multiple Access

WLAN Wireless Local Area Network

WWAN Wireless Wide Area Network

WiMAX Worldwide Interoperability for Microwave Access

Chapter 1

Introduction

The development of wireless terminals has facilitated the rapid growth of wireless communication and mobile computing. In wireless computing/communication, systems that offer high performance and availability/reliability are typically referred to systems that can provide a high "Quality of Service" (QoS). The user expectations of QoS makes the design and modelling of mobile communication systems a challenging task.

Mobility is one of the major issues in providing QoS in mobile and wireless communication systems. It is important to support services when mobile subscribers (MSs) move between the network technologies, which can be homogeneous or heterogeneous. The process of switching data transmission from one access point to another one is called *handover/handoff*. Handovers can be classified according to different providers and technologies. Handovers can be further divided into *horizontal handover* and *vertical handover* according to different mobility behaviour.

Horizontal handover is mobility between the same network technologies. This is the traditional definition of handover for homogeneous network systems. On the other hand, vertical handover occurs between different network technologies. Mobile users may be interested in changing their network operator or are forced to change the serving access point (AP) or base station (BS) in order to achieve a given level of QoS. Therefore, it is very important for the users to seamlessly access various technologies in heterogeneous environments (e.g. mobility between cellular network and WLAN). More specifically, heterogeneous networks aim at providing guaranteed QoS and Quality of Experience (QoE) (i.e. what quality a network provider can offer (or claims) and what quality the user experiences). It is already deployed that wireless and mobile networks such as cellular systems integrated with WLANs to provide better Internet access and multimedia services to mobile users (Xia and Shen, 2009), (Hasib and Fapojuwo, 2007), (Hasib and Fapojuwo, 2008), (Saravanan et al., 2006). Integrated systems give the advantages of their complementary features in terms of coverage area, capacity and mobility support. Therefore, it is obvious that the heterogeneity of network technologies brings new concepts of designing new handover algorithms and developing a suitable model in order to evaluate realistic QoS measures of such systems.

One of the aims of this research work is to present new analytical models with related performance and/or performability modelling and evaluation for wireless and mobile communication systems. Queuing theory has proved a very powerful tool in modelling such systems. Analytical modelling is used because it is computationally more efficient as compared to other performance evaluation methods (e.g. benchmarking, simulation). Analytical methods are employed together with certain assumptions in order to achieve a certain degree of mathematical tractability. Techniques such as probabilistic analysis (to represent the behaviour of the system under study), Markov processes, and queuing models are extensively used.

The analysis in this thesis is based on using mathematical models developed, which represents the process of arrival of calls that join the system and queue, the rules by which they are allowed into service, and the time it takes to serve the calls. Queuing theory represents the full range of characteristics of such models covering all perceivable systems which incorporate characteristics of a queue.

Calls are identified as the unit of demanding service (Trivedi et al., 2003). The unit providing service is known as the channel (Trivedi et al., 2003) ,(Jindal et al., 2006). This terminology of calls and channels is commonly used in the literature and applicable for most of the wireless and mobile systems (Hasib and Fapojuwo, 2008), (Trivedi et al., 2002), (Agrawal and Qing, 2006). While analysing wireless and mobile networks some basic elements of the systems are identified. The first one is the arrival process. In most cases there are large amounts of independent sources arriving to the system at different times. Therefore, the best way to describe the arrival process in terms of random variables are either the number arrivals during a time interval or the time interval between successive arrivals. If calls arrive in groups, their size can be a random variable as well. The parameters involved in the service mechanism are the number of channels, the number of calls getting served at any time, and the duration and mode of service. Some of the proposed models consist of more than one channels arranged in series and/or parallel. In queuing theory, if the incoming calls are served in parallel (more than one call, packet etc. at a given time) this kind of a model corresponds to a multi-sever queuing system. Random variables are used to represent service times, and the number of channels, when appropriate. If service is provided for calls in groups, their size can also be a random variable. System storage capacity is also an important element for such systems. The number of calls that can wait at a time in the system is a significant factor for consideration. It gives good measurements in determining of the system's QoS, since the blocking and dropping probabilities as well as delay are quite significant.

The study of proposed models is intended to understand how proposed models behave under various conditions. Therefore, the proposed models can be applied to various scenarios. There are many valuable applications of modelling, most of which have been well documented in the literature of probability, operations research, management science and industrial engineering (Banks et al., 2005), (Agrawal and Qing, 2006). Examples include: traffic flow (vehicles, aircraft, people, communications), scheduling (patients at the doctor, programs on a computer), and facility design (banks, post offices, fast-food restaurants) apart from the wireless and mobile communication systems. There are existing studies which attempts to model interactions between wireless communication systems. For instance, in Bobarshad (2010) a packet transmission scheme of WLAN is shown considering the packet dropping mechanisms in the queue and presenting the concept of buffer length in WLAN. The proposed model also supports for certain characteristics of the traffic produced by an application, such as maximum service rate and burst size. Therefore, such proposed models need to be considered so that, the QoS requirements of systems can be defined in terms of performance measurements such as delay, queue capacity and blocking probability.

The main strength of this work is the ability to provide exact solution for two dimensional Markov models and the ability of incorporating factors such as mobility, buffer management, operational areas, sensitivity, and feedback/forward analysis.

Performance, availability and prediction play a major role in the modelling, design, development, testing and maintenance of systems in many application areas such as computer systems, telecommunication systems, wireless networks and manufacturing systems. In many cases, the following properties are characteristic for the real-world systems under investigation. Wireless and mobile communication systems encounter failures due to various reasons such as, software, hardware, human error, or a combination of these factors. System failure depends on the nature of the system and can have different impacts on the system performance. Because of failures, the system may not support its active users efficiently and performance may degrade. In order to overcome this problem, availability and performance of the system should be considered together for individual networks. From the designer's and operator's points of view, it is necessary to take failure and repair factors into account.

Traditional pure performance models ignore failure and recovery but consider the system's ability to perform (Trivedi and Ma, July 2002), (Trivedi et al., 2003). The system availability model, which is associated with failure and repairs, should be taken into account in order to obtain more realistic QoS measures for the systems. The proper modelling methodology which considers both performance and availability models together is called *performability modelling and evaluation* (Mitrani, 2001), (Trivedi and Ma, July 2002), (Trivedi et al., 2003). Performability modelling and evaluation include measures formulation, model specification, model construction, tool development, and application to wide variety of systems (Mitrani, 2001).

In this thesis, different handover schemes, mobility issues, composite performance and availability models are considered in wireless and mobile communication systems for QoS measures. An analytical framework is developed for integrated heterogeneous systems considering mobility of users, buffering and area of coverage, together with probabilities of having various handover decisions for sensitivity analysis in order to provide vertical handover policy management. Various approaches such as product form solutions (for pure performance evaluation), approximate Markov Reward Model (MRM) solution and the exact spectral expansion solution approaches are employed and discussed in detail. Simulation results are also obtained for validation of the analytical models. While analysing these systems, it was discovered that feedback loops can exist in common models. This was investigated and a guard channel system was implemented as a practical solution to this problem.

1.1 Objective and Scope of the Research

The aim of this thesis is to develop improved analytical models for evaluation of performance and availability measures of various mobile and wireless systems to improve the QoS in heterogeneous wireless networks with an emphasis on integrated system environment.

In this study, analytical modelling techniques are used to model various complex mobile and wireless systems. In order to achieve a certain degree of mathematical tractability, the proposed models are considered together with certain assumptions. Probabilistic analysis, Markov processes and queuing theoretic models are extensively used. One of the most frequently used techniques to model and evaluate the performance of such a system at the dynamic level is to represent it by a stochastic process which, roughly speaking; consists of all system states and all possible transitions between them. For an important class of stochastic processes known as Continuous Time Markov Chains (CTMCs)(Trivedi and Ma, July 2002), (Trivedi et al., 2003), standard numerical algorithms can be used to compute the state probabilities. From these, the interesting performance and availability measures of the system can then be derived.

For systems whose dynamics can be described by only a few states, an experienced modeller is able to deduce the underlying stochastic process directly by inspection. Unfortunately, this is impossible for many interesting real world systems since they tend to possess a large number of states at the stochastic process level. Thus each system (e.g. cellular BSs, WLAN APs) will be modelled as a queuing model to evaluate the QoS performance such as mean queue length (MQL) and blocking probabilities (P_B). Concepts and methods further developed from the proposed models (e.g. cellular network) can be used to facilitate the analysis of system characteristics and evaluation of interworking performance for heterogeneous wireless networks.

This thesis also attempts to develop analytical models for wireless and mobile systems to overcome the limits of a simulation model. Unfortunately, computer simulations of wireless cellular networks require considerable computation time, making them cumbersome to use for the purposes of network design. A good analytic model of a network would be attractive as it would be able to evaluate network performance under a wide range of conditions. Analytical models compute comparatively easily and are likely to be faster than a simulation model. It is also essential to highlight the principles and assumptions in such modelling to allow us to get a better understanding of what is happening. The proposed approaches lead to a comprehensive understanding of the handover behaviour and a faster computation compared to simulation techniques. The proposed mathematical models are validated by simulation results as well as the solution of the systems of linear equations.

The mathematical frameworks presented provide the basis for modelling new handover algorithms by incorporating the real operating behaviour of the systems and the parameters under the control of the handover designer such as: radius of the cells, speed of the users, handover decision management, sensitivity analysis and operational spaces. The focus of this thesis is only on cellular technologies and WLANs. However, the analytical models presented in this thesis can be easily extended to other wireless and mobile technologies.

1.2 Research Question

Can an analytical model be developed to obtain realistic QoS measurements in wireless an mobile systems?

In order to address the research question, This work will focus on key subquestions:

- How does mobility impact on the QoS in heterogeneous environments?
- How does the buffering strategies at different networks affect the overall system performance taking into account the mobility of users?

- How do we express our results in terms of operational spaces based on the handover probabilities to different networks in heterogeneous environments?

1.3 List of Publications

The work presented in this thesis has given rise to the following publications.

- **Y. Kirsal**, E. Ever, G. Mapp, O. Gemikonakli and A. Kocyigit, "Exploring Analytical Models to Study Vertical Handover in Heterogeneous Networks using Guard Channels and Buffering", The International Journal of Computer and Telecommunications Networking: Computer Networks (Submitted).

- **Y. Kirsal**, O. Gemikonakli, E. Ever, and G. Mapp, "Enhancing the Modelling of Vertical Handover in Integrated Cellular/WLAN Environments", *The* 27^{th} *IEEE International Conference on Advanced Information Networking and Applications (AINA-2013)*, Barcelona, Spain, March 2013.

- **Y. Kirsal**, O. Gemikonakli, E. Ever, and G. Mapp, "Performance Analysis of Handovers to Provide a Framework for Vertical Handover Policy Management in Heterogeneous Environments", *In* 45^{th} *Annual Simulation Symposium, (ANSS'12)*, Orlando, FL, USA, March 2012, pp. 1-8.

- **Y. Kirsal**, E. Ever, O. Gemikonakli and G. Mapp, "Modelling and Performability Analysis of WLANs as a Queuing Model with Channel/Access Point Failures and Reconfiguration", *In IEEE Proceedings of* 5^{th} *European Modelling Symposium (EMS 2011)*, Universidad Politecnica de Madrid, Madrid, Spain, November 2011, pp. 440-445.

- **Y. Kirsal**, E. Ever, O. Gemikonakli and G. Mapp, "Critical Review of Analytical Modelling Approaches for Performability Evaluation of the Handover Phenomena in Mobile Communication Systems", *The Proceeding of IEEE* 11^{th} *International Conference on Computer and Information Technology,* 2^{th} *International Workshop on Dependable Service-Oriented and Cloud computing (DSOC 2011)*, Paphos, Cyprus, September 2011, pp. 132-137.

- **Y. Kirsal**, E. Gemikonakli, E. Ever, G.Mapp and O. Gemikonakli, "An Analytical Approach for Performance Analysis of Handoff in the Next Generation Integrated Cellular Networks and WLANs", *In IEEE Proceedings of* 4^{th} *International Conference on Computer Communications Networks (ICCCN 2010), Workshop on Performance Modelling and Evaluation in Computer and Telecommunication Networks*, Zurich, Switzerland, August 2010, pp. 1-6.

- Ever, E., **Y. Kirsal** and O. Gemikonakli, "Performability Modelling of Handoff in Wireless Cellular Networks and the Exact Solution of System Models with Service Rates Dependent on Numbers of Originating and Handoff Calls", *In IEEE Proceedings of International Conference on Computational Intelligence, Modelling and Simulation (CSSim 2009)*, Brno, Czech Republic, September 2009, pp. 282-287.

- **Y. Kirsal** and O. Gemikonakli, "Approaches to Modelling and Analysis for Performability Evaluation of Handoff Schemes in Wireless Cellular Networks", 10^{th} *Annual Post-Graduate Symposium on the convergence of Telecommunications, Networking and Broadcasting*, Liverpool, UK, June 2009.

- **Y. Kirsal** and O. Gemikonakli, "Performability Modelling of Handoff in Wireless Cellular Networks with Channel Failures and Recovery", *In IEEE Proceedings of* 11^{th} *International Conference on Computer Modelling and Simulation (UKSim 2009)*, Cambridge, UK, March 2009, pp. 544-547.

1.4 Thesis Outline

The following is an outline of the final thesis:

Chapter 1 - Introduction: This chapter introduces the proposed research area and provides the essential background. In addition, this chapter presents the general performance and availability evaluation methods, as well as handover issues in future networks and highlights the advantages of analytical modelling.

Chapter 2 - Literature Review: This chapter introduces the domain of the research by providing a critical review of related literature. Handover classification and modelling issues for wireless and mobile systems are discussed in detail.

Chapter 3 - Existing Modelling Approaches and Solutions Techniques (Methodology Chapter): This chapter provides the fundamental approach to address the research question. Existing analytical modelling techniques are critically analysed and compared in this chapter. Existing performance, availability as well as composite performance and availability modelling (performability modelling) techniques for the systems under study are investigated and critically analysed. Solution methods for two dimensional state spaces are compared and a detailed explanation of spectral expansion method is given.

Chapter 4 - Performability Modelling of Handover in Wireless Cellular Networks: Performability modelling and evaluation of wireless cellular systems are considered in this chapter. Various handover schemes and mobility issues are investigated. Practical applicability of various failures and repairs strategies are also considered. Existing approaches are analysed and numerical results are presented for various systems. The handover schemes and failure/repair behaviour of wireless cellular networks are the main focus point of this chapter. An exact performability model is developed and solved for wireless cellular systems with failures and repairs. Numerical results are presented for both Markov Reward Model (MRM) and exact spectral expansion solution approach.

Chapter 5 - Performability Analysis of WLAN as a Queuing Model: In this chapter, performability evaluation of a WLAN is presented with the inherent wireless channel unreliability behaviour caused by the channel failure/recovery and access point failure/recovery phenomena. The mobility related issues and queuing capacities are also considered for various scenarios. The effects of the mobility and failure/recovery on the system performance measures are emphasised. Numerical results obtained from the exact spectral expansion solution approach are validated by simulation.

Chapter 6 - Performance Analysis of Handover in Cellular Mobile Networks: This chapter presents the analytical modelling study which can be used for integration of different cellular technologies (i.e 2G and 3G). Vertical handover and mobility issues are addressed for successful modelling and solution of integrated cellular networks. Numerical results are presented and critically analysed. Simulation results are also presented for validation purposes.

Chapter 7 - Performance Analysis of Handover to Provide a Framework for Vertical Handover Policy Management: This chapter presents the proposed model for the performance evaluation of an integrated communication system composed of cellular networks and WLAN. In this chapter, performance evaluation of both homogeneous and heterogeneous technologies are analysed and effective handover algorithms are designed by using an exact solution technique. The model covers handovers, mobility of users, buffer capacity of WLAN, as well as cellular networks, and area of coverage, together with probabilities of having various handover decisions. Numerical results are validated by simulation and critically analysed.

Chapter 8 - Enhancing the Modelling of Vertical Handover in Integrated Cellular/WLAN Environments: This chapter presents a new analytical model to enhance the modelling of vertical handover for integrated cellular/WLAN systems. Recent work on the modelling of these systems using a two-stage model with feedback encountered some complications in terms of handling certain streams. The possibility of having continuous streams which keep on switching between WLAN and cellular systems may reduce the practicality of the models. The new analytical model presented in this chapter restricts certain flows to make sure that the handover calls from the WLAN to the cellular system are not allowed back to the WLAN. A new model is presented with an extra random variable in order to ensure that Markovian property is not violated. Results are presented to show that the new model is a significant improvement but more work is needed to analyse more complex scenarios.

Chapter 9- Exploring Analytical Models to Study Vertical Handover in Heterogeneous Networks using Guard Channels and Buffering: In this chapter the proposed model builds a framework to manage QoS using guard channel and buffering in the future mobile systems. It gives detailed analysis for such an integrated communication system together with an exact analytical approach. Numerical results are depicted and critically analysed presenting vertical handover decision management, sensitivity analysis and operational spaces.

Chapter 10 - Conclusions and Further Work: This chapter summarizes and concludes the entire thesis. It also discusses the future work in order to ensure continued development in the related areas of research.

Chapter 2

Literature Survey

2.1 Introduction

The objective of this thesis is to develop efficient analytical models for wireless and mobile systems. In order to do that, many factors including but not limited to, network characteristics, handover schemes, decision management, mobility (i.e speed of the users), cell size, number of channels, queue capacities, call holding time and cell residence times have to be taken into consideration while designing the system. One of the main issues of characterising the performance measures is the construction of a suitable system with handover schemes. In such systems, when a user dials the number an originating call is generated. It holds the channel until the call terminates. The duration between the times that the user generates a call and leaves the system is called channel holding time. When a user holding a channel moves to a neighbouring cell coverage area or different technologies, a handover request is generated. When a handover is generated, the cell residence time is used to characterize the time of a user staying in a certain cell during the connection period of service delivery.

Efficient analytical handover schemes need to be designed to provide the required QoS in future systems while allowing seamless roaming among a multitude of access network technologies. In this chapter, a comprehensive literature survey of the analytical model algorithms designed to satisfy these requirements is presented. The analytical models of the future heterogeneous systems are mentioned and performance/performability modelling of wireless cellular systems is considered with handover issues and mobility management. Modelling the integration of heterogeneous systems is also considered.

2.2 System Under Study

This section describes the evolution of wireless and mobile systems, and their features. The key challenges of standalone performance models of cellular systems as well as the integration of cellular systems with WLAN is presented.

2.2.1 The Wireless and Mobile Systems

Telecommunication systems and wireless network technologies have been in a phase of rapid development (3G, Wi-Fi, WiMAX, and LTE). Wireless communications began in the early 1980s, when the First Generation (1G) wireless telecommunication technology provided voice transmissions using frequencies around 900 MHz and analogue modulation (Agrawal and Qing, 2006), (Mishra, 2006). The Second Generation (2G) of wireless networks that appeared in 1990 were based on low-band digital data signalling. The data services in Global System for Mobile Communications (GSM) is limited to very low data speeds, such as 9.6 Kbps, because of the limited transmission rate and the existing billing services. After 2G and before the 3G, Second and a Half Generation (2.5G) was used as a way to progress to describe 2G systems that had implemented a packet switched domain in addition to the circuit switched domain. 2.5G provided some of the benefits of 3G (since it was packet-switched) and could use some of the existing 2G infrastructure in GSM networks (Mishra, 2006). Therefore, in the middle of the 1990s, General Packet Radio Service (GPRS) was introduced and offered up to 128 Kbps data rates, depending on how many time slots were allocated to use. However, GPRS suffers from high rates of delay. Since 2003, mobile telecommunications networks have been upgraded to use 3G technologies, hence, Enhanced Data rates for GSM Evolution (EDGE) and Universal Mobile Telecommunications System (UMTS) offer higher data rate services up to 384 Kbps and 2 Mbps respectively (Agrawal and Qing, 2006), (Mishra, 2006). However, these services are expensive, especially compared to WLAN. In the late 1990s the use of WLANs became popular as they provided a fast broadband connection in office and home environments. In the beginning of the 2000s, their use became more common in public settings such as airports, cafes and hotels (Makela, 2008), (Khatib, 2003). Customers are now interested in integrated mobile services especially when the systems provide access and service any time and anywhere, requiring seamless heterogeneous environments.

Such integration combines various access networks into a seamless system, referred to as Fourth Generation (4G) wireless systems (Akyildiz et al., 2004), (Mishra, 2006), (Choi et al., 2007). 4G wireless systems provide significantly higher data rates, offer a variety of services and applications previously not possible due to bandwidth limitations, and allow global roaming among a diverse range of wireless and mobile networks (Akyildiz et al., 2004), (Mishra, 2006), (Choi et al., 2007).

In 4G networking scenario, MSs are able to choose the most appropriate access network among the available alternatives. These access networks include IEEE 802.11 WLAN access (Hua et al., 2004), (Bobarshad, 2010) the traditional cellular networks, IEEE 802.16 WiMAX (Wang et al., 2008), Bluetooth as well as satellite systems (Mishra, 2006). For a satisfactory user experience, MSs must be able to seamlessly transfer to the best available access network among all available candidates with no interruption to an ongoing call.

Future wireless networks are expected to provide users with global information access capabilities with multimedia wireless services (Agrawal and Qing, 2006), (Mishra, 2006). Growing interest in 4G networks is leading to a convergence of various wireless network technologies. Dealing with customers expectation by providing the best service facilities is a challenging task. Integrating various technologies is an effective way to achieve these goals. Integrating various technologies aims to support seamless roaming among a variety of wireless access network technologies, including GSM, UMTS, WiMAX, Bluetooth and WLAN, through different handover mechanisms. This integration of different cellular network technologies and integration of cellular networks with WLANs are the main system integrations that need to be addressed. Hence, the integration of such networks can provide better QoS.

Cellular networks have been well deployed around the world. Both circuit-switched and packet-switched services are supported by cellular networks hence, a well-defined infrastructure having a global coverage area is the main benefit. However, cellular networks usually have low data rates and capability to serve applications with potential high bandwidth requirements such as video conferencing and high-resolution image applications. The deployment cost is very high due to the nature of implementation complexity, coverage area, and expensive frequency spectrum licences (Makela, 2008), (Song and Zhuang, 2005).

In contrast, WLANs usually support packet-switched services, and these technologies have smaller coverage area while providing higher data rates at lower costs. IEEE 802.11n/g is the most common WLAN and it supports a data rate up to 11 Mbps. The subsequent versions of WLAN such as 802.11a, 802.11g, 802.11e and the latest version, 802.11n, offers maximum rate of 54 Mbps and 500-600 Mbps respectively (Hua et al., 2004), (Cavalcanti et al., 2005), (Bobarshad, 2010). WLAN technologies are efficient in serving higher data traffic but they have limitations such as poor mobility management and higher handover times (Hasib and Fapojuwo, 2008), (Saravanan et al., 2006), (Song et al., 2005). The main problem using multiple heterogeneous wireless networks over time is selecting when, and to what network, to perform a handover. Handover schemes may be reactive, where the target network is selected on instantaneous measures such as signal strength, or proactive, where extrinsic information concerning the networks is used. In particular, knowledge of the coverage areas of the many networks available can enable mobile clients to increase their QoS significantly (Mapp et al., 2012).

The overall system in the integrated networks should be designed effectively through efficient and accurate modelling. Eventually, the future systems provide integrated and seamless services. However, vertical handover is the main challenge. Vertical mobility interconnects both existing and also any new wireless radio technologies in heterogeneous systems. When there is an arrival request to a channel, either a new call or a handover call can be admitted to the network (e.g. cell or WLAN). If the arrival is rejected by either one of the networks, due to the various reasons (i.e. capacity and mobility), it can move across to the other network to request admission. The vertical handover decision depends on several factors such as user mobility, network condition, and network technologies. Therefore one of the main problems is the continuation of an ongoing call. The attention should therefore be given to finding a solution to the vertical mobility problem. The handover issues and mobility management for wireless and mobile networks are explained in the following section.

2.2.1.1 Handover/Handoff Issues

Customer expectations are same as traditional wired networks in the sense of availability and performance (Trivedi and Ma, July 2002), (Zeng and Agrawal, 2001), (Tekinay and Jabbari, 1991), (Trivedi et al., 2002), (Trivedi et al., 2003), (Ma et al., 2001), (Ma et al., 2000). In a cellular system, the service area is divided into multiple adjacent cells. Each cell has its own base station (BS). Mobile subscribers (MSs) communicate via radio links with BSs (Trivedi et al., 2002), (Agrawal and Qing, 2006). In WLANs, access points (APs) provide access to the network. Mobility is one of the major issues in the performance characterization in wireless and mobile communication systems. It is important to support services when MSs move from one BS/AP coverage area to another. Handover is an important function of mobility management. Handover in wireless networks can be divided into horizontal and vertical handover. Handovers can also be classified according to when the connection to the previous point of attachment (PoA) is broken in relation to when the connection to the next PoA is made, where mobile user is in charge of handover, and the coverage areas of the relevant networks.

- *Hard and Soft Handover:* In a hard handover, the connection to the previous PoA is broken before the connection to the new PoA is made. In soft handover, the connection to the next PoA is made before the connection to the previous PoA is broken. Hard handover is also known as break before make. The old radio link is broken before a new radio link is established. The MS always communicates with one BS at any given time. FDMA and TDMA employ hard handover. On the other hand, in soft handover, the old radio link is broken after the new radio link is established. MS can connect multiple BSs at any given time. CDMA employs soft handover (Agrawal and Qing, 2006), (Mapp et al., 2012).

- *Network-based and Mobile-Based Handover:* In network-based handover, the network makes the final handover decision while in client-based handover, it is the mobile device that decides when to handover. Though current cellular systems use network-based handovers, mechanisms such as Mobile IP and Fast Mobile IP are mobile-based. In fact, for heterogeneous environments mobile-based handover is favoured (Agrawal and Qing, 2006), (Mapp et al., 2006).

- *Upward and Downward Handover:* Where a network, for example A, is completely covered by network B, then if users make a handover from network A to network B, this is referred to as an upward handover because users are going from a smaller network with substantial bandwidth to a network of a much larger coverage with lower bandwidth. While a handover from network B to network A is referred to as a downward handover because users are going from a larger to a smaller network.

In addition, handovers can also be divided into two advanced types such as *imperative* and *alternative* handovers. Imperative handovers occur due to technological reasons only. Hence the mobile node changes its network attachment because it has determined by technical analysis that it is good to do so. This could be based on parameters such as signal strength, coverage, and the quality-of-service offered by the new network. These handovers are imperative because there may be a severe loss of performance or loss of connection if they are not performed. By contrast, alternative handovers occur due to reasons other than technical issues (Mapp et al., 2012). Hence there is no severe loss of performance or loss of connection if an alternative handover does not occur. The factors for performing an alternative handover include a preference for a given network based on price or incentives. User preferences based on features or promotions as well as contextual issues might also cause handover. Finally, there may be other network services that are being offered by certain networks. The trends in handover design, handover arrangement as well as the detailed informations regarding the classification of handovers can be found in Pollini (1996), Noerpel and Lin (1997), Mapp et al. (2012) and Agrawal and Qing (2006).

The integration of different network technologies, heterogeneous networks, is one of the future trends in modern telecommunications. In recent years, more emphasis has been put on the heterogeneous networks to provide a user with ubiquitous network access. In order to achieve that, vertical handover schemes play an important role in switching the access technology seamlessly without application disruption and maintaining the required QoS. hence, the handover procedure for heterogeneous networks has received considerable attention. The vertical handover system that allows users to roam between cells in wireless overlay networks has been implemented in Stemm and Katz (1998). Several studies of the vertical handover procedure, mobility management, common radio resources management schemes and architectural aspects of the vertical handover have been reported in the literature (Gowrishankar et al., 2009), (Hasib and Fapojuwo, 2007), (Hasib and Fapojuwo, 2008), (Lampropoulos et al., 2007), (Mishra, 2006), (Saravanan et al., 2006), (Shensheng and Wei, 2005), (Song et al., 2008a), (Song et al., 2008b), (Stevens-Navarro et al., 2007), (Usman et al., 2008), (Xia and Shen, 2007), (Xia and Shen, 2009) and (Sadia et al., 2006).

In addition, there are several dynamic factors that must be emphasized and considered in vertical handover decisions for effective network usage such as user policies determining the best network and whether vertical handover should take place or not. Policy parameters and factors that affect the vertical handover can be found in Wang et al. (1999), Gazis et al. (2005), Rutagemwa (2007), Saravanan et al. (2006), Cavalcanti et al. (2005) and Varshney and Jain (2001). In fact, no single wireless network technology can provide ubiquitous network coverage or service. Therefore, it will be necessary for a mobile user to employ various points of attachment to maintain connectivity to the system at all times. The common trend is to provide small-coverage area and high-bandwidth for users such as WLAN whenever available, and to switch to an existing cellular network such as GPRS, EDGE, UMTS etc. network with lower bandwidth when a WLAN is not available. Several experimental studies and the analysis of vertical handover between cellular/WLAN (Gowrishankar et al., 2009), (Saravanan et al., 2006), (Shensheng and Wei, 2005), (Song et al., 2008a), (Song et al., 2005), (Song and Zhuang, 2005), (Xia and Shen, 2007), (Xia and Shen, 2009), UMTS/WLAN (Benson and Thomas, 2002), (Gazis et al., 2005), (Lampropoulos et al., 2007), (Song et al., 2007) and GPRS/WLAN (Salkintzis et al., 2002), (Usman et al., 2008) have been reported in the literature. In these studies, the vertical handovers are envisaged to perform from a WLAN to a GPRS/UMTS only with a very low priority. However, a handover from a GPRS/UMTS to a WLAN should be performed whenever possible.

The following part reviews relevant handover related work within the scope of the thesis.

2.2.1.2 Handover in Cellular Networks

In Trivedi and Ma (July 2002) and Trivedi et al. (2003) the Erlang-B formula has been used to obtain the loss probabilities in wired networks, but this formula cannot be used in wireless and mobile networks because of the handover phenomena. Handover procedures are an important part of cellular network design, since the some of the users in these networks are inherently mobile.

Pollini summarizes the early research work related to handover in cellular networks (Pollini, 1996). However, the handover related scientific research started in 1980s. Transferring data or an on-going call from one channel (or cell) to another was investigated and analysed. The articles focused on the analysis of the handover problem are based on the channel assignment in the cellular and microcellular networks. In addition, channel assignment strategies have an important role to obtain QoS measures in wireless cellular networks during the handover process. In Beck and Panzer (1989), Noerpel and Lin (1997), Hong and Rappaport (1986), Tekinay and Jabbari (1991), Agrawal et al. (1996) and Zhong et al. (2002) the effects of handover on the performance of a cellular network by presenting channel assignment strategies and handover policies have been analysed. In Noerpel and Lin (1997), Lin et al. (1994) and Tekinay and Jabbari (1991) various channel assignment strategies have been classified. Fixed channel assignment (FCA), dynamic channel assignment (DCA), and flexible channel assignment are the most widely used schemes. FCA is assumed in this research where the set of channels are permanently assigned to each BS or cell (Zeng et al., 1994a). More advanced channel management policies for handovers are presented in Agrawal et al. (1996). In Mande (1990) and Murase et al. (1991) the experimental simulation studies have been presented to analyse the relationship between handover parameters and unnecessary handover delay. In Gudmundson (1991) the analytical expressions and the performance measures of handover parameters without hysteresis margin are presented. Further analysis has been carried out in Vijayan and Holtzman (1992) based on the handover algorithms between two cells with hysteresis. The handover rate is given in the proposed model that utilizes received signal strength between two BSs. The work is extended in Agrawal and Holtzman (1997) adding more multi-cell system in the analysis. In Zhang and Holtzman (1996) the discrete-time approach is presented and improved by making the analysis more accurate. On the other hand, handover topics in cellular systems have also attracted attention for example: handover criterion for macro and micro-cells (Murase et al., 1991), (Beck and Panzer, 1989), prioritization schemes (Fantacci, 2000), (Zeng et al., 1994a), the effect of traffic or mobility models (El-Dolil et al., 1989), (Hong and Rappaport, 1986), (Ioannou et al., 2002), (Steele and Nofal, 1992), soft handovers in CDMA cellular networks (Wen and Ying, 2004), handover for integrated voice and data (Zeng and Agrawal, 2000a), (Zeng and Agrawal, 2000b), have also attracted attention. Moreover, the 3G and LTE network architectures were designed with a focus on macrocells, microcells and femtocells being a relatively new addition to the existing components. The growth of cellular networks, such as 3G, LTE, resulted in more analytical studies of handover decision algorithms due to increased user mobility together with expected QoS requirements.

In addition, wireless and mobile communication systems are prone to failures. Handover failures and unavailability of an idle channel cause the degradation of such systems. The system may not be supporting the active MSs efficiently. Hence, the system performance may degrade. In order to overcome this problem, availability and performance of the system should be considered together. To obtain realistic composite performance and availability measures, performance changes are considered with channel failure and its recovery behaviour in Trivedi and Ma (July 2002), Trivedi et al. (2002), Trivedi et al. (2003), Ma et al. (2001) and Ma et al. (2000).

2.2.2 Modelling Wireless and Mobile Networks

In the last two decades, the advent of mobile and wireless networks has changed the whole lifestyle and work. Wireless and mobile technologies have had an enormous impact on the telecommunications market. The two most important characteristics of these technologies is the construction of suitable algorithms to support mobility and the radio coverage of the area in which service is offered. In cellular systems, the coverage area where the service is offered is divided into small regions called cells. Each cell is served by a different base station which provides the radio resources to carry the network traffic of the users in the cell. Depending on the considered cellular technologies, radio resources are shared between users according to different channel allocation schemes, based on frequency, time or code division allocation. Regardless of the adopted technique, it is possible to define the fundamental mathematical modelling unit of the systems, the *channel*, as the one allowing the calls to the system for serving purposes. The fact that channels can be shared among the users in a cell is quite significant for system improvement. On the other hand, the new generation systems already appeared even before the full deployment of 3G systems (Sadia et al., 2006), (Varshney and Jain, 2001), (Mishra, 2006). It is not very clear as to how these developments will influence an already complex scenario of mobile cellular systems.

The future systems must be able to seamlessly switch between different wireless technologies. The system must also support service portability and scalability for the users. There are some challenges in modelling and designing such systems since they are expected to operate with other networks and standards. The focus of this section has been literature on development of such models. In the next sections analytical models for performance/performability modelling and evaluation of cellular and heterogeneous networks are discussed.

2.2.2.1 Modelling Cellular Networks

Wireless and mobile communication systems with customer expectations are same as traditional wired networks in the sense of QoS. Various approaches exist in the literature for performance evaluation of mobile systems, especially the handover phenomena (Dharmaraja et al., 2002), (Liu et al., 2006), (Ma et al., 2001), (Madan et al., 2008), (Trivedi et al., 2002), (Trivedi et al., 2003), (Zeng and Agrawal, 2002). Modelling and design of handover schemes is one of the major issues in the performance characterization in wireless and mobile communication systems and is not a new topic to the wireless communication systems field.

The simplest way of assigning priority to handover requests is to reserve a fixed number of channels for them. In the literature, reserving certain number of channels is referred to as the guard channel scheme (Wang et al., 2003), (Ahmed, 2005), (Gowrishankar et al., 2009), (Liu et al., 2006), (Ma et al., 2001), (Madan et al., 2008), (Shensheng and Wei, 2005), (Tekinay and Jabbari, 1991). The guard channel concept was introduced in the middle 1980s for mobile systems (Guerin, 1988). However, policies based on guard channels have long been used in telecommunication systems. In addition, providing buffering is another prioritizing scheme that allows either the handover request to be queued (Tekinay and Jabbari, 1991) or new calls to be queued (Guerin, 1988) until free channels are obtained in the cellular system. Several variations of the basic guard channel scheme, with queuing of handover requests or of new call requests, have been discussed in the literature (Guerin, 1988), (Tekinay and Jabbari, 1991). The queuing of handover requests, with or without employing guard channels, is another prioritizing scheme which increases QoS of the systems (Ekiz et al., 2005).

It has been mentioned in Lin et al. (1994) that the newly selected system or base station must have an available free channel for handling the handed over calls. Otherwise, the handed over call will be blocked. Blocking an on-going voice call is highly undesirable. Blocked handover calls cause bandwidth consumption as well as wasting of time in the networks where the blocked call has to be re-established. Implications of blocking data packets may have even more serious consequences because of the need to re-transmit the data to ensure reliable connectivity. In Lin et al. (1994) the guard channel technique is suggested for minimizing the call blocking probability due to the unavailability of channels. In this technique, one or more channels are reserved for handover calls. Guard channels based reservation scheme has been analysed in Ahmed (2005) and Ma et al. (2001) blocking probabilities are calculated using CTMC models. In addition, queueing mobile calls with guard channels is another important scheme that frequently used in the literature. Hong and Rappaport (1986) is a frequently referenced paper which studies the system with voice handover queue. Handover schemes with two-level priority reservation have been proposed in Zeng and Agrawal (2001), Liu et al. (2006), Agrawal et al. (1996) and Zhong et al. (2002).

A traffic model is a primary requirement to develop a detailed analytical model for mobile and wireless cellular systems. The call arrival process, call holding time and cell dwell time are three main parameters of a traffic model. To define the most suitable traffic model for wireless networks, a lot of research has been done based on different assumptions about user mobility. Different models are derived based on different assumptions for the distribution of dwell times and call durations. One important study considered in such research is Fang et al. (1997a), where a general distribution of dwell times is assumed while different kinds of distributions are considered for the call duration. Call arrivals are Poisson; the blocking probability is given and the authors drive a number of performance metrics including the expected effective call duration. A similar approach is followed in Fang et al. (1997b) by considering a general distribution of the call duration. In Orlik and Rappaport (1998), combinations of hyper-exponentials are employed for the dwell time and a multi-dimensional Markov chain model is derived. Phase-type distributions are employed in Orlik and Rappaport (1998) by expressing the distribution of both dwell times and call durations as sums of hyper-exponential distributions. The hyper-Erlang distribution is used in Fang and Chlamtac (1999a) and Fang and Chlamtac (1999b) instead. In Li and Alfa (2000), the case of general call duration and dwell time is considered when the policy of channel reservation to handovers is adopted. The results are obtained by introducing some approximations based on the modified offered load techniques and an optimal channel reservation rule is derived.

The distribution of dwell times can also be characterized from the cell shape and the user mobility in terms of speed and radius. In Hong and Rappaport (1986), a traffic model is proposed for a hexagon cell which is approximated by a circle, where it is assumed that the vehicles are spread over the service area. The location of a vehicle is uniformly distributed in the cell when a call is initiated by the user. It is also assumed that a vehicle moves from current location in any direction with equal probability when the call is initiated. The vehicle direction does not change while the vehicle remains in the cell. An extension of Hong and Rapport‘s traffic model for the highway micro-cellular networks has been introduced in El-Dolil et al. (1989). The highway is divided into micro-cells with small BSs broadcasting radio signals along the highway with certain assumptions. In addition, in Steele and Nofal (1992), a traffic model has been studied considering the pedestrians making calls while walking along a street based on the city street micro-cells. However, Steele‘s traffic model is not convenient for an irregular cell and/or vehicular users. Xie and Kuek assume a uniform density of mobile users through an area (Xie and Kuek, 1994). Users are equally likely to move in any direction with respect to the cell border. Xie and Kuek focused on the probability density function (pdf) of the user speed by making biased sampling. However, Xie and Kuek‘s traffic model is relatively small when the blocking probability of originating and handover calls are small. The distribution of mobile speeds of handover calls used in Hong and Rappaport (1986) has been adjusted by using *f(v)* which is the pdf of a random variable V (speed of the user) and *E[V]* is the average of the random variable V.

$$f^*(v) = \frac{vf(v)}{E[V]} \tag{2.1}$$

$f^*(v)$ leads to the conclusion that the probability of handover in Hong and Rappaport (1986) traffic model is the most convenient one because the speed distribution of the handover calls are not same as the overall speed distribution of all mobile users.

In Hong and Rappaport (1986), the user's mobility, the shape, and size of the cell have been considered. The exponential distribution is assumed to determine the distribution of the cell dwell time and the channel holding time. This method has been extensively used by Zeng and Agrawal (2002), Wang et al. (2003), Liu et al. (2006) and Zeng and Agrawal (2001). All approaches have their own advantages. Since the focus of this thesis is looking wireless and mobile systems, the exponential distributions for channel holding time and cell dwell time are carefully chosen to develop an analytical model. The main reason is that the exponential distribution for the call holding time and the cell dwell time make the model tractable in the analytical modelling. This approach is based on the two-dimensional fluid flow model. The MSs are assumed to be uniformly distributed in the whole service area and a MS is equally likely to move in any direction. The calculation of the cell dwell time is explained as follows.

An active user may finish the communication because of either an outgoing handover or a call termination. The channel holding time is the time interval where a MS occupies the channel in a given cell. MS releases the channel due to several reasons such as mobility and call duration. Channel release can then happen because of completion of the calls or handover to other systems or to a neighbouring cell. Let T_{oc} be defined as an average call holding time in cellular network and follow an exponential distribution with mean $E[T_{oc}] = 1/\mu_{oc}$. On the other hand, cell dwell time is the duration that MS stays in a certain cell during the connection period of call. According to the fluid flow model, the average moving speeds of non-real-time MS and real-time MS are equal, i.e., $E[V_n] = E[V_r]$, where $E[V_n]$ and $E[V_r]$ are the average moving speeds of the non-real-time MS and real-time MS, respectively. The average cell outgoing rate of MSs can be given as (Liu et al., 2006), (Zeng and Agrawal, 2001), (Zeng and Agrawal, 2002):

$$\mu_{cdwell} = \frac{E[v]L}{\pi A} \tag{2.2}$$

where $E[V] = E[V_n] = E[V_r]$ is the average of the random variable, V which is the speed of the mobile users, L is the length of the perimeter of a cell, and A is the area of the cell. The cell dwell time T_{cdwell} (random variable) of an MS can be given as (Wang et al., 2003), (Liu et al., 2006):

$$E[T_{cdwell}] = \frac{\pi A}{E[v]L} \tag{2.3}$$

Therefore, the channel holding time of a call is equal to smaller one of average call holding time T_{oc} and the cell dwell time T_{cdwell}. The details of this approach can be found in Hong and Rappaport (1986).

Clearly, in Ma et al. (2001),Trivedi et al. (2002), Trivedi et al. (2003), Zeng and Agrawal (2001), Liu et al. (2006), Wang et al. (2003), Ma et al. (2000), Ever (2007) and Gemikonakli et al. (2007), the call arrival process is assumed to be a

Poisson process since the arrivals are expected from large numbers of independent sources. In Zeng and Agrawal (2001), a steady-state traffic model is provided and an analytical model of the system performance is presented for analysis of data services in mobile cellular radio systems. Originating calls as well as the handover requests are allowed to queue,when they find no idle channels upon arrival. In Liu et al. (2006), smart antennas are used in future generation wireless and mobile networks in order to provide high capacity service and quality. An analytical model for a novel multimedia wireless and mobile network is considered using smart antennas. The system is analysed and blocking probability of originating calls and blocking probability of handover calls are computed. Moreover, an analytical model for integrated real-time and non-real-time service in a wireless mobile network with priority reservation and pre-emptive priority handover schemes are considered in Wang et al. (2003). The service calls are classified into four different types; namely, real-time and non-real-time service originating calls, and real-time and non real-time handover service request calls. Accordingly, the channels in each cell are divided into three parts: one for real-time service calls only, the second for non-real-time service calls only, and the last one for overflow of handover requests that cannot be served in the first two parts. In the third group, several channels are reserved exclusively for real-time service handover so that higher priority can be given to them. The system is modelled using a multi-dimensional Markov chain and a numerical analysis is presented to estimate the blocking probabilities of originating calls, forced termination probability, and average transmission delay.

However, in all of the above studies, the research is focused on performance evaluation of wireless cellular networks. The models developed also considered the mobility issues in wireless cellular systems. The probability of an ongoing call being dropped due to a handover failure and the probability of a new call being blocked due to the temporary unavailability of an idle channel are main metrics that affect the performance of wireless cellular systems; it is therefore essential to consider these systems for performability evaluation. Although Zeng and Agrawal (2001) considers mobility issues such as the radius of the underlying cell structure and the velocity of the mobile subscribers, the models presented do not take availability issues into account. On the other hand, Trivedi provides hierarchical modelling approaches for performability evaluation of wireless cellular systems, but do not take mobility issues into account (Trivedi and Ma, July 2002), (Ma et al., 2001),(Trivedi et al., 2002), (Trivedi et al., 2003). In this thesis, an exact modelling and solution method is proposed with both mobility issues and availability models for performability evaluation of such systems. In addition, the above studies are considering the standalone systems. Vertical handover decision management and sensitivity analysis are presented considering the integration of different technologies which is explained in the next section.

2.2.2.2 Modelling Heterogeneous Wireless Networks

The heterogeneous wireless networks are expected to support different radio access technologies, high-rate multimedia services, and ubiquitous access any-time and anywhere. Heterogeneous wireless networks are evolving to provide diverse types of services and traffic such as data messaging, web browsing, file transfer, video, multimedia streaming, as well as traditional voice. In addition, heterogeneous wireless networks should provide a seamless service while a user is roaming, with high mobility and a network that should be able to guarantee predefined levels of QoS allowing for diverse service types. Dealing with QoS requirements, and providing the best configuration in order to maximize the performance characteristics of such systems is a challenging task. This is mainly because of the heterogeneous nature of the wireless and mobile communication systems. In order to achieve these goals, integrating various technologies is an effective way. Convergence technology is a mechanism that attempts to seamlessly integrate various existing wireless networks instead of developing a new uniform standard for wireless communication systems (Saravanan et al., 2006). Currently, WLAN and cellular networks are the most dominant wireless networks. The following are some possible integrated networks:

- A network integrating different types of cellular networks such as GSM/2.5G General Packet Radio Service (GPRS) networks and 3G cellular networks such as WCDMA and CDMA-2000 systems;

- A network integrating different wireless networks based on IEEE 802 standards such as WLAN implemented by 802.11b, 802.11a, 802.11n/g and

WMAN by 802.16 standards;

- A network integrating WLANs and 3G cellular networks.

Before proceeding with the developing QoS models and suitable tool for heterogeneous wireless networks, it is necessary to briefly investigate its main characteristics in each infrastructure. Two major access technologies for heterogeneous wireless networks are cellular networks such as GSM, GPRS, EDGE, UMTS and WLANs. Cellular networks have been well deployed around the world. Hence they have a global coverage area. Both circuit and packet switched services are supported by cellular networks. However, cellular networks usually have low data rates. On the other hand, WLANs usually support packet-switched services and WLANs provide higher data rates at lower cost, but they have limited coverage area and less support for high speed mobility. WLAN technologies are efficient in serving higher data traffic, but they have limitations such as poor mobility management (Saravanan et al., 2006), (Song et al., 2005), (Hasib and Fapojuwo, 2008). Providing end-to-end QoS support is one of the main purposes in the design of integrated networks. In order to develop efficient integrated networks, handover issues caused by user mobility and transmission through the unreliable wireless media between both homogeneous cells and heterogeneous cells are the main difficulties that need to be considered. For the successful integration of heterogeneous networks, designing of vertical handover between two heterogeneous networks is the one of the key elements.

Most of the previous research studies consider homogeneous cellular networks without considering the characteristics of different networks (Wang et al., 2003), (Trivedi et al., 2002), (Zeng and Agrawal, 2001), (Tekinay and Jabbari, 1991), (Jindal et al., 2006), (Zhong et al., 2002). In other words, the integrated 2G/3G or 3G/WLAN structure is not considered with an exact solution technique for performance evaluation. GSM is the dominant type of WWAN used for wireless services. The popular 3G cellular networks for data services include GPRS, EDGE and UMTS systems. These networks operate in reserved frequency bands and support data rates from 10Kbps to 2Mbps. They usually cover a large geographical area (from few hundred meters to several kilometres) and support mobile user mobility by using several techniques such as location updating, paging, and handover. Most of the previous research studies considered cellular network characteristics and analyse the systems based on this. Cellular systems support a mixture of platforms. Channels are allocated for each BS. The wireless links between a BS and users employ either time-frequency division multiple access (TDMA/FDMA) or spread-spectrum code division multiple access (CDMA) techniques (Dharmaraja et al., 2002), (Gowrishankar et al., 2009), (Hasib and Fapojuwo, 2008), (Shensheng and Wei, 2005), (Trivedi et al., 2002), (Tekinay and Jabbari, 1991), (Zhong et al., 2002). Different channel allocation schemes have been studied in Noerpel and Lin (1997) and Tekinay and Jabbari (1991).

However, the resource allocation policies in WLANs are different to those for cellular systems. Contrary to WLANs, cellular networks can provide guaranteed QoS. However, cellular network infrastructure is more expensive than WLAN infrastructure and hence the charges involved in using cellular networks are more expensive compared to WLANs. WLANs operate at the license-exempt industrial, scientific, and medical (ISM) frequency bands and can support high data rates. Generally, WLANs cover small geographic area and do not support terminal mobility over a wide area. In addition, they are mostly available within distinct areas called hotspots. The WLAN media access mechanisms are mostly based on the contention and therefore limited quality of services (QoS) are provided. Most of WLANs today are based on the IEEE 802.11x standards. The IEEE 802.11b (the most popular WLAN standard) can support data rates up to 11 Mbps and operate at 2.4 GHz band. The IEEE 802.11a and 802.11n/g can support data rates up to 54 Mbps and respectively operate at 5 GHz and 2.4 GHz 12 bands. The coverage of IEEE 802.11x standards is up to few hundreds of meters around each access point (AP) (Hua et al., 2004), (Jayaparvathy et al., 2007), (Khatib, 2003), (Liu et al., 2010), (Winands, 2003), (Winands et al., 2004), (Yin et al., 2004), (Yin et al., 2004). WLANs are now seen as a complementary technology that can be used to enhance the system capacity in integrated wireless networks. In order to support integrated cellular/WLAN systems, provide an acceptable level of QoS and maximize the system capacity, modelling and performance analysis of such systems has to be taken into account.

In the literature, various integrated schemes have been proposed and performance analysis is considered for such systems (Cavalcanti et al., 2005), (Hasib and Fapojuwo, 2007), (Hasib and Fapojuwo, 2008), (Lampropoulos et al., 2007), (Mapp et al., 2012), (Mishra, 2006), (Rutagemwa, 2007), (Saravanan et al., 2006), (Song et al., 2008a), (Song et al., 2007), (Stevens-Navarro et al., 2007), (Usman et al., 2008), (Xia and Shen, 2007), (Xia and Shen, 2009). However, the research emphasis has been on the interworking architecture, channel assignment schemes and/or mobility. In addition, different interworking and handover strategies of the integrated systems have been presented.

In Xia and Shen (2007) and Xia and Shen (2009) integrated service-based handover schemes are proposed. In addition, several radio resources management schemes and mobility management policies for integrated systems are presented in Hasib and Fapojuwo (2007) and Hasib and Fapojuwo (2008). However, issues such as the speed of mobile users and radius of the networks are not taken into account for both cellular and WLAN. Based on the mobility models of Zeng and Agrawal (2001) and Zeng and Agrawal (2000a), the speed of a user and radius of the networks are taken into account in this thesis. Previous research efforts have focused on the integration of cellular and WLANs on multi-channel basis (Hasib and Fapojuwo, 2007), (Hasib and Fapojuwo, 2008), (Saravanan et al., 2006), (Shensheng and Wei, 2005), (Song et al., 2005), (Xia and Shen, 2007), (Xia and Shen, 2009). However, the existing channel allocation schemes and assuming multi-channels in WLAN AP are not convenient for WLANs due to the system architecture and characteristics of WLANs. In WLANs, calls/packets assignment depends on the medium access. WLAN (the IEEE 802.11) MAC employs the carrier sense multiple access with collision avoidance (CSMA/CA) approach to accommodate multiple users and is designed for sharing a single channel (Bianchi, 2000), (Bobarshad, 2010), (Chen et al., 2003b), (Chen et al., 2003a), (Hua et al., 2004), (Jayaparvathy et al., 2007), (Khatib, 2003), (Liu et al., 2010), (Winands, 2003), (Winands et al., 2004), (Zhao and Leung, 2006), (Agrawal and Qing, 2006). The MAC has two basic access modes. The primary access mode of this standard is the distributed co-ordination function (DCF). The second channel accessing mechanisms is the point coordination function (PCF). Both DCF and PCF are slotted protocols. The DCF is basically carrier sense multiple access with collision avoidance (CSMA/CA) mechanism and PCF is based on the polling technique. Basic and four-way handshaking schemes are two schemes for DCF (Hua et al., 2004), (Shensheng and Wei, 2005), (Wang et al., 2003). In the basic scheme, an acknowledgement (ACK) is transmitted by destination to the sender. The four-way handshaking scheme involves the Request To Send and Clear To Send (RTS/CTS) mechanism. The details could be found in Hua et al. (2004) and Wang et al. (2003). In the literature, most of the existing WLANs schemes are still using the same DCF MAC protocol mechanism to channel access based on CSMA/CA (Bobarshad, 2010), (Hui and Shankaranarayanan, 2004), (Shensheng and Wei, 2005), (Khatib, 2003), (Zeng et al., 1994a). The CSMA/CA mechanisms assures fair sharing of WLAN channel stations since WLANs use a shared medium (Bobarshad, 2010), (Hui and Shankaranarayanan, 2004), (Shensheng and Wei, 2005), (Khatib, 2003), (Zeng et al., 1994a). The CSMA/CA protocol first listens the medium and determines the status of the channel. If it is idle, it starts to transmit. If the channel is busy, it defers its transmission for a random period of time. The APs usually have more traffic variations when the network operates in the infrastructure mode (Bobarshad, 2010), (Hui and Shankaranarayanan, 2004), (Shensheng and Wei, 2005), (Khatib, 2003), (Zeng et al., 1994a). Traffic load makes the channel unreliable and unavailable for transmission. This causes severe degradation in the performance of WLANs. In the literature, existing work has been done on WLANs, however most of the investigations are related to throughput and QoS analysis of applications (Bianchi, 2000), (Chen et al., 2003b), (Hui and Shankaranarayanan, 2004), (Jayaparvathy et al., 2007), (Liu et al., 2010), (Malone et al., 2007a), (Raul and Jung, 2011), (Shensheng and Wei, 2005), (Song et al., 2005), (Winands et al., 2004), (Winands, 2003). Several simulation and analytical studies have been done to evaluate the performance of WLANs based on DCF. Winands showed in Winands et al. (2004) and Winands (2003), that the WLAN considers the characteristics of CSMA/CA such as the contention and binary exponential back-off mechanisms (BEB) as a queuing model. It is shown that the behaviour of the DCF and BEB is well approximated and could be assumed for WLANs. In addition, an experimental work is shown in Khatib (2003). Based on their analysis and testbed results, the WLAN system could be modelled as a queuing system with a single queue and a single channel. Hence, the work in this thesis models WLAN AP as a single channel queuing network due to the fact that WLAN AP operates on a single channel.

Mobility is also very important in determining the network performance in integrated networks. In Wang et al. (2003), Trivedi et al. (2002), Trivedi and Ma (July 2002) and Zeng and Agrawal (2001), horizontal handovers, and mobility between homogeneous networks are considered (e.g. cellular to cellular). However, vertical handovers within an integrated (heterogeneous) system is the main concern in the future generation systems. Because of the heterogeneous nature of modern communication systems, the network engineers should consider the best configuration where vertical handovers are considered together with horizontal ones.

In addition, proposed models consider vertical handover decision management, sensitivity analysis as well as operational spaces. The proposed models show that such an integrated system can be modelled as a two stage open queuing system. Queuing networks have been successfully used in performance and availability modelling of computer and communication systems. They are especially suited for representing resource contention and queueing for service. The two stage tandem network is considered for modelling and evaluation of the integrated networks. In a model for heterogeneous system QoS and performance analysis, a service centre for the users and each channel and possibly others are required. A service centre may have one or more servers/channels associated with it. If a call requesting service finds all the channels at the service centre busy, it will join the queue associated with the centre. When one or more of the channels become idle, a call from the queue will be selected for service according to some scheduling discipline. After completion of service at service centre, the call may join to another service centre for further service, re-enter the same service centre or leave the system. The probabilities of transitions between service centres and the distribution of call service times at each centre characterize the behaviour of calls within the network. For each network, the number of channels, the scheduling discipline, and the size of queue must be specified. Consider the simple two stage tandem network shown in Figure 2.1.

The system consists of two nodes with respective service rates, μ_1 and μ_2. The external arrival rate is λ. The output of the node, labelled 1, is the first stage of the two stage tandem network as well as the input of the second stage, labelled 2. The service time distribution at both nodes is exponential, and the arrival processes to the node 1 is Poisson. W_1 and W_2 are queue capacities of stage 1 and 2 respectively.

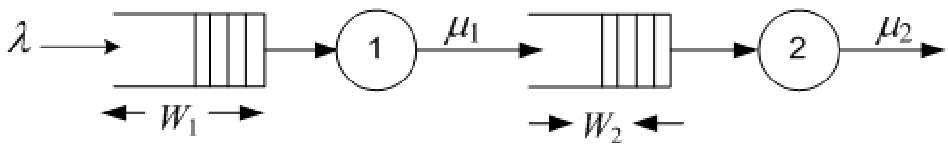

Figure 2.1: A two-stage tandem network

This system can be modelled with stochastic process whose states are specified by pairs (k_1,k_2), $k_1 \geq 0$,$k_2 \geq 0$, where k_i (i=1,2) is the number of calls at channel i in the steady state. The changes of states occur on a completion of service at one of the channels or with an external arrival. Since all the inter arrival times are exponentially distributed, the integrated systems can be introduced using this model. Currently in the literature on integrated heterogeneous systems, simulation appears to be preferred over application of queuing theory model for a number of reasons. Firstly, the difficulty in modelling mathematically complicated traffic features of such systems. However, complexity and the long simulation time make it difficult to employ simulations especially when studies such as operational space of parameters and sensitivity analysis are being performed. The use of simulation for such scenarios would inevitably introduce a large number of parameters and may require extensive amount of simulation runs. Furthermore, the analytical models employed are useful to understand the insight nature of stochastic processes and derive relevant conclusions by analysing the mathematical relations of parameters on a specific measure of interest. Thus with the advantage of being able to evaluate system performance under a wide range of conditions with relatively simple computation, an analytic model of a network would be still very attractive. Meanwhile, an analytic model is essential to the understanding of the underlying principles of the networks. So there is a need to extend the analytic models of such integrated systems. Using proposed models, the behaviour of the heterogeneous networks described above can be adopted for the performability evaluation of such systems. This can be done by multi-dimensional models with a Markovian framework. In this thesis multi-dimensional models are used for performance and performability evaluation for wireless and mobile communication systems. The spectral expansion method can be employed for both performance and availability model with well-known MRM.

2.3 Summary

In this thesis, analytical models are developed for performance/performability modelling and evaluation of wireless and mobile technologies. In this chapter, the domain of the research is given by providing a critical review of related literature. Handovers for the wireless and mobile systems are discussed with mobility and availability issues.

The mobility is one of the main challenges of modelling wireless and mobile systems. Mobility implies that users can roam in the service area while accessing the systems. An active user, who moves from a cell to another cell must execute handover procedure, transferring the call from the radio channel in the old cell to a channel in new cell without interruption of the call.

Since wireless and mobile networks are too complex to be analysed and studied as whole, their modelling and design are decomposed into two tasks. The first task is finding the suitable model to the system such as feasible allocation of the available channels, buffering and physicals constraints, i.e., radius and perimeter of the cell and the speed of the user. In addition, the following assumptions will be made for development of models. The dynamics which drive the user's behaviour are the generation of new service request (arrival rate) and the termination of service in progress (service rate). Radio channels are exclusively employed by one user at a time for the whole duration of the service. In the second task, cells are considered one by one, depending on the technology, by taking into account the user's need and behaviour.

The second task solves the system with a suitable solution approach to obtain acceptable performance while providing the desired QoS to users. The complete modelling and design procedure may require several iterations between these two tasks. Thus, the models for the cell analysis should be carefully considered to make sure the solution approach employed is not numerically cumbersome for the wireless system's modelling and analysis.

Analytical models presented in the literature for homogeneous standalone networks are considered with mobility issues and availability models for various handover schemes in order to obtain more realistic measures. In addition, handover failures and unavailability of an idle channel are also considered since they cause degradation of overall performance in such systems. Therefore, these factors are stochastically modelled in the next chapters for performance and availability evaluation of wireless and mobile systems in order to specify the thresholds and perform optimisation studies.

Generally, there is no single wireless network with all best characteristics to support wireless Internet application and services. However, the integration of WLAN and cellular networks presents complementary and attractive characteristics which can support a wide range of wireless Internet applications and services. Thus, the modelling and performance analysis focused on the deployment of WLAN AP which is located inside a cellular network (3G) to develop a generic approach. In other words, the integrated 2G/3G or 3G/WLAN structure is considered with an exact solution technique for performance evaluation. Mobility issues are considered for both cellular and WLAN. The proposed models take horizontal as well as vertical handovers, user mobility, queue capacity and characteristics of both cellular technologies and WLAN into account. Integration of cellular network technologies with WLANs is a common configuration especially when future generation cellular networks are considered. Eventually, such networks will provide integrated and seamless services.

The next chapter analyses the approach and constraints needed to address the research question. In the following chapter, existing performance/performability modelling techniques and solution approaches are studied together in order to develop exact and efficient analytical models for performance/performability evaluation of wireless and mobile communication systems.

Chapter 3

Existing Modelling Approaches and Solutions Techniques

Today's technologies and standards provide high speed, high capacity and multi-service for the users. To meet these requirements, the specification, design, verification, implementation, evaluation, and testing play an important role for developing models and analysis of wireless networks and communication systems. There are different approaches for modelling and analysis of computer networks and communication systems. Figure 3.1 depicts the picture of the general modelling approaches. These approaches and their analysis depend on the use of mathematical knowledge as well as computer aided tools and methodologies.

The most common approach is state space models. A state space model is a description of a configuration of states used as a simple model of the system under study (Bobbio and Trivedi, 1990). State space models provide the ability to model systems with transitions between states partially independent from each other. Many practical systems can be constructed and analysed using state space methods. The advantage of this approach is that it yields accurate results. However, the state space models are generally faced with two problems, the state space explosion problem, namely largeness, and stiffness (Trivedi and Ma, July 2002), (Trivedi et al., 2003), (Gemikonakli et al., 2009), (Ever et al., 2012). The transition probabilities or rates of the models widely vary and create stiffness. Aggregation techniques (Bobbio and Trivedi, 1990) and stiffness-tolerance (Malhotra et al., 1994) can be applied to stiffness problem. In order to solve the state space explosion problem, hierarchical modelling approach can be used (Trivedi and Ma, July 2002), (Ma et al., 1998), (Malhotra and Trivedi, 1993). This approach has several advantages, such as the ability to decompose large systems into several sub-models and hence avoid their largeness problem.

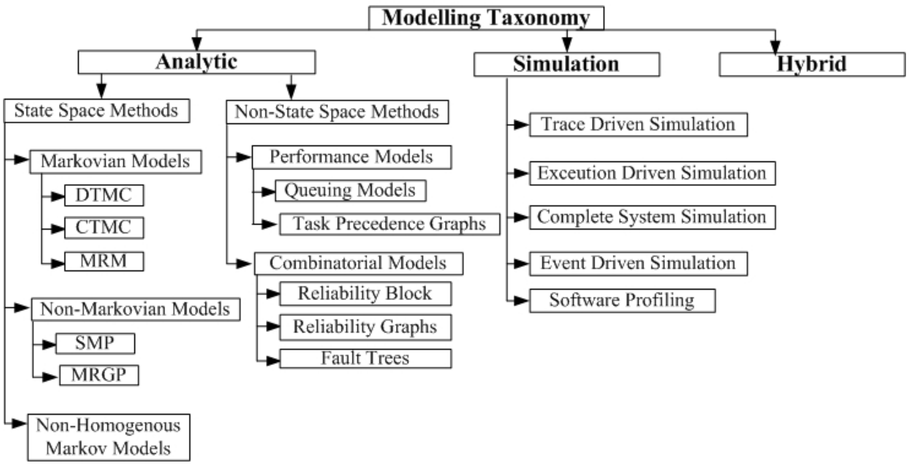

Figure 3.1: Modelling Approaches

In Trivedi and Ma (July 2002), Ma et al. (1998) and Malhotra and Trivedi (1993), hierarchical modelling approaches are used and performability models for wireless cellular networks based on MRM are presented. The behaviour of availability and performance models has been described as a CTMC. The availability models are used in MRM and reward rates are obtained from the performance measure of the associated state. To allow a Markov chain analysis, it is assumed that failure and repair times used in performability modelling of wireless cellular networks have exponential distributions.

The product form solution approach, two-dimensional representations of steady states with hierarchical modelling techniques and two stage tandem networks are used in this thesis. State variables ($I(t)$ and $J(t)$) are specified according to the characteristics of the model and the behaviour of system under study. In the following part, merits and demerits of pure performance evaluation, availability model and composite performance and availability modelling are investigated. In addition, the exact spectral expansion solution which is extensively used in thesis is explained.

3.1 Pure Performance Evaluation

The main scope in this thesis is performance and/or performability modelling and evaluation of wireless and mobile communication systems. Therefore, it is necessary to be aware of various performance evaluation approaches. Benchmarking, simulation and analytical modelling are three different and well known performance evaluation approaches. In this section, strengths and weaknesses of these approaches are discussed.

In the benchmarking approach, the system measures a standard set of tasks under normal conditions (Banks et al., 2005). The accurate and direct results can be obtained. This is the main advantage of using this method. On the other hand, testing takes large amounts of time and obtaining the system performance is costly in terms of personnel and equipment etc. Simulation is another well known approach used for performance evaluation. There are existing simulation packages such as OPNET, OMNeT NS-2/NS-3, NetSIM, JMT etc. or it is possible to write programs using well known programming languages (C++, Java) for lower level simulation models. Simulation models have excellent flexibility and ranges that can be simulated and ability of handling detailed scenarios are the advantages. It is faster than benchmarking but not as fast as analytical modelling. Usually large number of long simulation runs are required for an elaborate and correct analysis. Since the system's parameters can have various characteristics and since complex systems may have large numbers of parameters, constructing simulation software also takes times. The last approach is mathematical modelling. It is the most suitable approach and this approach provides exact or approximate solutions using cutting-edge mathematical techniques that are computationally efficient. Fast computations and the formula can be obtained and this is the main advantage of analytical modelling. On the other hand, high level mathematical skills is one of the requirements of analytical modelling. Several assumptions may be required for approximate solution. When it is compared with simulation, it is not as flexible as simulation and it is not as accurate as benchmarking. However, analytical modelling is ideal for quick and once validated relatively accurate computations. Therefore, a lot of complex and useful analytical models are in the research area and they are evolving quickly. Hence, understanding the mathematical models is essential. Queuing theory and Markov processes are mathematical representations of systems with stochastic processes. They are fundamental to the mathematical modelling. Terminologies and theories of queuing theory are widely used in performance modelling of computer networks and communication systems. These concepts are briefly discussed and explained as follows.

The term of queue is expressed as a waiting line in mathematical theory. Basically, theory of waiting lines can be explained as queuing theory. Queuing theory is used for describing the behaviour of queues in a system. So the realistic measures of interest such as performance and reliability/availability measures can be computed in mathematical terms. In queuing theory, the jobs/data/calls arriving at random points in time, queue for service and depart from the system after service completion. This is typical definition of queuing system. In addition, Kendall's notation is commonly used for representation of a typical queuing system. It has six parameters to specify a queuing system. A/B/c/K/m/Z is the notation where A is the distribution of arrival process, B is the service process or the service time distribution, c is the number of servers, K is the queue capacity, m is the population size and Z is the queuing discipline. In queuing theory, usually when all servers are busy and a new customer enters the queuing network, that customer must wait until any of the servers becomes free. The queues are divided into two categories: Unbounded (queue capacity $\Rightarrow \infty$) and bounded (queue capacity number of servers). Nevertheless, if the number of arrivals is larger than the queuing capacity (arrivals queue capacity) blocking will occur. Because the queue is restricted for a finite number of jobs and if arrivals jobs exceed that queue capacity, there will be no places for coming jobs and blocking will occur. In order to deal with some queuing problems, a number of random variables are considered for probability distributions. The main common discipline in operating systems for serving the arriving customer is first in first served (FIFS) or first in first out (FIFO). In queuing systems jobs/data/calls can arrive according to a Poisson distribution. Additionally, it is necessary to measure a system's ability in order to find effective service for customers. This is known as traffic intensity ρ, and defined as ratio of mean service time *E(s)* and mean inter-arrival time *E(τ)*. The service times are exponentially distributed. On the other hand, a Markov process is one in which the next state of the process depends only on the present state, irrespective of any previous states taken by the process. This means that knowledge of the current state and transition probabilities from this state allows us to predict the possible next state independent of any past state. A Markov chain is a discrete state Markov process. A Birth-Death process is a special type of Markov process often used to model the number of jobs in a queue or a population. If at any time the queue has n number of jobs/data/calls, then birth of another entity (arrival of another jobs/data/calls) causes the state to change to n+1. On the other hand, in case of a death process, jobs/data/calls are removed from the queue for service and this would cause the state to change to n-1. It is obvious that, in any state, transitions can be made only to one of the two neighbouring states.

Theories and terminology given in queuing theory and Markov processes are extensively used in both performance and availability modelling. The pure performance model assumes that the servers under study never fail. However, in practical life, failures are expected and they impact on the system performance. Nevertheless, when the systems are prone to failures and repairs, the systems do not have a simple product form solution. This is because of the irregularities caused by the system failures/repairs and affects performance and availability of the system significantly. Hence, the availability modelling in wireless and mobile networks is briefly mentioned in the next section.

3.2 Availability Modelling

Availability is defined as a fraction of time the system is providing service to its users (Trivedi et al., 2002), (Trivedi and Ma, July 2002). Availability models capture failures and repairs behaviour of systems and their components. Hence, from the designer's and operator's points of view, it is necessary to take failure and repair factors into account. Traditional pure performance models ignore breakdowns and recoveries. Thus, such an approach is optimistic since it assumes that the underlying structure never fails. The system availability model should be taken into account in order to obtain more realistic performability measures of the systems. Trivedi (2002) considers availability under three main titles as:

combinatorial model types, state space models and hierarchical models (shown in Figure 3.1). Combinatorial model types capture conditions that make a system fail, in terms of the structural relationships between the system components. In this approach, the models are solved without generating a state space. Typical examples for these types of models are reliability block diagrams, reliability graphs, and fault trees. Combinatorial type models can be solved using fast algorithms only by assuming stochastic independence between system components. They assume that the failure or repair of a component does not affect other components. Also, in combinatorial type models, it is assumed that there are as many repair facilities as necessary. In many practical systems, dependencies have been observed among system components, and repair facilities can be restricted (Trivedi et al., 1990), (Chakka, 1995), (Trivedi et al., 1996), (Chakka, 1998), (Chakka et al., 2002). To be able to model more complicated interactions between components, state space models can be used. Although non-state space models are efficient, to specify and solve the solution of these models assumes that components of the system are independent. A failure of one component does not affect the operation of other components, and components cannot share a repair facility. State space models provide the ability to model systems with components not completely independent from each other. The main drawback of this type of model is the state space explosion problem (Gemikonakli et al., 2009), (Trivedi et al., 2002), (Ma et al., 2001). It is possible to overcome the state space explosion problem by using hierarchical models and handling different levels independently (Trivedi et al., 2003).

Mitrani (2005) uses a general availability model for heavily loaded traffic. This study considers an M/M/N queue with independent random breakdowns and repairs. Jobs arrive in a Poisson stream at rate λ and join a single, unbounded queue. The required service time is modelled as independent and identically distributed (i.i.d) random variables. It is distributed exponentially with parameter μ. There are N identical parallel processors, each of which goes through alternating periods of being operative and inoperative, independently of the others. The operative periods are i.i.d random variables distributed exponentially with parameter η, and the inoperative ones are i.i.d random variables distributed exponentially with parameter ξ. Jobs are taken for service from the front of the queue, one at a time, by available operative processors. A job cannot occupy more than one processor simultaneously, and a processor cannot serve more than one job simultaneously. No operative processor can be idle if there are jobs waiting to be served. If a service is interrupted by a processor breakdown an operative processor becomes again available for it, the service is resumed from the point of interruption. The evolution of the system state is represented by the irreducible Markov process X = (I(t), J(t)) ; t⇒0, where I(t) is the number of operative processors and J(t) is the number of jobs present at a time t. The state-state distribution of that process is denoted by:

$$p_{i,j} = lim_{t\to\infty}\ P(I(t) = i, J(t) = j); i = 0, 1, 2, \ldots, N, j = 0, 1, \ldots \tag{3.1}$$

The marginal distribution of the number of operative processors is easily seen to be binomial:

$$p_{i,.} = \sum_{j=0}^{\infty} p_{i,j} = \binom{N}{i}(\frac{\eta}{\eta+\xi})^{i}(\frac{\xi}{\eta+\xi})^{N-i}, i = 0, 1, \ldots, N. \tag{3.2}$$

Hence, the processing capacity of the system, which is defined as the average number of operative processors, is equal to E(I)=Nη/η+ξ. The process X is ergodic if and only if, the offered load is less than the processing capacity λ/μ <Nη/η+ξ (Mitrani, 2001),(Mitrani, 2002). An ergodic process means that, the system is irreducible and the corresponding balance equations of the state probabilities have a unique solution which can be normalised. Steady states do exist for such processes (Chakka, 1995).

The availability model used in this thesis, that considers failures and repairs behaviour is presented in this section. The combined evaluation of availability and performance models is useful to obtain more realistic measures for computer networks and communication systems. In the following section, several examples of such combined performance and availability analysis are discussed in detail.

3.3 Composite Performance and Availability Analysis

The availability and performance models that are commonly used in computer networks and communication systems are discussed and analysed in previous sections. The analysis of computer networks and communication systems from pure performance view tends to be optimistic because it ignores the failures and repairs behaviour of the system. On the other hand, performance consideration is not taken into account in pure availability analysis and tends to be too conservative. In real systems, capacity, availability, and performance are important QoS requirements which should be considered together.

Significant amount of research has been performed in the development of techniques for evaluating the performance and availability/reliability of computer communication systems in an unified manner (Boxma et al., 1994), (Malhotra and Trivedi, 1993), (Malhotra et al., 1994), (Trivedi et al., 1994), (Trivedi et al., 1996). The concept of performability is the combined aspects of performance and availability (or reliability) models defined by Meyer in order to get more realistic results (Meyer, 1980). The author also proposed a general framework for model-based performability evaluation. Extensive research has been carried out to construct an appropriate model and solution in performability modelling. Moreover, in order to develop exact and efficient analytical models for performability evaluation of wireless and mobile communication systems, the tool development and applications have been studied. Several approaches in Gemikonakli et al. (2009), Ever et al. (2012), Trivedi et al. (2002), Trivedi and Ma (July 2002), Bobbio and Trivedi (1990),Ma et al. (1998) and Malhotra and Trivedi (1993) have been found to be useful for performability modelling and evaluation. Composite performance and availability modelling considers two basic steps. The construction of the suitable model and the solution of the model. The composite CTMC and two level hierarchical models are two common approaches that have been used in analysis for composite performance and availability modelling (Trivedi and Ma, July 2002), (Ma et al., 2001). In Trivedi et al. (1994), a unified performability and reliability analysis by using MRM was presented. The MRM approach leads to a separation of the performance and availability models of the system.

The MRM approach is an approximate approach and the most commonly used. The author often uses Stochastic Petri Net (SPN) for the construction and solution of CTMCs or MRMs. SPN is an extension of the Petri Nets (PNs). These approaches have rarely been used for modelling and analysis of networks and communication systems. The philosophy of PNs is the main reason. These models depend on the computational manner with few primitives and as a consequence, makes the developing of models for complex and large scale systems very difficult (Calzarossa and Marie, 1998). Also the BT approach is proposed by Bobbio and Trivedi (Bobbio and Trivedi, 1986), (Ma et al., 2001). This approach separates the transition rates into two as fast and slow rates. The fast rates are generally larger than the slow rates as a magnitude. The state space of CTMC is divided into fast and slow states. Fast states are the ones with at least one fast outgoing transition. Slow states do not have any fast outgoing transition. The fast states can be further divided into several recurrent subsets and at most one transient subset. States in the fast recurrent subset are connected with each other via fast transitions, but are connected to any outside state only by slow transitions. The fast transient subset is connected to an outside state by means of at least one fast transition. The key idea of the BT method is to separate the state space of a CTMC into fast recurrent subsets and/or a fast transient subset. The resulting Markov chain is small and non-stiff. Conventional numerical techniques can be discussed to analyse this Markov chain. However, BT method is an approximate approach (Bobbio and Trivedi, 1986), (Ma et al., 2001).

It is possible to develop an effective and accurate analytical model for performability modelling and evaluation of wireless networks and communication systems by using a two dimensional representation of the states of the system, in a Markovian framework. This approach has been extensively used in Chakka (1995), Chakka and Mitrani (1992), Chakka and Mitrani (1996), Chakka (1998), Chakka et al. (2002), Ever et al. (2005) and Ever (2007). For mathematical tractability, semi finite and finite lattice strips of Markov states are used in this approach. These lattice strips of Markov states, with certain regularity, are also known as Quasi Birth and Death (QBD), Quasi Simultaneous Multiple Births (QSMB) and Simultaneous Multiple Deaths (SMD) processes (Chakka, 1995). This approach has given rise to a number of useful research contributions for performability evaluation of various queuing systems (Chakka, 1995), (Chakka, 1998), (Chakka and Mitrani, 1996). When two-dimensional state space representation is employed, the state of a system at time t is described by a pair of integer valued state variables, *I(t)* and *J(t)*, specifying the server configuration (can also be termed, operative state of a multi-server system) and the number of jobs in the system, respectively. In general, if there are $N + 1$ server configurations, represented by the values *I(t)* $= 0,1,\ldots,N$, these $N + 1$ configurations can be used to represent the possible operative states of the model. The model assumptions are assumed to ensure that *I(t)*, t$\geq$0, is an irreducible Markov process. *J(t)* ($\leq$L) is the total number of jobs in the system at time t, including the one(s) in service. Then, Z = *[I(t),J(t)]; t$\geq$0* is an irreducible Markov process on a lattice strip (a QBD process), that models the system. Its state space is, $\{0,1,\ldots,N\}$ x $\{0,1,\ldots,L\}$ (where L can be finite or infinite). Once the steady state probabilities of such a system are computed, various performance and/or performability measures (mean queue length, blocking probabilities etc.) can easily be driven by using well-known queuing theory knowledge. Similar models in Ever (2007) were analysed for exact performability modelling of homogeneous and heterogeneous multi-server systems and some repair strategies.

Probabilistic models, queuing networks and Markov chains have been largely used in the design of complex systems. In addition, a variety of related applications have been employed to develop exact and efficient analytical models for performability evaluation of wireless and mobile communication systems considering queuing theory and Markov processes. The existing performance and availability modelling approaches can be adapted to wireless networks and communication systems. It is possible to expand and composite the existing performance and availability approaches for performability modelling of such systems. However, the performance modelling of multi-server systems with multiple queues usually leads to multi-dimensional models. The adaptation is only possible under certain assumptions in queuing theory. Most of the performance and availability modelling approaches are not directly applicable to multi-dimensional or complex systems. Therefore, various analytical approaches have been developed to solve multi-dimensional queuing systems. In the following section, existing solution methods are briefly discussed.

3.4 Existing Solution Methods for Multi Dimensional Models

The performance/performability evaluation of computer networks and communication systems leads to the development of multi-dimensional queuing models. The two-dimensional models with multiple components such as multiple queues and/or multiple servers are effectively used. Such models can be represented by certain two-dimensional Markov processes on finite or semi-infinite lattice strips (Gemikonakli et al., 2009). The QBD processes are the subset of Markov processes. Once the QBD Markov model is obtained, the steady state probabilities can be obtained by using several different methods. The basic hierarchical modelling is called MRM. The MRM approach is the one most commonly used, but gives approximate results. On the other hand, generating functions (Mitrani, 2001), matrix-geometric representation (Neuts, 1981) and spectral expansion (Chakka, 1995), (Ever, 2007), (Gemikonakli et al., 2009) are the most important and well known analytic approaches. However, the exact spectral expansion solution is explained for each proposed models since it is the solution approach used in this thesis.

It is possible to develop an effective and exact analytical model for performance and performability modelling of wireless cellular systems by using a multi-dimensional representation of the states of the system, in a Markovian framework. In Ever (2007), Chakka (1995), Mitrani (2001) and Mitrani (2002), this approach has been effectively used for various multi-server systems.

MRM is a hierarchical modelling approach and provides an unified framework for performance and availability modelling and evaluation. In this approach, rewards rates come from performance model as a reward rate and add to the availability model. Using unified manner of performance and availability models defines approximation in MRM. This approximation depends on the differences in the rates which both availability and performance models occur. More information for the MRM can be found in Trivedi et al. (1994).

The generating functions method generates functions to solve the set of balance equations. A number of unknown probabilities appear in the equations while generating functions. Those unknown probabilities are determined by exploiting the singularities of the coefficient matrix. In Gail et al. (1992) a detailed analysis can be found for this approach.

The Matrix-geometric solution method is based on a representation of the steady state probabilities. The probabilities are expressed in terms of the powers of a matrix and formed from the system parameters. R, is the minimal positive solution, and is computed by using an iterative algorithm (Ever, 2007). This method has probabilistic interpretation for each step of the computations (Neuts, 1981). In the Matrix-geometric method, some values of certain parameters are uncertain and relatively large. Therefore, the number of iterations for computing R cannot be predetermined and there is a great computational requirement to obtain R.

Spectral expansion is another exact solution technique which is mainly used in this thesis. In the spectral expansion method, the eigenvectors and eigenvalues are computed to obtain a system of linear equations. These processes are mostly arising in performance and dependability problems of computing and communication systems. Chakka has applied spectral expansion method in Chakka (1995) to several non-trivial and complicated modelling problems, occurring in computer and communication systems (Chakka, 1995), (Chakka and Mitrani, 125), (Chakka and Mitrani, 1996), (Chakka, 1998). An exact performability model for homogeneous and heterogeneous multi-server systems with breakdowns and repairs was solved by using this solution approach in Ever (2007). Performance measures are evaluated and optimisation issues are addressed. In the Matrix-geometric method, computation of R has a much greater computational requirement than computational requirement for eigenvalues and eigenvectors in spectral expansion solution. However, in the spectral expansion method, accurate eigenvalues and eigenvectors are necessary since the performance measures can be quite sensitive to these. In Mitrani (2002), Mitrani and Chakka (1995) and Haverkort and Alexander (1997), a comparative study is performed to show that the spectral expansion algorithm has an edge over the Matrix-geometric method in computational efficiency, accuracy and ease of use. It is also stated that the spectral expansion method is a better solution method, especially when more heavily loaded systems are studied and when batch arrivals (or departures) are included in the model.

In this research, the spectral expansion method has extensively been used for steady state solution of QBD Markov models. It is used to develop a performability model and steady state solution approach for wireless and mobile communication systems. The following section briefly explains the algorithm that spectral expansion method follows.

3.4.1 Spectral Expansion Method

Spectral expansion is an emerging solution technique which is useful in performance and dependability/availability modelling of discrete event systems. It solves certain types of Markov models that arise in several practical system models. The reported applications and results include, performability modelling of several types of multiprocessors, multi-task execution models, networks of queues with unreliable servers (Chakka, 1995), (Chakka, 1998). In this approach, matrix A is defined as the matrix of instantaneous transition rates from state *(i,j)* to state *(k,j)* with zeros on the main diagonal. These are the purely lateral transitions of the model. Matrices B and C are transition matrices for one-step upward and one-step downward transitions, respectively. The transition rate matrices do not depend on j for j≥M, where M is a threshold having an integer value. The process Z evolves with the following instantaneous transitions:

$A_j(i, k)$: Purely lateral transition rate, from state *(i,j)* to state *(k, j)*, *(i=0, 1, . . . , N; k=0, 1, . . . , N; i≠k; j=0, 1, . . . , L)*, usually caused by a change in the operative state (i.e., a change in random variable $I(t)$).

$B_j(i, k)$: One-step upward transition rate, from state *(i,j)* to state *(k, j+1)*, *(i=0, 1, . . . , N; k=0, 1, . . . , N; and j=0, 1, . . . , L)*, usually caused by a job arrival into the queue.

$C_j(i, k)$: One-step downward transition rate, from state *(i,j)* to state *(k, j-1)*, *(i=0, 1, . . . , N; k=0, 1, . . . , N; and j=0, 1, . . . , L)*, usually caused by the departure of a serviced job.

The spectral expansion method is applicable for systems with unbounded queuing capacities (i.e. K$\leq$L$<\infty$) as well as systems with bounded queuing capacities (i.e. finite L $\geq$ K). However, bounded queuing capacities are used in this thesis. The solution presented is also valid for steady states of multi-server systems (Ever, 2007).

Following the spectral expansion solution, the steady-state probabilities of the system considered can be expressed as:

$$p_{i,j} = lim_{t\to\infty}\ P(I(t) = i, J(t) = j); \quad 0 \leq i \leq N, \quad 0 \leq j \leq L \tag{3.3}$$

where L can be finite or infinite. Let's define certain diagonal matrices of size (N+1) x (N+1) as follows:

$$D_j^A(i,i) = \sum_{k=0}^{N} A_j(i,k); \quad D^A(i,i) = \sum_{k=0}^{N} A_j(i,k); \tag{3.4}$$

$$D_j^B(i,i) = \sum_{k=0}^{N} B_j(i,k); \quad D^B(i,i) = \sum_{k=0}^{N} B_j(i,k); \tag{3.5}$$

$$D_j^C(i,i) = \sum_{k=0}^{N} C_j(i,k); \quad D^C(i,i) = \sum_{k=0}^{N} C_j(i,k); \tag{3.6}$$

and Q_0 = B, Q_1 = A-D^A-D^B-D^C, Q_2 = C.

For both bounded and unbounded queuing systems, all state probabilities in a row can be defined as:

$$v_j = (p_{0,j}, p_{1,j}, \cdots, p_{N,j}); j = 0, 1, 2 \ldots \tag{3.7}$$

Here, for a bounded system, j is limited by finite L. In this case, when the queue is full, the arriving jobs are lost. The matrices given above are used in the spectral expansion solution for both bounded and unbounded queuing systems. As mentioned before, only bounded queuing capacities are used in this thesis. Thus, the steady-state balance equations for bounded queuing systems (0 $\leq$ j $\leq$

L) can now be written as:

$$v_0[D_0^A + D_0^B] = v_0 A_0 + v_1 C_1 \tag{3.8}$$

$$v_j[D_j^A + D_j^B + D_j^C] = v_{j-1}B_{j-1} + v_j A_j + v_{j+1}C_{j+1}; \quad 1 \leq j \leq M-1 \tag{3.9}$$

$$v_j[D^A + D^B + D^C] = v_{j-1}B + v_j A + v_{j+1}C; \quad M \leq j \leq L \tag{3.10}$$

$$v_L[D^A + D^C] = v_{L-1}B + v_L A \tag{3.11}$$

and the normalizing equation is given as follows:

$$\sum_{j=0}^{L} v_j e = \sum_{j=0}^{L}\sum_{i=0}^{N} P(i,j) = 1.0 \tag{3.12}$$

From equation 3.11 one can write

$$v_j Q_0 + v_{j+1}Q_1 + v_{j+2}Q_2 = 0 \quad (M-1) \leq j \leq (L-2) \tag{3.13}$$

Furthermore, the *characteristic matrix polynomial* Q(λ) can be defined as:

$$Q_\lambda = Q_0 + Q_1\lambda + Q_2\lambda^2; \quad \bar{Q}_\beta = Q_2 + Q_1\beta + Q_0\beta^2; \tag{3.14}$$

where

$$\Psi Q_\lambda = 0; \quad |Q_\lambda| = 0; \quad \phi\bar{Q}_\beta = 0; |\bar{Q}_\beta| = 0; \tag{3.15}$$

β and ϕ are eigenvalues and left-eigenvectors of $\bar{Q}_\beta$, respectively. Note that, ϕ is a vector defined as $\phi = \phi_{\,0}, \phi_{\,1}, \ldots, \phi_{\,N}$; $\beta = \beta_0, \beta_1, \ldots, \beta_N$.

Furthermore, v$_j$ $= \sum_{k=0}^{N}(a_k\Psi_k\lambda_k^{j-M+1} + b_k\phi_k(i)\beta_k^{L-j})$, M-1 $\leq$ j $\leq$ L and in the state probability form,

$$p_{i,j} = \sum_{k=0}^{N} (a_k \Psi_k \lambda_k^{j-M+1} + b_k \phi_k(i) \beta_k^{L-j}), \quad M-1 \leq j \leq L \tag{3.16}$$

where λ_k(k=0, 1, ..., N) and β_k(k=0, 1, ..., N) are N+1 eigenvalues each, that are strictly inside the unit circle (Chakka, 1995), and b_k (k=0, 1,..., N) are arbitrary constants which can be scalar or complex-conjugate, just like $a_k s$. $\boldsymbol{v}_j s$ can be obtained as explained in the previous case (Ever, 2007). From the $p_{i,j}$, a number of steady-state availability, reliability, performability measures can be computed quite easily. For example mean queue length (MQL) can be obtained as:

$$MQL = \sum_{j=0}^{L} j \sum_{i=0}^{N} P_{i,j} \tag{3.17}$$

where L can be finite or infinite depending on whether the case concerned is bounded or unbounded. For cases where L is finite, the blocking probability of the system (P_B) can be obtained by using the following equation:

$$P_B = \sum_{i=0}^{N} P_{i,L} \tag{3.18}$$

In addition to MQL and P_B, once the steady state probabilities are obtained, some other system performance measures such as mean response time, throughput and utilisation can be calculated. Systems can be planned, monitored and evaluated by achieving the system performance results. Therefore, performance measurement is primarily a management tool for many systems. In Ever (2007), Chakka (1995) and Mitrani (2005) more details of the spectral expansion method can be found.

Chapter 4

Performability Modelling of Handover in Wireless Cellular Networks

4.1 Introduction

Customer expectations of high QoS from wireless and mobile systems are the same as for the traditional wired networks. The availability and performance models are commonly used to obtain QoS results in computer networks and communication systems. In real systems, availability and performance models of the system are important QoS requirements which should be considered together. Performability models are obtained by combining performance and availability models in order to get more realistic results. In addition, the QoS of any mobile network is highly influenced by the mobility in the network. In fact, mobility behaviour is another important environmental factor that determines QoS measurements and influences system modelling. Thus, the proposed model in this chapter accurately captures the mobility factors.

In this chapter, performability modelling and evaluation of wireless cellular systems are considered. The failures and repairs behaviour of wireless cellular networks, various horizontal handover schemes and mobility issues are investigated. The approximate MRM is employed to obtain numerical results. The spectral expansion solution approach is also carried out to address the accuracy of the MRM and the efficiency of the proposed model since the spectral expansion is an exact solution approach. Both MRM and the spectral expansion approaches are explained in section 4.3 and section 4.4 respectively.

4.2 System Model

In a wireless cellular system, the service area is divided into multiple adjacent cells. Each cell has its own base station (BS). Mobile subscribers (MSs) communicate via radio links with BSs (Agrawal and Qing, 2006), (Trivedi et al., 2002). The proposed approach considers modelling a single cell in a wireless cellular network for performability evaluation. A homogeneous wireless system is assumed. Each cell has S channels with a queuing capacity H. The maximum number of calls allowed into the system is equal to the number of calls assigned with the channels in case of a fully operational system plus the queuing capacity. The maximum number of calls in the system is given by L where $L = S + H$. The Figure 4.1 depicts the system considered.

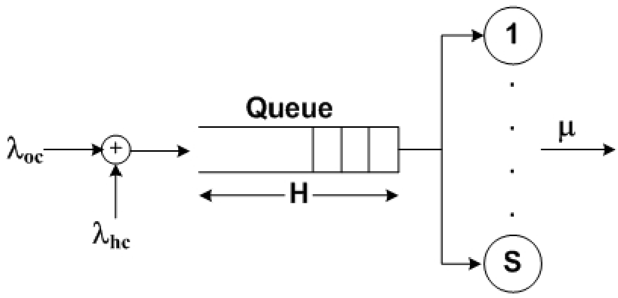

Figure 4.1: The model considered for a cellular network

Two different kinds of arrival rates are defined for originating calls and handover calls with mean arrival rates given as λ_{oc} and λ_{hc} respectively. Handover calls are the ongoing calls from one channel (or cell) to another and the originating calls are newly generated calls in the cell. The call arrivals can be assigned to any channel if it is available and idle in the cell. Otherwise, the incoming call request is added to the queue if the channels are unavailable or busy (Zeng and

Agrawal, 2001). The inter-arrival times of the incoming call requests are assumed to follow an exponential distribution. Let λ be defined as the total arrival rate of calls in the cell where, $\lambda = \lambda_{oc} + \lambda_{hc}$. A formula is given in Zeng and Agrawal (2001) for λ_{hc}. A similar formula is used in this chapter to calculate handover calls. However, to account for failures in neighbouring cells (and hence different handover rates), first $\lambda_{hc,i}$ (i=0,1,$\cdots$,S) is been calculated for all possible combinations of the number of operative channels. Then the expected handover rate is calculated as follows:

$$\lambda_{hc} = \sum_{i=0}^{S} q_i \cdot \lambda_{hc,i} \tag{4.1}$$

Where q_i is the probability that i channels are operative. Equation 4.3 is used to calculate these probabilities and is explained in subsection 4.3.1 (Mitrani, 2001).

An exponentially distributed call holding time, T_{ch} with mean rate of μ_{ch} is assumed. Dwell time, where the time MSs spend in the cell, is also needed. This is assumed to be exponentially distributed with mean rate of μ_{cd}. The equation 2.3 is used in literature chapter for the evaluation of the dwell time in cellular system hence μ_{cd} is the same as the equation 2.2 presented in the literature chapter and can be described as follows:

$$\mu_{cd} = \frac{E[V] \cdot L}{\pi \cdot A} \tag{4.2}$$

where $E[v]$ is the average of the random variable, V is the speed of mobile users, L is the length of the perimeter of cell (A cell with an arbitrary shape is assumed), and A is the area of the cell (Zeng and Agrawal, 2001). Let E[T] be the channel holding time of a call and is exponentially distributed with mean $1/\mu$ where, $\mu = \mu_{ch} + \mu_{cd}$. The MRM and the spectral expansion solution approaches are given in sections 4.3 and 4.4 respectively.

4.3 An Approximate Markov Reward Model Solution Approach for Cellular Networks

4.3.1 The Availability Model for Wireless Cellular System

The availability model indicates possible channel failures and repairs in the cellular system. The channel failure and repair behaviour are shown in Figure 4.2 for an S channel system. The distribution of time intervals between channel failures are exponential and given by mean $1/\xi$. At the end of the channel failures, the channels require an exponentially distributed repair time with mean $1/\eta$. A single repairman facility is assumed for all of the channels but the models provided can easily be extended for systems with multiple repairmen facilities.

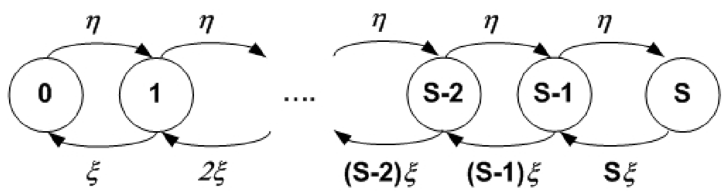

Figure 4.2: The state transition diagram for the availability model

Assume that q_i represents the steady state probabilities. The probability of having i operative channels, and since each operative state is reached through channel failures or repairs. The following equation 4.3 can be obtained as follows which is defined by Mitrani (2001) for independent random breakdowns and recovery.

$$q_i = \binom{S}{i} \left(\frac{\eta}{\eta + \xi}\right)^i \left(\frac{\xi}{\eta + \xi}\right)^{S-i} \quad i = 0, 1, \cdots, S \tag{4.3}$$

4.3.2 The Performance Model for Wireless Cellular Systems

The single cell is considered from a pure performance point of view. In addition, various issues such as handover calls, originating calls, mobility factors and queue capacity are taken into account. The state transition diagram of performance model of a single cell in wireless cellular systems is given in Figure 4.3. Let's define the states i (i=0,1,2,$\cdots$,S+H) as the number of calls in the system at time t.

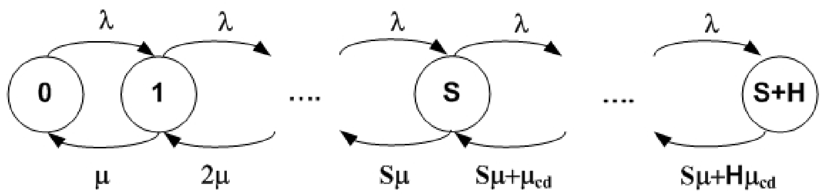

Figure 4.3: The state transition diagram for the performance model

To obtain performability measurements, the performance model is solved and the performance measures are passed as reward rates to the availability model. ρ is the traffic intensity in the system where ρ=λ/μ. Assuming a system in steady state, the state probabilities, P_i can be obtained and are given in equation 4.4 (Zeng and Agrawal, 2001).

$$P_i = \begin{cases} \frac{\rho^i}{i!} \cdot P_0 & 0 \leq i \leq S \\ \frac{\frac{\rho^S}{S!} \cdot \lambda^{i-S} \cdot P_0}{\prod\limits_{j=S+1}^{i} [S\mu + (j-S)\mu_{cd}]} & S < i \leq S+H \end{cases} \tag{4.4}$$

In equation 4.4, P_i is the probability that there are i calls in the system and P_0 can be defined as follows:

$$P_0 = \left[\sum_{i=0}^{S} \frac{\rho^i}{i!} + \sum_{i=S+1}^{S+H} \frac{\frac{\rho^S}{S!} \cdot \lambda^{i-S}}{\prod\limits_{j=S+1}^{i} [S\mu + (j-S)\,\mu_{cd}]} \right]^{-1} \tag{4.5}$$

The average number of calls in the system, N_S can then be calculated as $N_S = \sum_{i=0}^{S+H} i \cdot P_i$ which gives:

$$N_S = \left[\sum_{i=0}^{S} i \cdot \frac{\rho^i}{i!} + \sum_{i=S+1}^{S+H} \frac{i \cdot \frac{\rho^S}{S!} \cdot \lambda^{i-S}}{\prod_{j=S+1}^{i} \left[S\mu + (j - S)\,\mu_{cd} \right]} \right] \cdot P_0 \tag{4.6}$$

Equation 4.6 gives the mean queue length (MQL) assuming that all channels are operative. However, since only i channels are operative at any time, the MQL can now be represented by N_i where i is the number of operative channels. Considering that time between failures and repairs times follow exponential distributions, a MRM approach can be used to obtain the overall mean queue length. MQL is as follows; $MQL = \sum_{i=0}^{S} q_i \cdot N_i$. Similarly, the blocking probability P_B can be calculated as $P_B = \sum_{i=0}^{S} q_i \cdot P_{B,i}$ where,

$$P_{B,i} = \frac{\frac{\rho^i}{S!} \cdot \lambda^H \cdot P_0}{\prod_{j=i+1}^{i+H} \left[i\mu + (j - i)\,\mu_{cd} \right]} \tag{4.7}$$

4.3.3 Numerical Results

In this section, an approximate MRM solution approach is employed for the performability measures of horizontal handover scheme in wireless cellular systems assuming that S=10 and H=50. The expected call holding time $E[T_c]$, is 60 seconds. The average failure time $(1/\xi)$ for each channel is 1000 hours. The expected repair time $(1/\eta)$ is 30 minutes. The handover arrival rate, λ_{hc} is obtained through equation 4.1. The P_B and MQL results are shown as a function of originating calls in Figures 4.4 and 4.5 respectively.

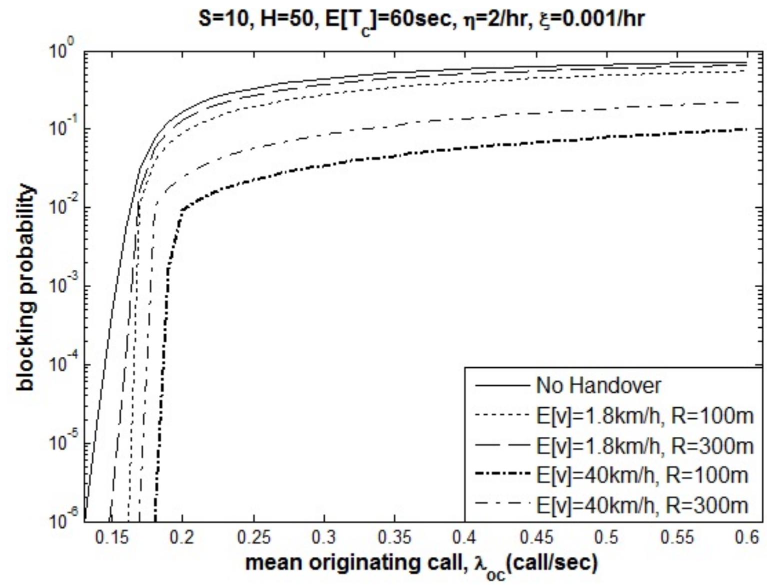

Figure 4.4: The blocking probability results as a function of originating calls for MRM

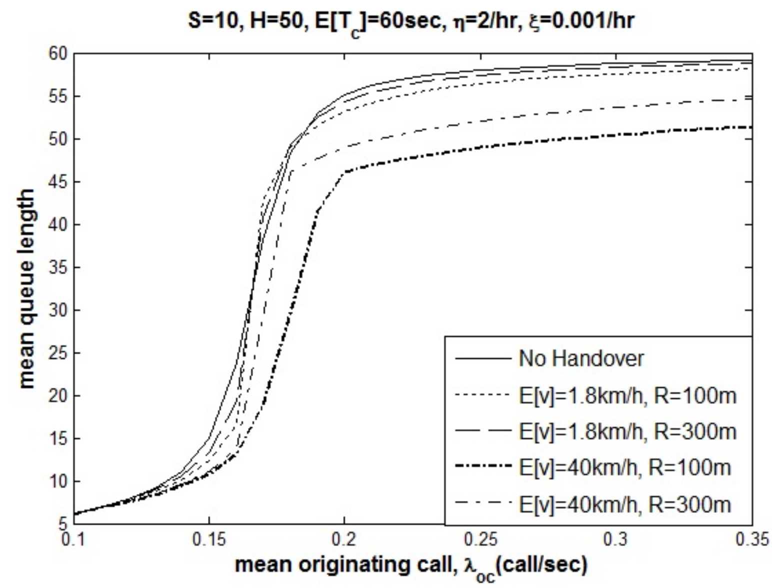

Figure 4.5: The mean queue length results as a function of originating calls for MRM

Clearly, Figures 4.4 and 4.5 show that the no handover case gives the worse performance. The quicker a user moves out of a cell coverage area, the better the system performance gets. It should be noted that, channels receive handover calls from the adjacent cells as well as initiating handover. Results also show that the wireless cellular systems with smaller cell radius and greater MS speeds perform better.

As a summary, the resulting MRM model represents an S channel per cell for homogeneous cellular system with handovers (due to the mobility), failures and repairs. The performance model of Zeng and Agrawal (2001) to account for the effects of handovers is combined with the availability model of Mitrani (2001). Computations have been carried out to evaluate the performability of the system under consideration giving mean queue length and blocking probability of the system. The findings clearly emphasise that mobility is a major issue for mobile communication systems. In addition, the speed of a user, *E[v]* and the radius of a cell, R are important parameters in determining the effects of handover. The handovers are also influenced by the rate of originating calls and number of operative channels. Mobile channels are prone to failures. Failures result in reduced number of operative channels, and this gives rise to increased handover.

4.4 The Exact Spectral Expansion Solution Approach for Cellular Networks

This exact solution approach can be defined by a pair of integer valued state variables, *I(t)* and *J(t)*, which present the operative state of the channels and the number of the calls in the system on horizontal and vertical directions of a lattice strip respectively. When the spectral expansion solution is considered, usually the service rates become independent of the number of jobs/calls in the system and a threshold value M is used in order to identify this point. However, for wireless cellular systems, the service rates are dependent on the number of calls in all states, since the number of handover arrivals affect the overall service rate ($S\mu + i\mu_{cd}$, i=0,1,2, $\cdots$, S+H). The details of the spectral expansion solution approach for the proposed model are explained in the following subsection.

4.4.1 Two Dimensional Modelling

The state of the system at time t can be described by a pair of integer valued state variables, *I(t)* and *J(t)* as explained in Chapter 3. The two state variables specify the channel configuration and the number of calls present respectively. In general, let's assume that there are $N+1$ channel configurations, represented by the values *I(t)* $=\{0,1,\cdots,N\}$. These $N+1$ configurations are the possible operative states of the model. The model assumptions are assumed to ensure that *I(t)*, $t \geq 0$, is an irreducible Markov process. *J(t)* $\leq L$ is the total number of calls in the system at time t, including the one(s) in service. Then, $Z = \{[I(t), J(t)];\ t \leq 0\}$ is an irreducible Markov process on a lattice strip (a QBD process), that models the system. Its state space is, $\{0,1,\cdots,N\}$ x $\{0,1,\cdots,L\}$. Similar models are analysed for exact performability evaluation in Chakka (1998), Gemikonakli et al. (2005) and Mitrani (2001) considering some general multi-server systems with single repairman and for both finite and infinite L with some repair strategies. Let the possible operative states be represented in the horizontal direction (the availability model) and the number of calls in the vertical direction of a lattice strip (the performance model).

Here, A is the matrix of instantaneous transition rates from state *(i,j)* to state *(k,j)*, (i=0,1,$\cdots$,S; k=0,1,$\cdots$,S; i$\neq$k; j=0,1,$\cdots$,S+H), caused by a change in the operative state of the channels (i.e. a failure followed by a repair) with zeros on the main diagonal. These are the pure lateral transitions of the model Z. Matrix B is transition matrix for one-step upward, from state *(i,j)* to state *(k,j+1)*, (i=0,1, $\cdots$, S; k=0,1,$\cdots$,S; and j=0,1,$\cdots$,S+H), caused by arrival of an originating or a handover call into the queue. In addition, matrix C is transition matrix for one-step downward transitions, from state *(i,j)* to state *(k,j-1)*, (i=0,1, $\cdots$, S; k=0,1,$\cdots$,S; and j=0,1,$\cdots$,S+H), caused by the departure of a serviced call. The transition rate matrices do not depend on j for j$\geq$M, where M is a threshold having an integer value. However, in the case of the wireless cellular networks the transition rate matrices always depend on j as Figure 4.3 clearly shows.

Clearly, the elements of A depend on the parameters S, ξ, and η. The transition matrices of a system with S channels are of size (S+1) x (S+1). The state transition matrices A, A_j, B, and B_j can be given as follows;

$$A = A_j = \begin{pmatrix} 0 & \eta & 0 & 0 & 0 & 0 & 0 & 0 \\ \xi & 0 & \eta & 0 & 0 & 0 & 0 & 0 \\ 0 & 2\xi & 0 & \eta & 0 & 0 & 0 & 0 \\ 0 & 0 & 3\xi & 0 & \ddots & 0 & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 & \eta & 0 & 0 \\ 0 & 0 & 0 & 0 & (S-2)\xi & 0 & \eta & 0 \\ 0 & 0 & 0 & 0 & 0 & (S-1)\xi & 0 & \eta \\ 0 & 0 & 0 & 0 & 0 & 0 & S\xi & 0 \end{pmatrix}$$

$$B = B_j = \begin{pmatrix} \lambda & 0 & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & 0 & \lambda \end{pmatrix}$$

The elements of matrices C and C_j depend on the parameters S, H, μ, and μ_{cd}. As Figure 4.3 clearly shows, C matrix is dependent on the number of calls for j=0,1,$\cdots$,L. Therefore, the threshold M is taken as M=L. If the number of calls in the system is less than the number of available channels, a channel is assigned for each call. Therefore, the downward transition rate is chosen as the minimum of number of calls and number of available channels. On the other hand, if the number of calls is greater than the number of available channels, all of the available channels are assigned to incoming calls and the calls in the queue have the service rates as μ_{cd}. The matrix C is defined below, together with C_j matrices which are defined for two different regions as explained above:

$$C = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \mu + H\mu_{cd} & 0 & 0 & 0 & 0 \\ 0 & 0 & 2\mu + H\mu_{cd} & 0 & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & 0 & 0 & (S-1)\mu + H\mu_{cd} & 0 \\ 0 & 0 & 0 & 0 & 0 & S\mu + H\mu_{cd} \end{pmatrix}$$

for j>S

$$C_j = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \mu + (j-S)\mu_{cd} & 0 & 0 & 0 & 0 \\ 0 & 0 & 2\mu + (j-S)\mu_{cd} & 0 & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & 0 & 0 & (S-1)\mu + (j-S)\mu_{cd} & 0 \\ 0 & 0 & 0 & 0 & 0 & S\mu + (j-S)\mu_{cd} \end{pmatrix}$$

for j≤S

$$C_j = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & min(1,j)\mu & 0 & 0 & 0 & 0 \\ 0 & 0 & min(2,j)\mu & 0 & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & 0 & 0 & min(S-1,j)\mu & 0 \\ 0 & 0 & 0 & 0 & 0 & min(S,j)\mu \end{pmatrix}$$

This system can be solved and the steady state probabilities, $P_{i,j}$, can be obtained using the steady state solution presented in the next section.

4.4.2 Steady State Solution

The solution is given for a bounded queue (i.e., S≤L). The balance equations defined for (M-1)≤ j ≤ $(L-2)$ are not used since the threshold value is defined as M=L. From the state probabilities, $P_{i,j}$s, a number of steady-state availability, reliability, performability measures can be computed quite easily. The mean queue length (MQL) and the blocking probability (P_B) can be obtained as:

$$MQL = \sum_{j=0}^{L} j \sum_{i=0}^{N} P_{i,j} \quad and \quad P_B = \sum_{i=0}^{N} P_{i,L} \tag{4.8}$$

4.4.3 Numerical Results and Discussions

In this section numerical results are presented for both solution approaches since the MRM is approximate, it is important to address the accuracy of the models proposed by an exact spectral expansion approach. The numerical results of spectral expansion approach are provided and presented comparatively with results obtained by the MRM approach. The results show the effectiveness and accuracy of both models developed for wireless cellular networks. The MRM is a hierarchical modelling approach and some assumptions are required for the solution of the stochastic model in case of Markov reward rates. For instance, the operative states of the channels of the pure performance model are not stable (e.g. when there are no channels available). However, these assumptions are not required for the spectral expansion approach since the model provided is two dimensional.

The parameters used in Ma et al. (2001) and Zeng and Agrawal (2001) are employed for the following numerical results. The wireless cellular networks with S=10 and H=10 are considered in Figures 4.6 and 4.7. Other parameters are given as λ_{oc} calls/sec, $E[T_c]$=60sec, η=2/hr , ξ=0.001/hr and L=S+H=20. Figures 4.6 and 4.7 show the mean queue length and blocking probability as functions of originating calls for various MS velocities and cell radius values respectively. Figures 4.6 and 4.7 clearly show that the systems with smaller cell radius and greater MS velocities perform better. In other words, the quicker a user moves out of a cell coverage area, the better the system performance. When the results from the spectral expansion and MRM solution approaches are compared, the discrepancies are less than 0.17% and 2.2% for MQL and P_B respectively.

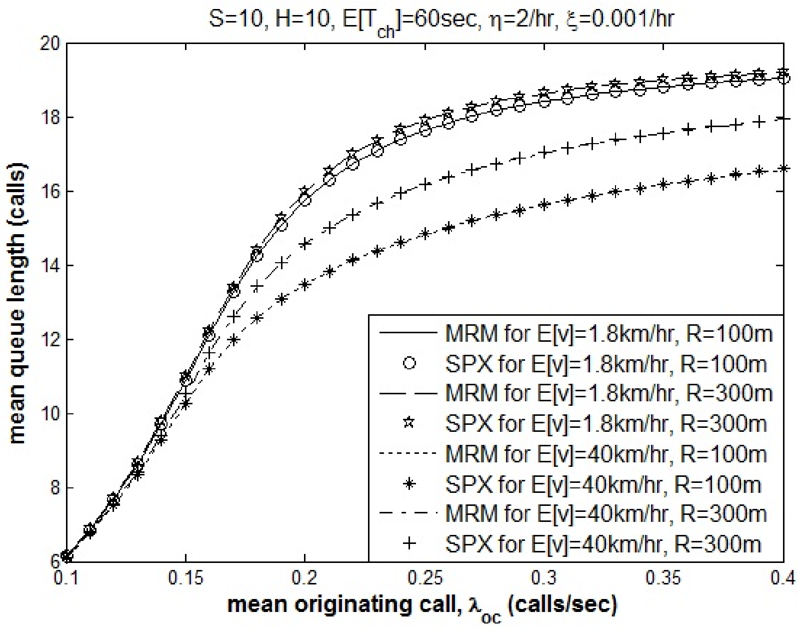

Figure 4.6: MQL results for systems with L=20

In Figures 4.8 and 4.9 wireless cellular networks with S=10 and H=50 are considered (L=S+H=60). Other parameters are the same as previous computations. When S=10 and H=50 are considered, the difference between the spectral expansion and MRM solution approaches are less than 4.1% and 2.3% for MQL and P_B respectively.

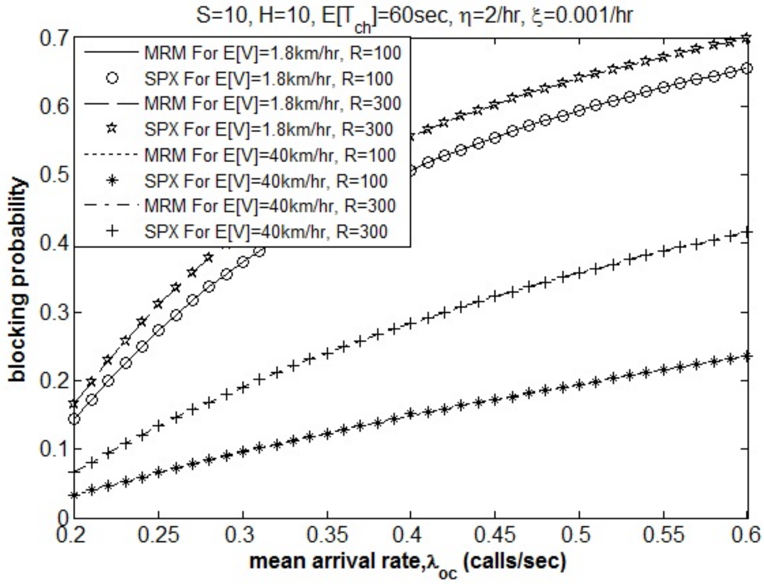

Figure 4.7: The P_B results for systems with L=20

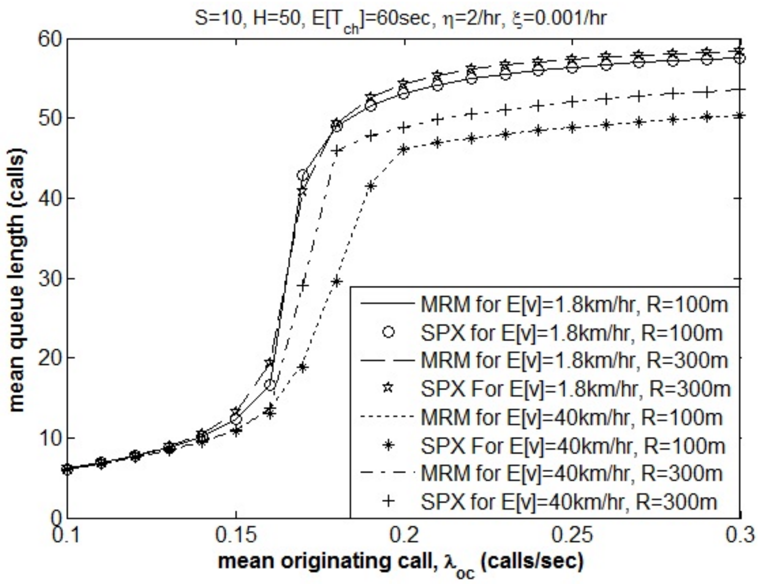

Figure 4.8: The MQL results for systems with L=60

Computations performed for Figures 4.6 and 4.8 are considered in Figures 4.10 for S=10 and H=200 (L=210). The difference between the spectral expansion and MRM solution approaches are less than 4.8%.

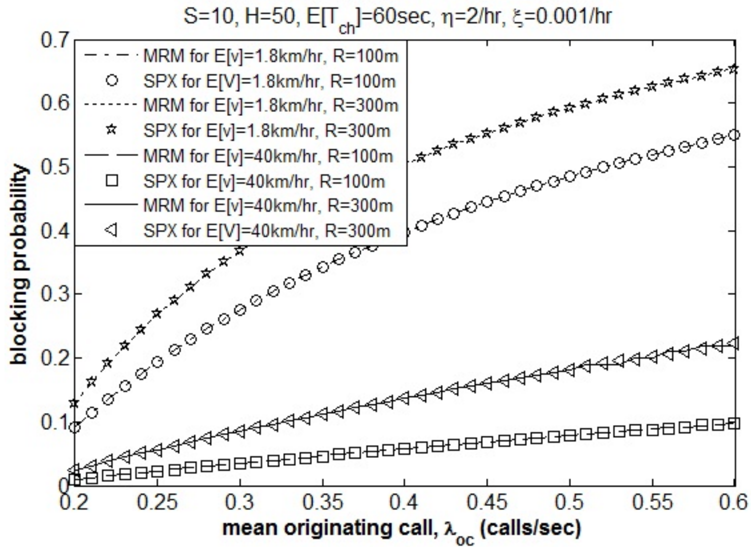

Figure 4.9: The P_B results for systems with L=60

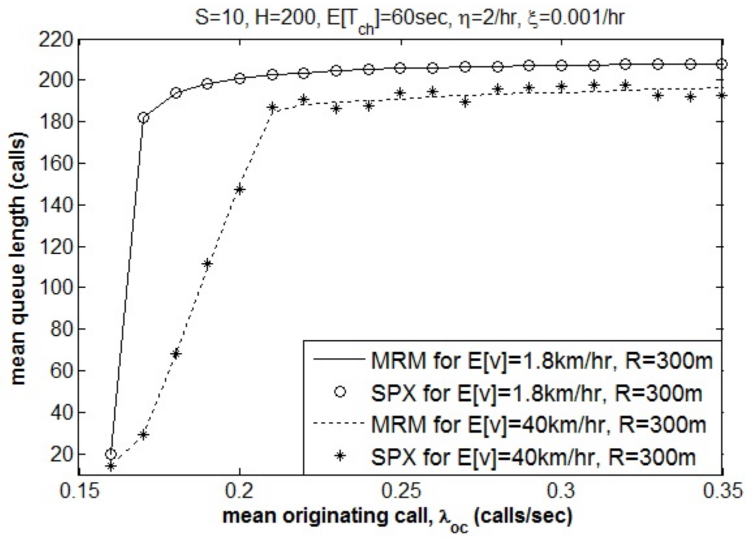

Figure 4.10: MQL results for systems with L=210

4.5 Conclusions

In this chapter, wireless cellular systems have been modelled using both approximate and exact solution approaches. A horizontal handover scheme is presented with channel failures and repairs. In addition, the mobility issues are considered such as speed of the MS and the radius of a cell. S channels per cell homogeneous cellular systems have been studied. The steady state probabilities of wireless cellular systems, with channel failures and repairs/recovery are derived using the QBD processes coupled with the spectral expansion method. An unusual threshold value M=L is used since the service rates of the channels are always dependent to the number of originating calls and the handover calls in the system. Having such an unusual threshold value has changed some of the equations in spectral expansion solution, since equations defined for (M-1)$\leq$ j $\leq$(L-2) can not be used. The results show that wireless cellular systems are prone to failures. The operative channels reduce because of the failures in the system and this gives rise to increased handover. In addition, the MS velocity and the radius of a cell are important parameters in determining the effects of handover, and originating calls as well as the handover requests. When light arrivals are expected, MQL performance, depends on values of R and E[v]. However, Figures 4.6-4.10 show, L is the main factor affecting the performability results with relatively large λ values for wireless cellular systems. According to the numerical results and performability requirements, the methodology could be modified and/or improved for the existing protocols and operational aspects effectively. The approximate MRM approach and the exact spectral expansion provided in this study give close results for smaller L values. For ten channels system with queuing capacities up to H=200 are considered in Figure 4.10, which is the case for most of the practical systems.

The approaches for the evaluation of the performability of wireless cellular systems presented in this chapter, lend themselves as powerful tools in order to make essential decisions for system design, and optimise mobility and queuing related parameters. Hence, the approach used in this chapter is also used with inherent WLAN behaviour caused by the access point failure and the recovery phenomena which is discussed in the next chapter.

Chapter 5

Performability Analysis of WLAN as a Queuing Model

5.1 Introduction

The demand of new multimedia services, increases in bandwidth available to users, high traffic load, and ubiquitous connectivity have facilitated the rapid growth of wireless and mobile communication systems. WLANs have gained widespread popularity mainly because of their low cost and relatively high data rates. In addition, the ease of installation and usage of access points (APs) have made WLANs ideal for hot spots such as the ones popularly used in homes, hotels, cafes, and airports.

IEEE 802.11 is the leading standard for WLAN (Hua et al., 2004). Currently, WLANs are considered as a complementary service offering for service providers. Providing the QoS requirements and maximizing the performance characteristics of WLANs is a challenging task. Although not a new topic, modelling and design of such systems are key issues in the QoS characterization of wireless communication systems.

It was mentioned in Chapters 2 and 4 that wireless systems are prone to failures. In wireless communication systems, failures may occur due to software, hardware, human error, or a combination of these factors (Capar et al., 2003), (Chen et al., 2003a), (Gowrishankar et al., 2009), (Jindal et al., 2006), (Ma et al., 2001), (Malone et al., 2007b), (Paramvir et al., 2004), (Trivedi et al., 2002). In practice, the signal strength, fading, reflection and refraction of signals, and man-made noise, etc., are some important factors that can cause channel unavailability in WLANs. From the designer's and operator's points of view, these factors have to be critically considered. In Paramvir et al. (2004), channel failures are considered to take place due to environmental interference as well as interference from other wireless networking devices. Similarly in Gowrishankar et al. (2009), the communication channels are defined as virtual channels and each virtual channel is subject to unavailability for data transmission due to frequency selective fading and multipath fading in WLANs. However, these approaches do not consider modelling and analysis of WLANs in an attempt to understand, and improve system performance in terms of system capacity, blocking and dropping probability of calls. Mobility related factors, and queuing capacity are not taken into account either.

In this chapter, performability evaluation of a WLAN is presented with the inherent wireless channel unreliability behaviour caused by the channel failure/recovery and AP failure/recovery phenomena. The mobility related issues and queuing capacities are also considered for various scenarios and effects on the system performance measures are emphasised.

5.2 Basic Architectural Concepts

5.2.1 Analysis of Failures in WLANs

Significant research has been done on analysis of failures for wireless systems (Capar et al., 2003), (Chen et al., 2003a), (Gowrishankar et al., 2009), (Jindal et al., 2006), (Ma et al., 2001), (Ma et al., 2000), (Malone et al., 2007b), (Paramvir et al., 2004), (Tay and Chua, 2001). Failures in telecommunication systems and wireless networks can be caused by physical faults and/or software faults (Capar et al., 2003), (Chen et al., 2003a), (Jindal et al., 2006), (Ma et al., 2001), (Ma et al., 2000). In this chapter, failures of end to end communication in WLAN are considered as a channel and AP failures. Channel failure is a temporary failure. The channel fails mainly due to software failure, noise, attenuation, multipath and fading (Capar et al., 2003), (Chen et al., 2003a), (Jindal et al., 2006), (Ma et al., 2001), (Ma et al., 2000), (Tay and Chua, 2001). In this situation, APs broadcast beacon signals and allow the arrival request into the system. However, the channel is broken and needs some time to recover. On the other hand, APs can be broken for a long time due to the hardware components failures, power outages, management failures and electro-mechanical equipment interferences (Capar et al., 2003), (Chen et al., 2003a), (Jindal et al., 2006), (Ma et al., 2001), (Ma et al., 2000), (Tay and Chua, 2001). In addition, natural disasters (e.g. floods, storms and earthquakes) can be considered as AP failures. These can be considered as permanent failures that need longer time to reconfiguration. In this situation, the APs could not allow incoming requests into the system. In other words, the AP is not reachable. For instance, typical AP failures may be introduced by a strong interference source (e.g., a microwave oven) placed near the AP and result in the incapability of the AP to receive data from other nodes (Capar et al., 2003), (Chen et al., 2003a). However, in the case of both channel and AP failure, the system could not provide its service to users.

5.3 WLAN Under Study

Before modelling a system, its characteristics and parameters must be well defined. To provide suitable service, an understanding of the behaviour of WLAN APs is essential.

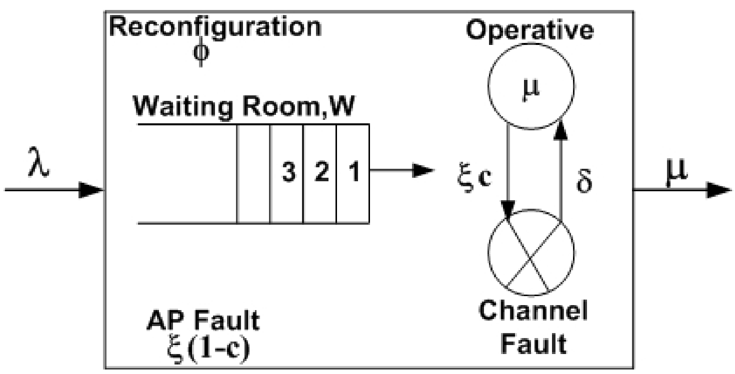

Figure 5.1: The system is considered for WLAN

To understand how the performance of a system could be enhanced, the first step is to define the system of interest. The model covers failure and reconfiguration behaviour related with mobility and queuing issues. The AP has one channel to operate. Allocation of incoming requests is usually done by the availability of channel and in this regard, it is well known that, in terms of efficiency, the common queue is more suitable than individual queues in the queuing theory. The system consists of a channel with a common queue. The common queue is bounded with a capacity of W ($W \geq 1$). The maximum number of requests in the system is equal to the number of requests assigned to the channel plus the queuing capacity. This is given by L. All arrival streams follow a Poisson process. Potential arrivals from other wireless technologies and origination calls in WLAN can be incorporated. In other words, inter-arrival times of the incoming requests are assumed to follow an exponential distribution with a mean rate of λ. The requests are assigned to the channel if the channel is available and idle in the WLAN. Otherwise, the incoming request is queued up in case of a channel failure and/or a busy channel. Figure 5.1 shows the proposed model considered for WLAN. The service times of the requests to be serviced by channel is distributed exponentially. In addition to this, the dwell time is assumed to be distributed exponentially with mean $1/\mu_{wd}$. T_{ch} is the average holding time in the WLAN and it follows an exponential distribution with mean E[T_{ch}]=$1/\mu_{ch}$. T can be defined as the total channel holding time of a request, which is exponentially distributed with mean E[T]=$1/\mu$, where μ=μ_{ch}+μ_{wd}. Similar studies are discussed and also dwell time equation is presented in the literature survey chapter.

5.4 Modelling WLAN with Channel/Access Point Failure and Reconfiguration

In this section, the model has been developed to evaluate the performability measures of WLANs subject to failures. All failures explained above are considered as AP failures that may be caused by power cut-off or resetting APs etc. or channel failure due to software, hardware, path and link loss etc. failures.

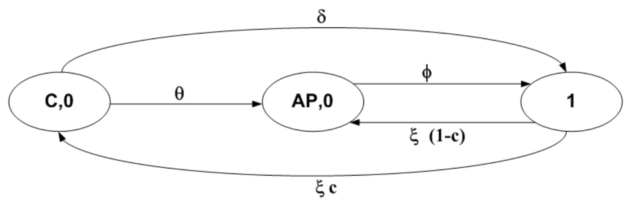

Figure 5.2: The operative states of the system

ξ is the failure rate of the system. The failures associated with the system may either be caused by AP failure with probability (1-c) or channel failure with probability c. In case of AP failure, AP cannot broadcast any beacon signals. Thus, mobile nodes cannot be connected to the system. In other words, there will be no packets in the queue and channel. AP recovery occur with an exponentially distributed failure rate of ϕ. On the other hand, when the channel fails with the probability c, the failure is related with software, link and path loss etc. Subsequent to a recovery, δ is the recovery rate for channel failure. $1/\delta$ is relatively shorter recovery time to AP failure. A longer time is needed to bring the AP up and $1/\phi$ is the recovery time for such failure. Both recovery behaviour of AP and channel are distributed exponentially.

The operative states of WLAN model considered is illustrated in Figure 5.2. The state *1* is the working state of the WLAN. In state (AP,0) the system is unreachable due to the AP failures and connection becomes unavailable. On the other hand, in state (C,0) the channel is unavailable but the system broadcasts and incoming requests are added to the queue. However, in both states, the system is unavailable for service. An inoperative period of a channel would also include the possible waiting time for a repair. No operative channel can be idle if there are requests waiting, and no repairman can be idle if the channel or AP waiting for repair. Since the failures of the AP (software or hardware related) are independent of the failures of the channel, clearly the transition rates from state *1* to state (AP,0) and state (C,0) to (AP,0) should be same. So θ is equal to ξ(1-c). In case of AP failure, ξ(1-c), the priority of repair is given to the AP because the channel cannot be used when the AP is inoperative. All failure/recovery of channels and AP failure and repair time random variables are distributed exponentially and are independent of each other.

5.4.1 Two Dimensional Markov Representation and Solution

As explained in Chapter 3, a pair of integer valued state variables, *I(t)* and *J(t)*, specifying the channel plus AP failure and recovery configurations and the number of arrivals present, respectively. Here, channel configuration, and hence the range of *I(t)*, refers to the operative states of the system. In general, let's assume that there are three configurations, represented by the values *I(t)*=(*C,0*), (*AP,0*), and (*1*). The first two configurations represent the states of the system where the system cannot serve to incoming requests ((*C,0*), (*AP,0*)). The third state is used to represent the working state of the system (*1*).

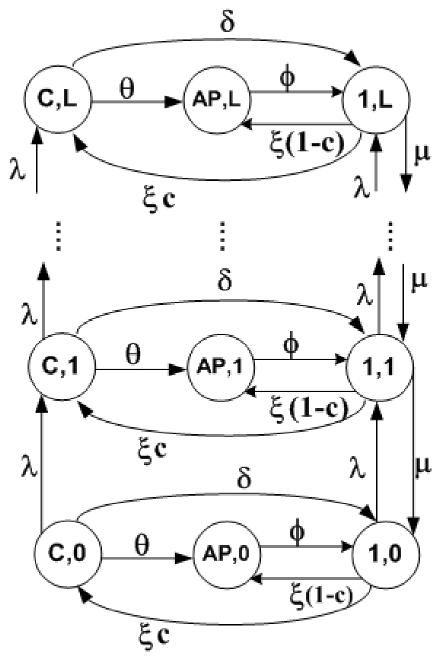

Figure 5.3: The state diagram of the proposed WLAN

With these assumptions *I(t)*, $t \geq 0$, is an irreducible Markov process. *J(t)* ($\leq$ L) is the total number of arrivals in the system at time t, including the one(s) in service. Then, Z = [*I(t)*, *J(t)*]; $t \geq 0$ is an irreducible Markov process on a lattice strip (a QBD process), that models the system. Its state space is, {0, 1, 2} x {0, 1,$\cdots$, L}. In Chakka (1998), Gemikonakli et al. (2006) and Gemikonakli et al. (2007), similar models are analysed for exact performability evaluation of some general multi-server systems as well as cellular systems with single repairman (R=1). However, channel/AP failures and reconfiguration were not considered. Figure 5.3 depicts the state diagram of WLAN considered. Note that the mobility issues are considered in Section 5.4.3 with the availability behaviour of the WLAN.

Let the possible operative states of both channels and AP be represented in the horizontal direction and the number of arrival requests in the vertical direction of a lattice strip. Here, A is the matrix of instantaneous transition rates from state (i,j) to state (k,j), (i=0, 1, 2; k=0, 1, 2; i$\neq$k; j=0, 1,$\cdots$, L), with zeros on the main diagonal, caused by a change in the state. These are purely lateral transitions of the process Z. Matrices B and C are transition matrices for one-step upward and one-step downward transitions respectively (Chakka, 1998), (Gemikonakli et al., 2006), (Gemikonakli et al., 2007). B is the matrix of one-step upward transitions from state (i,j) to state (k,j+1), (i=0, 1, 2; k=0, 1, 2; and j=0, 1,$\cdots$, L), caused by arrival requests. C is the matrix of one-step downward transitions from state (i,j) to state (k,j-1), (i=0, 1, 2; k=0, 1, 2; and j=0, 1,$\cdots$, L), Clearly, the elements of A depend on the parameters ξ, c, δ, ϕ and θ. The transition matrices of a system are of size (3)x(3). Therefore, the transitions matrices A, A_j,B, B_j, C and C_j, for this model can be expressed as follows;

$$A = A_j = \begin{pmatrix} 0 & \theta & \delta \\ 0 & 0 & \phi \\ \xi c & \xi(1-c) & 0 \end{pmatrix} \quad B = B_j = \begin{pmatrix} \lambda & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \lambda \end{pmatrix} \quad C = C_j = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \mu \end{pmatrix}$$

with C_0=0, and threshold M=1. The transition rate matrices do not depend on j for $j \geq M$, where M is a threshold having an integer value (Chakka, 1998), (Gemikonakli et al., 2006), (Gemikonakli et al., 2007).

5.4.2 Numerical Results and Discussions for WLAN with Channel/Access Point Failure and Recovery

In this section numerical results are presented for performability measures of WLAN. The WLAN system with channel and/or access point failures and repairs is considered first without mobility factor. In Figure 5.4, the MQL results are computed as a function of mean arrival rates for different failure rates. The results show the pure performance model gives the most promising results as expected since failures are not considered. However, the system performance degradation becomes more evident as the failure rate increases. For instance, for ξ=0.1/hr, the MQL values become up to twice as much as the MQL values calculated for the other rates.

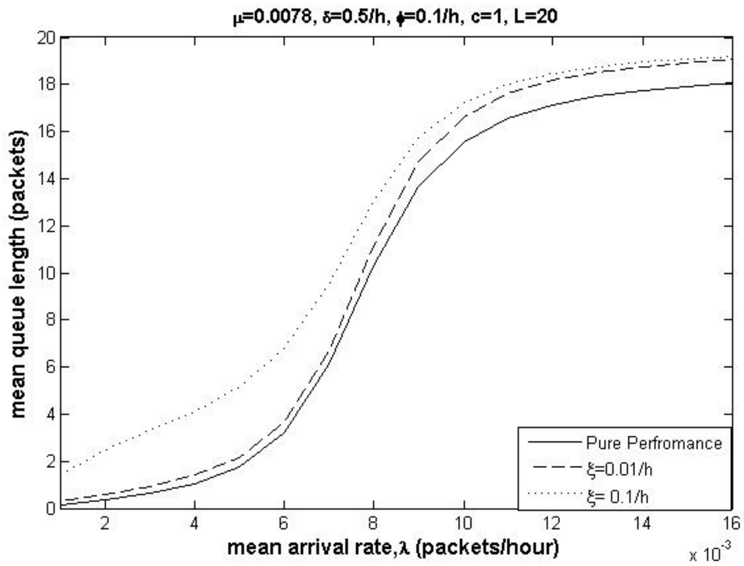

Figure 5.4: MQL results for systems with various channel failure rates

In addition, the effects of parameters $1/\delta$ are certainly important in determining the system performance. The system performance degradation caused by various channel recovery times are illustrated in Figure 5.5. The MQL results are computed as a function of mean failure rate ξ for different channel recovery times $1/\delta$. The results in Figure 5.5 indicate that as channel recovery time $1/\delta$ decreases, the change in MQL also decreases considerably. This is because of the quicker recovery time of the channel. The system can be repaired quickly and start to serve the packets. Moreover, increasing $1/\delta$ (e.g $1/\delta$=2hr) impacts the system performance significantly. The WLAN can allow the packets into the system until the queue capacity becomes full. However, the system must wait the channel to be fixed. Thus, the high the recovery time, the higher the MQL value.

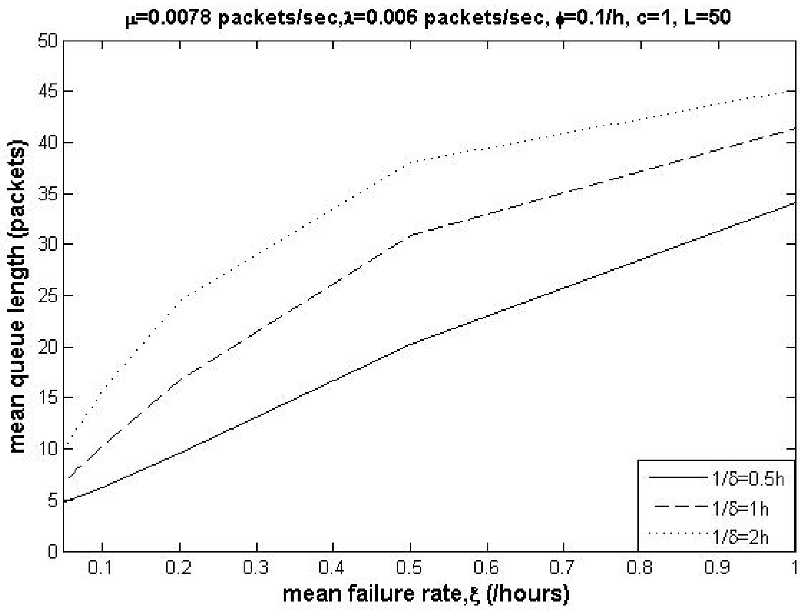

Figure 5.5: MQL results as a function of mean failure rate for different channel recovery time

On the other hand, the effects of the channel recovery times are also shown in Figures 5.6 and 5.7. The MQL and blocking probabilities are computed as a channel reconfiguration time for different arrivals rates in Figures 5.6 and 5.7 respectively. It is evident that an increase in $1/\delta$, results in increase in the MQL and blocking probability. For any value of λ, MQL increases because of the channel recovery time. For long channel recovery times, the system could be almost full and users wait for recovery and/or service. This causes an incremented number of packets in the system. For larger recovery times, blocking probability increases as well for all λ values, mainly due to the long channel recovery times. Results show that $1/\delta$ plays an important role in degrading the system's performance.

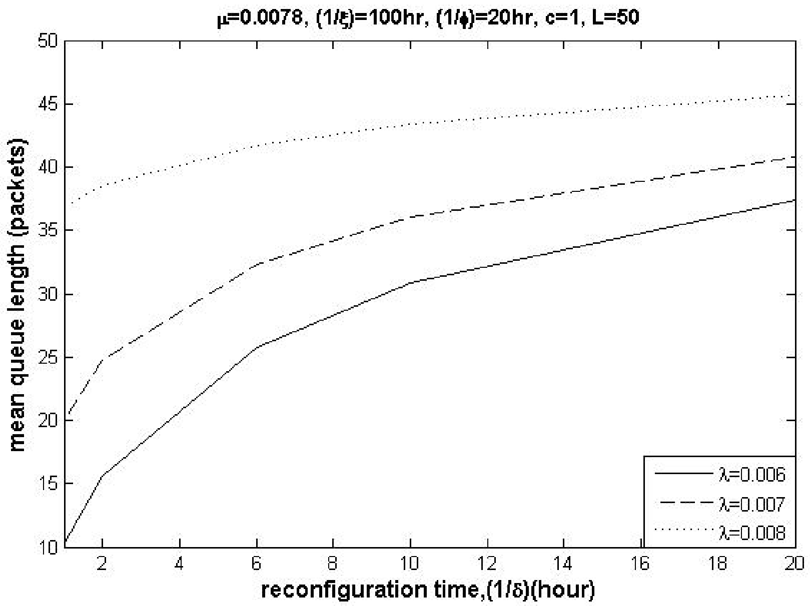

Figure 5.6: MQL results as a function of reconfiguration time for different arrival rates

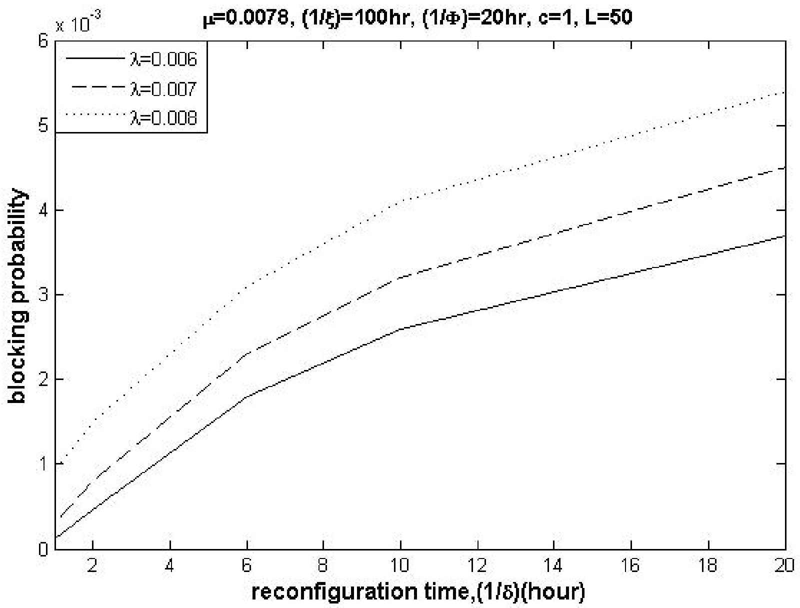

Figure 5.7: P_B results as a function of reconfiguration time for different arrival rates

5.4.3 Modelling of Channel/Access Point Failure and Reconfiguration in WLAN with Mobility Issues

This section presents the proposed model related with mobility issues for performability evaluation of WLANs. The mobility issues are considered for different scenarios. Thus, the state diagram of the proposed model is slightly different in respect of mobility factors from the Figure 5.3 presented in section 5.4.1. The new state diagram of WLAN model considered is illustrated in Figure 5.8.

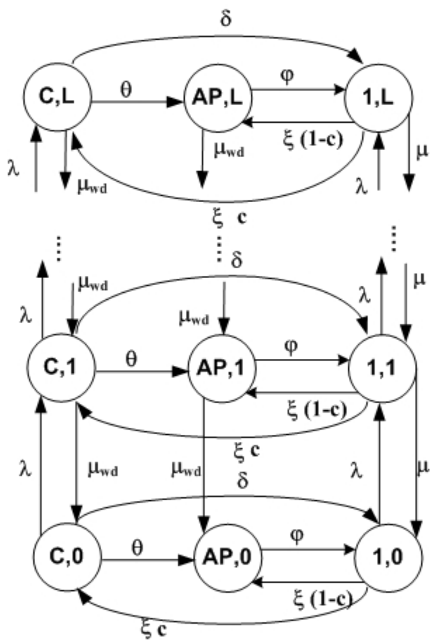

Figure 5.8: The state diagram of the proposed WLAN with μ_{wd}

μ_{wd} is the mean service rate of the mobile users. It can be obtained using equation 2.2 presented in the literature chapter which is related with the speed of the users *E[v]* and radius of the WLAN *R*. The mobile users can leave the system at any time of end to end communication. If the WLAN is operative due to the mobility, users can move in/out of the WLAN. Thus, the total system service rate μ can be obtained as $\mu_{ch}+\mu_{wd}$ as explained in section 5.3. In addition, in the case of both failure the mobile users can also leave the system due to the mobility.

The state of the system with the state variables, *I(t)* and *J(t)*, which is specifying the channel plus AP failure and recovery configurations and the number of arrivals will be the same as the model considered in section 5.4. Similarly, an irreducible Markov process $Z = [I(t),\ J(t)];\ t \geq 0$ models the system. Its state space is, {0, 1, 2} x {0, 1,$\cdots$, L}. However, the only differences will be in the C matrix due the mobility μ_{wd} which is matrix of one-step downward transitions. Therefore, the transitions matrices A, A_j, B, B_j, C and C_j for the model considered with mobility factor can be expressed as follows;

$$A = A_j = \begin{pmatrix} 0 & \theta & \delta \\ 0 & 0 & \phi \\ \xi c & \xi(1-c) & 0 \end{pmatrix} B = B_j = \begin{pmatrix} \lambda & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \lambda \end{pmatrix} C = C_j = \begin{pmatrix} \mu_{wd} & 0 & 0 \\ 0 & \mu_{wd} & 0 \\ 0 & 0 & \mu \end{pmatrix}$$

5.4.4 Numerical Results and Discussions on Mobility Factor for WLAN with Channel/Access Point Failure and Recovery

In order to analyse the effects of mobility together with channel/access point failure and recovery, numerical results are presented in this section. Figures 5.9 and 5.10 show the MQL results as a function of mean rate for incoming requests and mean velocity rates respectively. Figure 5.9 shows the mobility effect and two different scenarios are considered for computations. In the first scenario, the mobile stations are static, whereas in the second scenario the mobile stations are slightly mobile as pedestrians with E[v]=2km/h. The results show that, although the velocity of pedestrian users is relatively low, in case of high arrival rates, the mean number of users leaving the system due to mobility affects the performance quite significantly. For instance, MQL is 46.9667 for systems with static users in the system and 8.3968 for systems with both pedestrian and static users together in the system for λ=0.01. The other illustration of the effect of the mobility is evident in Figure 5.10. MQL results decrease when the average velocity of the user increases. Because of the mobility many users leave the system.

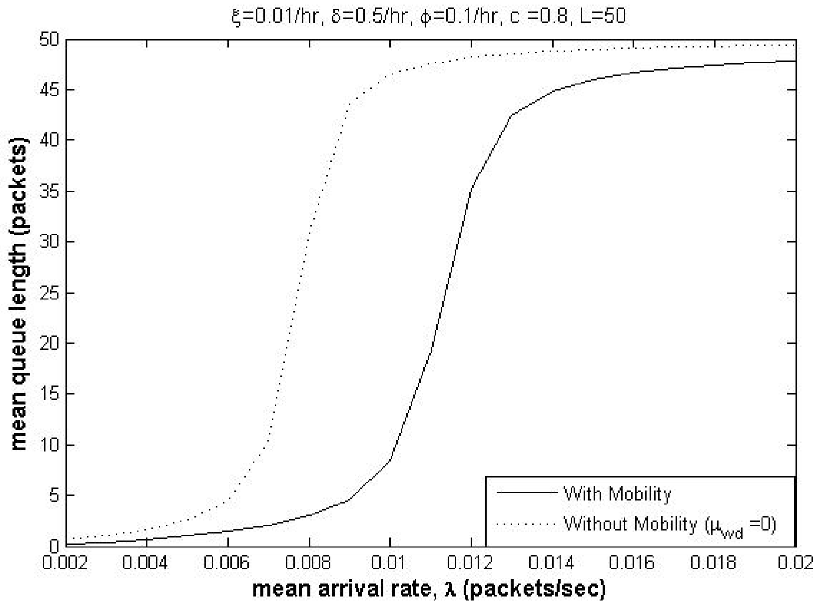

Figure 5.9: Mobility effects of the system considered with MQL results as a function of mean arrival rates

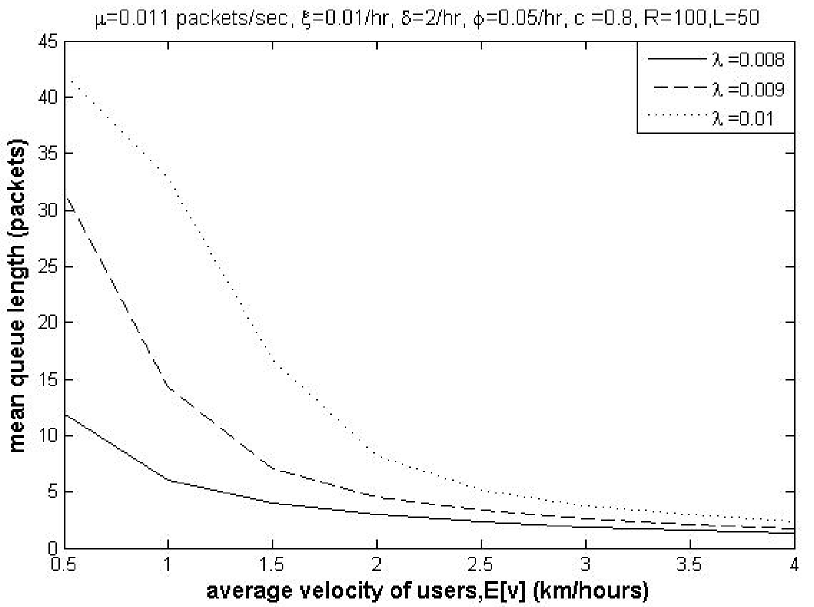

Figure 5.10: MQL results as a function of mean velocity rates for different arrival rates

5.4.5 Evaluation of the Accuracy of Failure/Reconfiguration and Mobility Factor in WLAN

In this section, the model has been validated by simulation to evaluate the performability measures of WLAN subject to failures as well as the mobility issues. The simulation is written in C++ language and is discrete event-based. The simulation model consists of the following components: An input model (arrivals) following a Poisson process to a single channel and mean service, failure, and repair rates as well as travelling time for repairman. Interarrival, service, repair, and operative times are all assumed to be exponentially distributed. For example, on an arrival event, the next arrival event is generated and the current arrival event is placed in the demand queue as well as the time of arrival is also stored. A service event indicates that the request at the head of the queue has been accepted and the time to deal with the request is calculated. Further, if the queue is empty the channel is set to free or the next service event is generated. This procedure continues until the required number of arrivals have been served. A failure event occurs with exponentially distributed time to failures (using mean time between failures (MTBF)). After an event of failure, the corresponding channel cannot provide service (arrivals continue) for an exponentially distributed repair time (using mean time to repairs (MTTR)). After the repair time, the channel continues normal operation until next failure.

The simulation tool is mainly used for validation of the model developed. It simulates the actual scenario rather than the Markov models. Well known queueing theory $M/M/1/L$ formulae as well as the well-known OPNET simulation tool are used for validating the simulation in development phase. The results obtained from the analytical model are presented comparatively with the simulation results for different performance measures to validate the proposed analytical model. Figure 5.11 shows the MQL results as a function of mean arrival rates for different failure rates as well as pure performance. The effect of failure rates on the system is shown on the diagram with simulation results comparatively. As expected, the failures affect the system performance significantly due to the availability of the system architecture. In addition, the other important QoS measurement, the channel recovery time $1/\delta$ is illustrated in Figure 5.12 in order to evaluate the accuracy of the results obtained for the systems with channel recovery times. The pure performance gives most promising results, since failures and repairs are not considered. Results increase in the MQL due to the long channel recovery times. Results from the proposed analytical approach and simulation show good agreement.

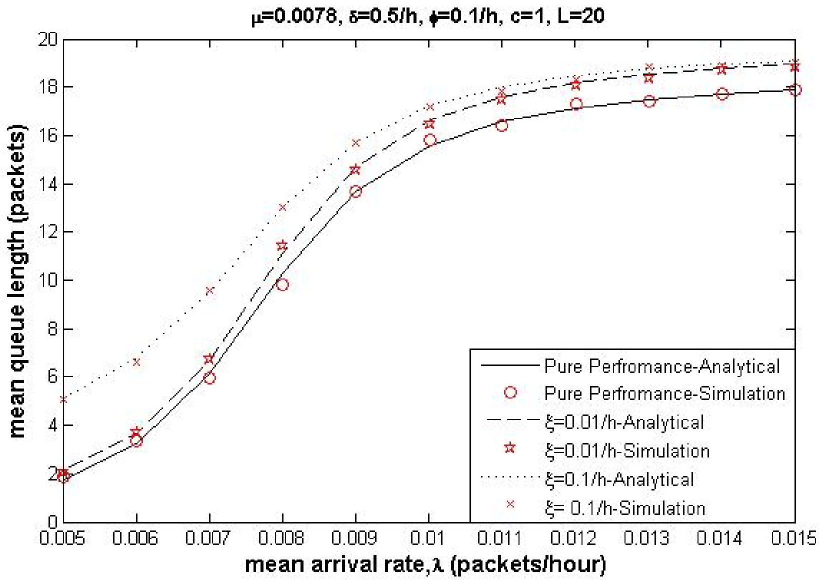

Figure 5.11: MQL results versus mean arrival rates with different failure rates

The system performability measures in terms of MQL due to the probability of having either access point or channel failure depending on c is shown with simulation results in Figure 5.13. It is evident an increase in c, results in increase in the MQL. This is mainly because c determines the system fault and is the marker for AP and channel failures. In Figure 5.14, the MQL results are given as a function of channel recovery time for different arrivals rates. In both Figures, the proposed analytical approach has been validated by simulation. The results show good agreement.

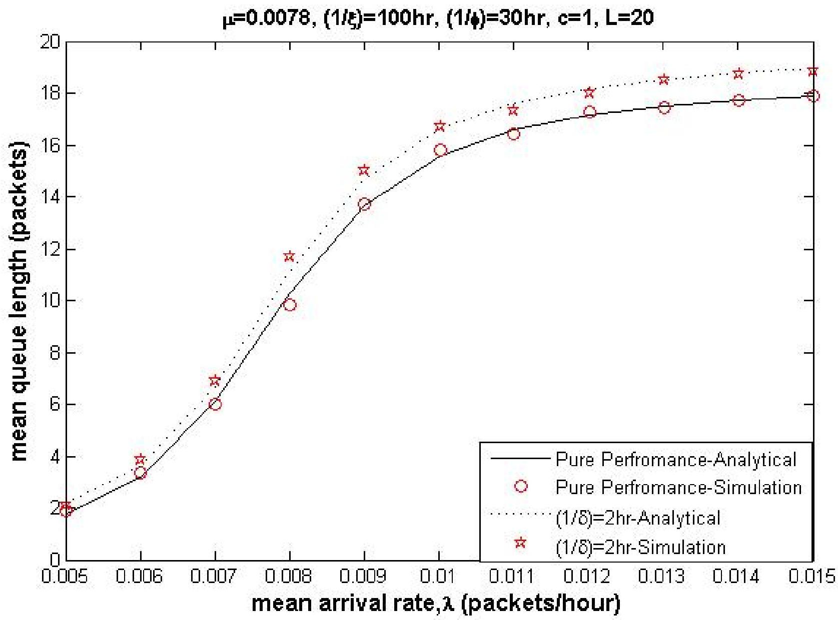

Figure 5.12: MQL results as versus mean arrival rates with different channel recovery time

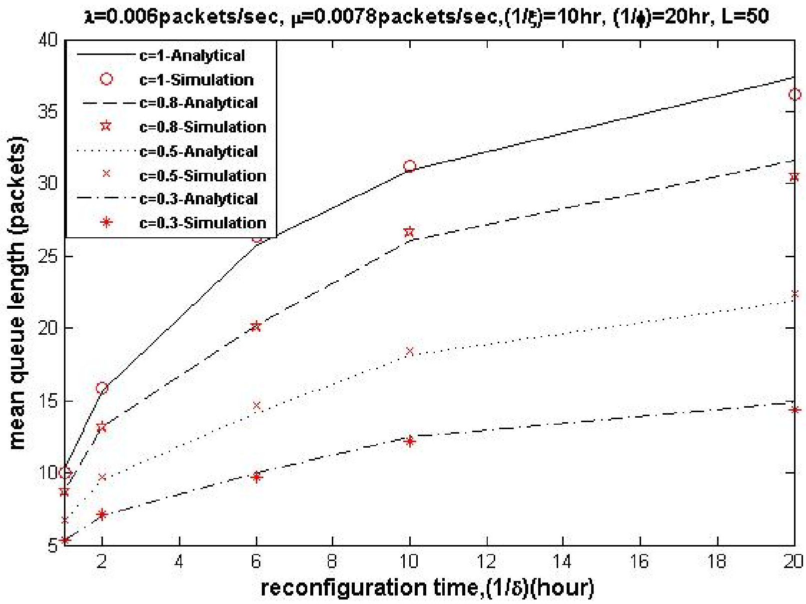

Figure 5.13: MQL results versus reconfiguration times with different c values

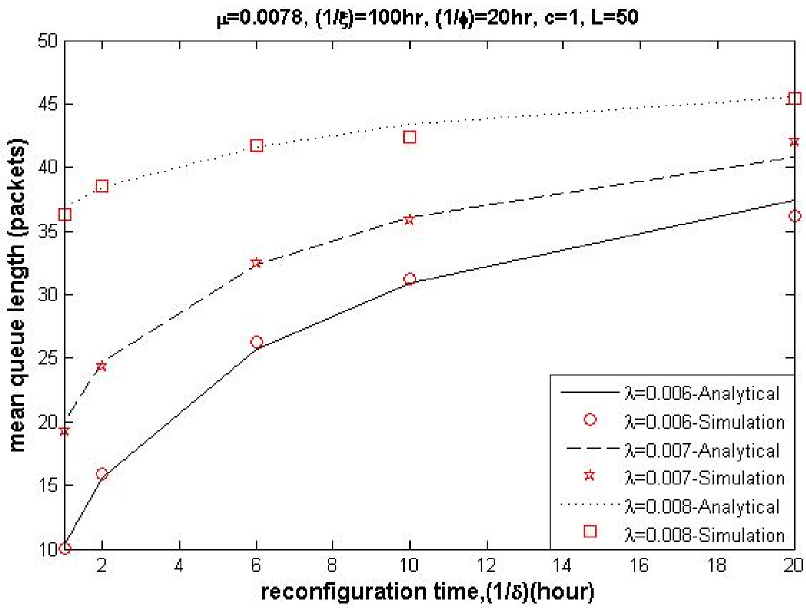

Figure 5.14: MQL results versus channel recovery time with different arrival rates

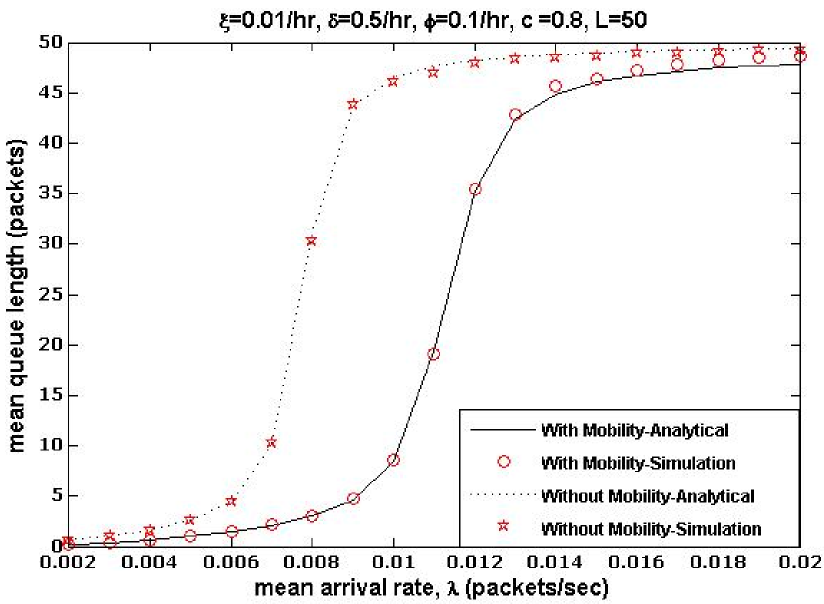

Figure 5.15: MQL results versus mean arrival rates with mobility factor

Finally, a comparative study is preformed in order to validate the proposed technique and evaluate the accuracy of the results obtained for the systems with mobility issues. Simulation and analytical results are presented comparatively for both mobile and static users in a WLAN in Figure 5.15. The proposed WLAN model explained in previous sections is considered for validation. The results from the proposed analytical approach are validated by simulation results. The maximum error rate for performability measures of WLAN are less then 4.61% when analytical and simulation results are compared. Each result obtained from the simulations is within the confidence interval of ± 5 of the mean value with a confidence level of 95%.

5.5 Numerical Results and Conclusions on Performability Modelling of WLAN

Numerical results are presented in this section to show the effectiveness of the model developed for WLAN as a queueing system with availability behaviour of the system. Performability measurements of the system are computed with channel/AP failure and recovery. System unavailability is also defined and presented. The other queuing system characteristics such as various queue sizes are considered to evaluate the performability measures of the systems. The parameters used in this numerical study are mainly taken from the literature. First, the system unavailability with different arrival rates are considered. Other parameters are given as μ_{ch}=0.0078packes/sec (Capar et al., 2003), (Winands et al., 2004), E[v]=2km/h, R=100m, $\rho = \lambda/\mu$ unless stated otherwise (So and Vaidya, 2004), (Song and Zhuang, 2005), (Trivedi et al., 2002), (Wang et al., 2003), (Xia and Shen, 2007), (Yin et al., 2004).

The unavailability of the AP is presented as a function of c in Figure 5.16. ζ is defined as the probability that the AP is not operative and does not broadcast any beacon signal. Therefore, while we are in state $(AP,0)$, the incoming requests are not allowed to enter the system. This makes ζ an important measure for evaluation of various systems and optimization of WLAN parameters. As expected, the availability of AP decreases as the c values increase reaching to one. This means that the probability of AP being operative increases when we have smaller c values (e.g. c=0). In other words, repair facilities for AP should be carefully considered since it is evident that, the AP is not reachable by the mobile users when the c values are higher.

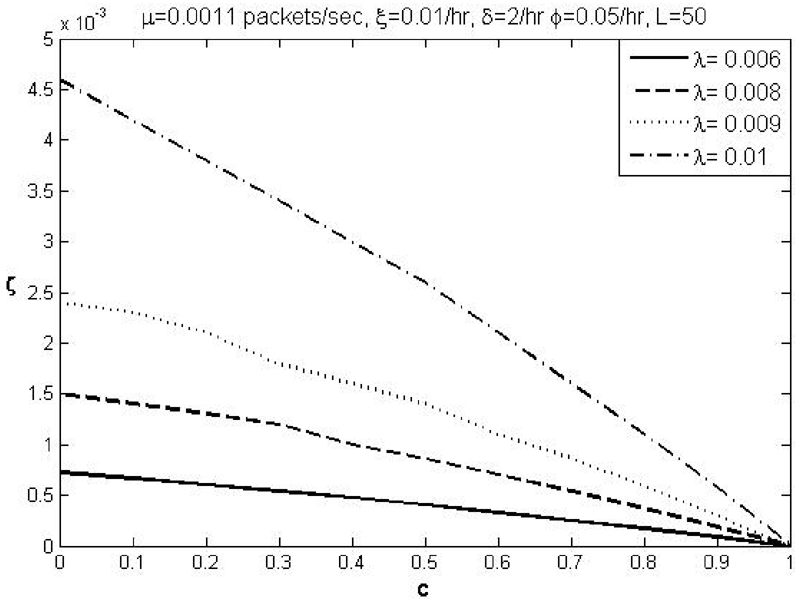

Figure 5.16: Unavailability of access point as a function of (c) for different arrival rates

The MQL results as a function of failure rate ξ for different AP recovery times are shown in Figure 5.17. Interestingly, the results indicate that the shorter the recovery time $1/\phi$, the higher is the MQL. This is mainly caused by the admission of the packets while the AP is available. The system can be repaired quickly and start accepting packets to the buffer. For $1/\phi$=10hr, the AP has smaller recovery time than others. Thus, the system can allow the new packets in the system to serve. When the AP recovery time increases, the system is down and no packets are allowed in the system for a much longer period which causes degradation in MQL results. This shows that ϕ plays an important role for the overall system performance.

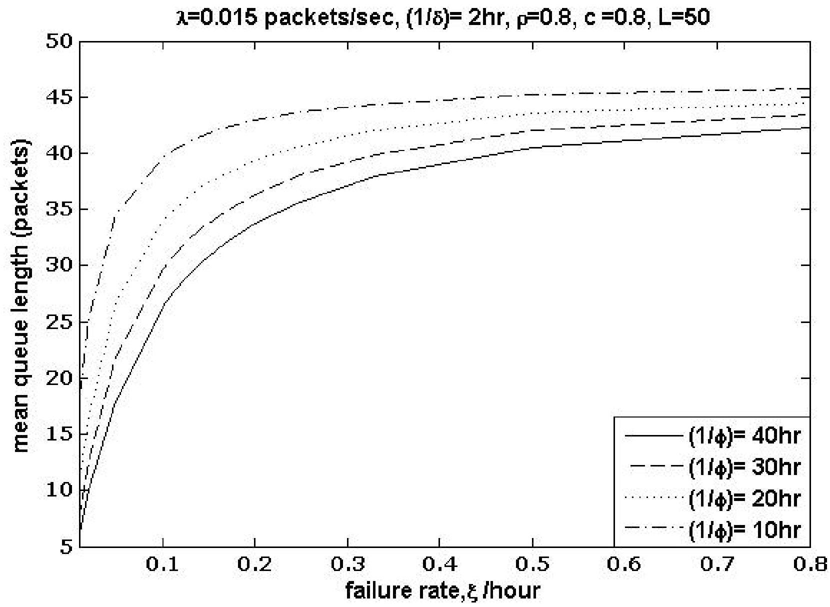

Figure 5.17: MQL results as a function of failure rate for different AP recovery time

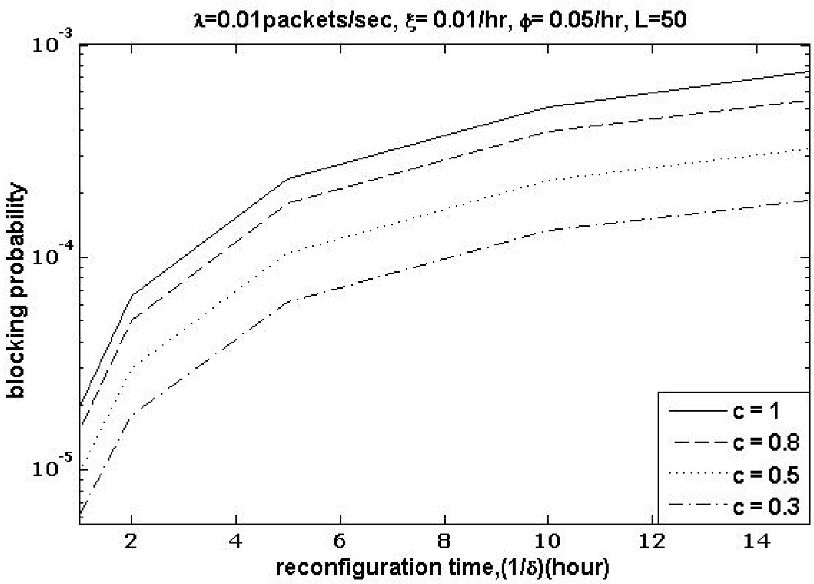

Figure 5.18: Blocking Probability results as a function of channel recovery time for different c values

The blocking probability and MQL measures due to the probability of having either are shown in Figures 5.18 and 5.19 as a function of reconfiguration time for various c values respectively. It is evident that an increase in c, results in increase in the MQL and blocking probability. The implementation issues of this configuration mainly involve two aspects: Firstly, decreasing c means, the AP can fail more frequently than the channel and mobile users do not receive the beacon signal from the AP.

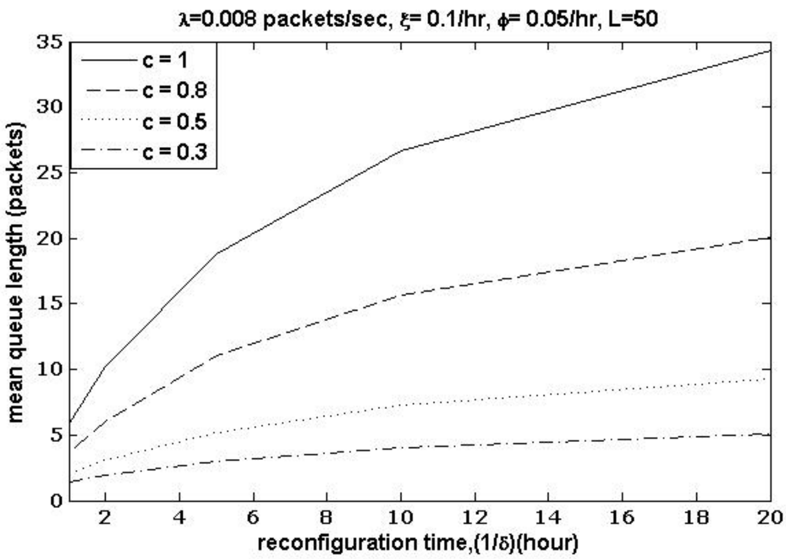

Figure 5.19: MQL results as a function of channel recovery time for different c values

Thus, for small c values mobile users cannot sense the system and cannot send the packets through the AP. Because of this, for smaller c values have less packets in the system. Since the mobile station cannot hear from the AP and cannot transmit, obviously the blocking probability decreases. The effect of this behaviour is of course not positive in terms of QoS and illustrated in Figure 5.16 as system unavailability. Secondly, for any value of the c, MQL increases because of the channel recovery time. For long channel recovery times, the packets queue up in the system and they wait for recovery and/or service. This causes an incremented number of packets in the system. For larger recovery times blocking probability increases as well for all c values, mainly due to the long channel recovery times.

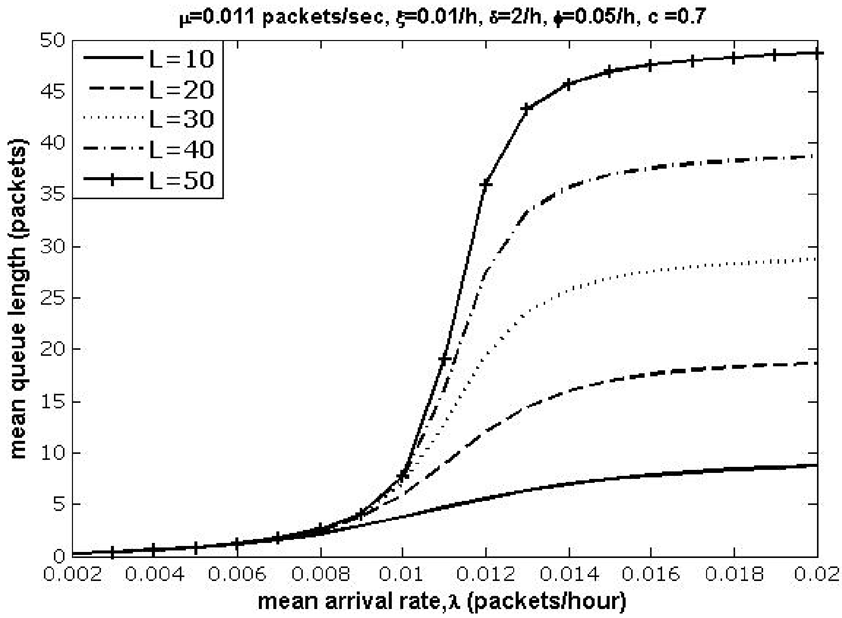

Figure 5.20: MQL results as a function of mean arrival rates for different queue capacities

Figure 5.20 shows the effects of queue capacity as an average number of packets in the system. Clearly, for low traffic loads, MQL results are almost the same regardless of queue length. This is due to reduced number of packets available for transmission in the system. However, when highly loaded systems are considered ($\lambda > 0.01$, traffic intensity, ρ=0.88), L is the main limiting factor. MQL results increase gradually and reach to the queue capacity. In other words, it becomes totally dependent on the value of L.

WLANs with channel failures and AP failures and reconfiguration have been modelled using exact solution approaches in this chapter. The steady state probabilities for WLANs with a single channel, channel/access point failures, and repairs/reconfigurations, are derived using the QBD processes coupled with the spectral expansion method. Numerical results have been obtained and presented for various performability measures. Results show that, the AP failures and reconfiguration affect the QoS more significantly than the channel failures of the system. The failures of access point cause significant system un-availabilities, and therefore the maintenance of the AP is important as well as the repair facility. Results also show that channel and AP recoveries play an important role in degrading the system's performance. In addition, the parameter c, which is the probability that associated with (1-c) and c of the AP and channel failure respectively, is an important parameter to take into consideration. The availability of AP is related with the c values. In other words, repair facilities for AP should be carefully considered. It is clearly shown in Figure 5.16 that, the AP is not reachable by the mobile users when the c values are high. Results provided in Figures 5.9 and 5.10 show the effects of mobility in the system. If the mobile stations with call requests waiting in the queue are not static, many of the users leave the system due to mobility. Because of this, the MQL of systems with static users in the queue is higher than systems with mobile users. The models presented in this chapter are flexible and useful for modelling systems with similar behaviour and queuing considerations.

As is widely agreed, future mobile communication systems, fourth-generation (4G) systems, will be heterogeneous networks, which provide ubiquitous access and seamless mobility across heterogeneous network technologies, such as cellular, WLAN, and broadcast access. Thus, in 4G wireless networks, mobility management is still one of the most important issues that needs to be considered. Mobility in wireless networks basically refers to a MS, changing its point of attachment to the network while its communication to the network remains uninterrupted. A change in the MS point of attachment between homogeneous systems is presented in the previous chapter and this one considering the failures and repairs. However, to achieve both seamless connectivity and continuous mobility, modelling of wireless and mobile systems is presented in the next chapter considering the vertical handover with the horizontal handover model used in the previous chapters.

Chapter 6

Performance Analysis of Handover in Cellular Mobile Networks

6.1 Introduction

The main feature of the cellular mobile networks is the ability to establish ubiquitous and seamless access to various radio access technologies (RATs) and standards. Dealing with QoS requirements and providing the best configuration in order to maximise the performance characteristics of such systems is a challenging task mainly because of the different technologies and standards. In order to achieve these goals, integrating various cellular mobile technologies is an effective way and has recently attracted significant interest (Hasib and Fapojuwo, 2007), (Hasib and Fapojuwo, 2008), (Saravanan et al., 2006), (Shensheng and Wei, 2005),(Song et al., 2005), (Xia and Shen, 2007). However, providing convergence technology is a very complex and challenging task. Analytical models are therefore required to deal with this complexity. In addition, modelling such systems for performance evaluation is essential to improve the architecture according to QoS requirements and performance characteristics. There have been no analytical models that could be used as part of an overall framework. In this chapter, a generic analytical model and solution approach is employed in order to analyse the interaction between different cellular technologies (e.g GSM, GPRS, UMTS etc.). In other words, it is assumed that, the network with the small coverage area namely micro-cell or pico-cell where located inside the network with large coverage area. It is expected that, the network with smaller coverage area will not be perfect; holes will exist within the coverage area. For instance, inside buildings, in deep and shadowed areas. Therefore, handover to larger coverage area will be required in order to avoid call dropping. In addition, handover between two technologies can also affect on capacity through the load sharing. On the other hand, leaving larger coverage area to micro or pico cell will avoid poor QoS. In other words, it provides an improvement in performance. Hence, MSs can seamlessly perform upward and downward vertical handovers between two technologies. Such an integrated system is modelled as a two stage open network and the spectral expansion solution is employed for exact solution. Note that the proposed model is applicable for most of the wireless and communication systems. The handover issues and the integration requirements, QoS characteristics and modelling of integrated cellular and WLAN is explained in Chapter 7. However, it must be noted that symbols and the common parameters are summarized and provided in Table 6.1. The common parameters and symbols for the remaining chapters are explained in detail in this chapter.

6.2 A Mathematical Framework for Vertical Handover in an Integrated Cellular Architectures

Queuing networks have been successfully used in performance and availability modelling of computer and communication systems. The two stage tandem network is considered for modelling and evaluation of the integrated cellular technologies. This system can be modelled as a stochastic process as explained in the Chapter 2. The system considers two stages. The two stage tandem network is an exponential service system. The arrival of requests into the system is modelled by a Poisson process and service times in each stage are exponentially distributed. Requests are served in each stage according to the FCFS basis. The blocking is a principal characteristic of the service in this network. When either the first or second stage queue capacities are full, the systems start to block the arrivals. The two stage tandem networks are not only important for modelling and performance evaluation of computer systems, but also in the field of production systems.

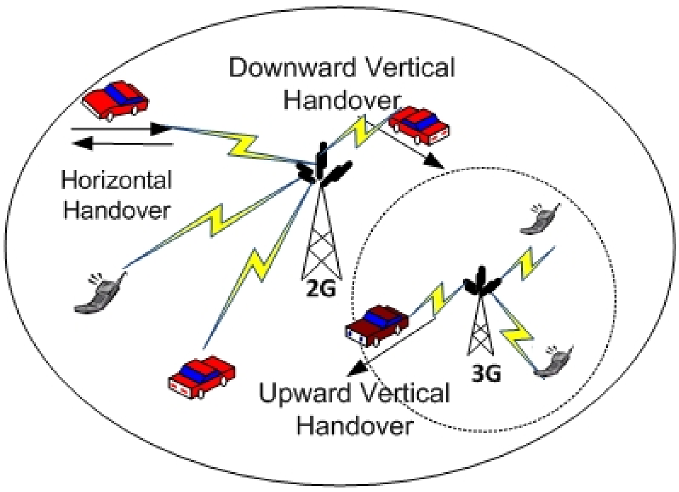

Figure 6.1: Integration of cellular technologies

Consider a global system for mobile communications consisting of a number of isolated 3G networks within the cell. The network diagram for integrated two cellular mobile systems is depicted in Figure 6.1. Assume that there are two user classes in the integrated network. However, potential user calls from all technologies can be incorporated. The 3G users generally have a relative high bandwidth requirement and limited mobility feature depending on the size of cell. On the other hand, the 2G users have a low bandwidth requirement and arbitrary mobility feature. The 2G users, the network that has a bigger coverage area, can always attempt a horizontal handover among neighbouring 2G cells. However, the 3G users can only attempt a vertical handover to the 2G network due to the mobility or if there is no channel and/or resources available in 3G network. No horizontal handover can be made since the 3G cell is isolated. If a 3G user moves out of the coverage area of the 3G cell, it can change its terminal's Ur-interface to Uu-interface to request a call, in such a way it will become another class of user (i.e. 2G user) and the request will become a new call request in the 2G cell. In other words, after successfully making a vertical handover to the 2G, the call may complete the call in the 2G cell, or continue to attempt a take-back vertical handover to the 3G cell.

Table 6.1: Summary of symbols used for integrated systems

Symbol	**Definition**
R	Radius of technology 1
r	Radius of technology 2
V	Speed of the mobile users
P_1	Length of perimeter of technology 1
P_2	Length of perimeter of technology 2
A_1	Area of technology 1
A_2	Area of technology 2
S	Total number of channels in technology 1
K	Total number of channel in technology 2
W_1	Queue capacity of technology 1
W_2	Queue capacity of technology 2
L_1	Maximum number of calls in technology 1
L_2	Maximum number of calls in technology 2
θ_1	Probability of a call to be transferred from technology 1 to technology 2
θ_2	Probability of a call to be transferred from technology 2 to technology 1
θ_3	Probability of a call to be transferred from technology 1 buffer to technology 2
$\lambda_{oc,1}$	Mean arrival rate of originating calls to technology 1
$\lambda_{hh,1}$	Mean arrival rate of horizontal handover calls between neighbouring cells of technology 1
λ_1	Total arrival rate of originating and horizontal handover calls in technology 1
$\mu_{vh,2}$	Mean arrival rate of upward vertical handover call from technology 2 to technology 1
$\lambda_{oc,2}$	Mean arrival rate of originating calls in technology 2
$\mu_{vh,1}$	Mean arrival rate of downward vertical handover call from technology 1 to technology 2
$\mu_{oc,1}$	Mean service rate of originating calls in technology 1
$\mu_{dw,1}$	Mean service rate of horizontal handover calls in technology 1
$\mu_{dw,2}$	Mean service rate of vertical handover calls in technology 1
$\mu_{oc,2}$	Mean service rate of originating calls in technology 2
μ_{dw}	Mean service rate of mobile calls in technology 2
$\lambda_{T,1}$	Total arrival rate of calls in technology 1
$\lambda_{T,2}$	Total arrival rate of calls in technology 2
$\mu_{T,1}$	Total service rate of calls in technology 1
$\mu_{T,2}$	Total service rate of calls in technology 2

6.2.1 The Proposed Model and The System Under Study

The proposed model considers the modelling between two different technologies (e.g 3G and 2G) in the cellular mobile systems for performance evaluation. The integrated two different technologies is shown in Figure 6.2 as a two stage open queuing system. In Gemikonakli et al. (2006), Gemikonakli et al. (2007) and

Gemikonakli et al. (2009) similar modelling approaches are introduced for performance and performability evaluation of two stage open queuing systems. However, open queuing systems with multiple servers in both stages are not considered. In addition, mobility issues such as speed of the users and radius of the cells are considered for both systems. Note that the proposed model is applicable for most of the wireless and communication systems. For instance, in two-tier hierarchical cellular structure, in which large macro-cell are overlaid with small size micro-cell.

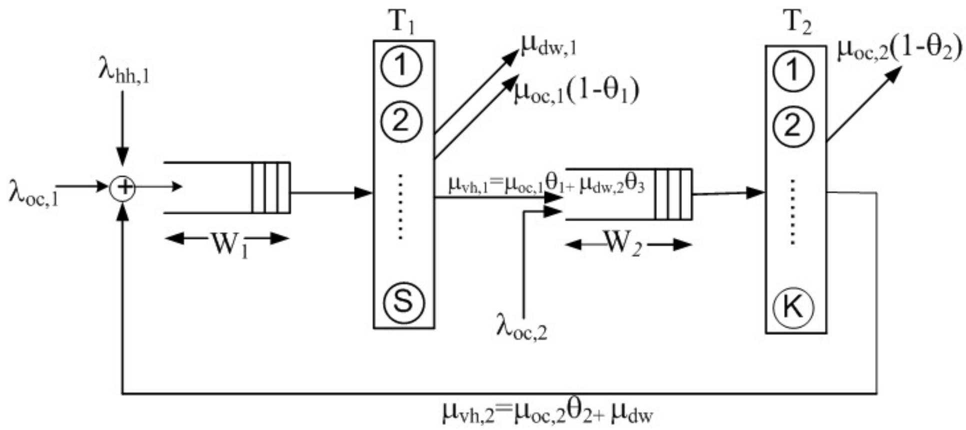

Figure 6.2: The two stage open queuing system considered

The proposed model considers two different cellular mobile technologies namely technology 1 (T_1) and technology 2 (T_2). T_1 is assumed to provide continuous coverage since it has the larger coverage area. T_1 and T_2 cells are assumed to be collocated. Therefore, T_2 cell is deployed inside the T_1 coverage area as shown in Figure 6.1. Thus, this provides the opportunity for both users to switch over to networks or remain connected to the current system. It is assumed that, in a particular time, each user can send one request to only one network for connection (Hasib and Fapojuwo, 2007), (Saravanan et al., 2006), (Song et al., 2005), (Xia and Shen, 2007). Both T_1 and T_2 have a fixed number of identical channels S and K respectively. The queuing capacities are limited with W_1 and W_2, which represent the maximum number of calls waiting for T_1 and T_2 respectively. Assuming the T_2 peak data rate is greater than that of the T_1. It is assumed that number of channels in T_2 is greater than number of channels in T_1 (K > S) and/or second system provides quicker service than the first one (E[$T_{oc,2}$]<E[$T_{oc,1}$]). For both networks, the maximum number of calls allowed into the system is equal to the number of calls assigned with the channels. The queuing capacity of the T_1 is given by L_1 where L_1=S+W_1. Similarly, in T_2 the capacity is given by L_2 where L_2=K+W_2. For analytical tractability, the hexagon cell shapes for both networks (T_1 and T_2) have circular shape of radii of R and r, respectively. In addition, it is important to introduce two types of vertical handovers according to the source and destination networks: Handover from T_1 to T_2 will be called *downward vertical handover* while handover from T_2 to T_1 is called *upward vertical handover*. The upward and downward vertical handover probabilities are explained in the following section.

6.2.1.1 Vertical Handover Decision Probabilities (θ_1,θ_2,θ_3)

There are three types of vertical handover decision probabilities being considered in this model (θ_1,θ_2,θ_3). As shown in Figure 6.2 at stage one, the calls may be transferred to T_2 (downward vertical handover) with a probability of θ_1 or they may leave T_1 following a call completion with a probability of (1-θ_1). In other words, θ_1 and (1-θ_1) are the probability of a call to be transferred from T_1 to T_2 and probability of a call to leave the T_1 after completion, respectively. In addition, if the channels are busy, calls are placed in a queue, then these calls may be transferred to T_2 due to the mobility of a caller with probability of θ_3. Users can stay in T_2 until they get a channel to be served and (1-θ_3) is the probability of a call to stay. On the other hand, no horizontal handover can be made from/to T_2 since there is only one T_2 located in the proposed system. The T_2 users may attempt vertical handover back to the T_1 due to mobility with a probability of θ_2. The T_2 users are served to completion with a probability of (1-θ_2) at the second stage.

6.2.1.2 The Arrival Rates

There are three types of arrivals to T_1. These are new originating calls, horizontal handover calls between neighbouring cells and upward vertical handover calls with arrival rates $\lambda_{oc,1}$, $\lambda_{hh,1}$, and $\mu_{vh,2}$ respectively. These arrivals can request admission from T_1 coverage area. Similarly, new originating calls and downward vertical handover calls from T_1 with rates $\lambda_{oc,2}$ and $\mu_{vh,1}$ respectively can be accommodated by T_2. For both T_1 and T_2, Poisson streams of arrivals are assumed. If the channels are busy in T_1, then the calls are queued. Similarly if a channels are busy in T_2 the requests join the queue. The queuing discipline is assumed to be FCFS for both stages. Mobile users with on-going calls in T_1 may move within T_2 coverage area or move to the neighbouring cells while they are either in the queue or being served in the system. If a user moves to T_2 coverage area, a vertical handover may take place. If T_2 queue is full, the new arrival request access will be denied and hence stay in T_1. Let $\lambda_{T,1}$ and $\lambda_{T,2}$ be defined as the total arrival rate of calls in T_1 and T_2 respectively. Then,

$$\lambda_{T,1} = \lambda_1 + \mu_{vh,2} \quad where \quad \lambda_1 = \lambda_{oc,1} + \lambda_{hh,1} \tag{6.1}$$

$$\lambda_{T,2} = \lambda_{oc,2} + \mu_{vh,1} \tag{6.2}$$

Similarly on-going calls in T_2 can make a vertical handover when they are moving out of T_2 coverage area. When users in T_1 with low mobility are admitted to T_2, i.e. a downward vertical handover proceeds, this happens at a rate of $\mu_{vh,1}$. The probability of vertical handover (T_1 to T_2 or T_2 to T_1) is taken as generic uniformly distributed probabilities and used as input to the queuing system in order to investigate the effects of vertical handover on overall system performance. The T_2 users try to find channels to use in T_2. They may move to T_1 due to mobility. When the T_1 queue is full, and a user being served is out of T_2 coverage area, then the request to handover to T_1 will fail and the call will be terminated. In other words, blocking takes place only if T_1 and/or T_2 cannot offer any buffering for the incoming requests (when the queues are full).

6.2.1.3 Service Characteristics and Channel Occupancy Time

$T_{oc,1}$ and $T_{oc,2}$ are the average call holding time in T_1 and T_2 which follows an exponential distribution with means E[$T_{oc,1}$]=$1/\mu_{oc,1}$ and E[$T_{oc,2}$]=$1/\mu_{oc,2}$ respectively. The T_1 dwell time, which is the time MSs spend in T_1 before moving to neighbouring cells and passing through to in T_2 are assumed to be exponentially distributed with means E[$T_{dw,1}$]=1/$\mu_{dw,1}$ and E[$T_{dw,2}$]=1/$\mu_{dw,2}$, respectively. The residence time in T_2 is also assumed to be exponentially distributed with mean E[T_{dw}]=1/μ_{dw}. $\mu_{dw,1}$ and $\mu_{dw,2}$ are the departure rates of the calls from the T_1 to the neighbouring cells and T_1 to T_2 respectively. Similarly, the departure rate of the residential calls from T_2 to T_1 is μ_{dw}. $\mu_{dw,1}$, $\mu_{dw,2}$, and μ_{dw} assumed to be exponentially distributed and they can can be calculated similarly by using the Equation 2.2 presented in Chapter 2 as follows:

$$\mu_{dw,1} = \frac{E_1[v]P_1}{\pi A_1}, \mu_{dw,2} = \frac{E_1[v]P_1}{\pi(A_1 - A_2)}, \mu_{dw} = \frac{E_2[v]P_2}{\pi A_2} \tag{6.3}$$

E[v] is the average of the random variable, V which is the speed of the mobile users, P_1, P_2 are the length of the perimeter of T_1 and T_2 and A_1, A_2 are the area of T_1 and T_2 respectively. As shown in Figure 6.2 and explained in above section 6.2.1.1, if the calls occupy the channels, they may be transferred to the T_1 with a probability of θ_1, be completed or they may leave T_1 with a probability of (1-θ_1). As stated above, if the calls are in the queue and waiting a channel to be served, they may be transferred to T_2 due to the mobility with probability of θ_3. (1-θ_3) is assumed to be the probability of a call to stay in the T_1 until it gets a channel to be served. Therefore the total service rate of completed call departures from the T_1 can be defined as μ_1 where,

$$\mu_1 = \mu_{oc,1}(1 - \theta_1) + \mu_{dw,1} \tag{6.4}$$

Then, $\mu_{T,1}$ can be defined as the total service rate of T_1 where,

$$\mu_{T,1} = \mu_1 + \mu_{vh,1} \quad where \quad \mu_{vh,1} = \mu_{oc,1}\theta_1 + \mu_{dw,2}\theta_3 \tag{6.5}$$

Upward vertical handover from T_2 to T_1 takes place with a probability of θ_2 with a mobility issue where $\mu_{vh,2}$= $\mu_{oc,2}$ θ_2+ μ_{dw}. In addition, the calls served in the T_2 leave the system with a probability of (1-θ_2) with a mean rate of $\mu_{oc,2}$ where μ_2=$\mu_{oc,2}$(1-θ_2). Hence the total service rate of T_2 can be defined as $\mu_{T,2}$ and is given as follows:

$$\mu_{T,2} = \mu_2 + \mu_{vh,2} \tag{6.6}$$

6.2.2 Two Dimensional Modelling and Steady State Solution

The proposed model is two dimensional Markov process and the steady state probabilities of the proposed system can be obtained using the spectral expansion method. Consider a discrete time, two dimensional Markov process on a finite lattice strip. The Markov process can be defined as $Z = [I(t), J(t)]; t \geq 0$ with an state space of $\{0,1,\cdots,L_1\}$ x $\{0,1,\cdots,L_2\}$. As explained in Chapter 3, the state of the system at time t can be described by a pair of integer valued state variables, $I(t)$ and $J(t)$, specifying the number of calls present at a time t for the T_1 and the T_2, respectively. The model assumes that $I(t)=0,1,\cdots,L_1$, $t \geq 0$, is an irreducible Markov process, which represents the number of calls in the T_1 including the one(s) in service. Moreover, $J(t)=0,1,\cdots,L_2$, is the total number of calls in the T_2 at time t, including the one(s) in service.

The irreducible Markov process is used for performance evaluation of the two stage open network considered. Let the number of calls at the T_1, $I(t)$, be represented in the horizontal direction and possible number of calls at the T_2, $J(t)$, be represented in the vertical direction of a lattice strip. Here, A is the matrix of instantaneous transition rates from state (i,j) to state (k,j) with zeros on the main diagonal. These are the purely lateral transitions of the model Z. Matrices B and C are transition matrices for one-step upward and one-step downward transitions respectively. The transition rate matrices do not depend on j for $j \geq M$, where M is a threshold having an integer value. However, in case of the T_1, the transition rate matrices always depend on j, since the calls in waiting room may depart the system due to mobility as shown in Figure 6.3.

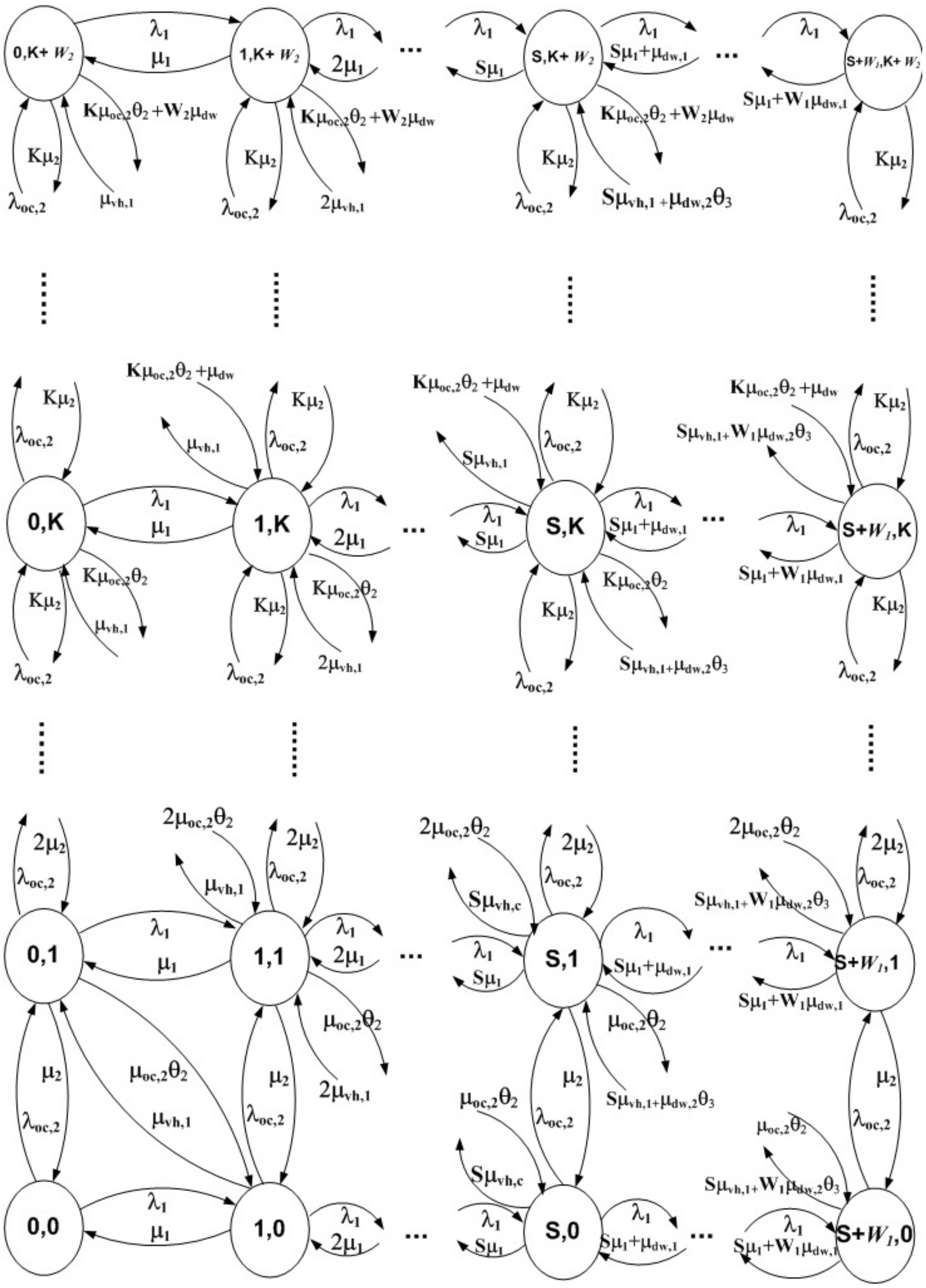

Figure 6.3: The state diagram of the irreducible Markov process of the integration of two different technologies

The transition matrices A, A_j, B, B_j, C and C_j are used for spectral expansion solution can be summarised as follows:

A_i(i,k): Purely lateral transition rate, from state *(i,j)* to state *(k,j)*, (i=0,1, $\cdots$, L_1; k=0,1,$\cdots$, L_1; i$\neq$k; j=0,1,$\cdots$,L_2), caused by a changing of the state in T_1 (i.e. a call has arrived at or departed from the T_1). Clearly, the elements of A depend on the parameters S, W_1, λ_1, $\mu_{dw,1}$, and μ_1. The transition matrices of a system with S channels are of size (Ł$_1$+1) x (Ł$_2$+1).

$$A = A_j = \begin{pmatrix} 0 & \lambda_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \mu_1 & 0 & \lambda_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2\mu_1 & 0 & \ddots & 0 & 0 & 0 & 0 \\ 0 & 0 & \ddots & 0 & \lambda_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & S\mu_1 & 0 & \lambda_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & S\mu_1+\mu_{dw,1} & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & \ddots & 0 & \lambda_1 \\ 0 & 0 & 0 & 0 & 0 & 0 & S\mu_1+W_1\mu_{dw,1} & 0 \end{pmatrix}$$

B_j(i,k): One-step upward transition rate, from state *(i,j)* to state *(k,j+1)*, (i=0,1, $\cdots$, L_1; k=0,1,$\cdots$,L_1; and j=0,1,$\cdots$,L_2), caused by arrival of an originating or a vertical handover call into the T_2.

$$B = B_j = \begin{pmatrix} \lambda_{oc,2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \mu_{vh,1} & \lambda_{oc,2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2\mu_{vh,1} & \ddots & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \ddots & \lambda_{oc,2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & S\mu_{vh,1} & \lambda_{oc,2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & S\mu_{vh,1}+\mu_{dw,2}\theta_3 & \ddots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \ddots & \lambda_{oc,2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & S\mu_{vh,1}+W_1\mu_{dw,2}\theta_3 & \lambda_{oc,2} \end{pmatrix}$$

C_i(i,k): One-step downward transition rate, from state *(i,j)* to state *(k,j-1)*, (i=0,1, $\cdots$, L_1; k=0,1,$\cdots$,L_1; and j=0,1,$\cdots$,L_2), caused by the departure of a serviced call. The elements of matrices C and C_j depend on the parameters K, μ_2, μ_{dw}, $\mu_{oc,2}$, L_2 and θ_2. The matrix C is dependent on the number of calls for j=0,1,$\cdots$,L_2. Therefore, the threshold M is taken as M=L_2. If the number of calls in the system is less than the number of available channels, a channel is assigned for each call. Therefore the downward transition rate is chosen as the minimum of number of calls and number of available channels. On the other hand, if the number of calls is greater than the number of available channels, all of the available channels are assigned to incoming calls and the calls in the queue have the service rates as μ_{dw}. Thus, the matrix C is defined below:

$$C = \begin{pmatrix} K\mu_2 & K\mu_{oc,2}\theta_2 + (L_2 - K)\mu_{dw} & 0 & 0 \\ 0 & K\mu_2 & \ddots & 0 \\ 0 & 0 & \ddots & K\mu_{oc,2}\theta_2 + (L_2 - K)\mu_{dw} \\ 0 & 0 & 0 & K\mu_2 \end{pmatrix}$$

The matrix C_j is defined below for two different regions as explained above:

for $0 \leq$ j $\leq$ K

$$C_j = \begin{pmatrix} j\mu_2 & j\mu_{oc,2}\theta_2 & 0 & 0 \\ 0 & j\mu_2 & \ddots & 0 \\ 0 & 0 & \ddots & j\mu_{oc,2}\theta_2 \\ 0 & 0 & 0 & j\mu_2 \end{pmatrix}$$

for K $<$ j $\leq$ K+W_2

$$C_j = \begin{pmatrix} K\mu_2 & K\mu_{oc,2}\theta_2 + (j - K)\mu_{dw} & 0 & 0 \\ 0 & K\mu_2 & \ddots & 0 \\ 0 & 0 & \ddots & K\mu_{oc,2}\theta_2 + (j - K)\mu_{dw} \\ 0 & 0 & 0 & K\mu_2 \end{pmatrix}$$

The balance equations defined for (M-1) $\leq$ j $\leq$ (L_2-2) are not used in this solution since the threshold value is defined as M=L_2. The steady state probabilities $P_{i,j}$, can be obtained using the steady state solution approach. The solution is given for a bounded queue (i.e. S $\leq L_1$ and K $\leq L_2$). Spectral expansion solution is employed in order to obtain performance measures. After having the $P_{i,j}$, a number of steady-state performance measures can be computed quite easily. For numerical results and discussions, the mean queue length (MQL_1/MQL_2) and blocking probabilities ($P_{B,1}/P_{B,2}$) of both technologies (T_1 and T_2) are computed respectively which can be obtained as:

$$MQL_1 = \sum_{i=0}^{L_1} i \sum_{j=0}^{L_2} P_{i,j} \quad and \quad MQL_2 = \sum_{j=0}^{L_2} j \sum_{i=0}^{L_1} P_{i,j} \tag{6.7}$$

$$P_{B,1} = \sum_{i=0}^{L_1} P_{i,L_2} \quad and \quad P_{B,2} = \sum_{j=0}^{L_2} P_{L_1,j} \tag{6.8}$$

6.2.3 Solution of system of linear equations for the steady state probabilities

Two stage tandem networks and multi-dimensional models are effectively used for analytical modelling of various communication and computer systems. Analytical solutions for two-dimensional Markov processes suffer from the state space explosion problem. Performance evaluation of tandem systems with feedback can be handled with these models. However, because of the numerical difficulties caused by large state spaces, the state explosion problem is a serious issue for numerical techniques such as the spectral expansion. Due to the state explosion problem, some of the results presented in this chapter are calculated by using the solution of system of linear equations. Considering the proposed model the advantage of the well-known system of balance equations is taken. MATLAB package is used for solution with increased number of states. When the balance equations are derived properly, it is possible to represent the system of state probabilities in the form of Ax=B as shown in the matrices below, where A is a matrix of size nxn, x is a column vector of n unknowns, and B is a column vector of n values.

$$\begin{pmatrix} A_{0,0} & A_{0,1} & \dots & A_{0,L_1.L_1} \\ A_{1,0} & A_{1,1} & \dots & A_{1,L_1.L_2} \\ \vdots & \vdots & \vdots & \vdots \\ A_{L_1.L_2,0} & A_{L_1.L_2,1} & \dots & A_{L_1.L_2,L_1.L_2} \end{pmatrix} \times \begin{pmatrix} x_0 \\ x_1 \\ \vdots \\ x_{L_1.L_2} \end{pmatrix} = \begin{pmatrix} B_0 \\ B_1 \\ \vdots \\ B_{L_1.L_2} \end{pmatrix}$$

In proposed system, A is of size L_1xL_2. x is a column vector of unknown state probabilities (P_i) where i=0,1,. . .,L_1xL_2. B consists of the scalars in the balance equations. Redundancy is a problem amongst the global balance equations. Thus the additional information of the normalisation condition is needed. This problem is resolved by replacing one of the balance equations by the normalisation condition.

If A is symmetric and has real, positive diagonal elements, MATLAB attempts a Cholesky factorization. If the Cholesky factorization fails, MATLAB performs a symmetric, indefinite factorization. If A is upper Hessenberg, MATLAB uses Gaussian elimination to reduce the system to a triangular matrix. If A is square but is neither permuted triangular, symmetric and positive definite, or Hessenberg, then MATLAB performs a general triangular factorization using LU factorization with partial pivoting. Similarly B is a column vector with all zeros except from the last element which is 1. B_i vector is denoted, where B={0,0,...,0,1}. Hence the resulting equation can be expressed as follow: $A_i.P_i = B_i$ where i=0,1,...,L_1xL_2.

In steady state, $\pi_{i,j}$ is the proportion of time that the process spends in a state $x_{i,j}$. Recall that the transition rates are the instantaneous rates that the model makes a transition from a state $x_{i,j}$ to a state $x_{i+1,j}$ or from a state $x_{i+1,j}$ to a state $x_{i,j}$. Thus, in an instant of time, the probability that a transition will occur from a state $x_{i,j}$ to a state $x_{i+1,j}$ is the probability that the model was in state $x_{i,j}$, $\pi_{i,j}$ multiplied by the transition probabilities. This is called the transition rates from a state $x_{i,j}$ to a state $x_{i+1,j}$. When the model is in steady state, transition rates used to obtain the matrix A considering the balance equations are in an equilibrium. The transition rates and the balance equations can be found in Appendix A.1.

6.2.4 Numerical Results and Discussions on Handovers in Integrated Heterogeneous Networks

Performance measures have been computed using spectral expansion solution approach as well as solution of system of linear equations numerical solution approach for the proposed model. Moreover, it demonstrates the applicability of the proposed model using simulation. The numerical study focuses on MQL and P_B of vertical handovers in the integrated heterogeneous networks. The values of mean arrival and mean service rates are mainly application dependent. In Figures 6.4 and 6.5 the first system (T_1) S=10 and W_1=10 are considered (L_1=S+W_1 = 20). The other parameters used are E[$T_{oc,1}$]=E[$T_{oc,2}$]=60sec, $\lambda_{oc,1}$=0.01 calls/sec, E_1[v]=40km/h, E_2[v]=2km/h, R=1000m, r=100m θ_1=0.1, θ_2=0.01 and θ_3=0.1. Performance measure, MQL, has been computed for both first (T_1) and second (T_2) systems as MQL_1 and MQL_2 respectively. The thinner line is used to represent MQL_1 where the thicker one is used for MQL_2.

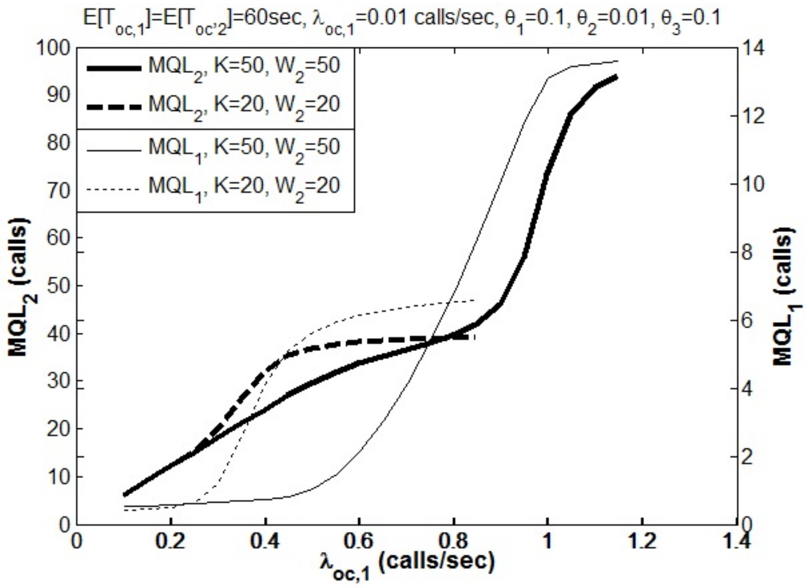

Figure 6.4: The MQL results of T_1 and T_2 with various K and W_2 values

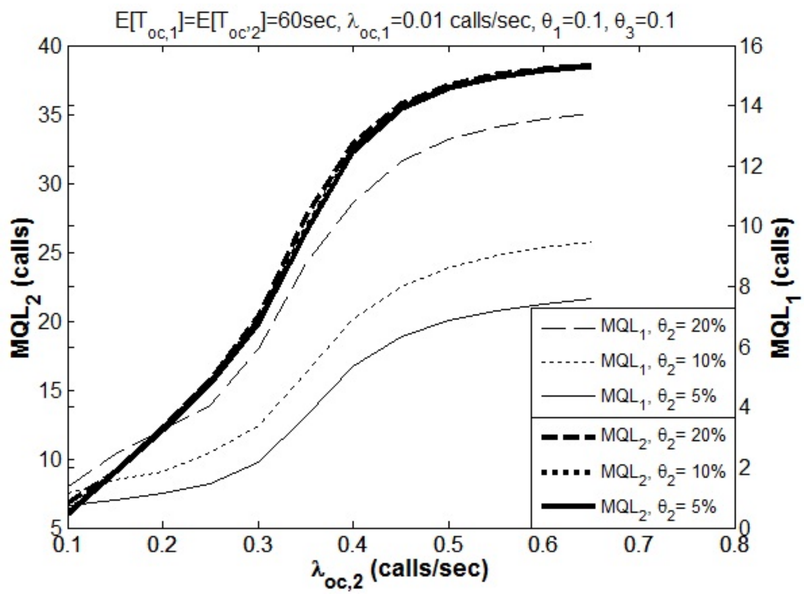

Figure 6.5: Affects of θ_2 for an integrated heterogeneous network where K=W_2=20

Parameters used for Figure 6.4 are used for computation of results presented in Figure 6.5 as well. This time K=20 and W_2=20 values are taken as constant (L_2=K+W_2=40) and effects of various θ_2 values are analysed. Please note that the θ_2 value does not affect MQL_2 significantly and the thick lines are on top of each other. The results provided in Figure 6.4 illustrate an extreme case. It shows that, upward vertical handover from T_2 to T_1 can become quite difficult when the nodes in T_2 are mobile and the capacity of T_2 is significantly greater than the capacity of the T_1. The mobile nodes accumulated in the queue of the T_2 leave the system depending on the radius of T_2 and the velocity of the nodes. Apart from the handovers coming from the channels of the T_2, the handovers due to mobility are fed into the T_1. In this case, since the capacity of the first system (T_1) is much smaller (S=10, W_1=10 compared to K=50, W_2=50) the T_1 can become full quite rapidly, making the overall integration unstable. It is essential to provide a better configuration, or develop new algorithms for similar cases since the system is stable when θ_2 value is around 0.01. This means that most of the calls switching to the T_2 actually need to be completed in the T_2 without the need to the T_1. This is more likely with data traffic which can use the aggregated bandwidth of the T_2 to complete its transfer. On the other hand, the results presented in Figure 6.5 illustrate that the T_1 is stable and θ_2 value can have values up to 0.2 when K=20 and W_2=20. Please note that the MQL of the T_2 is not affected significantly by the changes in the θ_2 value. The calls in the queue of the T_1 may be transferred to the T_2 and the transition rates are increased dramatically. This behaviour can also affect the stability of the entire model. Please also note that user can keep on switching from one network to other continuously when required, hence the effective vertical handover between two technologies will be imperative to ensure continuity of service.

To show the effectiveness and the flexibility of the integrated system, different parameters are considered in Figures 6.6, 6.7 and 6.8. This time the channel numbers and buffer capacities are taken same for the both systems S=K=12 and W_1=W_2=12. However, the channel holding times are different and taken as E[$T_{oc,1}$]=120sec and E[$T_{oc,2}$]=60sec. In addition, other parameters used are $\lambda_{oc,2}$=0.25 calls/sec, E_1[v]=80km/h, E_2[v]=40km/h, R=2000m and r=1000m. These parameters are taken as constant and affects of various θ_1 values are analysed in Figure 6.6.

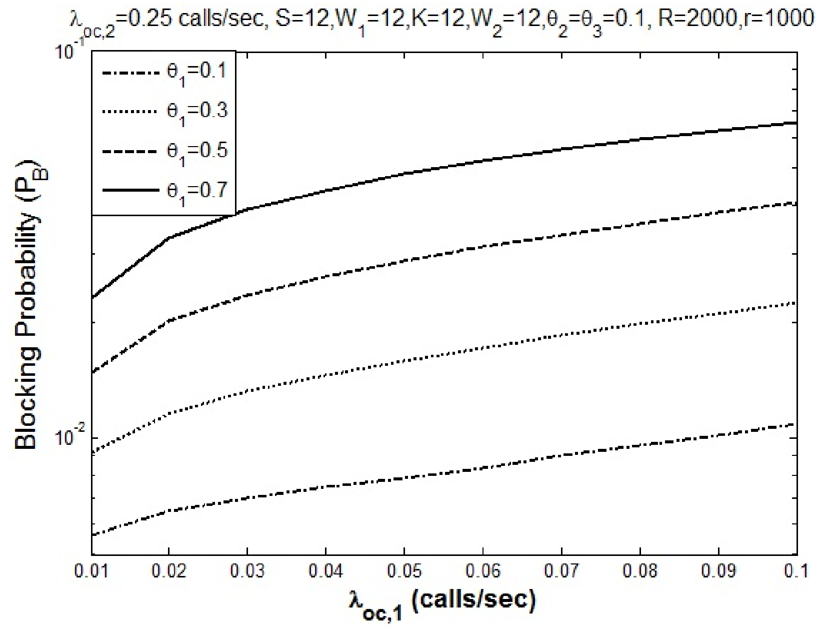

Figure 6.6: The P_B results for the second network ($P_{B,2}$) as a function of $\lambda_{oc,1}$ with various θ_1 values.

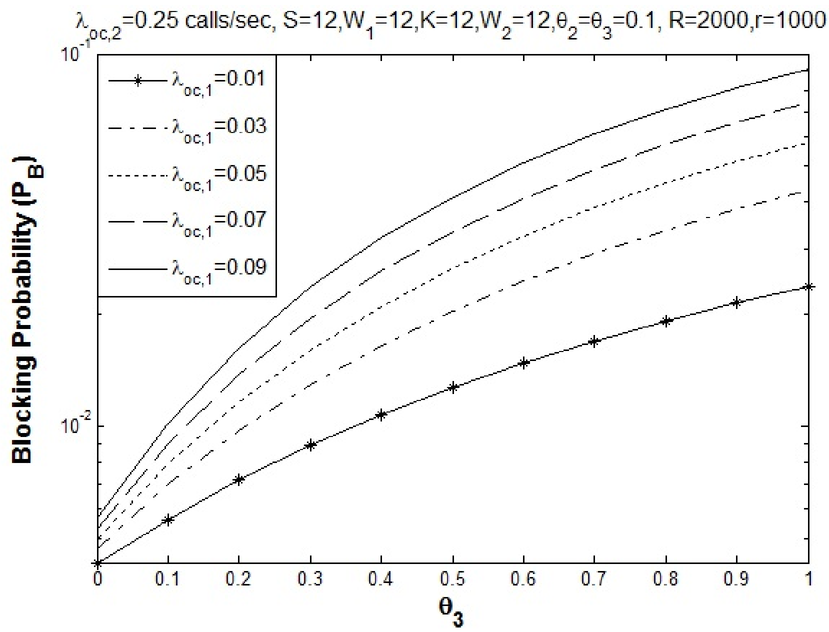

Figure 6.7: The P_B results for the second network ($P_{B,2}$) as a function of θ_3 with various $\lambda_{oc,1}$ values.

As shown in Figure 6.6, the blocking probability for the second network ($P_{B,2}$) increases as the θ_1 increase due to the vertical handover. This is mainly because a greater number of calls attempt to make downward vertical handover. Hence this causes more calls in the second systems. Thus the blocking probability in T_2 increases. In addition, increasing θ_3 affects the integrated system considerably. As shown in Figure 6.7, as θ_3 increases the rate of calls waiting in the buffer of T_1 can transfer to the T_2 without being served due to the mobility. Thus blocking probability ($P_{B,2}$) in the second network increases accordingly. Figure 6.8 shows that the ($P_{B,1}$) results decreases as θ_3 increases. This is mainly due to the mobility in the buffering. In other words, the calls waiting in the buffer of first system switch to the second system very quickly as θ_3 increases. This also leads the redaction of the blocking probability ($P_{B,1}$) in the first system. For instance, for heavy traffic loads (e.g $\lambda_{oc,1}$=0.09) the ($P_{B,1}$) deceases from 0.0268 to 0.001 when θ_3 increases.

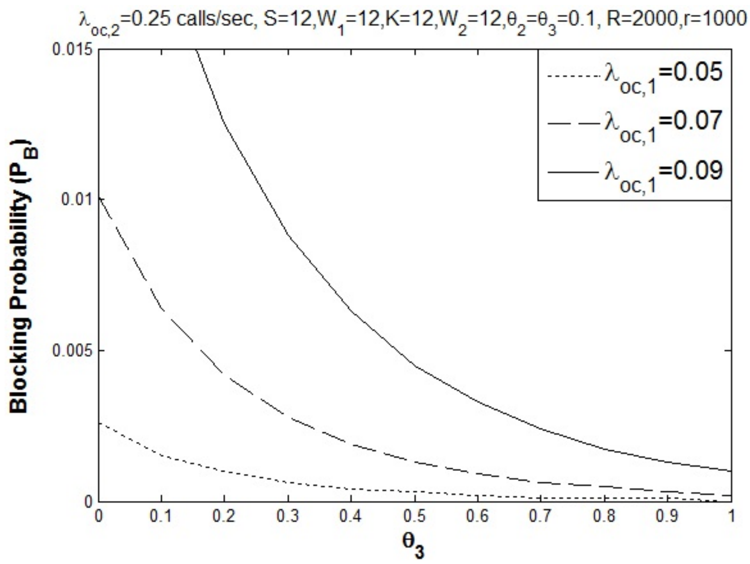

Figure 6.8: The P_B results for the first network ($P_{B,1}$) as a function of θ_3 with various $\lambda_{oc,1}$ values.

The results shown in Figures 6.6, 6.7 and 6.8 illustrate how important the θ_1 and θ_3 is for downward vertical handover from T_1 to T_2. The θ_1 and θ_3 values are important parameters for design and optimisation of integrated systems considered.

Figures 6.9-6.11 and Table 6.2 validate the proposed model of MQL and P_B results with simulation for the different cases. It can be observed that the results obtained from the proposed model are in good agreement with those obtained via simulation. The simulation tool is mainly used for validation of the proposed models, however it can be used for performance evaluation of various scenarios since it simulates the actual scenario rather than the Markov models. Similarly in the Chapter 5, a discrete event simulation (DES) approach is implemented for the two-stage tandem processes. A DES model is defined as one in which the state variables change only at those discrete points in time at which events occur. Two-stages tandem networks with different number of channels at both stages are used for validating the simulation in development phase (Chakka, 1998). The simulation developed considers the stochastic processes; all type of call arrivals and departures occur one at a time in a random discrete event triggered fashion when a arrival enters the system and service is completed respectively. The calls waiting in the queue are served in order of their arrival, first in first out (FIFO), by single and/or multi channel systems at the both stages in the proposed models. When the service event is completed, the channel (and/or channels) becomes idle or remains busy with requests which were stored in the queue. While a particular event is handled, the next event is generated.

The detailed simulation model has been developed of the system written in C++ language. In order to validate the proposed analytical model, the results obtained from the analytical model are presented comparatively with the simulation results for different performance measures such as MQL and P_B of the both systems.

Parameters used for Figure 6.9 are used for the results presented in Table 6.2 as well. The parameters are as follows: S=K=10, W_c=W_w=10, E[$T_{oc,1}$]=120sec, E[$T_{oc,2}$]=60sec, $\lambda_{oc,1}$=0.03 calls/sec, E_1[v]=40km/h, E_2[v]=20km/h, R=2000m. The same set of results are shown for validation purposes in Figure 6.9 and Table 6.2 with respect to two different r values. The plots for MQL_2 in Figure 6.9 and the MQL results of the second system in Table 6.2 show a high degree of conformity.

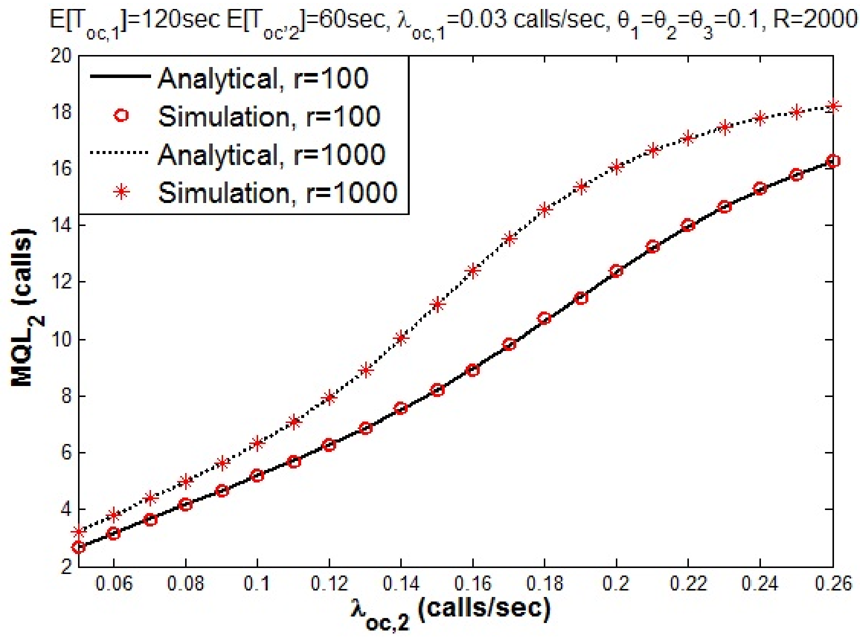

Figure 6.9: The MQL results of T_2 with various r values.

Table 6.2: The validation of the analytical and simulation MQL results as a function of $\lambda_{oc,2}$ with two different r values. (D is Discrepancy).

	MQL_2, r=100			MQL_2, r=1000		
$\lambda_{oc,2}$	**Analytical**	**Simulation**	**D(%)**	**Analytical**	**Simulation**	**D(%)**
0.05	2.6659	2.6786	0.4739	3.1906	3.2168	0.8211
0.06	3.1654	3.1423	0.7304	3.787	3.8023	0.4040
0.07	3.6659	3.643	0.6262	4.3891	4.3924	0.0751
0.08	4.1690	4.1903	0.5099	5.0039	5.021	0.3417
0.09	4.6773	4.6845	0.1539	5.6439	5.6683	0.4323
0.1	5.1949	5.2024	0.1429	6.3284	6.319	0.1485
0.11	5.7283	5.6846	0.7636	7.0837	7.0689	0.2089
0.12	6.2860	6.2695	0.2635	7.9394	8.0268	1.1008
0.13	6.8792	6.86	0.2791	8.9171	8.878	0.4384
0.14	7.5206	7.5615	0.5432	10.0153	9.9392	0.7598
0.15	8.2232	8.2106	0.1536	11.196	11.2055	0.0848
0.16	8.9545	8.9069	0.5315	12.3896	12.4212	0.2550
0.17	9.7643	9.8275	0.6472	13.5181	13.4183	0.7382
0.18	10.6192	10.7119	0.8729	14.5221	14.5072	0.1026
0.19	11.4958	11.4536	0.3670	15.3738	15.2696	0.6777
0.2	12.3647	12.3893	0.1989	16.0737	16.0549	0.11697
0.21	13.196	13.2446	0.3682	16.6386	16.6009	0.2265
0.22	13.9656	14.0276	0.4439	17.0915	17.144	0.3071
0.23	14.6586	14.6793	0.1412	17.4551	17.4963	0.2360
0.24	15.269	15.2848	0.1034	17.7489	17.7166	0.1819
0.25	15.7984	15.786	0.0784	17.9885	18.0002	0.0650
0.26	16.2528	16.2654	0.0775	18.1859	18.1523	0.1847

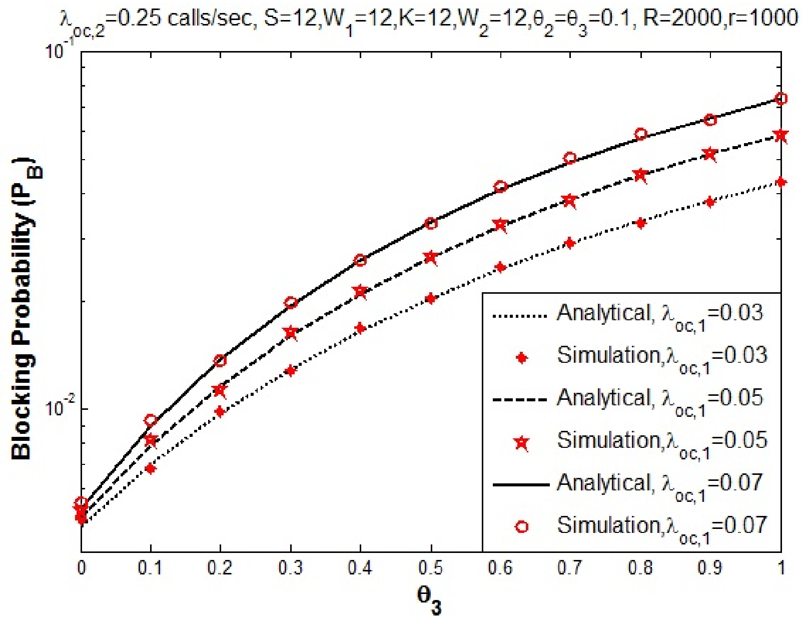

Figure 6.10: The analytical and simulation P_B results for the second network ($P_{B,2}$) as a function of θ_3 with various $\lambda_{oc,1}$ values.

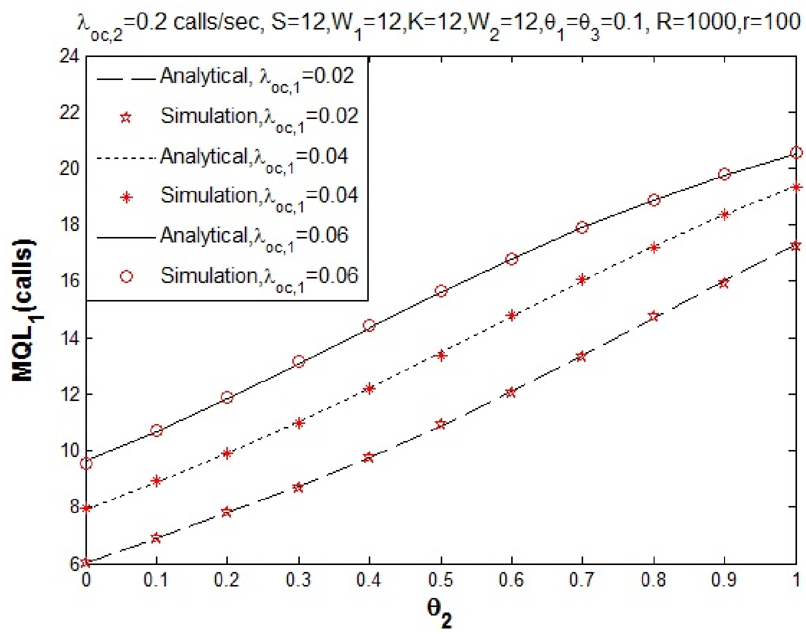

Figure 6.11: The analytical and simulation MQL_1 results for the first network as a function of θ_2 with various $\lambda_{oc,1}$ values.

In Figure 6.10 the P_B results for the second network ($P_{B,2}$) as a function of θ_3 with various $\lambda_{oc,1}$ values are validated using simulation. In Figure 6.10 different parameters are used to show the effectiveness and validity of the proposed model. The parameters are taken as S=K=12, W_c=W_w=12, E[$T_{oc,1}$]=120sec, E[$T_{oc,2}$]=60sec, $\lambda_{oc,2}$=0.25 calls/sec, E_1[v]=80km/h, E_2[v]=40km/h, θ_1=θ_2=0.1, R=2000m and r=1000m. It can be observed that the proposed model results show very good accuracy with those obtained via simulation.

Finally the MQL_1 results are validated using simulation in Figure 6.11. S=K=12, W_c=W_w=12, E[$T_{oc,1}$]=120sec, E[$T_{oc,2}$]=60sec, $\lambda_{oc,2}$=0.2 calls/sec, E_1[v]=40km/h, E_2[v]=5km/h, θ_1=θ_3=0.1, R=1000m and r=100m are used to obtain the numerical results in Figure 6.11. As shown in the figure, the MQL_1 results increases as the θ_2 increases. This is due to the rate of the calls that handed over from T_2 and T_1. The simulation and proposed model results are within the confidence interval of ±5 of the mean value, with a confidence level of 95%.

6.3 Conclusions

The analysis of the handover is an important issue in order to get good performance from the integrated wireless and mobile networks. In this chapter a generic analytical model in different cellular technologies (e.g 2G and 3G) is considered. Unlike the previous studies, for such systems, an exact spectral expansion solution approach is used for the performance evaluation. The mobility issues such as, mobile user's velocity in both systems are considered with different velocities as well as various radius values separately.

In this chapter the analytical framework of the integrated two different technologies modelled as two stage open queuing network. It is shown the fundamental of the modelling the heterogeneous systems and applicable for other types of networks. Simulation results have been obtained to establish a certain degree of accuracy for the models considered. In addition, the solution approach of the system of linear equations is used for some cases. The advantage of the well-known solution approach is taken to obtain some results. The solution approach can also be used for validation of the proposed system. The proposed model specifies QoS measurements of the systems from different application areas. The presented examples were kept simple due to the introductory nature of the proposed model. Obviously, the analysis should be expanded upon for more informed decisions, since the interaction between WLAN and cellular system is essential for optimization studies of heterogeneous wireless communication systems which are considered in the next chapter. Nevertheless, the generic model can easily be extended to express the structure and behaviour of the real system in more detail. Current solution approaches suffer from the well-known state space explosion problem when such complicated scenarios are considered. The solution method provided here looks promising, however because of the multiple servers used at both of the stages of the open network considered, it is desirable to develop alternative approaches to cope with the state space explosion problem.

Chapter 7

Performance Analysis of Handover to Provide a Framework for Vertical Handover Policy Management

7.1 Introduction

The integration of cellular network technologies with WLANs is a common configuration in heterogeneous environments. In this chapter the integration of cellular network and WLAN is considered for performance evaluation. This is done by considering the work presented in Chapter 6 where a generic model is proposed to represent the interactions between different cellular technologies (2G and 3G). The work in Chapter 6 is extended by considering the real operating behaviour of APs in WLAN. The proposed model builds a framework to manage QoS in the future mobile systems and gives detailed analysis for such an integrated communication system together with an exact analytical solution approach. The proposed model considers mobility of users, buffer capacity of WLAN as well as cellular networks, and area of coverage, together with probabilities of having various handover decisions (θ_1, θ_2, θ_3) in order to provide vertical handover policy management. To do this efficiently, WLAN processes are taken into account when considering the real operational behaviour of APs (e.g. single queue, single server). Vertical handover decision management, sensitivity analysis and operational spaces are presented with an analytical modelling and solution in this chapter. Calls are therefore considered as discrete jobs which are served by the 3G and WLAN networks. Using the proposed model, important performance measures, such as mean queue length and blocking probability are computed. The spectral expansion method is employed for the exact solution of the analytical model considered. The results of proposed model are validated by the simulation in order to show effectiveness of the proposed analytical model.

7.2 Performance Evaluation of Integrated Cellular/WLAN Networks

The proposed model considers the modelling of the integrated cellular/WLAN systems for performance evaluation. As shown in Figure 7.1 the WLAN is deployed inside a single cellular network. The cellular network has a larger coverage area than the WLAN. Since the WLAN is within the coverage area of the cellular network, this provides the opportunity for users on the cellular to switch over to the WLAN or remain connected to the cellular network.

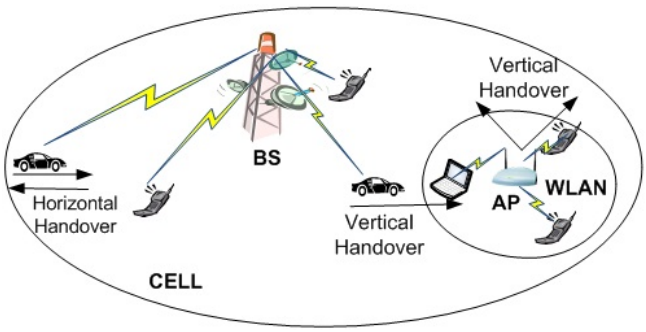

Figure 7.1: 3G Cellular/WLAN integration

The model has been developed to evaluate the performance measures of both cellular systems and WLAN. The model covers handover, various queue capacities and mobility issues (e.g. speed of the users) as well. The cellular network has multiple identical channels. However, a single channel and a single queue is considered for WLAN. Allocation of calls is based on FCA scheme in cellular system.

7.2.1 Model Description

The proposed model described in Chapter 6 can be mapped into the integration of cellular network and WLAN as shown in figure 7.2. In Chapter 6 similar modelling approaches were introduced for performance evaluation of different cellular technologies as a two stage open queuing system. However, a single server (K=1) and a single queue for WLAN AP at second stage are considered in this proposed model. The system consists of S identical channels in the first stage, numbered 1,2,3,$\cdots$,S with a common queue. The common queue has a finite capacity of W_c ($W_c \geq$ S). The maximum number of calls in the system is equal to the number of calls assigned to the channels plus the queuing capacity, $L_c = S + W_c$. On the other hand, a single server and a bounded queue with a capacity of W_w are considered at the second stage. This can be expressed as $L_w = 1 + W_w$. Similarly to the previous chapters, for analytical tractability, the hexagon cell shape and WLAN coverage area are assumed to be circular with radiuses of R and r respectively.

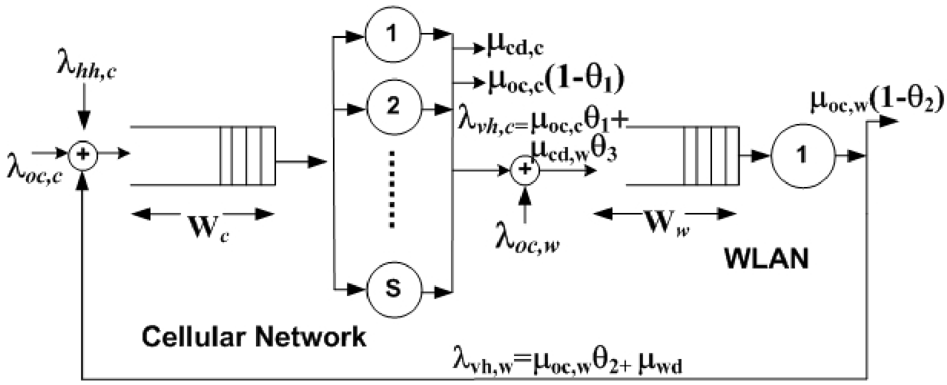

Figure 7.2: The two stage open queuing system considered

As shown in figure 7.2 at stage one, the calls may be performed downward vertical handover to the WLAN with a probability of θ_1 or they may leave the cell following a call completion with a probability of (1-θ_1). As stated in previous chapter, θ_1 and (1-θ_1) are the probability of a call to be transferred from the cell to the WLAN and probability of a call leaves the cell after completion respectively. In addition, if the channels are busy and there are calls in the queue, then these calls may be transferred to the WLAN due to the mobility of a caller with probability of θ_3. Users can stay in the cell until they get a channel to be served and (1-θ_3) is the probability of a call to stay. On the other hand, the WLAN users do have a low mobility speed and no horizontal handover can be made since there is only one WLAN AP in the proposed system. The WLAN users may attempt vertical handover back to the cell due to mobility with a probability of θ_2. The WLAN users are served to completion with a probability of (1-θ_2) at the second stage.

Based on the model introduced in Chapter 6, some of the symbol notations and parameters have changed in this chapter. Let us firstly define and recall some symbols to facilitate the further analysis. $\lambda_{oc,c}$ (new originating calls), $\lambda_{hh,c}$ (horizontal handover calls between neighbouring cells) and $\lambda_{vh,w}$ (upward vertical handover calls) are three types of arrival rates to a cellular system. These arrivals can be accommodated by a cellular system. Similarly, $\lambda_{oc,w}$ (new originating calls) and $\lambda_{vh,c}$ (downward vertical handover calls from cell) are the arrival rates to a WLAN. $T_{oc,c}$ is the average call holding time in cellular network which follows an exponential distribution with mean E[$T_{oc,c}$]=1/$\mu_{oc,c}$. In addition, the cell dwell time for MSs moving to neighbouring cells is assumed to be exponentially distributed with mean E[$T_{cd,c}$]= 1/$\mu_{cd,c}$. On the other hand, in WLAN, the service time of the requests $T_{oc,w}$ is distributed exponentially as reported in the literature and follows an exponential distribution with mean E[$T_{oc,w}$]=1/$\mu_{oc,w}$. In Winands (2003) and Winands et al. (2004) it is shown that service time is independent of the number of the active users. In addition, the dwell time of WLAN users is assumed to be distributed exponentially with mean 1/μ_{wd}. $\mu_{cd,c}$ and $\mu_{cd,w}$ are the departure rates of the calls from the cell to the neighbouring cells and cell to the WLAN respectively. $\mu_{cd,c}$, $\mu_{cd,w}$ and μ_{wd} can be calculated similarly by using the Equations 2.2 and 6.3 presented in Chapters 2 and 6 respectively.

7.3 Modelling for Integrated Cellular/WLAN Networks

The state of the system at time t can be described as explained in Chapter 3 by a pair of integer valued state variables, *I(t)* and *J(t)*, specifying the number of calls present at a time t for the cellular and the WLAN respectively.

7.3.1 The Two Dimensional Representation

The model assumes that $I(t)$=0,1,$\cdots$,L_c, $t \geq 0$, is an irreducible Markov process, which represents the number of calls in the cellular network including the one(s) in service. $J(t)$=0,1,$\cdots$,L_w is the total number of calls in the WLAN system at time t, including the one(s) in service. Then, $Z = [I(t),J(t)];t \geq 0$ is an irreducible Markov process on a lattice strip (a QBD process), that models the system. Its state space is, { 0,1,$\cdots$,L_c } x { 0,1,$\cdots$,L_w }. The irreducible Markov process is used for performance evaluation of the two stage open network considered. Let the number of calls at the cellular network *I(t)*, be represented in the horizontal direction and possible number of calls at the WLAN, *J(t)*, be represented in the vertical direction of a lattice strip. Here, A is the matrix of instantaneous transition rates from state (i,j) to state (k,j) with zeros on the main diagonal. These are the purely lateral transitions of the model Z. Matrices B and C are transition matrices for one-step upward and one-step downward transitions respectively. The transition rate matrices do not depend on j for j $\geq$ M, where M is a threshold having an integer value. However, in case of the cellular networks the transition rate matrices always depend on j, since the calls in waiting room may depart the system due to mobility as shown in figure 7.3.

Clearly, the elements of A depend on the parameters S, W_c, λ_c, μ_c, and $\mu_{cd,c}$. The transition matrices of a system with S channels are of size (L_c+1) x (L_c+1). The state transition matrices A, A_j, B, and B_j, can be given as follows;

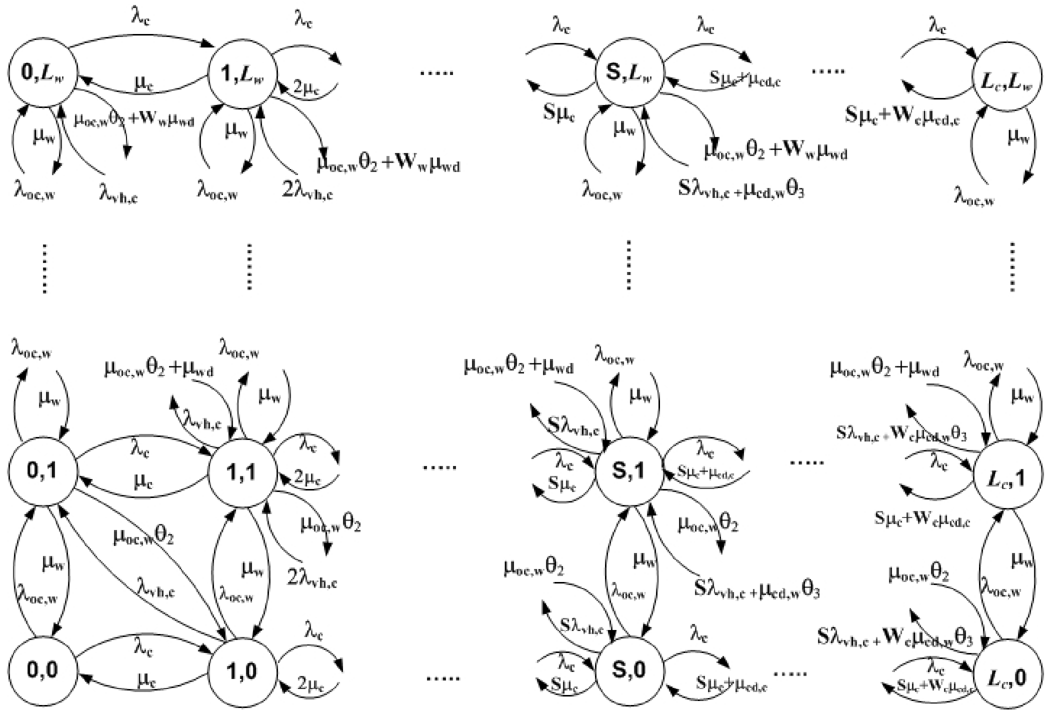

Figure 7.3: The state diagram of the irreducible Markov process of an integrated Cellular/WLAN networks

$$A = A_j = \begin{pmatrix} 0 & \lambda_c & 0 & 0 & 0 & 0 & 0 & 0 \\ \mu_c & 0 & \lambda_c & 0 & 0 & 0 & 0 & 0 \\ 0 & 2\mu_c & 0 & \ddots & 0 & 0 & 0 & 0 \\ 0 & 0 & \ddots & 0 & \lambda_c & 0 & 0 & 0 \\ 0 & 0 & 0 & S\mu_c & 0 & \lambda_c & 0 & 0 \\ 0 & 0 & 0 & 0 & S\mu_c + \mu_{cd,c} & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & \ddots & 0 & \lambda_c \\ 0 & 0 & 0 & 0 & 0 & 0 & S\mu_c + W_c\mu_{cd,c} & 0 \end{pmatrix}$$

$$B = B_j = \begin{pmatrix} \lambda_{oc,w} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \lambda_{vh,c} & \lambda_{oc,w} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2\lambda_{vh,c} & \ddots & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \ddots & \lambda_{oc,w} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & S\lambda_{vh,c} & \lambda_{oc,w} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & S\lambda_{vh,c} + \mu_{cd,w}\theta_3 & \ddots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \ddots & \lambda_{oc,w} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & S\lambda_{vh,c} + W_c\mu_{cd,w}\theta_3 & \lambda_{oc,w} \end{pmatrix}$$

The elements of matrices C and C_j depend on the parameters μ_w, μ_{wd}, $\mu_{oc,w}$, L_w and θ_2 where $\mu_w = \mu_{oc,w}$ (1-θ_2). The matrix C is dependent on the number of calls for j=0,1,$\cdots$,L_w. Therefore, the threshold M is taken as M=L_w. If the number of calls in the system is less than the number of available channels, a channel is assigned for each call. Therefore the downward transition rate is chosen as the minimum of number of calls and number of available channels. On the other hand, if the number of calls is greater than the number of available channels, all of the available channels are assigned to incoming calls and the calls in the queue have the service rates as μ_{wd}. Thus, the matrix C is defined below:

$$C = \begin{pmatrix} \mu_w & \mu_{oc,w}\theta_2 + (L_w - 1)\mu_{wd} & 0 & 0 \\ 0 & \mu_w & \ddots & 0 \\ 0 & 0 & \ddots & \mu_{oc,w}\theta_2 + (L_w - 1)\mu_{wd} \\ 0 & 0 & 0 & \mu_w \end{pmatrix}$$

The matrix C_j is defined below for two different regions as explained above: for j $\leq$ 1

$$C_j = \begin{pmatrix} j\mu_w & j\mu_{oc,w}\theta_2 & 0 & 0 \\ 0 & j\mu_w & \ddots & 0 \\ 0 & 0 & \ddots & j\mu_{oc,w}\theta_2 \\ 0 & 0 & 0 & j\mu_w \end{pmatrix}$$

for 1 $<$ j $\leq$ L_w

$$C_j = \begin{pmatrix} \mu_w & \mu_{oc,w}\theta_2 + (j - 1)\mu_{wd} & 0 & 0 \\ 0 & \mu_w & \ddots & 0 \\ 0 & 0 & \ddots & \mu_{oc,w}\theta_2 + (j - 1)\mu_{wd} \\ 0 & 0 & 0 & \mu_w \end{pmatrix}$$

This system can be solved and the steady state probabilities, $P_{i,j}$, can be obtained using the steady state solution presented in the next section.

7.3.2 Steady State Solution

The spectral expansion solution approach is employed as in previous chapters for the steady state solution. The balance equations defined for (M-1) $\leq$ j $\leq$ (L_w-2) are not used since the threshold value is defined as M=L_w. In addition, both systems are considered as bounded queues for the solution (i.e. S $\leq$ L_c and 1 $\leq$ L_w).

A number of steady-state performance measures can be computed after the steady state probabilities, $P_{i,j}$, obtained. The mean queue length (MQL_c/MQL_w) and blocking probability ($P_{B,c}/P_{B,w}$) of both cellular and WLAN networks are computed respectively which can be obtained as follows:

$$MQL_c = \sum_{i=0}^{L_c} i \sum_{j=0}^{L_w} P_{i,j} \quad and \quad MQL_w = \sum_{j=0}^{L_w} j \sum_{i=0}^{L_c} P_{i,j} \tag{7.1}$$

$$P_{B,c} = \sum_{i=0}^{L_c} P_{i,L_w} \quad and \quad P_{B,w} = \sum_{j=0}^{L_w} P_{L_c,j} \tag{7.2}$$

Please note that the solution of the systems of linear equations used in Chapter 6 could be also used in this chapter. Balance equations in Appendix A.1 could be used when K=1. The further details can be found in Chapter 6 and Appendix A.1.

7.4 Numerical Results and Discussion

Numerical results are presented in this section to show the effectiveness of the proposed framework for vertical handover decision policy. The model provided is two dimensional, and an exact spectral expansion solution is employed for steady state solution. In addition, the proposed model is also solved by using the solutions of the system of linear equations when necessary. The results obtained from this approach as well as simulation are also used for validation of the proposed system. Numerical results are presented for performance measures of both cellular system and the WLAN (e.g. MQL_c and MQL_w respectively) considered. The parameters used are mainly taken from the relevant literature and are given as E_c[v]=40km/h, E_w[v]=2km/h, R=1000m, r=100m (Wang et al., 2003), (Zeng et al., 1994a), (Zeng and Agrawal, 2001),(Hasib and Fapojuwo, 2008), (Hasib and Fapojuwo, 2007), unless stated otherwise. The values of mean arrival and mean service rates are application dependent. Please note that since the WLAN peak data rate is much greater than that of the cellular network, it is assumed that channel holding time in the WLAN is shorter than channel holding time in the cellular network and given as follows: E[$T_{oc,c}$]=60 sec and E[$T_{oc,w}$]=12 sec (Bekker et al., 2004), (Bianchi, 2000), (Khatib, 2003), (Xia and Shen, 2007). In the figures, the thicker line is used to represent the performance measures of the cellular system where the thinner one is used for the WLAN.

7.4.1 Performance for Different System Configuration

First of all, the system with different numbers of channels in the cellular system is considered. The effects of having various numbers of channels in the cellular system with various mean incoming call request rates are shown in Figures 7.4 and 7.5. The L_c is taken as 10. Figures 7.4 and 7.5 show that the number of channels affects the overall performance of the system significantly in the integrated system. When $\lambda_{oc,c}$=0.03 is considered, the MQL_c takes values such as *7.96* for two channel systems and it is around *3.4* for systems with six channels. Systems with different number of available channels may perform similarly for heavy loads and light loads, however, it is obvious in Figure 7.5 that as the number of channels employed increases, the blocking probability of the cellular network decreases due to the service provided by cellular channels. In other words, the higher the number of channels in the cellular system, the better the integrated system performs as expected. On the other hand, when WLAN is considered, MQL_w results slightly change (e.g when $\lambda_{oc,c}$=0.03 is considered, the MQL_w=3.93 and MQL_w=3.72 for S=2 and S=6 respectively for such systems). This has two main reasons. The first one is due to the service provided by more channels (most of the calls are completed). The second reason is related to θ_1 and θ_3 provided. The amount of handover calls coming from the cellular system either being served by channels of the cellular system or waiting in the queue, are not fed in to the WLAN due to the small values of the θ_1 and θ_3.

The effects of queue capacity of the WLAN which is certainly important in determining the overall system performance is shown in Figures 7.6 and 7.7. In this sense, the MQL and $P_{B,w}$ results are shown in Figures 7.6 and 7.7 respectively.

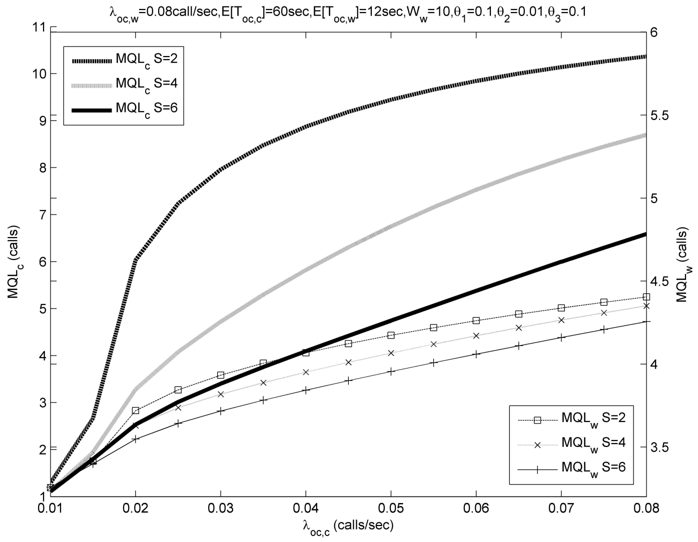

Figure 7.4: The MQL results as a function of mean arrival rates $\lambda_{oc,c}$ for both Cellular (MQL_c) and WLAN (MQL_w) with various S values.

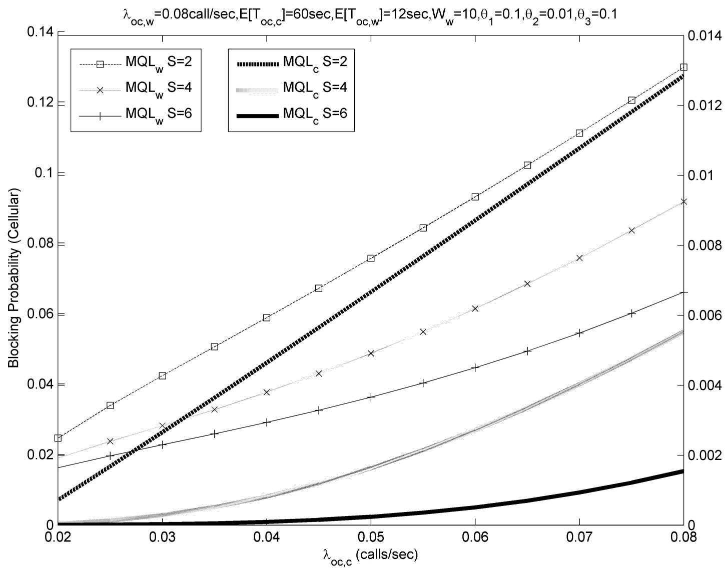

Figure 7.5: The P_c results as a function of mean arrival rates $\lambda_{oc,c}$ for both Cellular ($P_{B,c}$) and WLAN ($P_{B,w}$) with various S values.

As shown in Figure 7.6, the MQL_w results increase gradually and reach to the queue capacity W_w when highly loaded systems are considered ($\lambda_{oc,w} > 0.16$).

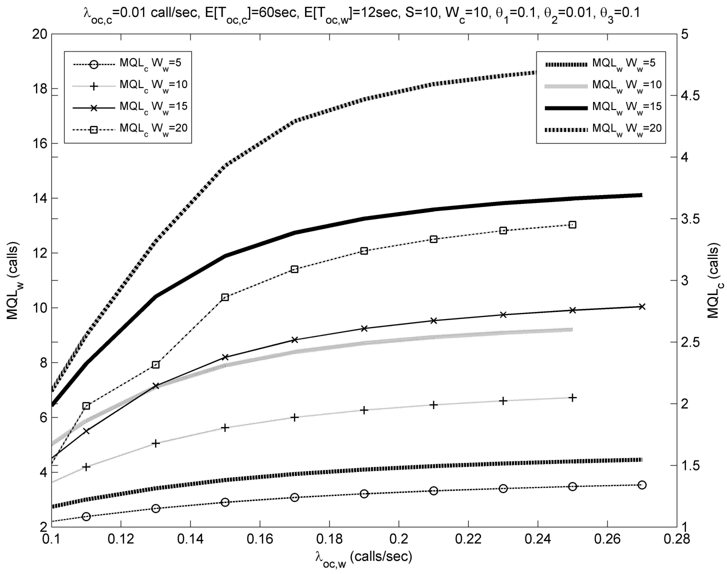

Figure 7.6: The MQL results as a function of mean arrival rates $\lambda_{oc,w}$ for both Cellular (MQL_c) and WLAN (MQL_w) with various W_w values.

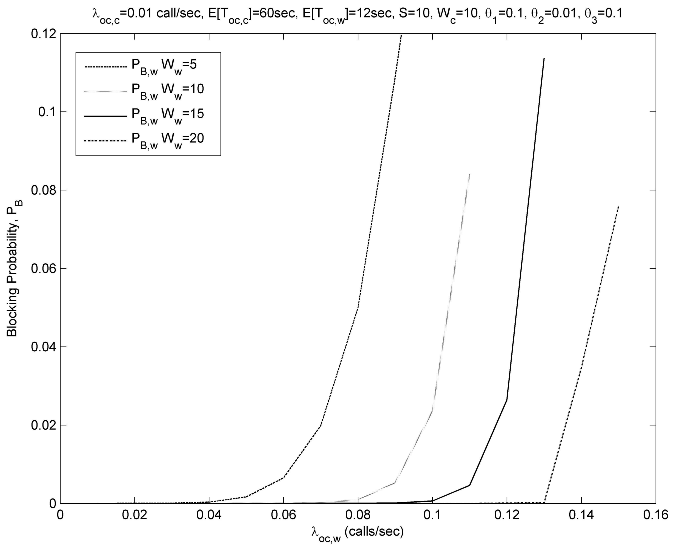

Figure 7.7: The P_B results for the WLAN ($P_{B,w}$) as a function of mean arrival rates with $\lambda_{oc,w}$ various W_w values.

Clearly it is observed that W_w is the main limiting factor in the WLAN and system performance becomes totally dependent on the value of W_w. This is due to the traffic and calls available for transmission in the system. The results show that, with finite queue in the WLAN, it is possible to derive practical QoS results such as blocking probability which is presented in Figure 7.7. The system behaviour in case of increased number of buffers is obvious. Here, it is important to note that results show that the cell blocking probability is zero. The main reason of this is the low rate of handover from the WLAN to the cell. Hence, it is important to look at the effect of increasing the rate of upward handover, for which results are presented in turn.

7.4.2 Effects of Vertical Handover on System Performance

In this section results are shown for various parameters (θ_1, θ_2, θ_3) of the system which are certainly important to show effects of both downward and upward vertical handovers in determining the overall system performance. The results shown in Figures 7.8 and 7.9 illustrate how important the decision policy management is for vertical handover from the cellular system to the WLAN. Figure 7.8 shows that, for lighter traffic loads (e.g. $\lambda_{oc,c}$=0.02) the MQL results slightly change for both systems. This means that, for lighter traffic a vertical handover from the cellular network to the WLAN should not take place due to mobility when the cellular network can provide a good service. Furthermore, increasing downward vertical handover can affect the WLAN performance significantly. For instance, the WLAN can become full quite rapidly, and this makes the overall system unstable for $\theta_1 > 0.78$. θ_3 is another important variable in determining the effect of vertical handover from cellular system to the WLAN due to the mobility in the queue. Figure 7.9 depicts the MQL results for both systems with various $\lambda_{oc,c}$ values as a function of θ_3. As shown in Figure 7.9, as θ_3 increases the rate of calls waiting in the queue of the cellular network can switch to the WLAN without being served in the cell. This causes more calls in the WLAN. The calls move from the cellular network due to mobility. In other words, for $\lambda_{oc,c}$=0.09, MQL_c decreases from *9.97* to *7.97* and MQL_w increases from *1.97* to *5.35*, when θ_3 is considered as *0.1* and *0.9* respectively. However, it happens gracefully. Therefore the results are presented for θ_1 and θ_3 force to focus on the WLAN. Another handover decision probability is the vertical handover from the

WLAN to the cellular network (θ_2).

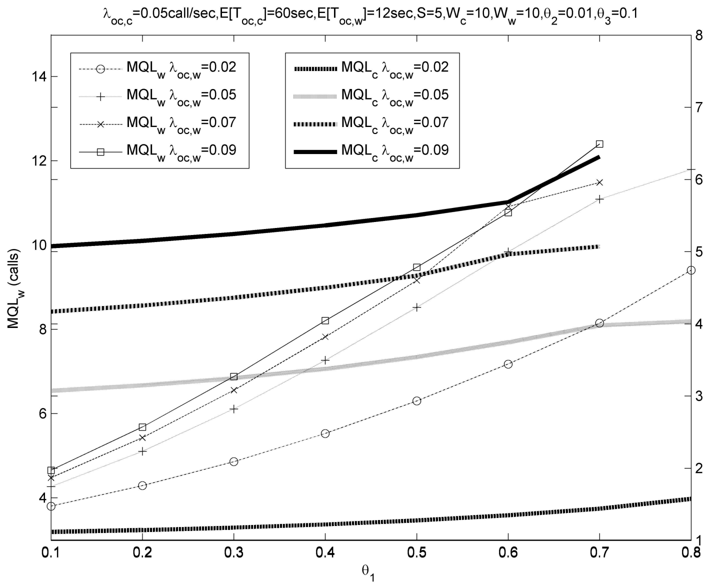

Figure 7.8: The MQL results for both Cellular (MQL_c) and WLAN (MQL_w) as a function of θ_1 with various $\lambda_{oc,c}$ values.

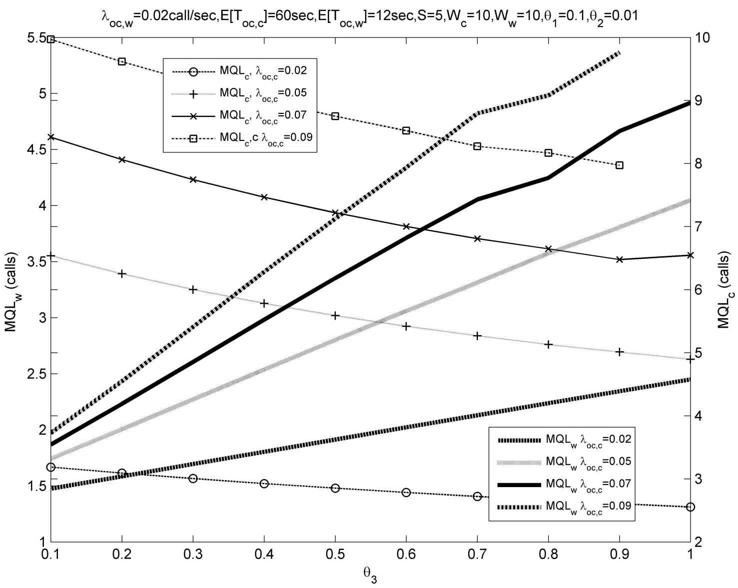

Figure 7.9: The MQL results for both Cellular (MQL_c) and WLAN (MQL_w) as a function of θ_3 with various $\lambda_{oc,c}$ values.

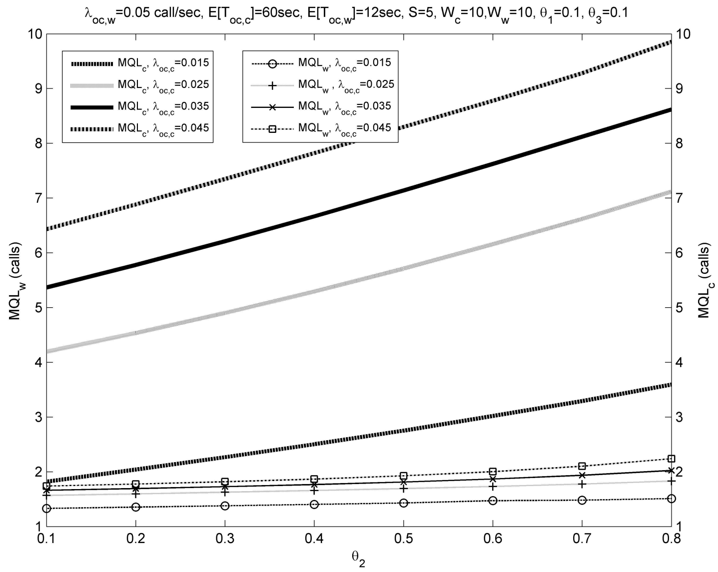

Figure 7.10: The MQL results for both Cellular (MQL_c) and WLAN (MQL_w) as a function of θ_2 with various $\lambda_{oc,c}$ values.

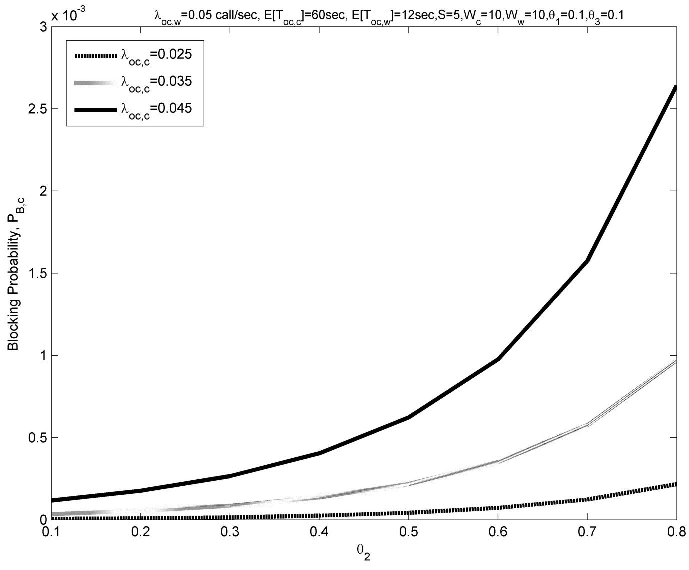

Figure 7.11: The P_B results for the Cellular network ($P_{B,c}$) as a function of of θ_2 with various $\lambda_{oc,c}$ values.

The results are shown in Figures 7.10 and 7.11 for MQL and $P_{B,c}$ respectively. It is observed that the most sensitive handover decision probability is the θ_2. As θ_2 increases, the MQL as well as the P_B of cellular network increases dramatically. This is mainly because of the higher service rate provided by the WLAN. Also, please note that when the buffer of cellular system is full, the WLAN stops transmitting to the cellular network, and the calls to be transmitted stay in the queue even if it is loaded.

Table 7.1: Validation of MQL_c results as a function of $\lambda_{oc,c}$ for different channel numbers. (D is Discrepancy).

	$MQL_c, S = 2$			$MQL_c, S = 4$			$MQL_c, S = 6$		
$\lambda_{oc,c}$	**Analytical**	**Simulation**	**D(%)**	**Analytical**	**Simulation**	**D(%)**	**Analytical**	**Simulation**	**D(%)**
0.015	2.666 6	2.6549	0.4405	1.9459	1.9499	0.2038	1.8013	1.8246	1.2898
0.02	6.0342	5.9824	0.8589	3.2816	3.3152	1.0224	2.5337	2.5491	0.6055
0.025	7.2378	7.2885	0.6992	4.0787	4.0409	0.9268	3.0134	2.9822	1.0371
0.03	7.9614	7.98	0.2326	4.7234	4.689	0.7298	3.4039	3.4407	1.0785
0.035	8.4771	8.4758	0.0161	5.2974	5.287	0.1967	3.7578	3.7536	0.1135
0.04	8.8720	8.8005	0.8064	5.8223	5.7961	0.4515	4.0942	4.0792	0.3674
0.045	9.1870	9.1843	0.0302	6.3056	6.3176	0.1892	4.4216	4.4191	0.0575
0.05	9.4453	9.4451	0.0025	6.7501	6.6687	1.2061	4.7440	4.7577	0.2880
0.055	9.6613	9.6645	0.0330	7.1573	7.1675	0.1420	5.0630	5.0779	0.2942
0.06	9.8447	9.8351	0.0975	7.5288	7.5217	0.0946	5.3787	5.3477	0.5770
0.065	10.0024	10.022	0.1959	7.8663	7.8377	0.3648	5.6905	5.6964	0.1022
0.07	10.1394	10.1187	0.2042	8.1721	8.161	0.1367	5.9973	5.9868	0.1760
0.075	10.2595	10.2439	0.1522	8.4485	8.4582	0.1145	6.2976	6.3309	0.5282
0.08	10.3656	10.3587	0.066	8.6979	8.6751	0.2623	6.5899	6.5635	0.4007

For validation purposes, the numerical results obtained from the analytical models are validated using simulation. Figure 7.12 shows the system with different numbers of channels in the cellular system. The parameters used here are the same parameters used in Figure 7.4. As the number of channels employed increases, the cellular system gives better QoS in the integrated cellular/WLAN. It is obvious in Figure 7.12 that effects of having various numbers of channels in the cellular system plots obtained from proposed model show exact agreement with the result obtained from simulation. The numerical results are also shown in Table 7.1 as a tabular format to show the effectiveness of the proposed model.

The P_B results for the WLAN ($P_{B,w}$) as a function of mean arrival rates ($\lambda_{oc,w}$) with various W_w values are shown in Figure 7.13 and validated by simulation. The results show that buffering (W_w) in the WLAN is the main limiting factor and needs to be considered to design practical QoS. The proposed analytical approach results and simulation results show good agreement.

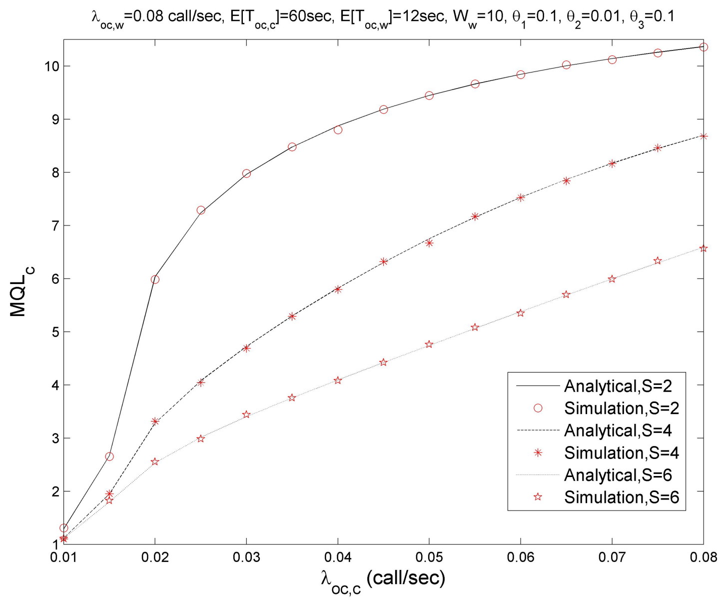

Figure 7.12: The MQL results as a function of mean arrival rates $\lambda_{oc,c}$ for Cellular (MQL_c) network with various S values.

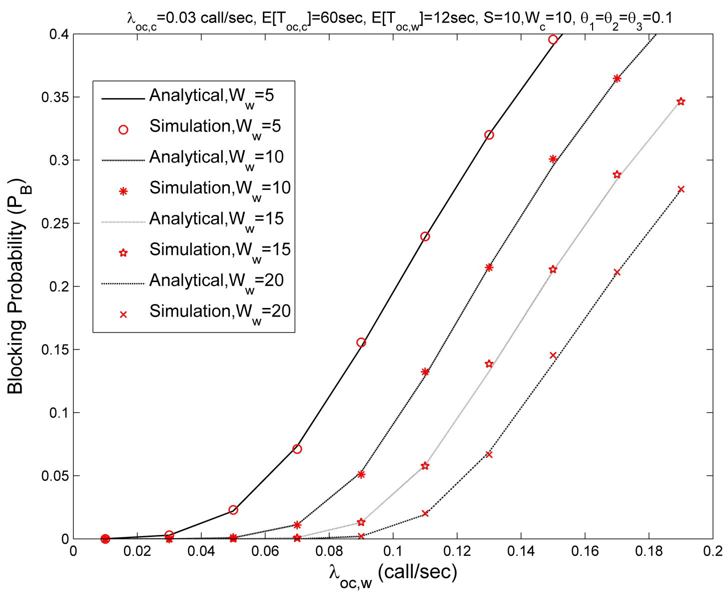

Figure 7.13: The P_B results for the WLAN ($P_{B,w}$) as a function of mean arrival rates with $\lambda_{oc,w}$ various W_w values.

Table 7.2: MQL results for both systems

	Analytical	Simulation		Analytical	Simulation	
$\lambda_{oc,c}$	MQL_c	MQL_c	**Discrepancy(%)**	MQL_w	MQL_w	**Discrepancy(%)**
0.02	3.5824	3.5561	0.7395	2.318	2.2984	0.8455
0.04	5.2395	5.2559	0.3120	3.1583	3.1646	0.1994
0.06	6.3534	6.3358	0.2777	3.8448	3.8823	0.9753
0.08	7.3951	7.357	0.5178	4.5558	4.512	0.9614
0.1	8.4556	8.4663	0.1263	5.3105	5.3494	0.7325
0.12	9.5887	9.6076	0.1967	6.1039	6.0199	1.3761
0.14	10.831	10.7811	0.4628	6.9183	6.7594	2.2968
0.16	12.1903	12.2716	0.6625	7.7266	7.5591	2.1678
0.18	13.6291	13.6602	0.2276	8.4966	8.4649	0.3730
0.2	15.0676	15.1299	0.4117	9.1982	9.1809	0.1880

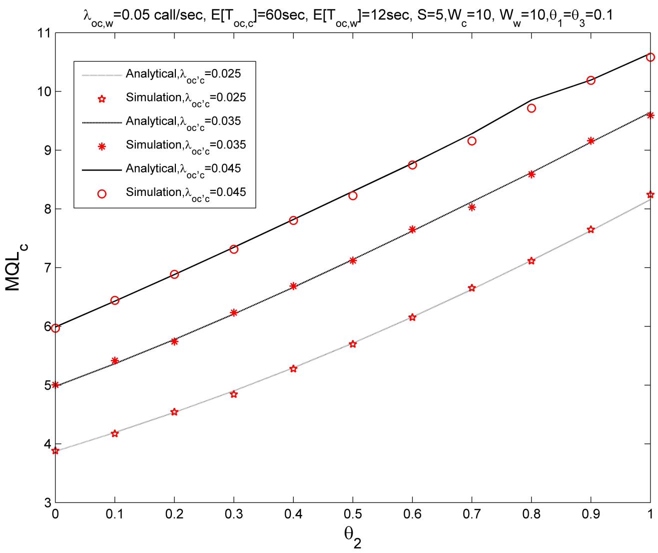

Figure 7.14: The MQL results for Cellular system (MQL_c) as a function of θ_2 with various $\lambda_{oc,c}$ values.

Numerical results are presented in Table 7.2 are the validation of MQL results of both cellular system and the WLAN with simulation (e.g. MQL_c and MQL_w). The parameters used are as $\lambda_{oc,w} = 0.05 calls/sec$, S=12, W_c=12, W_w=15, θ_1=0.2, θ_2=0.1, θ_3=0.3. It is clear from the table that the MQL results for both systems as a function of $\lambda_{oc,w}$ show very good accuracy. The MQL results for cellular network (MQL_c) as a function of θ_2 and MQL results for WLAN network (MQL_w) as a function of θ_1 with various $\lambda_{oc,c}$ values are validated and presented in Figures 7.14 and 7.15 respectively. The both figures show the results from simulation with the 95% confidence intervals shown explicitly.

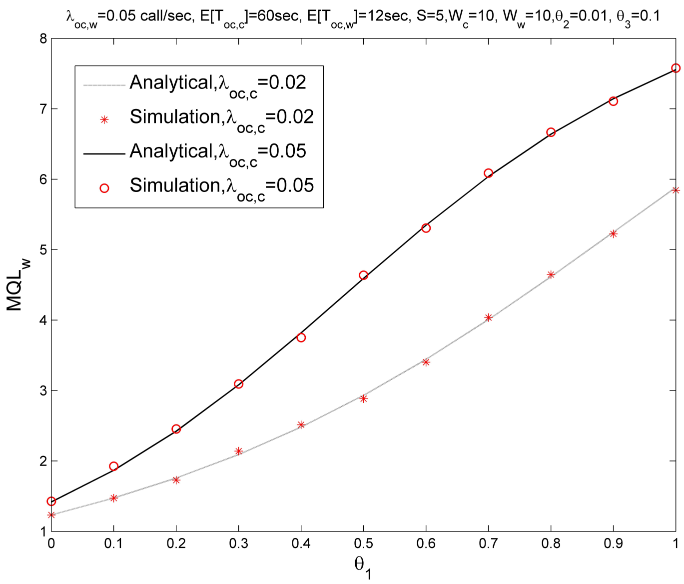

Figure 7.15: The MQL results for WLAN (MQL_w), as a function of of θ_1 with various $\lambda_{oc,c}$ values.

7.5 Conclusions

In this chapter, integrated heterogeneous networks, namely cellular networks and WLAN are considered for performance evaluation. Simulation results and the solution of the system of linear equations have been obtained to establish a certain degree of accuracy for the proposed model considered. The analysis of the handover is an important issue in order to get good performance from the integrated cellular/WLAN networks. Considerable amount of focus is given in specification of operative area while homogeneous as well as heterogeneous handover processes are investigated with the aid of an exact analytical solution approach. Although there are similar works used in explanation of handover phenomena, the analytical approach presented here goes one step further and considers mobility of the users, buffer capacity of the WLAN as well as the cellular networks, and area of coverage, together with probabilities of having various handover decisions.

The results show that increasing the number of channels in cellular network would increase the performance of the overall system. The MQL as well as blocking probability measures decrease for both WLAN and cellular network as the number of channels increases. This shows that, having a full buffer for the cellular network and pending the calls to be transmitted from the WLAN to the cellular network, affects the performance of the WLAN severely. Handover decisions considered are: transferring a call from the cellular network to the WLAN (θ_1), having a vertical handover from the WLAN to the cell due to the mobility (θ_2), and having a vertical handover from the cell to the WLAN (without being served by cellular network) once the mobile station gets in the coverage area of the WLAN due to the mobility (θ_3). The results show that the most sensitive decision to be considered should be for the vertical handover from the WLAN to the cellular network (θ_2). As the probability of having a vertical handover from the WLAN to the cellular system increase, the queue length as well as the blocking probability of cellular network increases dramatically. This is mainly because of the higher service rate provided by the WLAN. When the probability of having a handover from the cellular network to the WLAN is increased (θ_1), the queue size of the WLAN increases, however this happens more gracefully since the WLAN is able to deal with the incoming requests from the cellular network. A similar behaviour is observed for the parameter θ_3, as the rate of calls waiting in the buffer of the cellular network decides to handover to the WLAN without being served in the cell, the blocking probability and queue size decrease for the cellular network, and increase for the WLAN, however it happens gracefully. In addition, θ_3 highly depends on the mobility of the station, it is important to have the correct decision to hand the mobile stations over to the WLAN which provides better service rates. The analytical tool presented is very useful in that sense.

The results mainly show that when it comes to decision of handover, generally the handover process should take place towards the WLAN. However, proposed models need to avoid handovers being handed over to 3G cellular system from WLAN should not be allowed back to the WLAN for QoS enhancement. Therefore a new concept could be proposed in handover management called *handover-to-completion* where calls that have been handed to the cellular network from the WLAN must be completed in the cellular network. These calls are not returned to the WLAN. This means that, such handovers may demand a greater QoS while in the cellular network to complete its task (such as downloading video). In order to achieve this, new analytical models are considered and presented in the following Chapters 8 and 9. Obviously the analysis presented in this chapter is very important, since the interaction between WLAN and cellular system is essential for optimization studies of heterogeneous wireless communication systems.

Chapter 8

Enhancing the Modelling of Vertical Handover in Integrated Cellular/WLAN Environments

8.1 Introduction

In this chapter, the characteristics of the feedback loops are investigated based on the model presented in Chapter 7. The proposed model in Chapter 7 considers the modelling of the integrated cellular/WLAN systems for performance evaluation. Similar modelling approaches are introduced for performance evaluation of integrated 3G cellular network and WLAN systems as a two stage open queuing systems in Balsamo et al. (2003), Hasib and Fapojuwo (2008), Hasib and Fapojuwo (2007), Shensheng and Wei (2005), Song et al. (2008a), Song et al. (2008b), Xia and Shen (2009) and Xia and Shen (2007).

In practical mobile systems, due to different coverage areas for different networks, WLAN networks tend to be totally covered by a cell in a 3G network. Mobile users may arrive into a 3G cell with a horizontal handover from another 3G network. When a WLAN hotspot is discovered by the mobile user (the mobile user gets into the coverage area of WLAN), a downward vertical handover can be performed to the hotspot. A mobile node within the hotspot may leave the coverage area and perform a upward vertical handover back to a cellular network.

A mobile user leaving the coverage area of WLAN due to mobility is not likely to get back to the coverage area of the same WLAN. Therefore, when the flow of users is modelled as a stream of users arriving at the cellular system from WLAN, they should not be allowed back to the WLAN since this may cause an infinite loop of transitions between cells. In other words, such a behaviour may be modelled by closed queuing systems rather than open ones. In practice, once there is an upward vertical handover from the WLAN to the cellular system, there is no handover back to the WLAN, hence there is no feedback loop. In order to capture this adequately a model that prevents handover hysteresis between the WLAN and the cellular network is needed. In this chapter an improved analytical model approach which differentiates the requests coming from WLAN and requests originated in cellular system for the analysis of integrated 3G cellular network and WLAN is presented. Once these streams are differentiated, the calls in cellular system which are coming from WLAN are not allowed back to WLAN. In order to do that, an additional state variable is introduced to differentiate the number of requests originated in cellular network from the ones coming from WLAN. The system and the parameters considered are the same as used in the Chapter 7. More information about the model and the parameters can be found in previous chapter.

8.2 Markov Models for Vertical Handover

Queueing theory and Markov processes are widely used in performance modelling of wireless and mobile communication systems. Markov models are alternative analytical methodologies for the analysis of such systems based on the actual real life system behaviour, leading to both credible and cost-effective approximations for the performance prediction and optimisation of mobile systems. The continuous time Markov chains (CTMCs) are employed to analyse and show the effects of unseen feedback loop for an integrated cellular/WLAN queuing system. All the models demonstrated are irreducible Markov processes. Please note that for simplicity, we consider simple scenarios to analyse the systems. For instance, one WLAN originating call is considered for analysis in first scenario. In addition, the model presented in this chapter and the model in the Chapter 7 are considered comparatively for a simple scenario where there are up to two calls in cellular system and a single call in WLAN.

8.2.1 The Handover Model

The Markov models used in the Chapter 7 do not differentiate the requests originated within the cell from the requests coming from WLAN. For such a configuration, two random variables are sufficient and the Markov chains are shown in Figures 8.1 and 8.5. The number of calls in the cellular network including the one(s) in service is represented using state variable $I(t)$. $J(t)$ is the total number of calls in the WLAN system at time t, including the one(s) in service. Then, $Z = \{[I(t), J(t)]; t \geq 0\}$ is an irreducible Markov process on a lattice strip (a QBD process), that models the system. Once the system is solved for the steady state probabilities $P_{i,j}$, various performance measures can be obtained. In this study, the mean queue length of the cellular system (MQL_c) and the WLAN (MQL_w) are considered respectively which can be obtained as follows:

$$MQL_c = \sum_{i=0}^{L_c} i \sum_{j=0}^{L_w} P_{i,j}, MQL_w = \sum_{j=0}^{L_w} j \sum_{i=0}^{L_c} P_{i,j} \tag{8.1}$$

The model is two-dimensional and an exact spectral expansion solution is employed in order to obtain a steady state solution. The details of the exact spectral expansion solution method for the systems can be found in the Chapter 7. Each call may experience a number of handovers during its connection lifetime in cellular network. However, when a WLAN is located inside the coverage area of the cellular system, calls are expected to hand over from the WLAN to the cellular network and terminate in the cellular network or have a handover to neighbouring cells. As shown in Figures 8.1 and 8.5, the systems used in the Chapter 7 are unable to differentiate between cellular calls originated in the cellular networks from calls that have already gone through the WLAN network. Therefore the system allows calls handed over from WLAN to cellular network to hand over back to the WLAN causing an infinite loop of call requests circulating in the system similar to the behaviour of closed queuing systems.

8.2.2 Introducing and Additional Random Variable

The new model presented in this section is able to differentiate the cellular calls originated in cellular systems from the calls coming from the WLAN. An additional state variable is used to be able to differentiate the number of requests originated in the cellular network from the ones coming from the WLAN. The notation of the new model is accordingly given by the variables (C, C_w, W), where C, represents the number of calls in the cellular system. This can be either the originating calls or horizontal handover calls from neighbouring cells. C_w is the new state variable that represents the number of vertical handover calls from a WLAN to a cellular system. The third variable W, shows the number of calls in a WLAN system.

8.3 Scenario 1

8.3.1 Handover Model for Scenario 1

The first part of the analysis considers one originating call in the WLAN. A two-dimensional Markov chain is used and the analysis is performed using the method in the Chapter 7. Figure 8.1 considers two random variables (i,j), where i denotes the number of calls in the cellular system when λ_c=0 and j represents number of originating calls in the WLAN. In this scenario, for simplicity and ease of explanation, it is assumed that the WLAN can have one call in the system and there are no originating calls in the cellular system as shown in Figure 8.1. There are two types of arrivals to the WLAN system. There are new originating calls with rate λ_w or there may be a vertical handover from the cellular network with a rate of $\lambda_{vh,c}$. A new originating call first tries to get admission to the WLAN. The request is accepted if there is enough bandwidth to accommodate the call. The call in the WLAN may stay in the system until it gets a channel to be served. Since the users considered are mobile, the call may leave the WLAN. In this case the call can be handed over to the 3G cellular network, requesting an admission. In addition, when a Markov chain with two random variables is considered, the call can be sent back to the WLAN coverage area at a rate of $\lambda_{vh,c}$ or can be served by cell with a service rate of μ_c.

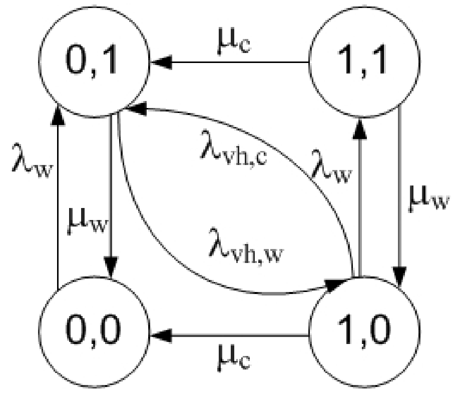

Figure 8.1: The state diagram with an originating call in WLAN

The WLAN system can generate an originating call in the system while the handover calls wait for service from the cellular system. However, according to the scenario 1 there is a single call allowed into the system. When the system is busy, the following requests are blocked. Please note that it is not possible to differentiate the originating calls from the calls coming from WLAN within cellular system, when the model in Figure 8.1 is employed. This is clearly shown in Figure 8.1 because the transitions from the WLAN to the cellular network is shown as (0,1) to (1,0) and the transitions from the cellular to the WLAN is shown as (1,0) to (0,1) hence the handover hysteresis is formed which represents the feedback loop that is observed in obtained results. Therefore, a new model is presented in this study to be able to separate the calls handed over from the WLAN.

8.3.2 The New Model for Scenario 1

The new model presented uses one additional random variable to represent the number of originating calls within the cellular network, and the number of calls handed over from WLAN separately. Considering the first scenario, state (i, k, j) shows that there are i number of requests in the cellular system originated in the cell, k number of requests in cellular system handed over from WLAN, and j number of originating calls in the WLAN. This model for scenario 1 is shown in Figure 8.2. With careful modelling and mature admission control algorithms, both 3G cellular network and WLAN are assumed to support both vertical handovers with QoS provisioning. However, due to handover hysteresis, requests being handed over to 3G cellular system from WLAN should not be allowed back to the WLAN for QoS enhancement. The new model presented clearly differentiates calls in the cell from the calls that have already been handed over to the cell ($\lambda_{vh,w}$) by adding the new state (0,1,0). When the calls are handed over to the cellular system, they terminate in cellular network with service rate of μ_c or the integrated system can accommodate one originating call to the WLAN hence the (0,1,1) state is generated.

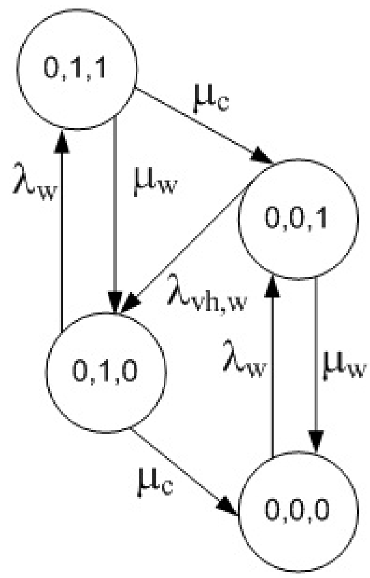

Figure 8.2: The state diagram of new analytical model with a call in WLAN

The vertical handover from cell to WLAN and originating call in cellular system do not appear in the state diagram since only an originating call in WLAN is considered for the first analysis. Hence the initial transition is from (0,0,0) to (0,0,1) which represents a call originating in the WLAN. Although $\lambda_c = 0$, please note that it is possible for the calls in the cellular system to do a downward vertical handover to the WLAN as long as they are in the coverage area of the WLAN (can be for better QoS especially in terms of bandwidth). The state probabilities $P_{i,j}$ can be obtained using the balance equations and solving the resultant system of simultaneous equations. From $P_{i,j}$, a number of steady-state performance measures can be computed. For illustration, the obtained results have concentrated on the mean queue length (MQL_{new_c}/ MQL_{new_w}) of both cellular system and WLAN respectively. MQL values can be obtained as:

$$MQL_{new_c} = P_{0,1,0} + P_{0,1,1} \tag{8.2}$$

$$MQL_{new_w} = P_{0,0,1} + P_{0,1,1} \tag{8.3}$$

8.3.3 Numerical Results

In this section, the performance results of scenario 1 for both handover model used in the Chapter 7 and new proposed model are presented. For a fair comparison, the system parameters used are mainly taken from Chapter 7 based on the relevant literature and are given as $\mu_c = 0.0221$, $\mu_w = 0.0667$, $\lambda_w = 0.07$ calls per second, E_c[v]=40km/h, E_w[v]=2km/h, R=1000m, r=100m, $\theta_1 = \theta_2 = \theta_3 = 0.1$ unless stated otherwise which correspond for specific dwelling times and velocities in the integrated system.

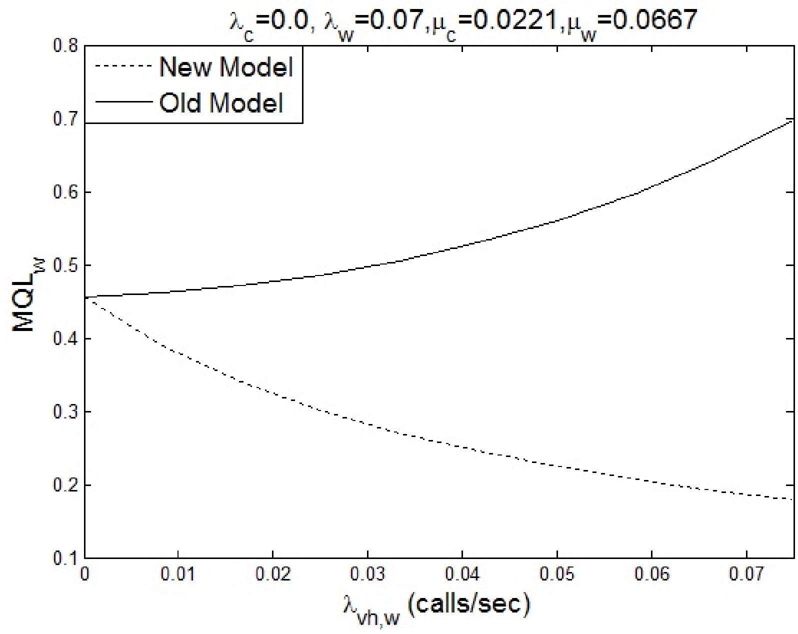

Figure 8.3: Effect of the feedback streams on MQL_w for scenario 1

The MQL_w results are shown as a function of upward vertical handover from WLAN to the cellular system in Figure 8.3 for scenario 1. A call tries to get admission to the cellular system. However, the request is rejected if there is not enough bandwidth to accommodate. Once the calls move to the cellular system, they should not do a vertical handover back to the WLAN in order to avoid infinite loop. With the additional random variable, C_w it is possible to differentiate the handed over calls from the WLAN and avoid to re-transmission back to WLAN. Because the WLAN is within the coverage area of the cellular network, this provides an opportunity for users to remain connected to the 3G cellular network for better performance. Thus the results in Figure 8.3 show that MQL_w for old model increases when the amount of upward vertical handover increases. This is mainly because the calls sent from WLAN to cellular system are allowed to do a downward vertical handover back to the WLAN. The MQL_w results obtained from the new model decreases as the amount of upward vertical handover increases because there will be less calls in the WLAN system. In other words, the results obtained from the new model shows the calls handed over from WLAN to the cellular system are not allowed back to the WLAN.

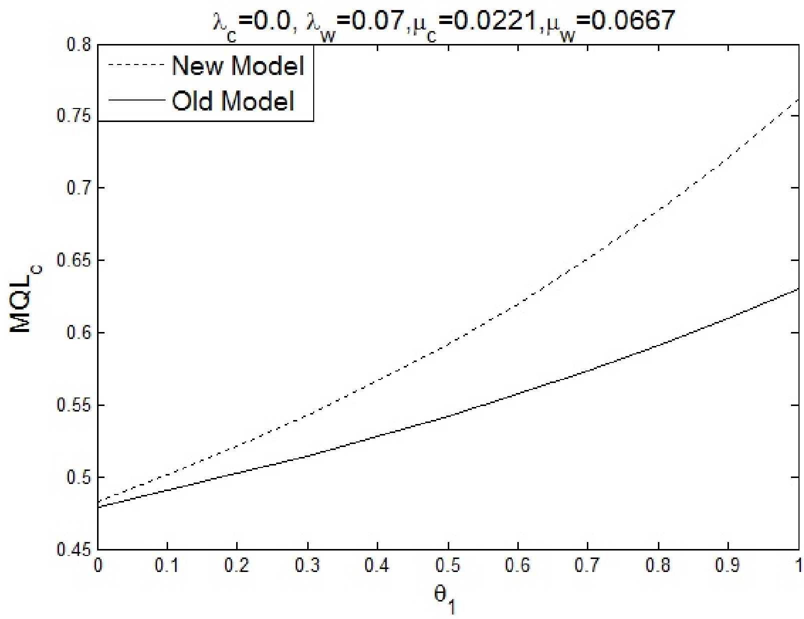

Figure 8.4: Effect of the feedback streams on MQL_c for scenario 1

In addition, MQL_c results are shown as a function of θ_1 in Figure 8.4 for scenario 1. The results in Figure 8.4 show that MQL_c results for new model increase more rapidly compared to the old model. This is because calls coming back from the WLAN must be serviced by the slower cellular network because they cannot again enter the faster WLAN, hence MQL_c increases in the new model compare to the old model.

8.4 Scenario 2

8.4.1 Handover Model for Scenario 2

In this section a more complicated scenario is explored, where there are up to two calls in the cellular system and a single call in the WLAN is considered in this section. The system is also analysed using two-dimensional Markov chain considered in the Chapter 7 which contains only two state variables for state (i,j), where i and j are the numbers of calls in the cellular and WLAN, respectively. The two-dimensional Markov chain model with two originating calls in the cellular system and an originating call in WLAN is shown in Figure 8.5.

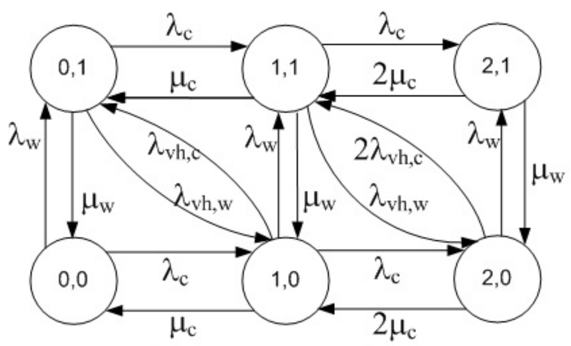

Figure 8.5: The state diagram with two calls in cell and an originating call in WLAN

In the second scenario, the cellular system can accommodate up to two calls in the cell coverage area. The calls in the cellular system may stay in the cell until they get a channel to be served or leave the system to either WLAN or neighbouring cells due to mobility. In addition, the calls handed over to the WLAN can do a vertical handover to the cellular system. On the other hand, similar to the scenario 1, a call in WLAN may stay in the system until it gets a channel to be served, or can leave the system to the coverage area of cell.

It is shown in both Figures 8.1 and 8.5 that the calls already handed over to the cellular system from WLAN are allowed to return back to WLAN. In other words, the system is unable to differentiate between the originating calls in cellular network and the calls which have already been forwarded from WLAN. In the next section a new vertical handover model is presented for scenario 2 to differentiate vertical handover calls coming from WLAN to cellular system from the other calls. Thus, the integrated 3G cellular/WLAN system could behave as a practical system. Even though the scenario 2 is not much more complicated than the scenario 1, the new handover model considered for scenario 2 becomes a much more complex Markov chain.

8.4.2 The New Model for Scenario 2

The new analytical model with two calls in cell and a call in WLAN system is shown in Figure 8.6.

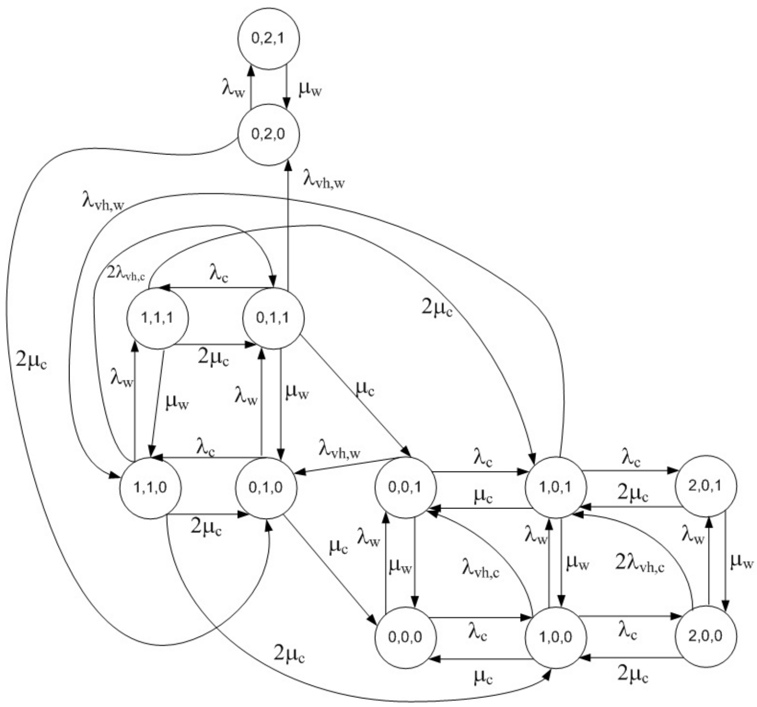

Figure 8.6: The state diagram of new analytical model with two calls in cell and a call in WLAN

A new model which is able to differentiate between originated calls in cellular systems and calls coming from WLAN as vertical handover is presented. A new state diagram is introduced to restrict the unexpected feedback streams from the WLAN to cellular and back to the WLAN. Applying the proposed model described in the previous sections, it is important to note that introducing a new random variable creates a new state for the system. In addition, new state-variable should be automatically generated which is very difficult. This kind of behaviour makes the analytical solution of the model rather complex. The new analytical model is solved by using balance equations together with a system of simultaneous equations. The state probabilities are then used to calculate the mean queue length of the proposed model for both cellular and WLAN systems (MQL_{new_c}/ MQL_{new_w}) respectively as follows:

$$MQL_{new_c} = \sum_{i=0}^{2} iP_{i,0,0} + \sum_{i=0}^{2} iP_{i,0,1} + \sum_{i=0}^{1} P_{0,1,i} + \sum_{i=0}^{1} 2P_{1,1,i} + \sum_{i=0}^{1} 2P_{0,2,i} \tag{8.4}$$

$$MQL_{new_w} = \sum_{i=0}^{2} P_{0,i,1} + \sum_{i=0}^{1} P_{1,i,1} + P_{2,0,1} \tag{8.5}$$

In order to evaluate the performance of the scenario 2 using the proposed model, a new Markov model is proposed by adding new state-variable to the system. The main modelling difficulty herein stems from the fact that the new state significantly increases the complexity of the system description and solution.

8.4.3 Numerical Results

As previously stated, the numerical results are given for a specific scenario (i.e. $I(t) \leq 2$ and $J(t) \leq 1$). The parameters used for numerical results are taken from the literature (Gowrishankar et al., 2009), (Hasib and Fapojuwo, 2008), (Hasib and Fapojuwo, 2007), (Khatib, 2003), (Shensheng and Wei, 2005), (Song et al., 2008a), (Song et al., 2008b), (Xia and Shen, 2009), (Xia and Shen, 2007). Results are presented for $\mu_c = 0.0221$, $\mu_w = 0.0667$, $\lambda_c = 0.027$, $\lambda_w = 0.05$ calls per second, E_c[v]=40km/h, E_w[v]=2km/h, R=1000m, r=100m $\theta_1 = \theta_2 = \theta_3 = 0.1$ unless stated otherwise. The results presented clearly show the effects of feedback streams looping between WLAN and cellular systems.

In Figures 8.7 and 8.8, the effect of the unseen feedback streams of the second scenario on the cellular system and WLAN are presented for the both handover models (systems with two and three state variables respectively). Figure 8.7 depicts the MQL_c results for scenario 2 as a function of vertical handover calls from WLAN to cellular system. It is clearly shown in Figure 8.7 that, MQL_c results for new model increase more rapidly compared to the model with only two state variables. This is mainly because the vertical calls handed over are not re-transmitted back to WLAN. The calls stay in the cellular system and wait for service by the cellular network hence MQL_c increases.

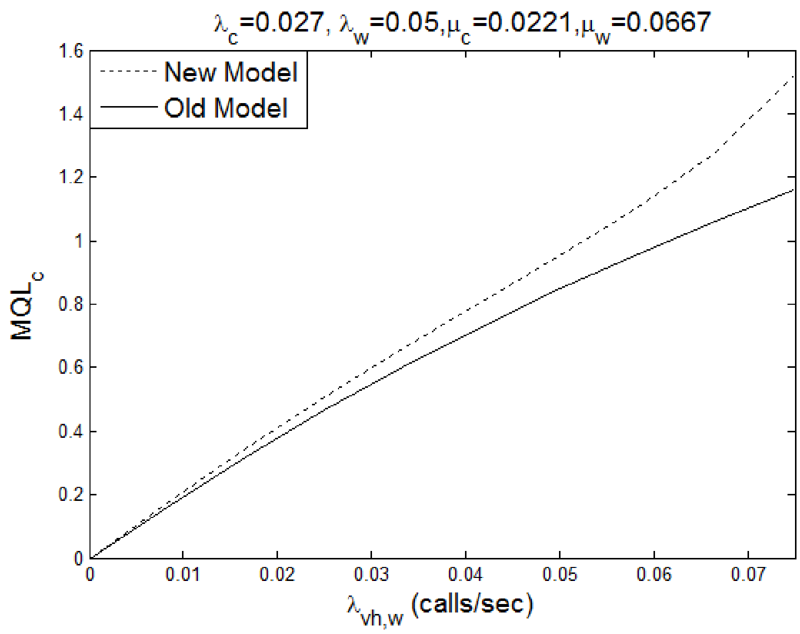

Figure 8.7: Effect of the feedback streams on MQL_c for scenario 2

As the vertical handover rates from WLAN to cellular system increase, the MQL_w is also affected considerably. More significantly, Figure 8.8 shows that increasing the amount of vertical handover calls increases the MQL_w in both scenarios when existing models are employed. This is mainly due to the stream coming back from the cellular system to WLAN. However, in practice, once the requests are transferred to the cellular system from WLAN, they are not expected to be re-admitted to WLAN with a vertical handover from cellular system. This is accurately captured by the new model, which shows that as the upward vertical handover rate increases, there will be less calls in the WLAN system because there is no longer any feedback from the cellular network. Therefore as the results obtained from the new model shows, the MQL_w is expected to decrease as the amount of vertical handover from WLAN to cellular system increase.

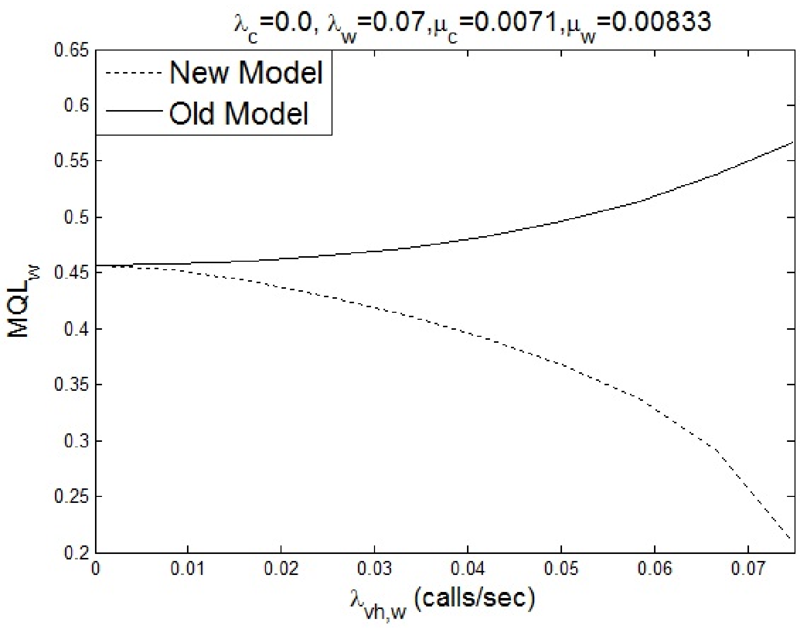

Figure 8.8: Effect of the feedback streams on MQL_w for scenario 2

Table 8.1: MQL results for both systems and models (D is difference)

$\lambda_{vh,w}$	MQL_{new_w}	MQL_w	$D_{wlan}(\%)$	MQL_{new_c}	MQL_c	$D_{cell}(\%)$
0	0.457	0.4565	0	0	0	0
0.0083	0.452	0.458	1.26	0.174	0.161	7.142
0.0166	0.442	0.461	4.204	0.344	0.317	7.735
0.0249	0.429	0.466	8.635	0.506	0.465	8.178
0.0332	0.412	0.473	14.768	0.660	0.602	8.741
0.0415	0.392	0.483	23.080	0.884	0.729	17.422
0.0498	0.369	0.496	34.681	0.953	0.847	11.028
0.0581	0.338	0.514	52.34	1.105	0.958	13.289
0.0664	0.292	0.538	84.265	1.279	1.06	17.087

The results in Table 8.1 clearly show the effects of the feedback streams considering the second scenario when $\lambda_w = 0.07$ calls/sec. Originating calls in cellular system $\lambda_c = 0$ and effects of vertical handover is maximised since it is the only source of incoming requests to the cellular system. In other words MQL_{new_w} decreases as expected because there is no stream coming back from the cellular system to the WLAN. On the other hand, MQL_c results for new model increase more rapidly compared to the old model. This is because calls coming back from the WLAN must be serviced by the slower cellular network since they cannot re-enter the faster the WLAN. Hence MQL_{new_c} increases in the new model compared to the old model. Even for smaller values of $\lambda_{vh,w}$ the difference in the results between the two models can be high. For instance $\lambda_{vh,w} = 0.0664$ the difference between MQL_{new_w} and MQL_w is up to 84.265%.

The existing modelling approach uses one state variable in order to represent the number of calls in cellular system. Due to the restrictions of Markov property, they fail to classify the source of requests stored in the cellular system. The results indicate that the proposed model solves the unseen feedback streams problem.

8.5 Conclusions and Future Work

In order to analyse the integration of vertical handover in future heterogeneous wireless networks, two stage queuing system models such as Balsamo et al. (2003), Hasib and Fapojuwo (2008), Hasib and Fapojuwo (2007), Shensheng and Wei (2005), Song et al. (2008a), Song et al. (2008b), Xia and Shen (2009) and Xia and Shen (2007) have been proposed. Feedback during the vertical handover procedure in integrated cellular/WLAN system is a critical issue in analytical modelling which affects the QoS of the system.

A new Markov model is proposed for cellular/WLAN integration. The new model clearly differentiates requests originating in the cellular system, from requests being handed over from WLAN to cellular system, therefore the calls handed over from WLAN to cellular network are not allowed back to the WLAN. This prevents a handover hysteresis which the obtained results show will severely affect system performance. The new model therefore is an improved model for studying handover in real systems. The proposed model is therefore important for the design of practical vertical handover schemes.

This new model is definitely more accurate. However, parameters and the generation of state-space should be automatically performed, and this process can become quite complicated. The model considered here gets complicated as the number of calls allowed in the systems increases. Hence, more work is needed to automate the generation of the state diagrams and to solve the equations for the state probabilities in order to make this new model useful for more complex mobile systems. Furthermore, since the states of the system is quite complicated, it is desirable to use the existing methods for the automatic generation of the states similar to the approaches used in Inverardi et al. (2006) and Inverardi et al. (2005).

As future research, exploring the use of guard channels as a way of benefiting from having reserved channels in the 3G cellular networks should be taken into account. Such an approach would also allow us to differentiate the originating calls from vertical handover calls in system level. Modelling and solution of the proposed system using the guard channels scheme and buffering are presented in the next chapter where upward vertical handover calls only associated with reserved channels.

Chapter 9

Exploring Analytical Models to Study Vertical Handover in Heterogeneous Networks using Guard Channels and Buffering

9.1 Introduction

This chapter presents a new analytical model from the analysis and results described in Chapters 7 and 8 to define the operational spaces that need to be explored. Using the proposed model and analysed constraints, it explores the operational spaces where the QoS of the integrated cellular/WLAN networks could be provided and obtains optimal operational points.

This chapter mainly focuses on QoS of the integrated systems for future networks considering the system's behaviour as well as architecture in a heterogeneous environment from an analytical point of view. Similarly to the previous chapters, calls are therefore considered as discrete jobs which are served by the 3G and WLAN networks. The proposed model builds a framework to manage QoS in the future mobile systems using guard channels and buffering. The old modelling approach in Chapter 7 uses one state variable in order to represent the number of calls in the cellular system. Due to the restrictions of Markov property, they fail to classify the source of requests stored in the cellular system. The results indicate that the new model presented in Chapter 8 solves the unseen feedback streams problem. However, when the proposed model with three state variables is considered, it is important to note that introducing a new random variable creates additional states for the system. In addition, new state variables should be automatically generated which can be a challenging task. The main modelling difficulty herein stems from the fact that the new state significantly increases the complexity of the system description and solution because state transitions occur due to both service and handover mechanisms. In order to solve the handover hysteresis, the use of guard channels as a way of benefiting from having reserved channels in 3G cellular networks is taken into account in this chapter. The simplest way of giving priority to handover calls is to reserve some channels in cellular networks for handover calls going from the WLAN back to the 3G cellular network explicitly in cellular network. Therefore, it is important to provide channels to existing calls that goes through the handover process so that on-going calls can be continued. Thus, the guard channel approach also allows the system to differentiate the originating calls from vertical handover calls in system level.

9.2 The System Model

This section presents the proposed model for the performance evaluation of an integrated communication system composed of cellular network and WLAN AP. The model has been developed to evaluate the performance measures of both cellular systems and WLAN. The system considered is similar to systems considered in Chapters 7 and 8. However, in this new model a guard channel allocation scheme and buffering for upward vertical handover calls have been used to practically solve the infinite loop in the system. The cellular network has a larger coverage area than the WLAN. As shown in Figure 7.1 in Chapter 7 the WLAN is deployed inside a single cellular network. The model covers handover, various queue capacities and mobility issues (e.g. speed of the users) as well. The system model is shown in Figure 9.1. As many symbols are used in the analysis, Table 9.1 summarises and recalls the important ones with extra parameters used for further analysis.

Table 9.1: Summary of symbols

Symbol	Definition
S_c	Number of channels for originating and horizontal handover calls in the cell
S_g	Number of guard channels in the cell
W_c	Queue capacity for originating and horizontal handover calls in the cell
W_g	Queue capacity for upward vertical handover calls in cell
W_w	Queue capacity of the WLAN
L_c	Maximum number of originating and horizontal handover calls in the cellular system
L_w	Maximum number of calls in the WLAN
θ_1	Probability of a call to be transferred from the cell to the WLAN
θ_2	Probability of a call to be transferred from the WLAN to the cell
θ_3	Probability of a call to be transferred from the cell buffer to the WLAN
$\lambda_{oc,c}$	Mean arrival rate of originating calls to the cell
$\lambda_{hh,c}$	Mean arrival rate of horizontal handover calls between neighbouring cells
λ_c	Total arrival rate of originating and horizontal handover calls in the cell
$\lambda_{vh,w}$	Mean arrival rate of upward vertical handover call from WLAN to the cell
$\lambda_{oc,w}$	Mean arrival rate of originating calls to the WLAN
$\lambda_{vh,c}$	Mean arrival rate of downward vertical handover call from cell to the WLAN
$\mu_{oc,c}$	Mean service rate of originating calls in the cell
$\mu_{cd,c}$	Mean service rate of horizontal handover calls in the cell
$\mu_{cd,w}$	Mean service rate of vertical handover calls in the cell
$\mu_{oc,w}$	Mean service rate of originating calls in the WLAN
μ_{wd}	Mean service rate of mobile calls in the WLAN
μ_g	Mean service rate of calls for the guard channel
μ_c	Total service rate of completed call departures in the cell
μ_w	Total service rate of completed call departures in the WLAN
γ_c	Throughput of the cell (without guard channel)
γ_w	Throughput of the WLAN

The system consists of S identical channels in the first stage, numbered 1,2,3,…,S where S_g channels are reserved as guard channels and the rest S_c=S-S_g is shared by originating and horizontal handover calls. These calls are assumed to arrive independently. When there are no channels available in the system, the calls start to queue up with buffer size W_c ($W_c \geq S_c$). However, no more than L_c=S_c+W_c calls are allowed to make horizontal handover and/or to be generated in the cell. On the other hand, upward vertical handover calls use the reserved channels (S_g) for the service. W_g is the buffer capacity of the upward vertical handover calls when there are no channels available. No more than L_g=S_g+W_g calls are allowed to perform upward vertical handover in the cell. On the other hand, WLAN has one channel as explained in previous chapters with buffer capacity W_w.

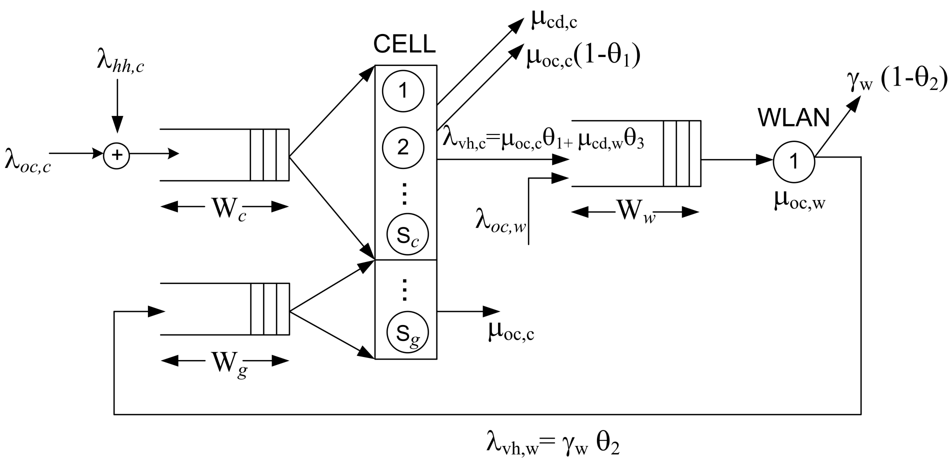

Figure 9.1: The two stage open queuing system considered with guard channel and buffering

The three types of vertical handover decisions probabilities considered in previous chapters (θ_1,θ_2,θ_3) play an important role in determining the operational space in this chapter. Hence, the details of vertical handover decisions probabilities can be found in Section 6.2.1.1. θ_1 and (1-θ_1) are the probability of a call to be transferred from the cell to the WLAN and probability of a call to leave the cell after completion, respectively. In addition, if the channels in the cellular network are busy, calls are placed in a queue, then these calls may be transferred to the WLAN due to the mobility of a caller with probability of θ_3. Users can stay in the cell until they get a channel to be served and (1-θ_3) is the probability of a call to stay. On the other hand, the WLAN users may attempt vertical handover back to the cell due to mobility with a probability of θ_2. The WLAN users are served to completion with a probability of (1-θ_2) at the second stage. γ_w is the throughput of the WLAN. In other words, the calls served in the WLAN leave the system with a probability of (1-θ_2) where μ_w=γ_w(1-θ_2). In addition, upward vertical handover from WLAN to cellular system takes place with a probability of θ_2 where $\lambda_{vh,w} = \gamma_w\ \theta_2$. Similarly the arrival, service and departures rates used in Chapters 7 and 8 are used in this chapter considering the guard channels and buffering for further analysis. It must be noted that, the total service rate of completed call departures from a WLAN, μ_w can be obtain unlike the previous chapters where μ_w=γ_w(1-θ_2).

The following assumptions have been made in order to further clarify the analysis. These assumptions are taken from the literature and are reasonable for the analytical results obtained to remain as accurate as shown in previous chapters. When a vertical handover occurs, the velocity of the mobile node is based on the initial velocity that is predetermined for the new network. In addition, the mobile velocity does not change during its call lifetime. In reality, mobiles do not move with constant speeds. A speed change occurs when a mobile moves from a more crowded area to a less crowded area or if a mobile encounters a traffic signal. However, usually a speed change is small considering the fact that a call normally lasts a short period of time (Liu et al., 2006), (Zeng and Agrawal, 2001), (Zeng and Agrawal, 2002). Similarly to previous analysis, the cellular network has multiple identical channels and FCA scheme is considered. The proposed system considered is a system with homogeneous traffic loads in statistical equilibrium state.

9.3 Modelling Integrated Cellular/WLAN Network with Guard Channels

The cellular/WLAN interaction can be considered in two dimensions when the guard channel and buffering implementation is isolated. Please note that once a request is transferred to cellular system from WLAN, it is allocated to a guard channel and guard channels also have allocated queuing mechanisms. These calls should not be transferred back to the WLAN. Therefore, the guard channel system can be handled in isolation from two stage open networks since the main interaction with the system depends on the throughput of the WLAN only.

The state of the system at time t can be described by a pair of integer valued state variables, $I(t)$ and $J(t)$, specifying the number of calls present at a time t for the cellular and the WLAN, respectively. The model assumes that $I(t)$=0,1,...,L_c, $t \geq 0$, is an irreducible Markov process, which represents the number of calls in the cellular network including the one(s) in service. $J(t)$=0,1,...,L_w, is the total number of calls in the WLAN system at time t, including the one(s) in service. Then, $Z=[I(t),J(t)]$; $t \geq 0$, is an irreducible Markov process on a lattice strip (a QBD process), that models the system. Its state space is, $\{0,1,\ldots,L_c\}$ x $\{0,1,\ldots,L_w\}$. The irreducible Markov process is used for performance evaluation of the two stage open network considered. Let the number of calls at the cellular network *I(t)*, be represented in the horizontal direction and possible number of calls at the WLAN, *J(t)*, be represented in the vertical direction of a lattice strip. Here, A is the matrix of instantaneous transition rates from state (i,j) to state (k,j) with zeros on the main diagonal. These are the purely lateral transitions of the model Z. Matrices B and C are transition matrices for one-step upward and one-step downward transitions respectively. The transition rate matrices do not depend on j for $j \geq M$, where M is a threshold having an integer value. However, in the case of the cellular networks the transition rate matrices always depend on j, since the calls in waiting room may depart the system due to mobility as shown in Figure 9.2.

The state diagram clearly shows the calls handed over from WLAN to cellular network are not allowed back to the WLAN. In other words, there is no transitions from *(i,j)* to $(i+1,j-1)$. This is done to solve the feedbacks during the vertical handover procedure in the integrated cellular/WLAN system. Clearly, the elements of A depend on the parameters S_c, W_c, λ_c, $\mu_{cd,c}$ and μ_c. The transition matrices of a system with S_c channels are of size (L_c+1) x (L_c+1). The state transition matrices A, A_j, B and B_j can be given as follows;

$$A = A_j = \begin{pmatrix} 0 & \lambda_c & 0 & 0 & 0 & 0 & 0 & 0 \\ \mu_c & 0 & \lambda_c & 0 & 0 & 0 & 0 & 0 \\ 0 & 2\mu_c & 0 & \ddots & 0 & 0 & 0 & 0 \\ 0 & 0 & \ddots & 0 & \lambda_c & 0 & 0 & 0 \\ 0 & 0 & 0 & S_c\mu_c & 0 & \lambda_c & 0 & 0 \\ 0 & 0 & 0 & 0 & S_c\mu_c + \mu_{cd,c} & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & \ddots & 0 & \lambda_c \\ 0 & 0 & 0 & 0 & 0 & 0 & S_c\mu_c + W_c\mu_{cd,c} & 0 \end{pmatrix}$$

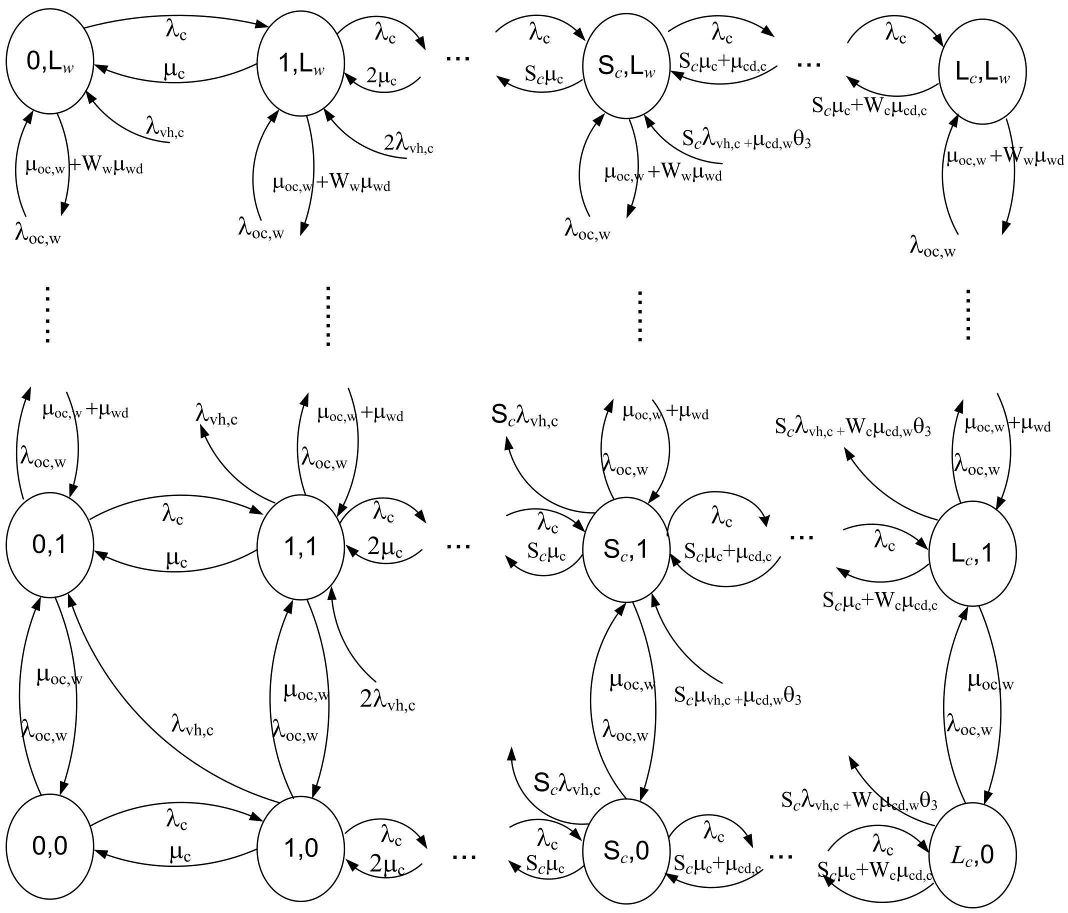

Figure 9.2: The stage diagram queuing system considered for downward vertical handover

$$B = B_j = \begin{pmatrix} \lambda_{oc,w} & 0 & 0 & 0 & 0 & 0 & 0 \\ \lambda_{vh,c} & \ddots & 0 & 0 & 0 & 0 & 0 \\ 0 & \ddots & \lambda_{oc,w} & 0 & 0 & 0 & 0 \\ 0 & 0 & S_c\lambda_{vh,c} & \lambda_{oc,w} & 0 & 0 & 0 \\ 0 & 0 & 0 & S_c\lambda_{vh,c}+\mu_{cd,w}\theta_3 & \ddots & 0 & 0 \\ 0 & 0 & 0 & 0 & \ddots & \lambda_{oc,w} & 0 \\ 0 & 0 & 0 & 0 & 0 & S_c\lambda_{vh,c}+W_c\mu_{cd,w}\theta_3 & \lambda_{oc,w} \end{pmatrix}$$

The elements of matrices C and C_j depend on the parameters μ_{wd}, $\mu_{oc,w}$ and W_w. The matrix C is dependent on the number of calls for j=0, 1,…,L_w. Therefore, the threshold M is taken as M=L_w. If the number of calls in the system is less than the number of available channels, a channel is assigned for each call.

Therefore, the downward transition rate is chosen as the minimum of number of calls and number of available channels. On the other hand, if the number of calls is greater than the number of available channels, all of the available channels are assigned to incoming calls and the calls in the queue have the service rates as μ_{wd} (Liu et al., 2006), (Zeng and Agrawal, 2001), (Zeng and Agrawal, 2002).

The matrix C is defined below: (j=M)

$$C = \begin{pmatrix} \mu_{oc,w} + (W_w \mu_{wd}) & 0 & 0 & 0 \\ 0 & \mu_{oc,w} + (W_w \mu_{wd}) & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \mu_{oc,w} + (W_w \mu_{wd}) \end{pmatrix}$$

The matrix C_j is defined below for two different regions as explained above: for j ≤ 1

$$C_j = \begin{pmatrix} \mu_{oc,w} & 0 & 0 & 0 \\ 0 & \mu_{oc,w} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \mu_{oc,w} \end{pmatrix}$$

for $1 <$ j $\leq L_w$

$$C_j = \begin{pmatrix} \mu_{oc,w} + (j-1)\mu_{wd} & 0 & 0 & 0 \\ 0 & \mu_{oc,w} + (j-1)\mu_{wd} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \mu_{oc,w} + (j-1)\mu_{wd} \end{pmatrix}$$

The spectral expansion solution is employed for the steady state solution. The balance equations defined for (M-1) $\leq$ j $\leq$ (L_w-2) are not used since the threshold value is defined as M=L_w. In addition both systems are considered as bounded queues for the solution (i.e. $S_c \leq L_c$ and $1 \leq L_c$). A number of steady-state performance measures can be computed after the steady state probabilities, $P_{i,j}$, obtained. The mean queue length (MQL_c/MQL_w), blocking probability ($P_{B,c}/P_{B,w}$) and throughput (γ_c/γ_w) of both cellular and WLAN networks are computed respectively which can be obtained as follows:

$$MQL_c = \sum_{i=0}^{L_c} i \sum_{j=0}^{L_w} P_{i,j} \quad and \quad MQL_w = \sum_{j=0}^{L_w} j \sum_{i=0}^{L_c} P_{i,j} \tag{9.1}$$

$$P_{B,c} = \sum_{i=0}^{L_c} P_{i,L_w} \quad and \quad P_{B,w} = \sum_{j=0}^{L_w} P_{L_c,j} \tag{9.2}$$

$$\gamma_c = \sum_{i=0}^{L_c} \mu_i \sum_{j=0}^{L_w} P_{i,j} \quad and \quad \gamma_w = \sum_{j=0}^{L_w} \mu_j \sum_{i=0}^{L_c} P_{i,j} \tag{9.3}$$

Once the throughput of the WLAN is obtained, the downward vertical handover can be obtained accordingly. The guard channel system can be considered as one dimensional M/M/S_g/L_g system where L_g=S_g+W_g as shown in Figure 9.3.

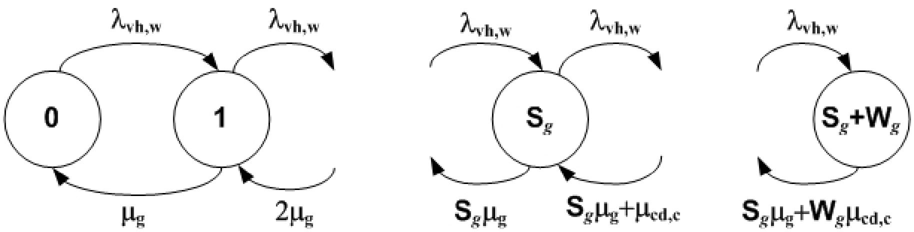

Figure 9.3: The stage diagram of the guard channel system considered for upward vertical handover

$\lambda_{vh,w}$ is the total arrival rate of calls allocated to the guard channels in the cellular system where $\lambda_{vh,w} = \gamma_w \ \theta_2$. As explained in the Chapters 2 and 6, E[$T_{oc,c}$] is the channel holding time of calls in the cellular system. However, the calls handed over from the WLAN to the cellular system will not hand back over to the WLAN. Thus the service rate of the upward vertical handover calls exponentially distributed with mean E[T_g]= $1/\mu_g$ where $\mu_g = \mu_{oc,c} + \mu_{cd,c}$. Please note that since there will be no vertical handover back to WLAN for the calls being served from the WLAN hence (1-θ_1) is not considered. Let's define the states i (i=0, 1, 2,. . .,$S_g + W_g$) as the number of calls in the system at time t. Assuming a system in steady state, the state probabilities, P_i can be obtained as follows where $\rho = \lambda_{vh,w}/\mu_g$.

$$P_i = \begin{cases} \frac{\rho^i}{i!} \cdot P_0 & 0 \le i \le S_g \\ \frac{\frac{\rho^{S_g}}{S_g!} \cdot \lambda_{vh,w}^{i-S_g} \cdot P_0}{\prod\limits_{j=S_g+1}^{i} \left[S_g\mu_g + (j-S_g)\mu_{cd,c}\right]} & S_g < i \le S_g + W_g \end{cases} \tag{9.4}$$

In above equation, P_i is the probability that there are i calls in the system and P_0 can be defined as follows:

$$P_0 = \left[\sum_{i=0}^{S_g} \frac{\rho^i}{i!} + \sum_{i=S_g+1}^{S_g+W_g} \frac{\frac{\rho^{S_g}}{S_g!} \cdot \lambda_{vh,w}^{i-S_g}}{\prod\limits_{j=S_g+1}^{i} \left[S_g\mu_g + (j-S_g)\,\mu_{cd,c}\right]} \right]^{-1} \tag{9.5}$$

The average number of calls in the system, MQL_g can then be calculated as $MQL_g = \sum\limits_{i=0}^{Sg+W_g} i \cdot P_i$ which gives:

$$MQL_g = \left[\sum_{i=0}^{S_g} i\frac{\rho^i}{i!} + \sum_{i=S_g+1}^{S_g+W_g} \frac{i\frac{\rho^{S_g}}{S_g!} \cdot \lambda_{vh,w}^{i-S_g}}{\prod\limits_{j=S_g+1}^{i} \left[S_g\mu_g + (j-S_g)\,\mu_{cd,c}\right]} \right] \cdot P_0 \tag{9.6}$$

Similarly, the blocking probability $P_{B,g}$ can be calculated as:

$$P_{B,g} = \frac{\frac{\rho^i}{S_g!} \cdot \lambda_{vh,w}^{W_g} \cdot P_0}{\prod\limits_{j=i+1}^{i+W_g} \left[i\mu_g + (j-i)\,\mu_{cd,c}\right]} \tag{9.7}$$

Similarly in Chapters 6 and 7, the solution of the systems of linear equations could be also used in this chapter. Table 9.2 shows the MQL and throughput results as a function of the originating call in the cell for both solution approaches. The results obtained using spectral expansion and system of balance equations, show very good agreement. The differences are shown in Table 9.2. Hence, the balance equations used to obtain the steady state probabilities using the solution of the systems of linear equations can be found in Appendix A.2. The further details can be found in Chapter 6.

Table 9.2: MQL and throughput results for both solution approaches. (D is the differences)

	MQL_c			γ_c			MQL_w			γ_w		
$\lambda_{oc,c}$	Spectral	Linear	D(%)	Spectral	Linear	D(%)	Spectral	Linear	D(%)	Spectral	Linear	D(%)
0.02	2.884	2.884	0	0.07	0.07	0	1.575	1.576	0.002	0.058	0.059	0.429
0.04	4.435	4.4352	0	0.108	0.108	0	1.797	1.797	0	0.062	0.062	0.753
0.06	5.496	5.496	0	0.133	0.133	0	1.958	1.958	0	0.064	0.065	1.01
0.08	6.503	6.503	0	0.156	0.156	0	2.113	2.113	0	0.066	0.067	1.224
0.1	7.546	7.546	0	0.177	0.177	0	2.268	2.268	0	0.068	0.069	1.532
0.12	8.672	8.672	0	0.198	0.198	0	2.425	2.425	0	0.07	0.071	1.885
0.14	9.893	9.891	0.013	0.218	0.218	0	2.575	2.579	0.1538	0.071	0.073	2.193
0.16	11.171	11.17	0.011	0.236	0.236	0	2.729	2.726	0.096	0.073	0.075	2.424
0.18	12.436	12.434	0.01	0.252	0.252	0	2.861	2.859	0.084	0.074	0.076	2.602
0.2	13.612	13.603	0.065	0.265	0.265	0	2.966	2.972	0.207	0.075	0.077	2.737

9.4 Numerical Results and Discussion

In this section, the numerical results of the proposed system with guard channel reservation and buffering for an integrated cellular/WLAN system are presented to make valid comparisons, the assumptions in Chapters 7 and 8 are employed in this section as well, unless stated otherwise.

9.4.1 Key Parameters

The number of channels allocated to the cellular network and WLAN are S=12 and K=1 respectively (Hasib and Fapojuwo, 2007), (Hasib and Fapojuwo, 2008), (Liu et al., 2006), (Ma et al., 2001), (Madan et al., 2008), (Shensheng and Wei, 2005), (Song et al., 2008a), (Xia and Shen, 2007), (Zeng and Agrawal, 2002). In the cellular network, the number of guard channels which are reserved only for upward vertical handover calls are 2. In other words, S_c=10 and S_g=2 thus S=S_c+S_g=12. The cell can accommodate an upward vertical handover call from

WLAN if there are idle guard channels available and/or the queue of the cell system is not full. The addition of guard channels and new buffering avoid the infinite loops (mentioned in the previous sections) in the integration of cellular/WLAN system. The proposed model differentiates the originating and horizontal handover calls in the cellular system from the upward vertical handover calls from WLAN. Therefore the calls handed over from WLAN to the cellular network are not allowed back to the WLAN. The system parameters used are mainly taken from Hasib and Fapojuwo (2007), Hasib and Fapojuwo (2008), Liu et al. (2006), Madan et al. (2008) and Zeng and Agrawal (2002) based on the relevant literature (Gazis et al., 2005), (Khatib, 2003), (Ma et al., 2001), (Rutagemwa, 2007), (Saravanan et al., 2006), (Song et al., 2008a), (Song et al., 2008b), (Song et al., 2005), (Song and Zhuang, 2005), (Stemm and Katz, 1998), (Trivedi et al., 2002), (Xia and Shen, 2009), (Xia and Shen, 2007), (Zeng and Agrawal, 2001), (Zeng et al., 1994b). The threshold value of call blocking probability is defined as $P_{B_{th}}$. $P_{B_{th}}$ is the probability that a call is blocked. In various studies the main target of network, communication system engineering, optimisation studies is specified as building as high quality networks as possible (Mishra, 2006). In order to achieve this, the success rate of handovers should be greater than 95% (Mishra, 2006). Therefore the threshold value of call blocking probability $P_{B_{th}}$ is expected to be less than 0.05 in order to obtain high QoS identified in previous studies (Mishra, 2006). In this chapter the operational area is presented for various θ_1, θ_2 and θ_3 values. The figures present values these parameters may have to achieve a threshold value of call blocking probability $P_{B_{th}}$ of 5% or less. In other words, in order to obtain high performance systems, all values of the blocking probabilities are taken less than the given threshold value ($P_{B_{th}}$) in this chapter.

9.4.2 Results

Figure 9.4 shows possible downward vertical handover probabilities for different originating calls in the cellular system. The parameters are $E_c[v] = 40km/h$, E_w[v]=2km/h, R=1000m, r=100m, W_c=W_w=W_g = 10, μ_c=0.0221, μ_w=0.066, $\lambda_{oc,w}$=0.05, and the originating call arrival rates per user for cellular system varies from 0.01 to 0.05 calls per second. As mentioned above, the results are illustrated considering $P_{B_{th}}$.

In the proposed model, θ_2 does not affect θ_1 and θ_3 in the cellular system since θ_2 is only associated with guard channels in the cellular system. However, θ_1 and θ_3 values have an impact on the WLAN system and indirectly affect the θ_2 values. It is clear from the figure that θ_1 and θ_3 affect the system performance significantly. θ_1 and θ_3 are important parameters for the decision management of the downward vertical handover from the cellular system to the WLAN.

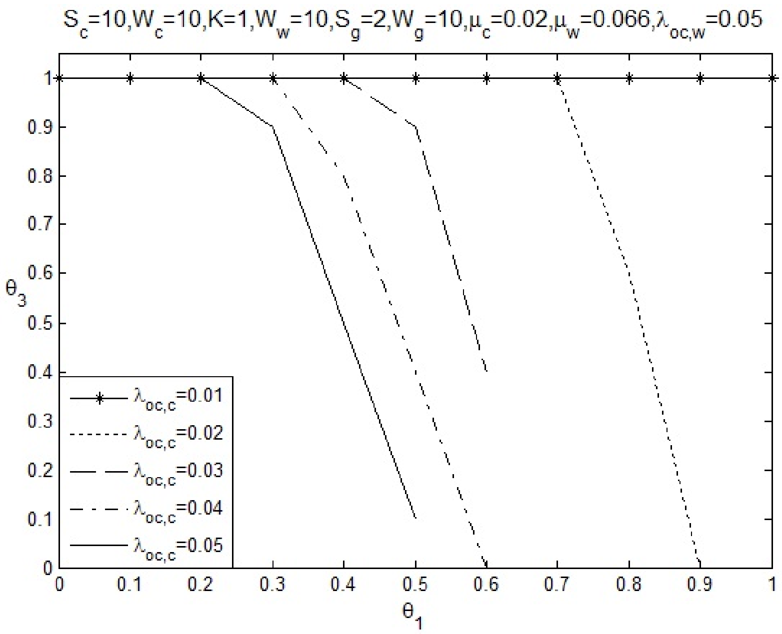

Figure 9.4: Downward vertical handover probabilities for different originating calls in the cell, $W_c = 10$

Figure 9.4 shows that, for lighter traffic loads e.g $\lambda_{oc,c}$=0.01 the users can do downward vertical handover regardless of θ_1 and θ_3 values (it is a straight line in the figure which means that θ_1 and θ_3 can get any values between 0 and 1 where $P_{B_{th}}$ of WLAN < 0.05). When the traffic load (originating calls) increases in the cellular system, it affects the WLAN system performance significantly. For instance when $\lambda_{oc,c}$=0.02, the system can perform downward vertical handover up to θ_1=0.7 regardless of θ_3. However, because of the traffic loads in the system and mobility θ_3 values decreases accordingly. In addition, when $\lambda_{oc,c}$=0.05 it is more difficult to perform downward vertical handover depending on the θ_1 and

θ_3 values. The possible downward vertical handover probabilities for different originating calls in the cellular system with a different buffer capacity W_c=20 is shown in Figure 9.5. All the parameters used in Figure 9.4 are also used in 9.5, however buffer capacity of cell is increased from 10 to 20. It is clear from the Figure 9.5 that increasing the buffer capacity of the cellular system affects the downward vertical handover decision policy considerably. The cellular network can accommodate increased number of calls in the buffer hence this causes increased number of calls to hand over to the WLAN affecting the blocking probability of the WLAN depending on θ_3 values.

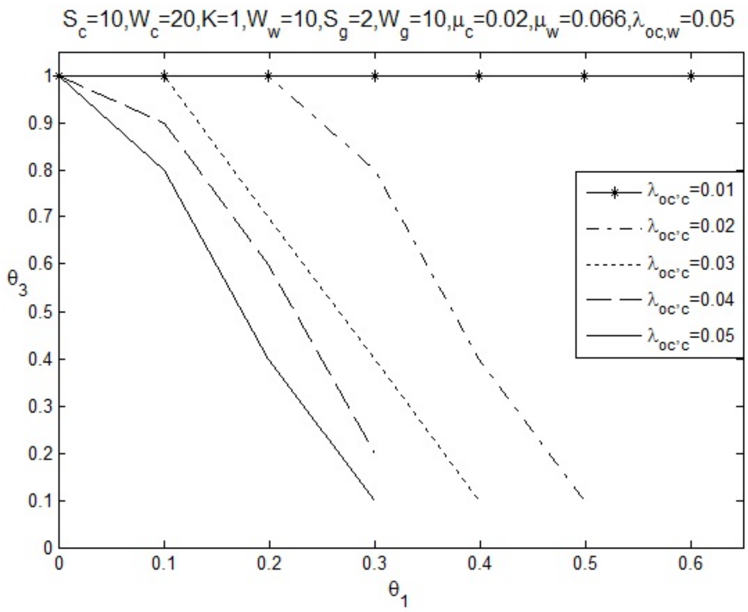

Figure 9.5: Downward vertical handover probabilities for different originating calls in the cell, $W_c = 20$

Figure 9.6 shows the effects of the buffering in the cellular system on the downward vertical handover in an integrated cellular/WLAN system. The results shown in Figure 9.6 depict how important the cellular buffering (W_c) is for downward vertical handover decision. For a busy and larger capacity system, the calls can join the system and wait to be served. However, θ_3 plays an important role in determining the downward vertical handover decision due to the mobility in the queue. In other words, the WLAN can become full quite rapidly. Increasing the buffer capacity in the cellular system e.g. W_c=40 increases the mean number of calls waiting in the queue to hand over to the WLAN without being served. This causes more calls to hand over to the WLAN. However, depending on the WLAN buffer capacity, the downward vertical handover calls may be blocked accordingly in the integrated cellular/WLAN system.

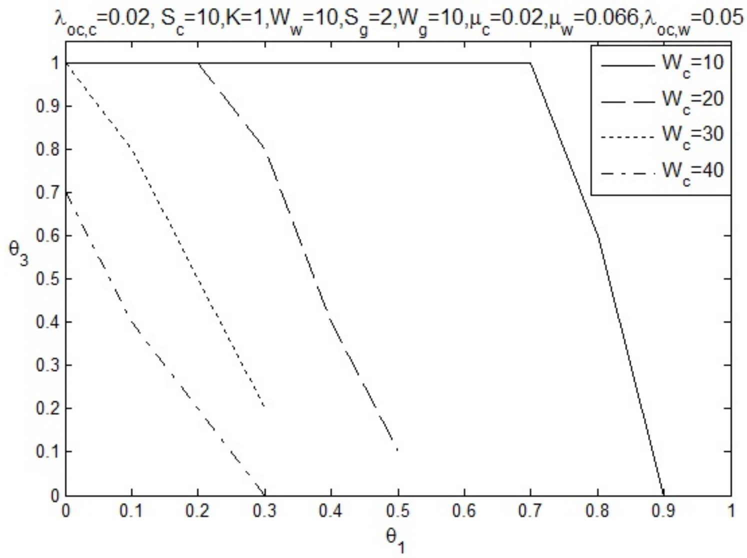

Figure 9.6: Downward vertical handover probabilities for different W_c

In Figures 9.7, 9.8 and 9.9 the effects of buffering in the WLAN are shown. These results are very important in management of the seamless downward vertical handover as well as in determining the overall system performance. The possible downward vertical handover probabilities for different queue capacities of WLAN are shown in Figure 9.7. When the higher buffer capacities are considered in the WLAN (e.g W_w=30), the downward vertical handover is possible almost regardless of values of θ_1 and θ_3. When the buffer capacity of the WLAN is small (e.g W_w=10) the calls being handed over from the cell will be blocked with a higher probability. In order to make seamless downward vertical handover possible with acceptable QoS, the buffer capacity of the WLAN should be sufficient. In other words, when θ_1 and θ_3 increase, the traffic load increases inside the WLAN due to the downward vertical handover. This causes increased number of calls in the WLAN. Thus, downward vertical handover becomes totally dependent on the value of W_w. The larger the buffer capacity, the better the downward vertical handover results.

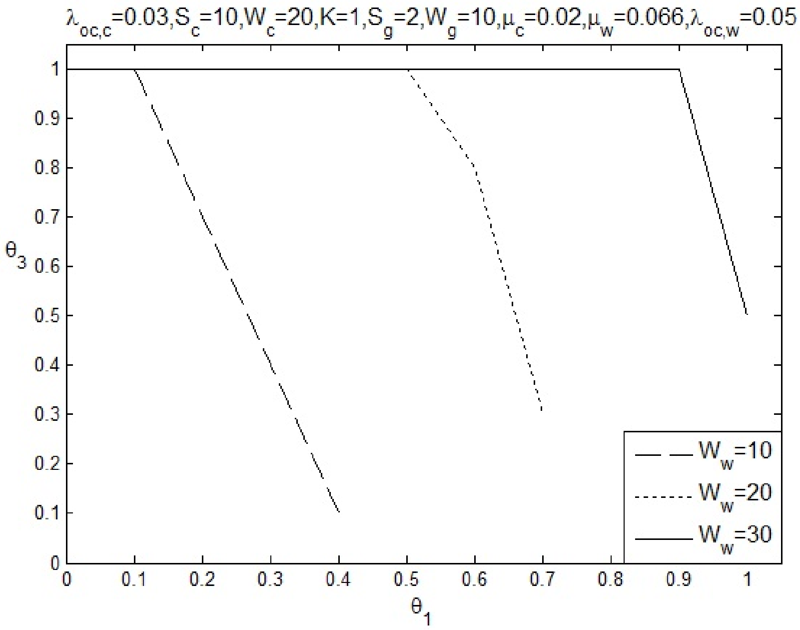

Figure 9.7: Downward vertical handover probabilities for different W_w

In addition, the results of possible downward vertical handover probabilities for different originating calls in the cell are shown in Figure 9.8. The parameters used in Figure 9.6, are used in Figure 9.8 as well for a fair analysis. Comparing the results in both figures clearly shows that systems with W_w=30 (Figure 9.8 results for $\lambda_{oc,c}$=0.02 and W_c=20) performs better than systems with W_w=10 (Figure 9.6 results for W_c=20). Observe that large buffer capacity in the WLAN can support a higher load of downward vertical handover calls. Moreover, possible downward vertical handover probabilities for different originating calls with large buffering capacities in the cell (e.g. $W_c = 30$) as well as in the WLAN (e.g. $W_w = 30$) are considered in Figure 9.9. It can be observed from the figure that, with larger buffer capacity of the WLAN can accommodate downward vertical handover calls from the cell. In other words, the WLAN buffering is the main limiting factor for the WLAN.

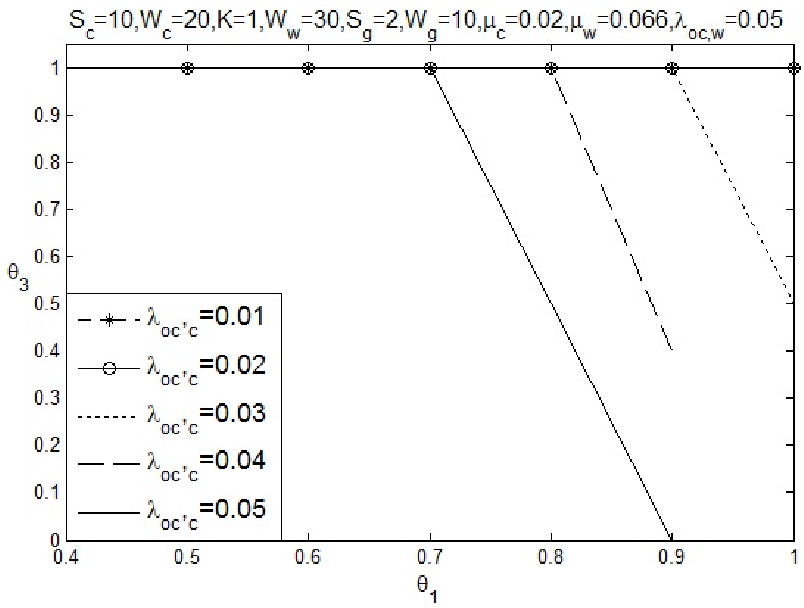

Figure 9.8: Downward vertical handover probabilities for different originating calls in the cell,W_c=20

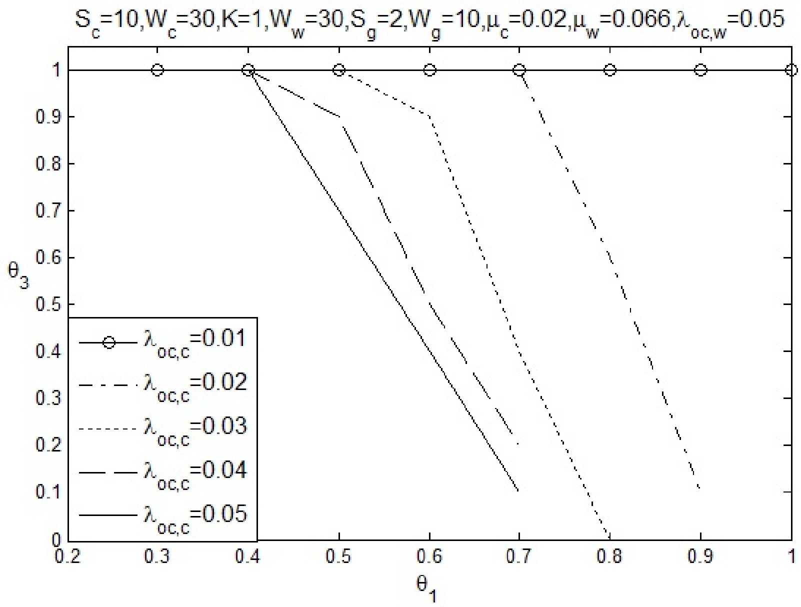

Figure 9.9: Downward vertical handover probabilities for different originating calls in the cell,W_c=30

In Figure 9.10 possible downward vertical handover probabilities for different service times of the cell are shown. It can be clearly seen from this figure that possible downward vertical handover probabilities and service rates of the cell should be carefully considered from the design and modelling point of view. The calls in the cellular system increase when the system has a longer service time. This causes increased number of calls in the queue of the cellular system.

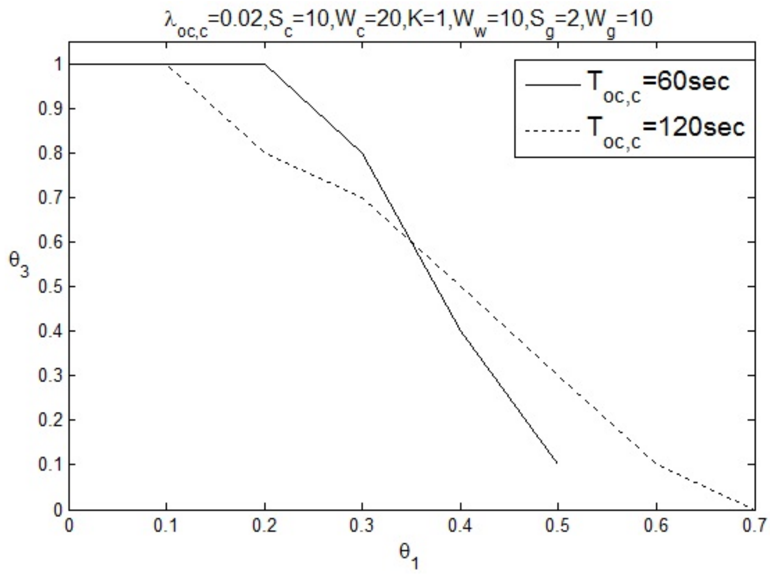

Figure 9.10: Downward vertical handover probabilities for different service times of the cell

These calls may be transferred to the WLAN due to the mobility with a probability of θ_3. However, it is obvious from the results that increasing θ_3 values with longer service time leads the degradation of the performance of downward vertical handover with the corresponding θ_3 values. However, users can stay in the system until they get a channel to be served. This can be achieved by obtaining smaller θ_3 values which will give the probability of users to stay connected the cellular system. They can wait and get a channel to be served. That is the reason that θ_1 can gets higher values when service time is longer as shown in Figure 9.10. When the service time T_c is 60sec, the calls in the cell will be serviced quicker compared to the case when T_c is 120sec. Thus, the calls are rapidly serviced by the cellular network hence the calls can be transferred to the WLAN after the completion faster. As the probability of having a vertical handover from the cellular system to the WLAN increase due to the mobility in the queue (θ_3), θ_1 decreases accordingly. This is mainly because of the higher service rate provided by the cellular system.

In this section, the focus has mainly been given to the downward vertical handover. In order to perform a seamless downward vertical handover in the integrated cellular/WLAN system, the results show that, θ_1, θ_3, W_c, W_w as well as service rate of the cellular system, play a vital role to enhance the QoS. θ_1 is the possible downward vertical handover decision probability which is considered when transferring calls from the cellular network to the WLAN after the completion of call. In addition, θ_3 is the possible downward vertical handover decision probability when the users get into the coverage area of the WLAN due to the mobility in the queue without being served by the cellular system. Thus, both downward vertical handover decision probabilities (θ_1 and θ_3) should be considered and planned together. These two parameters are mutually dependent. The results show that W_c and W_w are also important parameters in the integrated systems. As the downward vertical handover probability increases, the blocking probability of the vertical handover calls increases dramatically. This is mainly because of the small buffer capacity provided by the WLAN. Increasing the buffer capacity of the cellular system with a small buffer capacity of the WLAN causes the worst performance among the cases considered. When the buffer capacity of the WLAN increases, the WLAN becomes able to deal with the downward vertical handover calls. The designers and researchers have to take these parameters into account to drive and/or construct practical systems. In this context, the guard channel model presented in this chapter is also accurate and the computationally efficient.

In addition, the other important handover decision probability θ_2, has been investigated in the following part for the operational area as well as sensitivity analysis. Figures 9.11 and 9.12 show possible upward vertical handover probabilities for different θ_1 and θ_3 values in the integrated cellular/WLAN system respectively. The parameters are S_c=10, W_c=W_w=20, K=1, S_g=2, W_g=10, $E_c[v] = 40km/h$, E_w[v]=2km/h, R=1000m, r=100m, $\mu_{oc,c}$=0.0166, $\mu_{oc,w}$=0.0833, $\lambda_{oc,c}$=0.03 and $\lambda_{oc,w}$=0.05 calls per second. As mentioned above, the results are also illustrated for $P_{B_{th}}$. The calls can be transferred only to the cellular guard channel system with a probability of θ_2. As shown in the figures θ_1 and θ_3 values directly affect the upward vertical handovers in the system. This is mainly because the downward vertical handover calls switch from the cellular network to the WLAN. θ_1 and θ_3 increase the number of calls in the WLAN and the throughput in the WLAN increases accordingly. Thus, θ_1 and θ_3 values have an impact on the WLAN system and indirectly affect the upward vertical calls in the system. In other words, in practical mobile networks, after successfully performing a downward vertical handover to the WLAN, the call may be completed in the WLAN or try to perform an upward vertical handover to the cellular system. Figure 9.10 shows the mobile users in the queue of the cellular system affect the upward vertical handover significantly when θ_1 increases. This is mainly because an increased number of calls switch to the WLAN, hence the calls can perform an upward vertical handover to the cellular network depending on the values of θ_2.

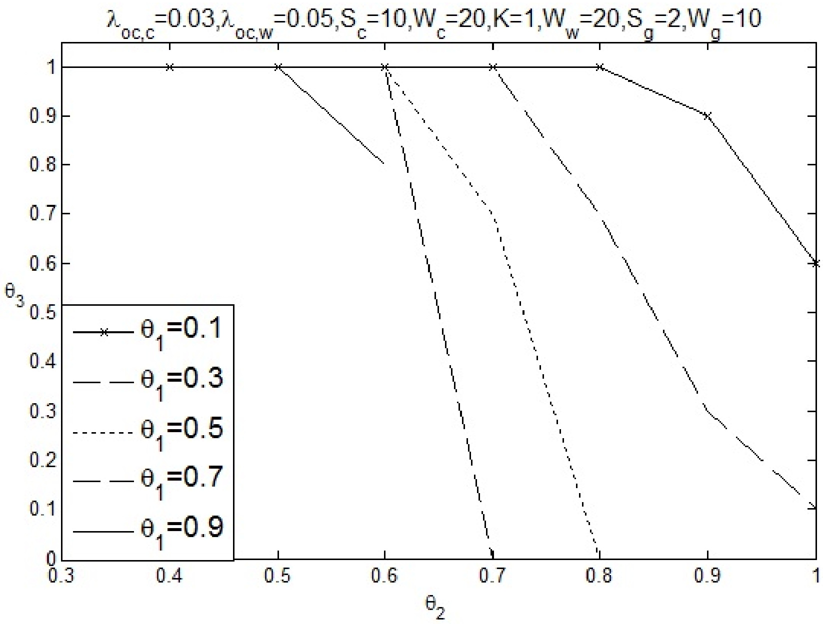

Figure 9.11: Upward vertical handover probabilities vs θ_3 for different θ_1

The number of upward vertical handover calls increases when the θ_2 increases as expected. However, Figures 9.11 and 9.12 show that θ_3 values make the cellular system become unstable more rapidly compared to the θ_1 values. This is mainly because of the mobility in the queue and the buffer capacity in the cellular system.

The mobile users in the cellular system can perform downward and/or upward vertical handover depending on the θ_1, θ_3 and θ_2 values respectively. Thus the system sensitivity increases. However, for different θ_3 values the calls transferred from the cell to the WLAN affects the system stability more gracefully as shown in Figure 9.12.

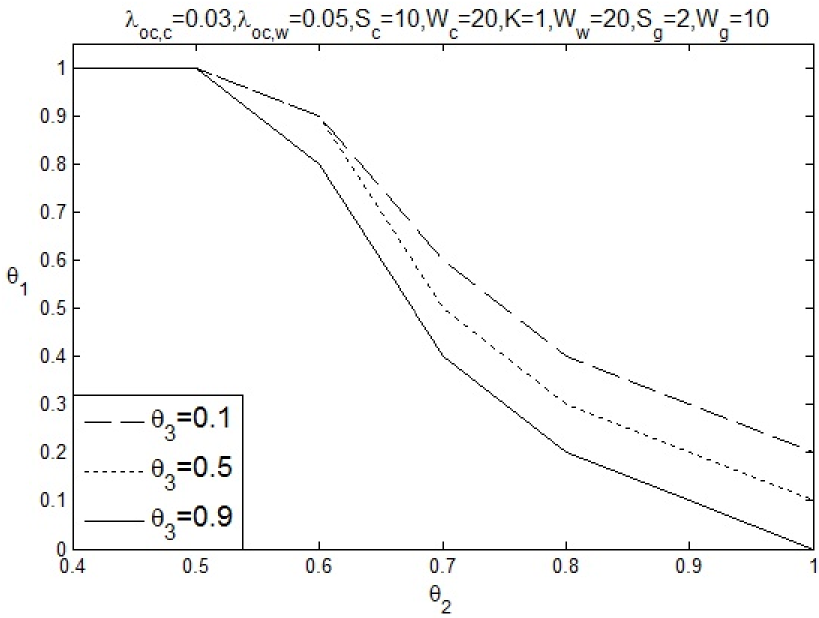

Figure 9.12: Upward vertical handover probability vs θ_1 for different θ_3

In addition, Figure 9.13 shows upward vertical handover rate as a function of θ_2 for different θ_1 values. The parameters used in Figure 9.11 and 9.12 are used in Figure 9.13 as well. It is clearly shown in the Figure 9.13 that the upward vertical handover rates increase as expected when θ_2 increases. Increasing θ_1 and θ_2 leads to increased amount of traffic in the cellular system. However, it is appropriate to reiterate that θ_2 is only associated with the guard channels in the cellular system. Thus, the mean queue length (MQL_g) of the guard channels system is shown as a function of θ_2 for different originating calls in the WLAN in Figure 9.14. The MQL_g increases in the cellular guard channel system as the upward vertical handover flow increases. This is due to calls transferred to the cellular system when the upward vertical handover probability increases.

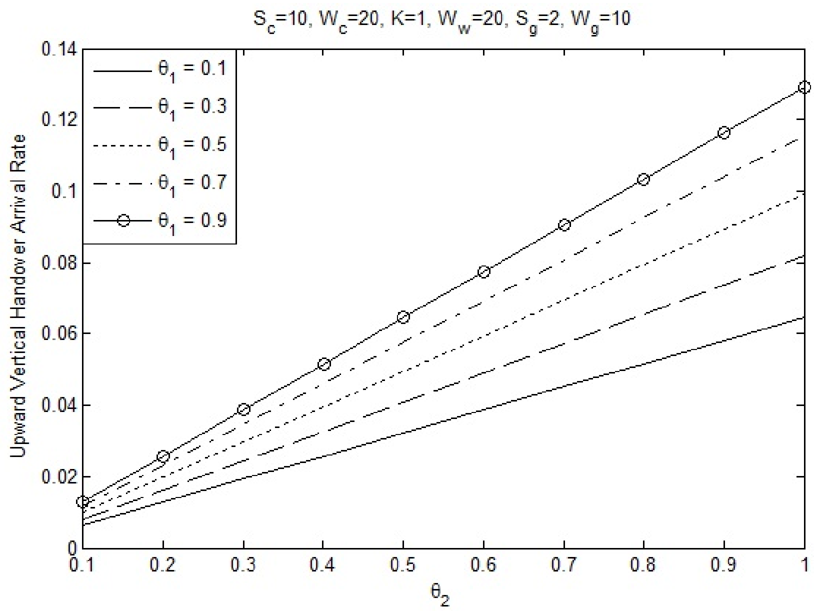

Figure 9.13: Upward vertical handover rate as a function of θ_2 for different θ_1 values

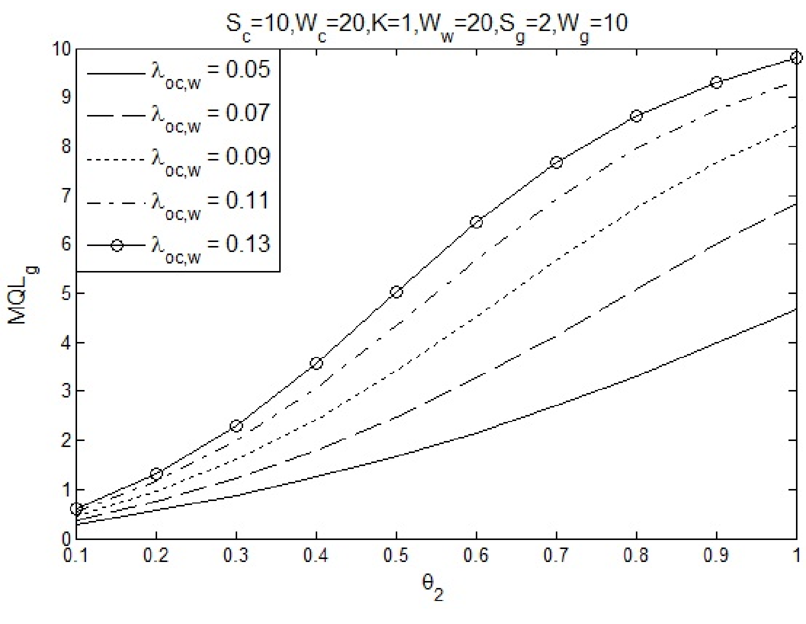

Figure 9.14: The MQL_g results as a function of θ_2 for different originating calls in the WLAN

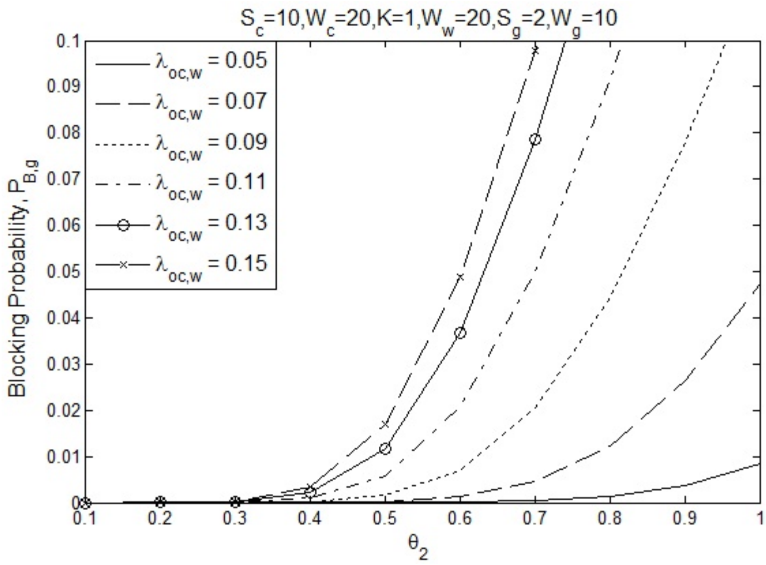

Figure 9.15: The blocking probability results as a function of θ_2 for different originating calls in the WLAN

Furthermore, the blocking probability in the guard channels increases accordingly due to the upward vertical handover flow. This is shown in Figure 9.15 as a function of θ_2 for different originating calls in the WLAN. Blocking probability of the guard channel system is presented in Figure 9.16 as a function of originating calls in the WLAN for different θ_2. It is clearly shown in Figure 9.16 that, similar to results presented in Figure 9.15, the blocking probability of the guard channel system increases when the traffic increases in the WLAN. In addition, the effects of θ_2 values are also shown in the Figure 9.16. The figures show that θ_2 is the main factor to make the decision on the upward vertical handover calls in integration of cellular/WLAN systems.

Finally, the blocking probability of guard channels ($P_{B,g}$) is presented as a function of upward vertical handover rate for different number of guard channels in the cellular system. These results are shown in Figure 9.17. The results show that increasing the number of guard channels in the cellular networks decreases the blocking probability in the guard channel system. Increased number of guard channels in the cellular networks would give a higher opportunity to the user to perform the seamless upward vertical handover regardless of the θ_2 values.

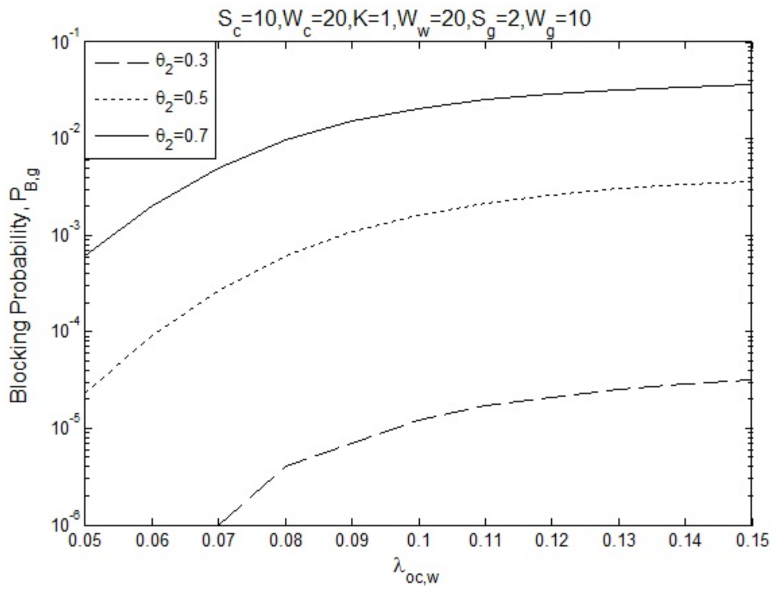

Figure 9.16: The blocking probability of guard channel ($P_{B,g}$) results as a function of originating calls in the WLAN for different θ_2

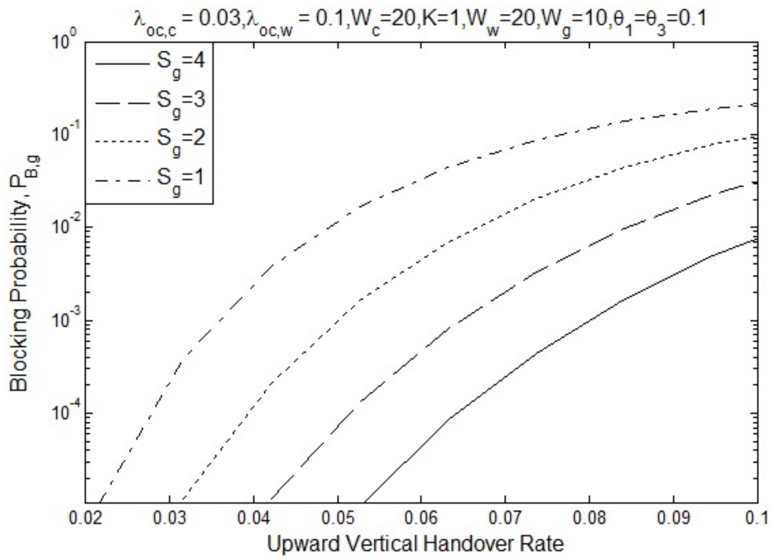

Figure 9.17: The blocking probability results as a function of upward vertical handover rate for different S_g

However, on the other hand the performance of the cellular system for originating and horizontal handover calls will decrease accordingly. This is because of the decreased number of channels provided for originating and horizontal handover calls in the system depending on the total number of channels for the entire cellular system ($S = S_c + S_g$).

9.5 Conclusions and Recommendations

The analysis of the vertical handover is an important issue in order to get good performance from the integrated cellular/WLAN networks. Two stage queuing system models have been proposed in the literature in order to analyse the integration of vertical handover in future heterogeneous wireless networks (Balsamo et al., 2003), (Gazis et al., 2005), (Gowrishankar et al., 2009), (Hasib and Fapojuwo, 2007), (Hasib and Fapojuwo, 2008), (Lampropoulos et al., 2007), (Rutagemwa, 2007), (Salkintzis et al., 2002), (Saravanan et al., 2006), (Shensheng and Wei, 2005), (Song et al., 2008a), (Song et al., 2008b), (Song et al., 2005), (Xia and Shen, 2007), (Xia and Shen, 2009). Feedback during the vertical handover procedure in integrated cellular/WLAN system is a critical issue in the analytical modelling since it affects the QoS of the system. Considerable amount of focus is given in specification of operative area while homogeneous as well as heterogeneous handover processes are investigated with the aid of an exact analytical solution approach. A new analytical model has been proposed using guard channel and buffering for cellular/WLAN system integration. The new model clearly differentiates requests originating in the cellular system, from requests being handed over from the WLAN to the cellular system, therefore the calls handed over from the WLAN to the cellular network are not allowed back to the WLAN. This prevents a handover hysteresis which the results show severely affect system performance. The new model therefore is an improved model for studying handover in real systems. In addition, mobility of the users, buffer capacity of the WLAN as well as cellular networks, and area of coverage are considered together with probabilities of having various handover decisions. Handover decisions considered are: transferring a call from the cellular network to the WLAN (θ_1), having a vertical handover from the WLAN to the cell due to the mobility (θ_2), and having a vertical handover from the cell to the WLAN (without being served by cellular network) once the mobile station gets in the coverage area of the WLAN due to the mobility (θ_3).

The results mainly show that θ_1 and θ_3 are important parameters for the decision management of the downward vertical handover from the cellular system to the WLAN. For lighter traffic loads in the cellular system the users can perform downward vertical handover regardless of θ_1 and θ_3 values. However, increasing traffic load in the cellular system makes it more difficult to perform downward vertical handover depending on the θ_1 and θ_3 values. In addition, it is clearly shown by the results that the buffer capacity of the cellular system affects the downward vertical handover decision policy considerably. Increasing the buffer capacity of the cellular system can have serious impact on the WLAN performance depending on the θ_3 values. Hence θ_3 plays an important role in determining the downward vertical handover decision due to the mobility in the queue. However, the WLAN buffering capacity plays an even more important role in determining good performance results in the integrated system. The downward vertical handover calls will be blocked with a higher probability when the small buffering capacity of the WLAN considered. On the other hand, considering the high buffering capacity in the WLAN the downward vertical handover is possible almost regardless of values of θ_1 and θ_3. In addition, the service time of the cellular network is another vital parameter that needs to be carefully taken into account. The calls can be serviced quicker and/or slower depending on the service time of the cellular networks. For instance, due to the higher service rate provided in the cellular system the probability of having a vertical handover from the cellular system to the WLAN increase because of the mobility in the queue θ_3 hence the θ_1 decreases accordingly. In addition, θ_2 is the other important handover decision probability. The calls in the WLAN can perform upward vertical handover only to the cellular guard channel system with a probability of θ_2. As it is stated above, θ_1 and θ_3 values directly affect the upward vertical handovers in the system. The results obtained from the analysis of θ_2 show that when it comes to decision of upward vertical handover, generally the handover process can take place towards the guard channel system depending on the number of guard channels and buffering. In other words, the handed over calls from the cellular network to the WLAN could make seamless upward vertical handover to the cellular system with the help of guard channels as well as the buffer provided.

Moreover, the guard channels and buffering avoid the calls which are handed over from the WLAN to have another downward vertical handover to the WLAN as practical systems.

The mobile users in the cellular system can perform seamless downward and/or upward vertical handover depending on various parameters. The θ_1, θ_2 and θ_3 values are most important parameters in the proposed systems in order to achieve optimum QoS measurements. Thus, the system sensitivity increases and operational areas could be obtained depending on the handover decision making parameters and both system buffering. The analytical tool presented is very useful in that sense. Obviously the analysis presented is very important, since the interaction between the WLAN and the cellular system is essential for optimization studies of heterogeneous wireless communication systems.

Chapter 10

Conclusions and Future Works

In this chapter, a summary of the work done in this thesis is given where the major contributions are highlighted. This is followed by conclusions resulting from this work and a discussion on the directions for future research is presented.

10.1 Summary of the Work Done

This thesis describes the performance and/or performability modelling of handovers in wireless and mobile networks. Two dimensional state space is used to represent the state of the system considered at a given time and the spectral expansion method is extensively used for the solution of the two dimensional state space. Numerical results are presented in order to show the effectiveness of the models developed. Some assumptions are made to make analytical modelling possible. In order to validate the accuracy of the developed models, simulations are performed for systems with certain assumptions. In addition, due to the state explosion problem, some of the results presented in this thesis are calculated by using the solution of system of linear equations. Considering the solution approach, we take advantage of system of linear equations to obtain as well as to validate the results of the proposed systems.

10.2 Contribution of the Thesis

The main contribution of this research is the development of new mathematical models to evaluate performance and performability measures of wireless and mobile communication systems with handover, in order to provide QoS to today's networks. The contributions of the thesis can be summarised as follows:

1. In Chapter 1, we began by motivating the need to look at performance and performability modelling techniques for wireless and mobile systems. In Chapter 2, the domain of the research is given by providing a critical review of related literature. In addition, the buffering, mobility and handover issues for the wireless systems and the detailed literature survey are also given. Handovers for the wireless and mobile systems are discussed with mobility and failure/repair which affects the performance measures efficiently.

2. In Chapter 3, performance/performability modelling and evaluation techniques for computer networks and telecommunication systems are analysed. In addition, pure performance, availability and composite performance and availability (performability) modelling techniques are studied. The existing modelling methods are critically analysed and compared. The systems under study and background information are also given. The main failings of existing models are underlined.

2. In Chapter 4, analytical models are presented for homogeneous cellular networks for performability evaluation with handover and mobility issues. Performability models considered in Trivedi et al. (2003), Trivedi et al. (2002), Jindal et al. (2006), and Gowrishankar et al. (2009) were extended by considering the mobility issues such as speed of the mobile users and radius of the cell etc. The state probabilities in the case of a homogeneous cellular system considering failures and repairs together with mobility issues are derived using MRM solution approach. The two dimensional representation of the proposed system is also presented and analysed in the second part of the Chapter 4 using the exact spectral expansion solution approach. Numerical results have been obtained and presented for various performability measures. The results show that, wireless and mobile cellular systems

are prone to failures. The failures of the operative channels in the system reduce the QoS and this gives rise to increased handover. In addition, the user velocity and the radius of cells are important parameters in determining the effects of handover, and originating calls as well as the handover requests.

3. In Chapter 5, performability modelling and the analytical models used in Chapter 4 are extended for modelling of WLAN as a queuing system with failures and repairs for exact solution. The state probabilities of such systems are derived using spectral expansion method. Simulation results were presented comparatively in order to validate the accuracy of the models presented. The behaviour of the WLAN was taken as a case study. The inherent wireless channel unreliability behaviour caused by the channel failure/recovery and access point failure/recovery phenomena have been modelled and presented. In addition, the mobility related issues and queuing capacities are also considered for various scenarios. Numerical results have been obtained and presented for various performability measures with and without mobility. Results show that, the AP failures and reconfiguration affect the performability of the system more significantly, compared to the channel failures of the system. The failures of AP cause significant system un-availabilities, and therefore the maintenance of the AP is important as well as the repair facility.

4. Chapter 6, presented an important framework for the research by presenting an analytical model which represents vertical handover in an integrated cellular architectures. The model was used to obtain QoS measures for vertical handover and mobility issues in integrated cellular technologies. In other words, the proposed analytical model could be used as part of an overall framework for such networks. The spectral expansion solution approach is used in order to obtain the state probabilities of the proposed model. The results from the analytical model were close to the results obtained by the simulation. The model was also shown to be effective over a wide operational range. The model itself can be applied to any other cellular technologies such as the integration of micro and pico cells and/or

hierarchical networks etc.

5. In Chapter 7, we used the optimal point derived in Chapter 6 and the model extended for the integration of the 3G cellular network and WLAN. The proposed model showed that it can be used as a key mechanism in order to obtain QoS of the integrated cellular/WLAN for future networks considering the systems behaviour as well as architecture in a heterogeneous environment in an analytical point of view. This showed that, this work can be directly applied to the current integrating environments. The spectral expansion method is employed for the exact solution of the analytical model considered. The results of proposed model are validated by the simulation.

6. In Chapter 8, a new approach is presented in order to prevent handover hysteresis in Chapter 7. The new approach used an additional state variable to differentiate the number of requests originated in cellular network from requests being handed over from WLAN to cellular system in two stage tandem networks. In order to provide required QoS, the proposed model was analysed considering simple scenarios but very important results were obtained. The chapter showed that there was a need for an analytical model which prevents a handover hysteresis in such systems. However, introducing a new random variable creates additional states for the system. Hence the new state significantly increases the complexity of the system description and solution. In order to prevent the handover hysteresis and the complexity the use of guard channels and buffering is considered in Chapter 9.

6. Chapter 9, described and explored the operational areas using the fundamental constraints discussed in Chapter 8. Using the guard channels and buffering in the proposed model showed how to obtain optimal operational areas by solving the fundamental constraints in Chapters 7 and 8. These were derived for different values of the θ_1, θ_2 and θ_3 values considering the given threshold value, $P_{B_{th}}$. Finally, the proposed model showed, which could be used in a real system, the system sensitivity increases and operational areas could be obtained depending on the handover decision making parameters and both system buffering.

10.2.1 Summary

The research presented in this thesis suggests future directions of interest. The proposed models perform an important role to enhance network performance, and handover decision probabilities can be optimized for different types of traffic and QoS requirements. Thus, the proposed models can be used to optimize the network deployment, network planning and other systems QoS characteristics such as capacity, delay etc. Experimental evaluation demonstrated that the proposed design can provide the required QoS. It also showed that it can be easily integrated into existing infrastructures. In addition, the results obtained from the proposed models in this thesis, lend themselves as powerful tools in order to make essential decisions for system design, optimise parameters to fit user requirements, and to examine optimal design trade-offs. Also, these models are helpful in understanding the complex interaction between systems components and vital effects of various factors such as failures/repairs, handover and mobility.

10.3 Conclusion

In conclusion, the thesis showed that based on the characteristics of the wireless and mobile systems, it is possible to improve QoS by developing new analytical models and solutions by calculating performance and performability measures for such systems considering handover phenomenon and mobility issues. The thesis gives a more balanced picture in order to obtain realistic QoS measures for wireless and mobile systems which answers the research question in this thesis.

10.4 Suggestions for future study

This thesis has raised various issues that need to be addressed. Hence some suggestions for the future study are given below:

1. The method given in Chapters 6, 7 and 9 can be further extended to obtain performability models for integrated heterogeneous environments. However,

developing performability models and evaluation for such systems is a challenging task. The performability analysis of integrated heterogeneous environments makes the system very complex and leads to multi-dimensional queuing models. It is possible to find approximate solution methods which give accurate results to ease such problems. An approximate three dimensional solution is shown in Ever (2007) for two stage tandem networks. It is possible to develop a similar approach (4-dimensional) for integrated heterogeneous systems as well. The 4-dimensional model may provide a large degree of flexibility and various scenarios can be handled as long as the system is stable and the spectral expansion method is able to handle the two dimensional lattice strips. Such an approach would be very useful especially for large scale and complex future networks.

2. Current solution approaches suffer from the well-known state space explosion problem when such complicated scenarios and multi-dimensional models are considered. The research work in Ever (2007) is attempting to combine performance model and availability model to obtain more realistic performability measures. However, these systems lead to develop a multi-dimensional model and create the state space explosion problem. Because of the state space explosion problem, it is desirable to develop an alternative solution approach to cope with this and consider performability measures. In Gemikonakli et al. (2009) and Ever et al. (2012) an approximate and hybrid solution approaches considered in order to minimise state space explosion problem for the solution of two stage tandem queues respectively. The similar approaches can be used in modelling various kinds of integrated systems.

3. The method presented in Chapter 8 with an additional variable can be extended for such systems with various characteristics. In the given approach, the states where the additional state is presented, systems with complex architecture may require a large amount of states. In this case, the existing methods can be used for the automatic generation of the states similar to the approaches used in Inverardi et al. (2005) and Inverardi et al. (2006) to implement the proposed model in order to ease the the numerical difficulties

caused by large number of states.

4. Researchers, at Middlesex University, have studied an analytical traffic model to analyse the effect on the performance due to the different traffic schemes for wireless and mobile networks. We would like to combine our network handover models with an analytical traffic model being developed by researchers at Middlesex which looks at supporting voice, data and video over heterogeneous networks using handover techniques.

10.5 Final Statement

With the fast development of networking technology, the amount of network resources increases rapidly. Due to the diversity of networks, access using heterogeneous network interconnection is becoming more and more important. In order to improve QoS and to satisfy demand of the users, we have proposed new analytical models and solutions by calculating performance and performability measures for wireless and mobile systems. However, this contribution represents a small step towards the fundamental goal of providing QoS support for future wireless and mobile networks.

Appendix A

A.1 These transitions lead to a set of balance equations in total in Chapter 6. All approximate steady state probabilities, $p_{i,j}$, can be calculated for different number of operative channels, S and K for both systems. The balance equations can be given as follows:

L_1 = S+W_1
L_2 = K+W_2
i=0 and j=0

$$p_{i,j} = \frac{\mu_1 p_{i+1,j} + \mu_2 p_{i,j+1}}{\lambda_1 + \lambda_{oc,2}} \tag{1}$$

i=0 and 0<j< L_w
If(j<K)

$$p_{i,j} = \frac{\lambda_{oc,2} p_{i,j-1} + min(j,K)\mu_2 p_{i,j+1} + \mu_{vh,1} p_{i+1,j-1} + \mu_1 p_{i+1,j}}{\lambda_1 + \lambda_{oc,2} + min(j,K)\mu_{oc,2}\theta_2 + min(j,K)\mu_2} \tag{2}$$

If(j≥K)

$$p_{i,j} = \frac{\lambda_{oc,2} p_{i,j-1} + min(j,K)\mu_2 p_{i,j+1} + \mu_{vh,1} p_{i+1,j-1} + \mu_1 p_{i+1,j}}{\lambda_1 + \lambda_{oc,2} + (K\mu_{oc,2}\theta_2) + (j-K)\mu_{dw} + K\mu_2} \tag{3}$$

i=0 and j=L_2

$$p_{i,j} = \frac{\lambda_{oc,2} p_{i,j-1} + \mu_{vh,1} p_{i+1,j-1} + \mu_1 p_{i+1,j}}{\lambda_1 + (K\mu_{oc,2}\theta_2) + (W_2\mu_{dw}) + K\mu_2} \tag{4}$$

0<i< L_1 and if(j=0)

if(i<S)

$$p_{i,j} = \frac{\mu_2 p_{i,j+1} + \lambda_1 p_{i-1,j} + \mu_{oc,2}\theta_2 p_{i-1,j+1} + i\mu_1 p_{i+1,j}}{\lambda_1 + \lambda_{oc,2} + min(i,S)\mu_1 + min(i,S)\mu_{vh,1}} \tag{5}$$

if(i>S)

$$p_{i,j} = \frac{\mu_2 p_{i,j+1} + \lambda_1 p_{i-1,j} + \mu_{oc,2}\theta_2 p_{i-1,j+1} + (S\mu_1 + (i-S)\mu_{dw,1}) p_{i+1,j}}{\lambda_1 + \lambda_{oc,2} + S\mu_1 + (i-S)\mu_{dw,1} + S\mu_{vh,1} + (i-S)\mu_{dw,2}\theta_3} \tag{6}$$

if(i=S)

$$p_{i,j} = \frac{\mu_2 p_{i,j+1} + \lambda_1 p_{i-1,j} + \mu_{oc,2}\theta_2 p_{i-1,j+1} + (S\mu_1 + \mu_{dw,1}) p_{i+1,j}}{\lambda_1 + \lambda_{oc,2} + S\mu_1 + S\mu_{vh,1}} \tag{7}$$

0<j< L_2

if(i<S) and **if(j<K)**

$$\begin{aligned} p_{i,j} = & \frac{min(j,K)\mu_2 p_{i,j+1} + min(i,S)\mu_1 p_{i+1,j} + min(i,S)\mu_{vh,1} p_{i+1,j-1}}{\lambda_1 + \lambda_{oc,2} + min(j,K)\mu_{oc,2}\theta_2 + min(j,K)\mu_2 + min(i,S)\mu_1 + min(i,S)\mu_{vh,1}} \\ & + \frac{\lambda_{oc,2} p_{i,j-1} + \lambda_1 p_{i-1,j} + min(j,K)\mu_{oc,2}\theta_2 p_{i-1,j+1}}{\lambda_1 + \lambda_{oc,2} + min(j,K)\mu_{oc,2}\theta_2 + min(j,K)\mu_2 + min(i,S)\mu_1 + min(i,S)\mu_{vh,1}} \end{aligned} \tag{8}$$

if(i<S) and **if(j=K)**

$$\begin{aligned} p_{i,j} = & \frac{K\mu_2 p_{i,j+1} + min(i,S)\mu_1 p_{i+1,j} + min(i,S)\mu_{vh,1} p_{i+1,j-1}}{\lambda_1 + \lambda_{oc,2} + K\mu_{oc,2}\theta_2 + K\mu_2 + min(i,S)\mu_1 + min(i,S)\mu_{vh,1}} \\ & + \frac{\lambda_{oc,2} p_{i,j-1} + \lambda_1 p_{i-1,j} + (K\mu_{oc,2}\theta_2 + \mu_{dw}) p_{i-1,j+1}}{\lambda_1 + \lambda_{oc,2} + K\mu_{oc,2}\theta_2 + K\mu_2 + min(i,S)\mu_1 + min(i,S)\mu_{vh,1}} \end{aligned} \tag{9}$$

if(i=S) and **if(j=K)**

$$p_{i,j} = \frac{K\mu_2 p_{i,j+1} + (S\mu_1 + \mu_{dw,1})p_{i+1,j} + (S\mu_{vh,1} + \mu_{dw,2}\theta_3)p_{i+1,j-1}}{\lambda_1 + \lambda_{oc,2} + K\mu_{oc,2}\theta_2 + K\mu_2 + S\mu_1 + S\mu_{vh,1}} + \frac{\lambda_{oc,2}p_{i,j-1} + \lambda_1 p_{i-1,j} + (K\mu_{oc,2}\theta_2 + \mu_{dw})p_{i-1,j+1}}{\lambda_1 + \lambda_{oc,2} + K\mu_{oc,2}\theta_2 + K\mu_2 + S\mu_1 + S\mu_{vh,1}} \tag{10}$$

if(i<S) and **if(j>K)**

$$p_{i,j} = \frac{K\mu_2 p_{i,j+1} + min(i,S)\mu_1 p_{i+1,j} + min(i,S)\mu_{vh,1}p_{i+1,j-1}}{\lambda_1 + \lambda_{oc,2} + K\mu_{oc,2}\theta_2 + (j-K)\mu_{dw} + K\mu_2 + min(i-S)\mu_1 + min(i-S)\mu_{vh,1}} + \frac{\lambda_{oc,2}p_{i,j-1} + \lambda_1 p_{i-1,j} + (K\mu_{oc,2}\theta_2 + (j-K)\mu_{dw})p_{i-1,j+1}}{\lambda_1 + \lambda_{oc,2} + K\mu_{oc,2}\theta_2 + (j-K)\mu_{dw} + K\mu_2 + min(i-S)\mu_1 + min(i-S)\mu_{vh,1}} \tag{11}$$

if(i=S) and **if(j>K)**

$$p_{i,j} = \frac{K\mu_2 p_{i,j+1} + (S\mu_1 + \mu_{dw,1})p_{i+1,j} + (S\mu_{vh,1} + (i-S)\mu_{dw,2}\theta_3)p_{i+1,j-1}}{\lambda_1 + \lambda_{oc,2} + K\mu_{oc,2}\theta_2 + (j-K)\mu_{dw} + K\mu_2 + S\mu_1 + S\mu_{vh,1}} + \frac{\lambda_{oc,2}p_{i,j-1} + \lambda_1 p_{i-1,j} + (K\mu_{oc,2}\theta_2 + (j-K)\mu_{dw})p_{i-1,j+1}}{\lambda_1 + \lambda_{oc,2} + K\mu_{oc,2}\theta_2 + (j-K)\mu_{dw} + K\mu_2 + S\mu_1 + S\mu_{vh,1}} \tag{12}$$

if(i=S) and **if(j<K)**

$$p_{i,j} = \frac{min(j,K)\mu_2 p_{i,j+1} + (S\mu_1 + \mu_{dw,1})p_{i+1,j} + (S\mu_{vh,1} + \mu_{dw,2}\theta_3)p_{i+1,j-1}}{\lambda_1 + \lambda_{oc,2} + min(j,K)\mu_{oc,2}\theta_2 + min(j,K)\mu_2 + S\mu_1 + S\mu_{vh,1}} + \frac{\lambda_{oc,2}p_{i,j-1} + \lambda_1 p_{i-1,j} + (min(j,K)\mu_{oc,2}\theta_2)p_{i-1,j+1}}{\lambda_1 + \lambda_{oc,2} + min(j,K)\mu_{oc,2}\theta_2 + min(j,K)\mu_2 + S\mu_1 + S\mu_{vh,1}} \tag{13}$$

if(i>S) and **if(j<K)**

$$p_{i,j} = \frac{min(j,K)\mu_2 p_{i,j+1} + (S\mu_1 + (i-S)\mu_{dw,1})p_{i+1,j} + (S\mu_{vh,1} + (i-S)\mu_{dw,2}\theta_3)p_{i+1,j-1}}{\lambda_1 + \lambda_{oc,2} + min(j,K)\mu_{oc,2}\theta_2 + min(j,K)\mu_2 + S\mu_1 + (i-S)\mu_{dw,1} + S\mu_{vh,1} + (i-S)\mu_{dw,2}\theta_3} + \frac{\lambda_{oc,2}p_{i,j-1} + \lambda_1 p_{i-1,j} + (min(j,K)\mu_{oc,2}\theta_2)p_{i-1,j+1}}{\lambda_1 + \lambda_{oc,2} + min(j,K)\mu_{oc,2}\theta_2 + min(j,K)\mu_2 + S\mu_1 + (i-S)\mu_{dw,1} + S\mu_{vh,1} + (i-S)\mu_{dw,2}\theta_3} \tag{14}$$

if(i>S) and **if(j=K)**

$$p_{i,j} = \frac{K\mu_2 p_{i,j+1} + (S\mu_1 + (i-S)\mu_{dw,1})p_{i+1,j} + (S\mu_{vh,1} + (i-S)\mu_{dw,2}\theta_3)p_{i+1,j-1}}{\lambda_1 + \lambda_{oc,2} + K\mu_{oc,2}\theta_2 + K\mu_2 + S\mu_1 + (i-S)\mu_{dw,1} + S\mu_{vh,1} + (i-S)\mu_{dw,2}\theta_3} + \frac{\lambda_{oc,2}p_{i,j-1} + \lambda_1 p_{i-1,j} + (K\mu_{oc,2}\theta_2 + \mu_{dw})p_{i-1,j+1}}{\lambda_1 + \lambda_{oc,2} + K\mu_{oc,2}\theta_2 + K\mu_2 + S\mu_1 + (i-S)\mu_{dw,1} + S\mu_{vh,1} + (i-S)\mu_{dw,2}\theta_3} \tag{15}$$

if(i>S) and **if(j>K)**

$$p_{i,j} = \frac{K\mu_2 p_{i,j+1} + (S\mu_1 + (i-S)\mu_{dw,1})p_{i+1,j} + (S\mu_{vh,1} + (i-S)\mu_{dw,2}\theta_3)p_{i+1,j-1}}{\lambda_1 + \lambda_{oc,2} + K\mu_{oc,2}\theta_2 + (j-K)\mu_{dw} + K\mu_2 + S\mu_1 + (i-S)\mu_{dw,1} + S\mu_{vh,1} + (i-S)\mu_{dw,2}\theta_3} + \frac{\lambda_{oc,2}p_{i,j-1} + \lambda_1 p_{i-1,j} + (K\mu_{oc,2}\theta_2 + (j-K)\mu_{dw})p_{i-1,j+1}}{\lambda_1 + \lambda_{oc,2} + K\mu_{oc,2}\theta_2 + (j-K)\mu_{dw} + K\mu_2 + S\mu_1 + (i-S)\mu_{dw,1} + S\mu_{vh,1} + (i-S)\mu_{dw,2}\theta_3} \tag{16}$$

j$= L_2$

if(i=S)

$$p_{i,j} = \frac{(S\mu_1 + \mu_{dw,1})p_{i+1,j} + \lambda_{oc,2}p_{i,j-1} + \lambda_1 p_{i-1,j} + (S\mu_{vh,1} + \mu_{dw,2}\theta_3)p_{i+1,j-1}}{\lambda_1 + (K\mu_{oc,2}\theta_2) + (W_2\mu_{dw}) + K\mu_2 + (S\mu_1)} \tag{17}$$

if(i<S)

$$p_{i,j} = \frac{min(i,S)\mu_1 p_{i+1,j} + \lambda_{oc,2}p_{i,j-1} + \lambda_1 p_{i-1,j} + min(i,S)\mu_{vh,1}p_{i+1,j-1}}{\lambda_1 + (K\mu_{oc,2}\theta_2) + (W_2\mu_{dw}) + K\mu_2 + min(i,S)\mu_1} \tag{18}$$

if(i>S)

$$p_{i,j} = \frac{(S\mu_1 + (i-S)\mu_{wd,1})p_{i+1,j} + \lambda_{oc,2}p_{i,j-1} + \lambda_1 p_{i-1,j} + S\mu_{vh,1} + (i-S)\mu_{dw,2}\theta_3 p_{i+1,j-1}}{\lambda_1 + (K\mu_{oc,2}\theta_2) + (W_2\mu_{dw}) + K\mu_2 + S\mu_1 + (i-S)\mu_{dw,1}} \tag{19}$$

i$= L_1$

if(j=0)

$$p_{i,j} = \frac{\mu_2 p_{i,j+1} + \lambda_1 p_{i-1,j} + \mu_{oc,2}\theta_2 p_{i-1,j+1}}{\lambda_{oc,2} + S\mu_1 + W_1\mu_{dw,1} + S\mu_{vh,1} + W_1\mu_{dw,2}\theta_3} \tag{20}$$

0<j$< L_2$

if (j<K)

$$p_{i,j} = \frac{min(j,K)\mu_2 p_{i,j+1} + \lambda_{oc,2}p_{i,j-1} + \lambda_1 p_{i-1,j} + min(j,K)\mu_{oc,2}\theta_2 p_{i-1,j+1}}{\lambda_{oc,2} + min(i,K)\mu_2 + S\mu_1 + W_1\mu_{dw,1} + S\mu_{vh,1} + W_1\mu_{dw,2}\theta_3} \tag{21}$$

if (j=K)

$$p_{i,j} = \frac{K\mu_2 p_{i,j+1} + \lambda_{oc,2} p_{i,j-1} + \lambda_1 p_{i-1,j} + (K\mu_{oc,2}\theta_2 + \mu_{dw}) p_{i-1,j+1}}{\lambda_{oc,2} + K\mu_2 + S\mu_1 + W_1\mu_{dw,1} + S\mu_{vh,1} + W_1\mu_{dw,2}\theta_3} \tag{22}$$

if (j>K)

$$p_{i,j} = \frac{K\mu_2 p_{i,j+1} + \lambda_{oc,2} p_{i,j-1} + \lambda_1 p_{i-1,j} + (K\mu_{oc,2}\theta_2 + (j-K)\mu_{dw}) p_{i-1,j+1}}{\lambda_{oc,2} + K\mu_2 + S\mu_1 + W_1\mu_{dw,1} + S\mu_{vh,1} + W_1\mu_{dw,2}\theta_3} \tag{23}$$

if(j= L_2**)**

$$p_{i,j} = \frac{\lambda_{oc,2} p_{i,j-1} + \lambda_1 p_{i-1,j}}{K\mu_2 + S\mu_1 + W_1\mu_{dw,1}} \tag{24}$$

A.2 The balance equations used in Chapter 9 given below:

L_c = S+W_c
L_w = 1+W_w
i=0 and j=0

$$p_{i,j} = \frac{\mu_c p_{i+1,j} + \mu_w p_{i,j+1}}{\lambda_c + \lambda_{oc,w}} \tag{25}$$

i=0 and 0<j< L_w
If(j>1)

$$p_{i,j} = \frac{\lambda_{oc,w} p_{i,j-1} + (\mu_w + (j-1)\mu_{dw}) p_{i,j+1} + \mu_{vh,c} p_{i+1,j-1} + \mu_c p_{i+1,j}}{\lambda_c + \lambda_{oc,w} + \mu_w + (j-1)\mu_{dw}} \tag{26}$$

If(j=1)

$$p_{i,j} = \frac{\lambda_{oc,w} p_{i,j-1} + (\mu_w + \mu_{dw}) p_{i,j+1} + \mu_{vh,c} p_{i+1,j-1} + \mu_c p_{i+1,j}}{\lambda_c + \lambda_{oc,w} + \mu_w} \tag{27}$$

i=0 and j=L_2

$$p_{i,j} = \frac{\lambda_{oc,w}p_{i,j-1} + \mu_{vh,c}p_{i+1,j-1} + \mu_c p_{i+1,j}}{\lambda_c + \mu_w + W_w\mu_{dw}} \tag{28}$$

0<i< L_c and if(j=0)

if(i< S_c)

$$p_{i,j} = \frac{\mu_w p_{i,j+1} + \lambda_c p_{i-1,j} + i\mu_c p_{i+1,j}}{\lambda_c + \lambda_{oc,w} + i\mu_c + i\mu_{vh,c}} \tag{29}$$

if(i= S_c)

$$p_{i,j} = \frac{\mu_w p_{i,j+1} + \lambda_c p_{i-1,j} + (S_c\mu_c + \mu_{cd,c})p_{i+1,j}}{\lambda_c + \lambda_{oc,w} + S_c\mu_c + S_c\mu_{vh,c}} \tag{30}$$

if(i> S_c)

$$p_{i,j} = \frac{\mu_w p_{i,j+1} + \lambda_c p_{i-1,j} + (S_c\mu_c + (i-S_c)\mu_{cd,c})p_{i+1,j}}{\lambda_c + \lambda_{oc,w} + S_c\mu_c + (i-S_c)\mu_{cd,c} + S_c\mu_{vh,c} + (i-S_c)\mu_{cd,w}\theta_3} \tag{31}$$

0<j< L_w

if(i<S) and **if(j<1)**

$$\begin{aligned} p_{i,j} = &\frac{\mu_w p_{i,j+1} + min(i, S_c\mu_c)p_{i+1,j} + min(i,S_c)\mu_{vh,c}p_{i+1,j-1}}{\lambda_c + \lambda_{oc,w} + min(j,1)\mu_w + min(i,S_c)\mu_c + min(i,S_c)\mu_{vh,c}} \\ &+ \frac{\lambda_c p_{i-1,j} + \lambda_{oc,w}p_{i,j-1}}{\lambda_c + \lambda_{oc,w} + min(j,1)\mu_w + min(i,S_c)\mu_c + min(i,S_c)\mu_{vh,c}} \end{aligned} \tag{32}$$

if(i<S) and **if(j=1)**

$$\begin{aligned} p_{i,j} = &\frac{(\mu_w + (j-1)\mu_{dw)}p_{i,j+1} + min(i,S_c)\mu_c)p_{i+1,j} + min(i,S_c)\mu_{vh,c}p_{i+1,j-1}}{\lambda_c + \lambda_{oc,w} + min(j,1)\mu_w + min(i,S_c)\mu_c + min(i,S_c)\mu_{vh,c}} \\ &+ \frac{\lambda_c p_{i-1,j} + \lambda_{oc,w}p_{i,j-1}}{\lambda_c + \lambda_{oc,w} + \mu_w + min(i,S_c)\mu_c + min(i,S_c)\mu_{vh,c}} \end{aligned} \tag{33}$$

if(i= S_c**)** and **if(j=1)**

$$p_{i,j} = \frac{(\mu_w + (j-1)\mu_{dw})p_{i,j+1} + (S_c\mu_c + \mu_{cd,c})p_{i+1,j}}{\lambda_c + \lambda_{oc,w} + \mu_w + S_c\mu_c + S_c\mu_{vh,c}} + \frac{(S_c\mu_{vh,c} + \mu_{cd,w}\theta_3)p_{i+1,j-1} + \lambda_c p_{i-1,j} + \lambda_{oc,w}p_{i,j-1}}{\lambda_c + \lambda_{oc,w} + \mu_w + S_c\mu_c + S_c\mu_{vh,c}} \tag{34}$$

if(i< S_c**)** and **if(j>1)**

$$p_{i,j} = \frac{(\mu_w + (j-1)\mu_{dw})p_{i,j+1} + S_c\mu_c p_{i+1,j}}{\lambda_c + \lambda_{oc,w} + \mu_w + (j-1)\mu_{dw} + min(i,S_c)\mu_c + min(i,S_c)\mu_{vh,c}} + \frac{min(i,S_c)\mu_{vh,c} + \lambda_c p_{i-1,j} + \lambda_{oc,w}p_{i,j-1}}{\lambda_c + \lambda_{oc,w} + \mu_w + (j-1)\mu_{dw} + min(i,S_c)\mu_c + min(i,S_c)\mu_{vh,c}} \tag{35}$$

if(i= S_c**)** and **if(j<1)**

$$p_{i,j} = \frac{min(j,1)\mu_w p_{i,j+1} + (S_c\mu_c + \mu_{cd,c})p_{i+1,j}}{\lambda_c + \lambda_{oc,w} + min(j,1)\mu_w + S_c\mu_c + S_c\mu_{vh,c}} + \frac{(S_c\mu_{vh,c} + \mu_{cd,w}\theta_3) + \lambda_c p_{i-1,j} + \lambda_{oc,w}p_{i,j-1}}{\lambda_c + \lambda_{oc,w} + min(j,1)\mu_w + S_c\mu_c + S_c\mu_{vh,c}} \tag{36}$$

if(i> S_c**)** and **if(j<1)**

$$p_{i,j} = \frac{min(j,1)\mu_w p_{i,j+1} + (S_c\mu_c + (i-S_c)\mu_{cd,c})p_{i+1,j}}{\lambda_c + \lambda_{oc,w} + min(j,1)\mu_w + S_c\mu_c + (i-S_c)\mu_{cd,c} + S_c\mu_{vh,c} + (i-S_c)\mu_{cd,w}\theta_3)} + \frac{(S_c\mu_{vh,c} + (i-S_c)\mu_{cd,w}\theta_3) + \lambda_c p_{i-1,j} + \lambda_{oc,w}p_{i,j-1}}{\lambda_c + \lambda_{oc,w} + min(j,1)\mu_w + S_c\mu_c + (i-S_c)\mu_{cd,c} + S_c\mu_{vh,c} + (i-S_c)\mu_{cd,w}\theta_3)} \tag{37}$$

if(i> S_c**)** and **if(j=1)**

$$p_{i,j} = \frac{(\mu_w + (j-1)\mu_{dw})p_{i,j+1} + (S_c\mu_c + (i-S_c)\mu_{cd,c})p_{i+1,j}}{\lambda_c + \lambda_{oc,w} + \mu_w + S_c\mu_c + (i-S_c)\mu_{cd,c} + S_c\mu_{vh,c} + (i-S_c)\mu_{cd,w}\theta_3)} + \frac{(S_c\mu_{vh,c} + (i-S_c)\mu_{cd,w}\theta_3) + \lambda_c p_{i-1,j} + \lambda_{oc,w}p_{i,j-1}}{\lambda_c + \lambda_{oc,w} + \mu_w + S_c\mu_c + (i-S_c)\mu_{cd,c} + S_c\mu_{vh,c} + (i-S_c)\mu_{cd,w}\theta_3)} \tag{38}$$

if(i> S_c**)** and **if(j>1)**

$$p_{i,j} = \frac{(\mu_w + (j-1)\mu_{dw})p_{i,j+1} + (S_c\mu_c + (i-S_c)\mu_{cd,c})p_{i+1,j}}{\lambda_c + \lambda_{oc,w} + \mu_w + (j-1)\mu_{dw} + S_c\mu_c + (i-S_c)\mu_{cd,c} + S_c\mu_{vh,c} + (i-S_c)\mu_{cd,w}\theta_3)} + \frac{(S_c\mu_{vh,c} + (i-S_c)\mu_{cd,w}\theta_3) + \lambda_c p_{i-1,j} + \lambda_{oc,w}p_{i,j-1}}{\lambda_c + \lambda_{oc,w} + \mu_w + (j-1)\mu_{dw} + S_c\mu_c + (i-S_c)\mu_{cd,c} + S_c\mu_{vh,c} + (i-S_c)\mu_{cd,w}\theta_3)} \tag{39}$$

j$= L_w$

if(i$= S_c$)

$$p_{i,j} = \frac{(S_c\mu_c + \mu_{cd,c})p_{i+1,j} + \lambda_{oc,w}p_{i,j-1} + \lambda_c p_{i-1,j} + (S_c\mu_{vh,c} + \mu_{cd,w}\theta_3)p_{i+1,j-1}}{\lambda_c + \mu_w + (W_w\mu_{dw}) + (S_c\mu_c)} \tag{40}$$

if(i$< S_c$)

$$p_{i,j} = \frac{(min(i,S_c)\mu_c)p_{i+1,j} + \lambda_{oc,w}p_{i,j-1} + \lambda_c p_{i-1,j} + (min(i,S_c)\mu_{vh,c})p_{i+1,j-1}}{\lambda_c + \mu_w + (W_w\mu_{dw}) + (min(i,S_c)\mu_c)} \tag{41}$$

if(i$> S_c$)

$$\begin{aligned} p_{i,j} = & \frac{(S_c\mu_c + (i-S_c)\mu_{cd})p_{i+1,j} + \lambda_{oc,w}p_{i,j-1}}{\lambda_c + \mu_w + (W_w\mu_{dw}) + (S_c\mu_c) + (i-S_c)\mu_{cd,c}} \\ & + \frac{\lambda_c p_{i-1,j} + (S_c\mu_{vh,c} + (i-S_c)\mu_{cd,w}\theta_3)p_{i+1,j-1}}{\lambda_c + \mu_w + (W_w\mu_{dw}) + (S_c\mu_c) + (i-S_c)\mu_{cd,c}} \end{aligned} \tag{42}$$

i$= L_c$

if(j$= 0$)

$$p_{i,j} = \frac{\lambda_c p_{i-1,j} + \mu_w p_{i,j+1}}{\lambda_{oc,w} + S_c\mu_c + W_c\mu_{cd,c} + S_c\mu_{vh,c} + W_c\mu_{cd,w}\theta_3} \tag{43}$$

j$< L_w$

if(j$= K$)

$$p_{i,j} = \frac{\lambda_{oc,w}p_{i,j-1} + \lambda_c p_{i-1,j} + (\mu_w + j\mu_{dw})p_{i,j+1}}{\lambda_{oc,w} + \mu_w + S_c\mu_c + W_c\mu_{cd,c} + S_c\mu_{vh,c} + W_c\mu_{cd,w}\theta_3} \tag{44}$$

if(j$> K$)

$$p_{i,j} = \frac{\lambda_{oc,w}p_{i,j-1} + \lambda_c p_{i-1,j} + (\mu_w + j\mu_{dw})p_{i,j+1}}{\lambda_{oc,w} + \mu_w + (j-1)\mu_{dw} + S_c\mu_c + W_c\mu_{cd,c} + S_c\mu_{vh,c} + W_c\mu_{cd,w}\theta_3} \tag{45}$$

if(j= L_w**)**

$$p_{i,j} = \frac{\lambda_{oc,w} p_{i,j-1} + \lambda_c p_{i-1,j}}{\mu_w + W_w \mu_{dw} + S_c \mu_c + W_c \mu_{cd,c}} \tag{46}$$

References

D. P. Agrawal and A. Z. Qing. *Introduction to Wireless and Mobile Systems*. Bill Stenquist, 2006.

P. Agrawal, D. K. Anvekar, and B. Narendran. Channel management policies for handovers in cellular networks. *Bell Labs Technical Journal*, 1(2):97–110, 1996.

S. Agrawal and J. M. Holtzman. Modelling and analysis of handoff algorithms in multi-cellular systems. *IEEE* 47^{th} *Vehicular Technology Conference*, 1:300–304, 1997.

M.H. Ahmed. Call admission control in wireless networks:a comprehensive survey. *IEEE Communication Surveys and Tutorials*, 7(1):50–69, 2005.

I. F. Akyildiz, J. Xie, and S. Mohanty. A survey of mobility management in next-generation all-ip-based wireless systems. *IEEE Wireless Communications*, 11 (4):16–28, 2004.

S. Balsamo, V. D. N. Persone, and P. Inverardi. A review on queueing network models with finite capacity queues for software architectures performance prediction. *Performance Evaluation*, 51:269–288, 2003.

J. Banks, J. S. Carson, B. L. Nelson, and D.M. Nicol. *Discrete-Event System Simulation*. (4th. ed.) Prentice-Hall: Upper Saddle River, NJ., 2005.

R. Beck and H. Panzer. Strategies for handover and dynamic channel allocation in micro-cellular mobile radio systems. *IEEE* 39^{th} *Vehicular Technology Conference*, 1:178–185, 1989.

R. Bekker, S.C. Borst, O.J. Boxma, and O. Kella. Queues with workload-dependent arrival and service rates. *Queueing Systems*, 46:537–556, 2004.

M. Benson and H.J. Thomas. Investigation of the umts to gsm handover procedure. *Vehicular Technology Conference*, 4:1829–1833, August 2002.

G. Bianchi. Performance analysis of the ieee 802.11 distributed coordination function. *IEEE Journal on Selected Areas on Communications*, 18(3):535–547, March 2000.

H. Bobarshad. *Evaluation and Modelling of Real Time Video Transmission Over WLAN*. Phd thesis, Department of Electronic Engineering, King's College, London, 2010.

A. Bobbio and K. S. Trivedi. An aggregation technique for the transient analysis of stiff markov chains. *IEEE Transaction on Computer*, 39(13):1291–1298, 1986.

A. Bobbio and K. S. Trivedi. Computing cumulative measures of stiff markov chains using aggregation. *IEEE Transaction on Computer*, 39(13):1291–1298, 1990.

O. Boxma, G. Koole, and Z. Liu. Queueing-theoretic solution methods for models of parallel and distributed systems. In *Performance Evaluation of Parallel and Distributed Systems Solution Methods*, pages 1–24, 1994.

M. C. Calzarossa and R. Marie. Tools for performance evaluation. *Performance Evaluation*, 33:1–3, 1998.

F. Capar, T. Weiss, I. Martoyo, and F. Jondral. Analysis of coexistence strategies for cellular and wireless local area networks. *Vehicular Technology Conference*, 3:1812–1816, 2003.

D. Cavalcanti, D. Agrawal, C. Cordeiro, B. Xie, and A. Kumar. Issues in integrating cellular networks wlans, and manets: A futuristic heterogeneous wireless network. *IEEE Wireless Communications*, 12(3):30–41, 2005.

R. Chakka. *Performance and Reliability Modelling of Computing Systems Using Spectral Expansion*. PhD thesis, University of Newcastle Upon Tyne, UK, 1995.

R. Chakka. Spectral expansion solution for some finite capacity queues. *Annals of Operations Research*, 79:27–44, 1998.

R. Chakka and I. Mitrani. Heterogeneous multiprocessor systems with breakdowns: Performance and optimal repair strategies. *Theoretical Computer Science*, 1994:91–109, 125.

R. Chakka and I. Mitrani. A numerical solution method for multiprocessor systems with general breakdowns and repairs. *In Proceedings of the* 6^{th} *International Conference Modelling Techniques and Tools*, pages 289–304, 1992.

R. Chakka and I. Mitrani. *Stochastic Networks: Theory and Applications: Approximate Solutions for Open Networks with Breakdowns and Repairs.* Oxford University Press., 1996.

R. Chakka, O. Gemikonakli, and P. Basappa. Modelling multiprocessor systems with time or operation dependent breakdowns, alternate repair strategies, reconfiguration and rebooting delays. *In Proceedings of SPECTS 2002*, pages 266–277, 2002.

D. Chen, C. Kintala, S. Garg, and K. S. Trivedi. Dependability enhancement for ieee 802.11 wireless lan with redundancy techniques. *Proceedings of the International Conference on Dependable Systems and Networks*, pages 521–528, 2003a.

Y. Chen, Q.A. Zeng, and D.P. Agrawal. Performance of mac protocol in ad hoc networks. *Proceedings of Communication Networks and Distributed Systems Modeling and Simulation Conference*, pages 55–61, 2003b.

Y. J. Choi, K. B. Lee, and S. Bahk. All-ip 4g network architecture for efficient mobility and resource management. *IEEE Wireless Communications*, 14(2): 42–46, 2007.

S. Dharmaraja, Trivedi K. S., and D Logothetis. Performance analysis of cellular networks with generally distributed handoff interarrival times. *Symposium on Performance Evaluation of Computer and Telecommunication Systems, SPECTS 2002*, pages 1747–1755, 2002.

N. Ekiz, T. Salih, S Kkner, and K. Fidanboylu. An overview of handoff techniques in cellular networks. *International Journal of Information Technology*, 2(2): 132–136, 2005.

S.A. El-Dolil, W. Wai-Choong, and R. Steele. Teletraffic performance of highway microcells with overlay macrocell. *IEEE Journal on Selected Areas in Communications*, 7(1):71 –78, 1989. ISSN 0733-8716.

E. Ever. *Performability Modelling of Homogeneous and Heterogeneous Multiserver Systems with Breakdowns and Repairs.* PhD thesis, School of Computing Science, Middlesex University, London, UK., 2007.

E. Ever, O. Gemikonakli, R. Chakka, and T.V. Do. The use of markov modulated poisson process (mmpp) in open queuing systems with breakdowns and repairs. *In Proceedings of PREP*, 2005.

E. Ever, O. Gemikonakli, A. Kocyigit, and E. Gemikonakli. A hybrid approach to minimize state space explosion problem for the solution of two stage tandem queues. *Journal of Network and Computer Applications*, pages 1–19, 2012.

Y. Fang and I. Chlamtac. A new mobility model and its application in the channel holding time characterization in pcs networks. *Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies. IEEE INFOCOM*, 1:20–27, 1999a.

Y. Fang and I. Chlamtac. Teletraffic analysis and mobility modeling of pcs networks. *IEEE Transactions on Communications*, 47(7):1062–1072, 1999b.

Y. Fang, I. Chlamtac, and Y. B. Lin. Modeling pcs networks under general call holding time and cell residence time distributions. *IEEE/ACM Transactions on Networking*, 5(6):893–906, 1997a.

Y. Fang, I. Chlamtac, and Y. B. Lin. Call performance for a pcs network. *IEEE Journal on Selected Areas in Communications*, 15(8):1568–1581, 1997b.

R. Fantacci. Performance evaluation of prioritized handoff schemes in mobile cellular networks. *IEEE Transactions on Vehicular Technology*, 49(2):485–493, 2000.

H.R. Gail, S.L. Hantler, and B.A. Taylor. Spectral analysis of m/g/1 type markov chains. Technical report, IBM Research Division, 1992.

V. Gazis, N. Alonistioti, and L. Merakos. Toward a generic "always best connected" capability in integrated wlan/umts cellular mobile networks (and beyond). *IEEE Wireless Communications*, 12(3):20–29, June 2005.

O. Gemikonakli, T. V. Do, R. Chakka, and E. Ever. Numerical solution to the performability of a multiprocessor system with reconfiguration and rebooting delays. *In Proceedings of ECMS 2005*, pages 766–773, 2005.

O. Gemikonakli, G. Mapp, D. Thakker, and E. Ever. Modelling and performability analysis of network memory servers. *Annual Simulation Symposium*, pages 127–134, 2006.

O. Gemikonakli, G. Mapp, E. Ever, and D. Thakker. Modelling network memory servers with parallel processors, break-downs and repairs. *Annual Simulation Symposium*, pages 11–20, 2007.

O. Gemikonakli, E. Ever, and A. Kocyigit. Approximate solution for two stage open networks with markov-modulated queues minimizing the state space explosion problem. *Journal of Computational and Applied Mathematics*, 223: 519–533, 2009.

Gowrishankar, G.N. Sekhar, and P.S. Satyanarayana. Analytic performability model of vertical handoff in wireless networks. *Journal of Computer Science*, 5(6):445–450, 2009.

M. Gudmundson. Analysis of handover algorithms [microcellular radio]. *41st IEEE Vehicular Technology Conference*, pages 537–542, 1991.

R. Guerin. Queueing blocking system with two arrival streams and guard channels. *IEEE Trans. on Commun.*, 36:153–163, 1988.

A. Hasib and A.O. Fapojuwo. Performance analysis of common radio resource management scheme in multi-service heterogeneous wireless networks. *IEEE WCNC*, pages 3298–3302, March 2007.

A. Hasib and A.O. Fapojuwo. A mobility model for heterogeneous wireless networks and its application in common radio resource management. *The Institution of Engineering and Technology, IET Communications Journal*, 2(9):1186 – 1195, October 2008.

R. B. Haverkort and O. Alexander. Steady-state analysis of infinite stochastic petri nets: Comparing the spectral expansion and the matrix-geometric method. *In Proceedings of the Seventh International Workshop on Petri Nets and Performance Models*, pages 36–45, 1997.

D Hong and S. S. Rappaport. Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and nonprioritized handoff procedures. *IEEE Transactions on Vehicular Technology*, 35(3):77–92, 1986.

Z. Hua, M. Li, I. Chlamtac, and B. Prabhakaran. A survey of quality of service in ieee 802.11 networks. *In IEEE Wireless Communications Journals*, 11:6–14, 2004.

L. Hui and N. K. Shankaranarayanan. A distributed dynamic channel allocation technique for throughput improvement in a dense wlan environment. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 345–348, May 2004.

P. Inverardi, Mostarda L., M. Tivoli, and M. Autili. Synthesis of correct and distributed adaptors for component-based systems: an automatic approach. *20th IEEE/ACM international Conference on Automated software engineering*, pages 405–409, 2005.

P. Inverardi, Mostarda L., and A. Navarra. Distributed idss for enhancing security in mobile wireless sensor networks. *International Conference on Advanced Information Networking and Applications*, pages 116–120, 2006.

K. Ioannou, S. Louvros, I. Panoutsopoulos, S. Kotsopoulos, and G.K. Karagiannidis. Optimizing the handover call blocking probability in cellular networks with high speed moving terminals. *In IEEE Communications Letters*, 6(10): 422–424, 2002.

R. Jayaparvathy, S. Anand, S. Dharmaraja, and S. Srikanth. Performance analysis of ieee 802.11 dcf with stochastic reward nets. *International Journal of Communication Systems*, pages 273–296, 2007.

V. Jindal, S. Dharmaraja, and K. S. Trivedi. Analytical survivability model for fault tolerant cellular networks supporting multiple services. *International Symposium on Performance Evaluation of Computer and Telecommunication Systems*, pages 505–512, 2006.

I. A. Khatib. *Performance Analysis of Wireless LAN Access Points*. Phd thesis, Department of Microelectronics and Information Technology, Royal Institute of Technology Stockholm, Sweden, May 2003.

L. Lampropoulos, N. Passas, A. Kaloxylos, and L. Merakos. A flexible umts/wlan architecture for improved network performance. *Wireless Personal Communications*, 43:889–906, 2007.

W. Li and A.S. Alfa. Channel reservation for handoff calls in a pcs network. *IEEE Transactions on Vehicular Technology*, 49(1):95–104, 2000.

B. Lin, S. Mohan, and Noerpel A. Pcs channel assignment strategies for hand-off and initial access. *IEEE Personal Comm.*, 3:47–56, 1994.

H. Liu, Y. Xu, and Q. A. Zeng. Modelling and performance analysis of a channel reservation handoff scheme for multimedia wireless and mobile networks using smart antennas. *International Journal of High Performance Computing and Networking*, 4(1/2):13–22, 2006. ISSN 1740-0562.

R. P. Liu, G. J. Sutton, and I. B. Collings. A new queuing model for qos analysis of ieee 802.11 dcf with finite buffer and load. *IEEE Transaction on Wireless Communications*, 9(8):2664– 2675, 2010.

Y. Ma, C. W. Ro, and Kishor S. T. Performability analysis of channel allocation with channel recovery strategy in cellular networks. *International Conference on Universal Personal Communications (ICUPC'98)*, pages 71–75, October 5-9 1998.

Y. Ma, J. J. Han, and K.S. Trivedi. Composite performance and availability analysis of communications networks: A comparison of exact and approximate approaches. *IEEE Global Telecommunications Conference, 2000. GLOBECOM '00.*, 3:1771–1777, 2000.

Y. Ma, J. Han, and K.S. Trivedi. Composite performance and availability analysis of wireless communication networks. *IEEE Transaction on Vehicular Technology*, 50(5):1216–1223, 2001.

B. B. Madan, S. Dharmaraja, and K. S. Trivedi. Combined guard channel and mobile-assisted handoff for cellular networks. *IEEE Transactions on Vehicular Technology*, 57:502–510, 2008.

J.P. Makela. *Effects of handoff algorithms on the performance of multimedia wireless networks*. PhD thesis, Faculty of Technology, Department of Electrical and Information Engineering, University of Oulu, Finland, 2008.

M. Malhotra and K. S. Trivedi. A methodology for formal specification of hierarchy in model solution. *In Proceeding* 5^{th} *Int. Workshop Petri Nets and Performance Models (PNPM93)*, pages 256–267, 1993.

M. Malhotra, J. Muppala, and K. S. Trivedi. Stiffness-tolerant methods for transient analysis of stiff markov chains. *Microelectronics and Reliability*, 34(11): pp. 1825–1841, 1994.

D. Malone, P. Clifford, and D. J. Leith. Mac layer channel quality measurement in 802.11. *IEEE Communication Letters*, 11, February 2007a.

D. Malone, P. Clifford, and D. J. Leith. Mac layer channel quality measurement in 802.11. *IEEE Communication Letters*, 11(2):143–145, 2007b.

W. R. Mande. Evaluation of a proposed handover algorithm for the gsm cellular system. *IEEE* 40^{th} *Vehicular Technology Conference*, pages 264–269, 1990.

G. Mapp, D. Cottingham, and F. Shaikh. An architectural framework for heterogeneous networking. *In Proceedings of the International Conference on Wireless Information Networks and Systems (WINSYS 06).*, page 510, 2006.

G. Mapp, F. Katsriku, N. Chinnam, M. Aiash, R.L. Rigolin, E. Moreira, R.P. Vanni, and M. Augusto. Exploiting location and contextual information to develop a comprehensive framework for proactive handover in heterogeneous environments. *Journal of Computer Networks and Communications: Special Issue on Context-Enriched and Location-Aware Services*, July 2012.

J. F. Meyer. On evaluating the performability of degradable computing systems. *IEEE Transaction on Computer*, C-29:pp. 720–731, 1980.

Ajay R. Mishra. *Advanced Cellular Network Planning and Optimisation: 2G/2.5G/3G...Evolution to 4G.* John Wiley and Sons, 2006.

I. Mitrani. Approximate solutions for heavily loaded markov-modulated queues. *Performance Evaluation*, 62(1-4):117–131, 2005.

I. Mitrani and R. Chakka. Spectral expansion solution of a class of markov models: Application and comparison with the matrix-geometric method. *Performance Evaluation*, 23:pp. 241–260, 1995.

Isi Mitrani. *Queues with Breakdowns, Performability Modelling:Techniques and Tools*. Wiley, Chichester, 2001.

Isi Mitrani. Spectral expansion solutions for markov-modulated queues. pages 17–35, 2002.

A. Murase, I. Symington, and E. Green. Handover criterion for macro and microcellular systems. *Gateway to the Future Technology in Motion:* 41^{st} *IEEE Vehicular Technology Conference*, pages 524–530, 1991.

M.F. Neuts. *Matrix Geometric Solutions in Stochastic Model*. Johns Hopkins University Press, Baltimore, 1981.

A. Noerpel and Y. B. Lin. Handover arrangement for a pcs network. *IEEE Personal Communications*, 4(6):18–24, 1997. ISSN 1070-9916.

P.V. Orlik and S.S. Rappaport. A model for teletraffic performance and channel holding time characterization in wireless cellular communication with general session and dwell time distributions. *IEEE Journal of Selected Areas in Comminications*, 16(5):788–803, 1998.

B. Paramvir, A. Adya, J. Padhye, and A. Walman. Reconsidering wireless systems with multiple radios 2004. *SIGCOMM Computer Communications Review*, 35 (5):39–46, 2004.

G. P. Pollini. Trends in handover design. *IEEE Communications Magazine*, 34 (3):82–90, 1996.

E. H. Raul and L. G. Jung. Synchronization in a wireless node. *United States Patent Application*, September 2011.

H. Rutagemwa. *Performance Modeling, Design and Analysis of Transport Mechanisms in Integrated Heterogeneous Wireless Networks*. PhD thesis, Electrical and Computer Engineering, University of Waterloo, 2007.

H. Sadia, Z. Hamid, and N. S. Khattak. Mobility management challenges and issues in 4g heterogeneous networks. *In Proceedings of the first international conference on integrated internet ad hoc and sensor networks*, 14, 2006.

A. K. Salkintzis, C. Fors, and R. Pazhyannur. Wlan-gprs integration for ng mobile data netwoks. *IEEE Wireless Communications*, 9:112–124, 2002.

I. Saravanan, G. Sivaradje, and P. Dananjayan. Qos provisioning for cellular/wlan interworking. *IFIP International Conference on Wireless and Optical Communications Networks*, pages 50–55, 2006.

T. Shensheng and L. Wei. Performance analysis of the 3g network with complementary wlans. *Global Telecommunications Conference (GLOBECOM)*, 5: 2636–2641, December 2005.

J. So and N. Vaidya. Multi-channel mac for ad hoc networks: handling multi-channel hidden terminals using a single transceiver. *ACM International Symposium on Mobile Ad Hoc Networking and Computing (MOBIHOC)*, pages 222–233, 2004.

G. Song, L. Cuthbert, and J. Schormans. Modelling cellular/wireless lan integrated systems with multi-rate traffic using queueing network. *International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM)*, pages 1–4, 2008a.

G. Song, L. Cuthbert, and J. Schormans. Open networks of loss systems in the modelling of tow-tier heterogeneous hierarchical overlay wireless networks. *International Conference on Communications and Networking in China*, pages 148–152, 2008b.

J. Y. Song, H. J. Lee, S. H. Lee, S. W. Lee, and D. H. Cho. Hybrid coupling schemes for umts and wireless lan interworking. *AEU International Journal on Electronics and Communications*, 61:329–336, 2007.

W. Song and W. Zhuang. Qos provisioning via admission control in cellular/wireless lan interworking. *2nd International Conference on Broadband Networks*, 1:543–550, 2005.

W. Song, H. Jiang, W. Zhuang, and X. Shen. Resource management for qos support in cellular/wlan interworking. *IEEE Network*, 19(5):12–18, October 2005.

R. Steele and M. Nofal. Teletraffic performance of microcellular personal communication networks. *IEE Communications, Speech and Vision*, 139(4):448 –461, aug. 1992. ISSN 0956-3776.

M. Stemm and R.H. Katz. Vertical handoffs in wireless overlay networks. *Mobile Networks and Applications - Special issue: mobile networking in the Internet*, 3(4):335–350, 1998.

E. Stevens-Navarro, V. W.S. Wong, and Y. Lin. A vertical handoff decision algorithm for heterogeneous wireless networks. *IEEE Wireless Communications and Networking Conference (WCNC07)*, pages 3201–3206, 2007.

Y.C. Tay and K.C. Chua. A capacity analysis for the ieee 802.11 mac protocol. *ACM/Baltzer Wireless Networks*, 7:159171, 2001.

S. Tekinay and B. Jabbari. Handover and channel assignment in mobile cellular networks. *IEEE Communications Magazine*, 29(11):42–46, 1991.

K. S. Trivedi and X. Ma. Performability analysis of wireless cellular networks. *Symposium on Performance Evaluation of Computer and Telecommunication Systems*, SPECTS2002, July 2002.

K. S. Trivedi, A.S. Sathaye, O.C. Ibe, and R.C. Howe. Should i add a processor? *In Proceedings of 23rd Annual Hawaii International Conference on System Sciences, IEEE Computer Society Press*, pages pp. 214–221, 1990.

K. S. Trivedi, M. Malhotra, and R. M. Fricks. Markov reward approach to performability and reliability analysis. pages 7–11, 1994.

K. S. Trivedi, O. C. Ibe, R. C. Howe, and A. Agrawal. Availability and performance-based sizing of multiprocessor systems. *Communications in Reliability, Maintainability and Serviceablity:An International Journal*, 2(2):pp. 14–23., 1996.

K. S. Trivedi, S. Dharmaraja, and X. Ma. Analytic modeling of handoffs in wireless cellular networks. *Information Sciences, Applications: An international Journal*, 148:155–166, December 2002. ISSN 0020-0255.

K. S. Trivedi, S. Dharmaraja, and X. Ma. Performability modelling of wireless communication systems. *International Journal of Communication Systems*, 16: 561–577, 2003.

S. K. Trivedi. *Probability and Statistics with Reliability, Queuing, and Computer Science Applications*. 9814-12-604-7. Wiley, NY, USA,, 2002.

M. Usman, W. Noshairwan, M. Gilani, A. Irshad, and E. Irshad. Seamless vertical handoff using authentication certificate in gprs-wlan tightly coupled integrated networks. 4^{th} *International Conference on emerging Technologies, ICET*, pages 199–204, 2008.

U. Varshney and R. Jain. Issues in emerging 4g wireless networks. *IEEE Computer*, 34(6):94–96, 2001.

R. Vijayan and J.M. Holtzman. Analysis of handoff algorithms using nonstationary signal strength measurements. *IEEE Global Telecommunications Conference (GLOBECOM)*, pages 1405–1409, 1992.

F. Wang, A. Ghosh, C. Sankaran, P. J. Fleming, F. Hsieh, and S. J. Benes. Mobile wimax systems: performance and evolution. *IEEE Communications Magazine*, 46:41–49, 2008.

H.J. Wang, R.H. Katz, and J. Giese. Policy-enabled handoffs across heterogeneous wireless networks. *IEEE* 2^{nd} *Workshop on Mobile Computing Systems and Applications, (WMCSA)*, pages 51–60, 1999.

J. Wang, Q.A. Zeng, and D. P. Agrawal. Performance analysis of a preemptive and priority reservation handoff scheme for integrated service-based wireless mobile networks. *IEEE Transactions on Mobile Computing*, 2(1):65–75, March 2003.

Y. Wen and D. Ying. Hybrid-dynamic-system-based modeling and controlling of cdma soft handoff. 5^{th} *World Conference on Intelligent Control and Automation*, 2:1045–1050, 2004.

E.M.M. Winands. A finite-source feedback queueing network as a model for the ieee 802.11 distributed coordination function. Masters thesis, Department of Mathematics and Computer Science, Eindhoven University of Technology, 2003.

E.M.M. Winands, T.J.J. Denteneer, J.A.C. Resing, and R. Rietman. A finite-source feedback queueing network as a model of the ieee 802.11 distributed coordination function. *Proceedings of European Wireless*, pages 551–557, 2004.

W. Xia and L. Shen. Modeling and analysis of handoffs in cellular and wlan integration. *IEEE International Conference on Communications ICC '07I*, pages 385–391, June 2007.

W. Xia and L. Shen. Modelling and analysis of hybrid cellular/wlan systems with integrated service-based vertical handoff schemes. *IEICE Transactions on Communications*, pages 2032–2043, 2009.

H. Xie and S. Kuek. Priority handoff analysis. *International Journal of Wireless Information Networks*, 1:141–148, 1994. ISSN 1068-9605.

J. Yin, Q. A. Zeng, and D.P. Agrawal. Performance evaluation of ieee 802.11 dcf in error-prone channel and ideal channel. *Proceedings of the Communication Networks and Distributed Systems Modeling and Simulation Conference*, 2:1142–1145, January 2004.

Q. A. Zeng and D. P. Agrawal. Modeling and efficient handling of handoffs in integrated wireless mobile networks. *IEEE Transactions on Vehicular Technology*, 51(6):1469–1478, 2002.

Q. A. Zeng and D.P. Agrawal. An analytical modeling of handoff for integrated voice/data wireless networks with priority reservation and preemptive priority procedures. *IEEE International Workshop on Parallel Processing*, pages 523–530, 2000a.

Q. A. Zeng and D.P. Agrawal. Performance analysis of a handoff scheme in integrated voice/data wireless networks. *IEEE Vehicular Technology Conference*, 4:1986–1992, 2000b.

Q. A. Zeng, K. Mukumoto, and A. Fukuda. Performance analysis of mobile cellular radio system with priority reservation handoff procedures. *Proceeding of the IEEE VTC*, pages 1829–1833, June 1994a.

Q.A. Zeng and D. P. Agrawal. Modeling of handoffs and performance analysis of wireless data networks. *International Conference on Parallel Processing Workshops*, 2001.

Q.A. Zeng, K. Mukumoto, and A. Fukuda. Performance analysis of mobile cellular radio system with priority reservation handoff procedures. *IEEE Vehicular Technology Conference*, pages 1829–1833, 1994b.

N. Zhang and J.M. Holtzman. Analysis of handoff algorithms using both absolute and relative measurements. *IEEE Transactions on Vehicular Technology*, 45 (1):174–179, 1996.

Y. Zhao and K. K. Leung. Adaptive channel allocation for ieee 802.11 wireless lans. *The 12th European Wireless Conference (EW2006)*, 2006.

X. Zhong, Y. Zhenqiang, K. Srikanth, T. Satish, and K. Molle. A new adaptive channel reservation scheme for handoff calls in wireless cellular networks. *Proceedings of the Second International IFIP-TC6 Networking Conference on Networking Technologies, Services, and Protocols; Performance of Computer and Communication Networks; and Mobile and Wireless Communications*, pages 672–684, 2002.