Modeling Excess Life in Cellular Telecommunications Networks

Student: Pei-Chun Lee

Advisors: Dr. Yi-Bing Lin and Dr. Hui-Nien Hung

A Dissertation Submitted to

Department of Computer Science

College of Computer Science

National Chiao Tung University

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

in

Computer Science

Hsinchu, Taiwan, Republic of China

March, 2006

Abstract in Chinese

Modeling Excess Life in
Cellular Telecommunications Networks

Student: Pei-Chun Lee

Advisors: Dr. Yi-Bing Lin

Dr. Hui-Nien Hung

Department of Computer Science

National Chiao Tung University

ABSTRACT

Many performance evaluation issues for cellular telecommunications are related to excess life modeling. The period when a *mobile station* (MS) resides in a *cell* (the radio coverage of a base station) is called the *MS cell residence time.* The period between when a call arrives at the MS and when the MS moves out of the cell is called the *excess life* of the MS cell residence time for that MS. In performance evaluation of a cellular telecommunications network, it is important to derive the excess life distribution from the MS cell residence times. This distribution determines if a connected call will be handed over to a new cell, and therefore significantly affects the handover force-termination probability of the network. In simulation of cellular telecommunications networks, we need to generate random numbers for the excess life of non-exponential distributions. However, generating these random numbers for non-exponential distributions is not a trivial task, which has not been addressed in the literature. We show how to generate the random numbers from the excess life distribution, and develop the excess-life random number generation procedures for MS cell residence times with gamma, Pareto, lognormal and Weibull distributions. We use two examples to show how our excess life modeling techniques can be effectively utilized in performance evaluation of cellular networks.

In the first example, we study the *periodic location area update* (PLAU) scheme, which is utilized in cellular telecommunications networks to detect the presence of

an MS. In 3GPP Technical Specifications (TS) 23.012 and 24.008, a fixed PLAU scheme was proposed for *Universal Mobile Telecommunications System* (UMTS), where the interval between two PLAUs is of fixed length. We observe that MS presence can also be detected through call activities and *normal location area updates* (NLAUs). Therefore we propose a dynamic PLAU scheme where the PLAU interval is dynamically adjusted based on the call traffic and NLAU rate. An analytic model is developed to validate against the simulation model with the excess life modeling technique. Then we investigate the performance of dynamic and fixed PLAU schemes. Our study provides guidelines to select parameters for the dynamic PLAU scheme.

In the second example, we study the new call blocking, handover force-termination, and call incompletion probabilities for a small-scale cellular network. We show that the handover traffic to a cell depends on the workloads of the neighboring cells. Based on this observation, we derive the exact equation for the handover force-termination probability when the MS cell residence times are exponentially distributed. Then we propose an approximate model with general MS cell residence time distributions. We use the analytic model to validate against the simulation model with the excess life modeling technique. Then the analytic results are compared with a previously proposed model, where the simulation results are used as the baseline. Our comparison study indicates that the new model can capture the handover behavior much better than the old one for small-scale cellular telecommunications networks.

Key words: cellular telecommunications network, mobility management, periodic location area update (PLAU), Universal Mobile Telecommunications System (UMTS), channel assignment, handover, call holding time, MS cell residence time, excess life

Acknowledgment

I would like to express my sincere thanks to my advisors, Prof. Yi-Bing Lin and Prof. Hui-Nien Hung. Without their supervision and perspicacious advice, I can not complete this dissertation. Special thanks to Prof. Ming-Feng Chang, Prof. Ming-Syan Chen, Prof. Ge-Ming Chiu, Dr. Sheng-Lin Chou, Dr. Herman Chung-Hwa Rao, and Prof. Wen-Nung Tsai for their valuable comments. Thanks also to the colleagues in Laboratory 117.

In addtion, I am grateful to my family, church, and friends for their kind encouragement and support during these years.
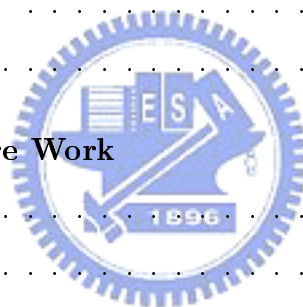
Last but not the least, I give my wholehearted thanks to God, my Heavenly Father, my Lord Jesus Christ and Holy Spirit. He has been breaking me and then strengthening me again and again. It is Him to make me stronger than ever before, as well as humbler. Thank Him for all He has done for me and in me during these years. In Jesus' name, amen.

Contents

List of Tables

List of Figures

Abbreviation

The abbreviations used in this dissertation are listed below.

3GPP: 3rd Generation Partnership Project

AB_DETACH: MS Abnormal Detach

BCCH: Broadcast Control Channel

BS: Base Station

CDF: Cumulative Distribution Function

CKPNT: Checkpoint

GPRS: General Packet Radio Service

GSM: Global System for Mobile Communications

ID: Implicit Detach

i.i.d.: independent and identically distributed

L3-RRC: Layer 3-Radio Resource Control

LA: Location Area

MS: Mobile Station

NLAU: Normal Location Area Update

PLAU: Periodic Location Area Update

TAM: Transform Approximation Method

TS: Technical Specification

UMTS: Universal Mobile Telecommunications System

VLR: Visitor Location Register

Chapter 1

Introduction

Emerging *cellular telecommunications network* technologies have attracted considerable attention in academic research as well as commercial deployment. A cellular telecommunications network supports telephony services when users are in movement [30]. A typical cellular phone service area is populated with several base stations (BSs). The radio coverage of a BS is referred to as a *cell*. Mobile users within a cell can connect to the corresponding BS via mobile stations (MSs). *Mobility management* plays an important role in cellular telecommunications networks, which consists of two parts: *location management* and *handover management*. Location management includes two tasks: *location update* and *paging*. Handover management is the process by which a communicating MS keeps its connection alive when it moves from one BS to another. We describe these two parts as follows.

Location management: In order to track the MSs, the cells in a cellular network service area are grouped into several location areas (LAs). To deliver services to an MS, the cells in the LA covering the MS will page the MS to establish the radio link. To identify the LA of an MS, location management is required. In location management, the MS informs the network of its location through the LA update procedure. The update procedure is executed in two situations.

Normal location area update (NLAU) is performed when the location of an MS has been changed. Location change of an MS is detected as follows.

The cells continuously broadcast their cell identities. The MS periodically listens to the broadcast cell identity, and compares it with the cell identity stored in the MS's buffer. If the comparison indicates that the LA has been changed, then the MS sends the LA update message to the network.

Periodic location area update (PLAU) allows an MS to periodically report its "presence" to the network even if the MS does not move. A *periodic LA update timer* (PLAU timer) is maintained in the MS. Corresponding to the PLAU timer, an *implicit detach* (ID) timer is maintained in the cellular network. When the PLAU timer expires, the MS performs PLAU.

Note that NLAU has been intensively investigated in the literature (see [5, 30] and the references therein). Most studies on PLAU focused on mobility database failure restoration [12, 26].

Handover management: When a call for a mobile user occurs, one radio channel of the BS is used for connecting the MS and the BS. If all radio channels are in use when a new call is attempted, the call will be blocked and cleared from the system. If the call is accepted, a radio channel will be occupied until the call completes, or until the MS moves out of the cell. When a communicating MS moves from one cell to another, the occupied channel in the old cell is released, and an idle channel is acquired in the new cell. This process is called *handover*. During this handover procedure, if no channel is available in the new cell, the call is forced to terminate before its completion. When the call is connected, the call may complete at the first cell or complete after several successful handovers. On the other hand, the call may also be forced to terminate due to a failed handover.

The above issues are related to *excess life* modeling. For example, when we evaluate the performance of the handover management, we need to consider how long a communicating MS would reside in a cell; that is, we need to consider the *MS cell*

residence times, which have significant impact on the performance of the handover management. Usually, we assume that the MS cell residence times are independent and identically distributed (i.i.d.) random variables with the same distribution. However, the period between when a call arrives at an MS and when the MS moves out of the "first" cell (where the call arrives at the MS) is actually the excess life of the MS cell residence time. If we assume that the MS cell residence times have an exponential distribution, the excess life distribution of the MS cell residence time is exactly the same as the MS cell residence time distribution. However, if the MS cell residence times follow a non-exponential distribution, then the excess life distribution of the MS cell residence time is not the same as the original MS cell residence time distribution, and generating random numbers for these excess life distributions is not a trivial work in cellular network simulations. This issue has not been addressed in the literature. Therefore, we show how to generate the random numbers from the excess life distribution, and develop the excess-life random number generation procedures for MS cell residence times with gamma, Pareto, lognormal and Weibull distributions. We use two examples to show how our excess life modeling techniques can be effectively utilized in performance evaluation of cellular networks.

In the first example, we investigate the PLAU scheme, which is utilized in cellular networks to detect the presence of an MS. Consider the circuit-switched domain of UMTS [2, 3]. In 3GPP TS 23.012 and 24.008, a fixed PLAU scheme was proposed for UMTS, where the interval between two PLAUs is of fixed length. We observe that MS presence can also be detected through call activities and NLAUs. Therefore, we propose a dynamic PLAU scheme where the PLAU interval is dynamically adjusted based on the call traffic and NLAU rate. An analytic model is developed in Chapter 3 to validate against the simulation model with the excess life modeling technique. Then we investigate the performance of dynamic and fixed PLAU schemes. Our study provides guidelines to select parameters for the dynamic PLAU scheme.

In the second example, we study the new call blocking, handover force-termination, and call incompletion probabilities for a small-scale cellular network. We show that the handover traffic to a cell depends on the workloads of the neighboring cells.
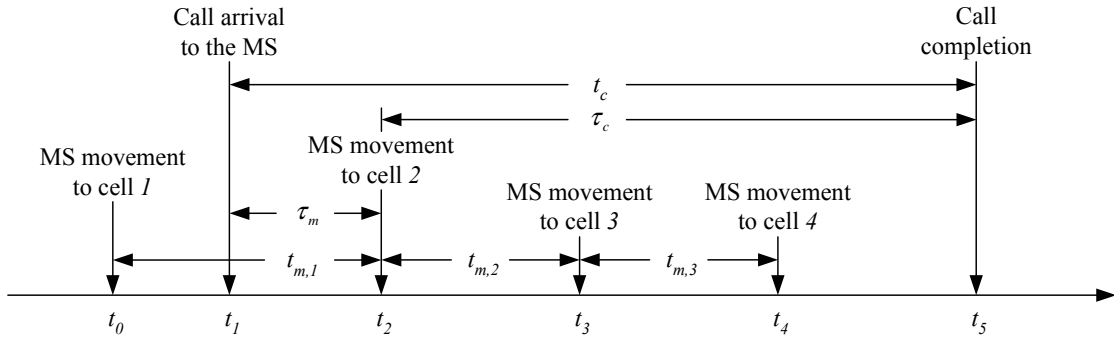
Figure 1.1: The Timing Diagram for MS Movement and Call Arrival

Based on this observation, we derive the exact equation for the handover force-termination probability when the MS cell residence times are exponentially distributed. Then we propose an approximate model with general MS cell residence time distributions. We use the analytic model to validate against the simulation model with the excess life modeling technique. Then the analytic results are compared with a previously proposed model [34], where the simulation results are used as the baseline. Our comparison study indicates that the new model can capture the handover behavior much better than the old one for small-scale cellular networks.

In the remainder of this chapter, we use the timing diagram in Fig. 1.1 (see Section 1.1 for more details) to describe the notation used in this and latter chapters. Then we briefly overview the contents in this dissertation. Finally, we present the organization of this dissertation.

The notation includes

- η: the MS mobility rate (i.e., $1/\eta$ is the mean MS cell residence time)

- λ: the call arrival rate (i.e., $1/\lambda$ is the mean inter call arrival time)

- p_{nc}: the call incompletion probability

- t_c: the call holding time (if the call successfully completes)

4

- t_i $(i > 1)$: the time when the MS moves to cell i after a call is connected to it (note that Fig. 1.1 is only an example and i can be any possible number other than 4)

- $t_{m,1}$: the time interval that the MS resides in cell 1 (i.e., $t_{m,1}$ is the MS cell residence time in cell 1; in Fig. 1.1, $t_{m,1} = t_2 - t_0$)

- $t_{m,i}$ $(i > 1)$: the time interval that the MS resides in cell i (i.e., $t_{m,i}$ is the MS cell residence time in cell i; in Fig. 1.1, $t_{m,i} = t_{i+1} - t_i$; note that Fig. 1.1 is only an example, and i can be any possible number other than 4)

- τ_c: the remaining call holding time (i.e., τ_c is the excess life of t_c; in Fig. 1.1, τ_c can be $t_5 - t_2$ or $t_5 - t_3$ or $t_5 - t_4$)

- τ_m: the period between when a call arrives at an MS and when the MS moves out of the first cell (i.e., τ_m is the excess life of the MS cell residence time; in Fig. 1.1, $\tau_m = t_2 - t_1$ is the excess life of $t_{m,1}$)

1.1 Random Number Generation for Excess Life of MS Cell Residence Time

Figure 1.1 illustrates the relationship between movement of an MS and a call session to that MS. The MS moves to cell 1 at time t_0, and then moves to cell i at time t_i for $i > 1$. A call for the MS arrives at time t_1. If the call is not blocked or forced to terminate, it completes at time t_5 (note that this is only an example; the call may successfully complete at any time point according to the call holding time t_c). At time t_1, if cell 1 does not have enough radio resources to accommodate this call (which can be a plain voice call or a multimedia call), the call is *blocked*. When the MS moves to cell i, the call is handed over from cell $i - 1$ to cell i (for $i > 1$). If no radio resources are available in cell i (for $i > 1$), the call is *forced to terminate*. Performance of a cellular network is typically evaluated by the *new call blocking probability* (a new call attempt is blocked), the *handover force-termination probability*

(a handover call is forced to terminate), and the *call incompletion probability* (a call is either blocked or forced to terminate).

Many studies [7, 8, 25, 15, 45] have been devoted to evaluate these probabilities for various radio resource allocation strategies exercised in cellular telecommunications networks. Most of them utilized analytic approaches that provide useful insights to cellular network modeling. However, analytic analysis has its limitations. For example, in Figure 1.1, if the *call holding time* $t_c = t_5 - t_1$ is non-exponential (which is probably true for multimedia calls) [6], then it is difficult to derive the remaining call holding time $\tau_c = t_5 - t_i$ after the MS moves into cell i (for $i > 1$). Furthermore, most analytic studies made an approximate assumption that the handover traffic to a cell is a fixed Poisson Process. This assumption is reasonable for large-scale cellular networks, but may result in significant inaccuracy for small-scale cellular networks [49, 20]. In reality, the handover traffic to a cell depends on the workloads of the neighboring cells (i.e., the numbers of busy channels in neighboring cells are not independent). Also, if the resource allocation policies under consideration are very complicate (which is probably true for wireless data sessions with QoS), it is impossible to find analytic solutions.

An alternative modeling technique to analytic analysis is discrete event simulation. There are two approaches to cellular network simulation: the *MS-based* simulation and the *call-based* simulation. In the MS-based simulation, the number of MSs are defined in the simulation, and the MS objects are actually simulated for their movements (even if there are no calls destinated at these MSs). Examples of MS-based simulation can be found in [29]. In the call-based simulation [32, 24], the call arrival rate to the network is considered as the input that drives the simulation progress. In this approach, after a call arrival event is processed, the corresponding MS movement and the call termination events are generated following the timing diagram illustrated in Figure 1.1 (details of the call-based simulation is described in Appendix A). When the number of MSs is small in a cellular network, the MS-based simulation will produce more accurate results than the call-based simulation. When the number of MSs is large, both approaches produce results with similar accuracies.

On the other hand, the execution time for the MS-based simulation is much longer than that for the call-based simulation (e.g., 100 times longer [29]). Since large MS population is expected in most third generation systems such as UMTS (Universal Mobile Telecommunications System) [1, 30], the call-based simulation will become more important in advanced cellular telecommunications studies.

In cellular network modeling, several random variables are defined. Two of them are elaborated here; others are described in Appendix A. In Figure 1.1, $t_{m,1} = t_2 - t_0$ is the time interval that the MS resides in cell 1, and $t_{m,i} = t_{i+1} - t_i$ (for $i > 1$) are the time intervals that the MS resides in cell i. These MS cell residence times are typically modeled by a random variable with a specific distribution such as gamma and mixed Erlang [32, 24, 10]. The interval $\tau_m = t_2 - t_1$ is the period between when a call arrives and when the MS moves out of the first cell, which is referred to as the excess life of the MS cell residence time. In the call-based simulation, it is required to generate the random numbers for the excess life τ_m (see Appendix A). Clearly, the τ_m distribution must be derived from the MS cell residence time distribution. The call arrivals are typically assumed to be random observers of the MS cell residence times. If the MS cell residence times have the exponential distribution, then τ_m also has the same exponential distribution [42]. On the other hand, if the MS cell residence times have an arbitrary distribution, generation of the τ_m random numbers is a non-trivial task. We describe how to generate the τ_m random numbers from the MS cell residence time distribution. For various MS cell residence time distributions, generation of τ_m random numbers need separate treatments. We show how to generate the excess-life random numbers for MS cell residence time random variables with gamma, Pareto, lognormal and Weibull distributions. Our study indicates that the generated random numbers closely match the true excess life distributions.

7

1.2 Dynamic Periodic Location Area Update

A major purpose of PLAU is to allow the network to detect if an MS is still attached to the network in the normal network operation situation (i.e., the mobility databases do not fail). To our knowledge, this aspect has not been investigated in the literature. An important issue for PLAU is the selection of the value for the PLAU/ID timers. In 3GPP TS 23.012 [2] and TS 24.008 [3], the value of the PLAU/ID timers is set/changed by the network and broadcasted to every MS in the LA through the L3-RRC SYSTEM INFORMATION BLOCK 1 message on the Broadcast Control Channel (BCCH). In this approach, the PLAU timer value is the same for all MSs in an LA. There are two issues regarding this *fixed PLAU* scheme:

- How is the value of the PLAU/ID timers determined when an MS first enters an LA?

- Is it appropriate to have a fixed value for PLAU/ID timers during the MS's stay in an LA?

The above two issues are addressed. The value of the PLAU/ID timers should be selected based on the call and movement activities of the MS. Therefore we propose a dynamic PLAU scheme and investigate its performance. Then we compare it with the fixed PLAU scheme. Our study provides guidelines to select parameters for dynamic PLAU scheme.

1.3 Modeling Channel Assignment of Small-Scale Cellular Networks

For billing and network planning purposes, the handover behavior and the probability of call completion need to be analyzed. Several analytic studies have contributed to cellular network performance evaluation [34, 10, 7, 48, 19, 13, 36, 39]. Most studies assume that the handover traffic to a cell is a fixed-rate Poisson process. This
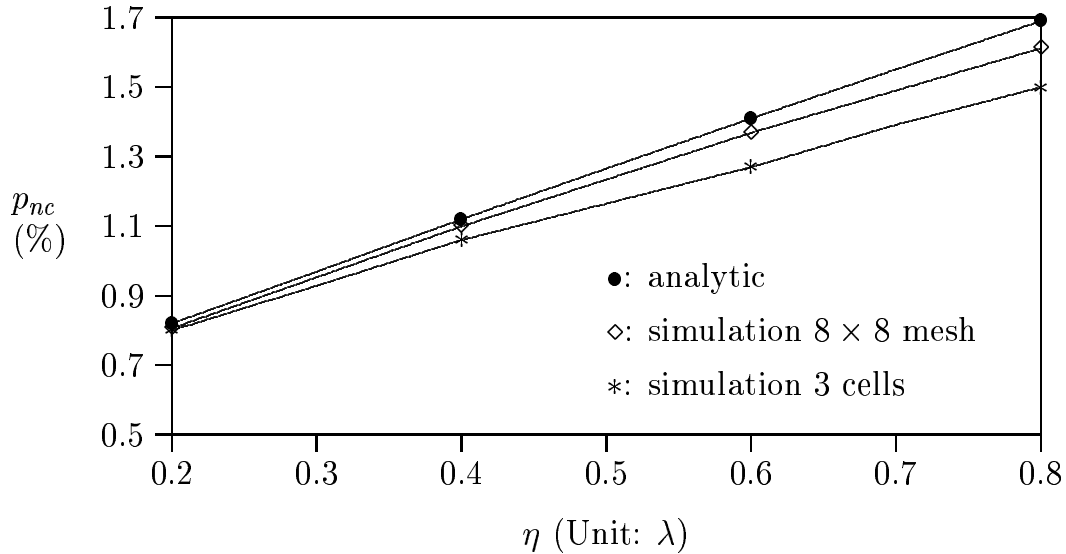
Figure 1.2: Effect of Network Size on the Call Incompletion Probability ($\mu = 0.3\lambda$)
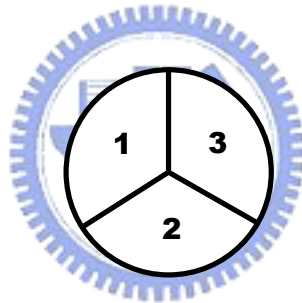
Figure 1.3: A Three-Cell Cellular System

assumption is reasonable for large-scale cellular networks, or when the networks experience light load traffic [9]. In reality, the handover traffic to a cell depends on the workloads of the neighboring cells. This fact has significant impact on modeling of small-scale cellular networks. Fig. 1.2 plots the call incompletion probability p_{nc} against the MS mobility rate η and the call arrival rate λ where the number of radio channels in a cell is 9. The "•" curve is generated from a previous analytic model that assumes fixed-rate handover traffic [34]. The "⋄" curve is generated from simulation of a 64-cell mesh configuration, and the "∗" curve is generated from simulation of a 3-cell configuration (illustrated in Fig. 1.3). Fig. 1.2 indicates that

9

the fixed-rate assumption is acceptable when the number of cells is reasonably large, but is inaccurate for small-scale cellular networks. We derive the exact equation for the handover force-termination probability when the MS cell residence times are exponentially distributed. Then we propose an approximate model with general MS cell residence times. We use the analytic model to validate against the simulation model with the excess life modeling technique. Then the analytic results are compared with the previously proposed model [34], where the simulation results are used as the baseline. Our comparison study indicates that the new model can capture the handover behavior much better than the old one for small-scale cellular networks.

This dissertation is organized as follows. Chapter 2 presents excess-life random number generation in cellular network simulation. Chapter 3 describes the first example: fixed and dynamic PLAU schemes in cellular networks. An analytic model is proposed to validate against the simulation model with the excess life modeling technique. Then the two PLAU schemes are compared. In Chapter 4, we describe the second example: channel assignment modeling of small-scale cellular networks. Chapter 5 concludes this dissertation and describes the future work.

Chapter 2

Random Number Generation for Excess life of MS Cell Residence Time

A cellular telecommunications network is populated with several base stations (BSs). Mobile users receive cellular telecommunications services by using mobile stations (MSs; or mobile terminals) connecting to the BSs. When an MS moves from the radio coverage (called *cell*) of a BS to the radio coverage of another BS, the MS is disconnected from the old BS and re-connected to the new BS. This process is called *handover*. The notation used in this chapter is listed below.

- α: the shape parameter of a gamma distribution

- β: the scale parameter of a gamma distribution

- a: the shape parameter of a Pareto distribution

- b: the scale parameter of a Pareto distribution

- (θ, σ): the parameters of a lognormal distribution

- γ: the shape parameter of a Weibull distribution

- ϕ: the scale parameter of a Weibull distribution

- t_c: the call holding time (if the call successfully completes)

- t_i $(i > 1)$: the time when the MS moves to cell i after a call is connected to it (note that Fig. 2.1 is only an example and i can be any possible number other than 4)

- $t_{m,1}$: the time interval that the MS resides in cell 1 (i.e., $t_{m,1}$ is the MS cell residence time in cell 1; in Fig. 2.1, $t_{m,1} = t_2 - t_0$)

- $t_{m,i}$ $(i > 1)$: the time interval that the MS resides in cell i (i.e., $t_{m,i}$ is the MS cell residence time in cell i; in Fig. 1.1, $t_{m,i} = t_{i+1} - t_i$; note that Fig. 1.1 is only an example, and i can be any possible number other than 4)

- t_m: an arbitrary MS cell residence time with the density function $f_m(t_m)$, the distribution function $F_m(t_m)$ and the mean μ

- $f_m(t_m)$: the density function of an MS cell residence time t_m

- $F_m(t_m)$: the distribution function of an MS cell residence time t_m

- μ: the mean of the MS cell residence time t_m (i.e., $\mu = \int_0^\infty t_m f_m(t_m) dt_m$)

- τ_c: the remaining call holding time (i.e., τ_c is the excess life of t_c; in Fig. 2.1, τ_c can be $t_5 - t_2$ or $t_5 - t_3$ or $t_5 - t_4$)

- τ_m: the period between when a call arrives at an MS and when the MS moves out of the first cell (i.e., τ_m is the excess life of the MS cell residence time t_m with the density function function $r_m(\tau_m)$ and the distribution function $R_m(\tau_m)$; in Fig. 2.1, $\tau_m = t_2 - t_1$ is the excess life of $t_{m,1}$)

- $r_m(\tau_m)$: the density function of the excess life of an MS cell residence time t_m

- $R_m(\tau_m)$: the distribution function of the excess life of an MS cell residence time t_m

- T: a random variable with the density function $f_T(t)$, where $f_T(t) = \dfrac{t f_m(t)}{\mu}$
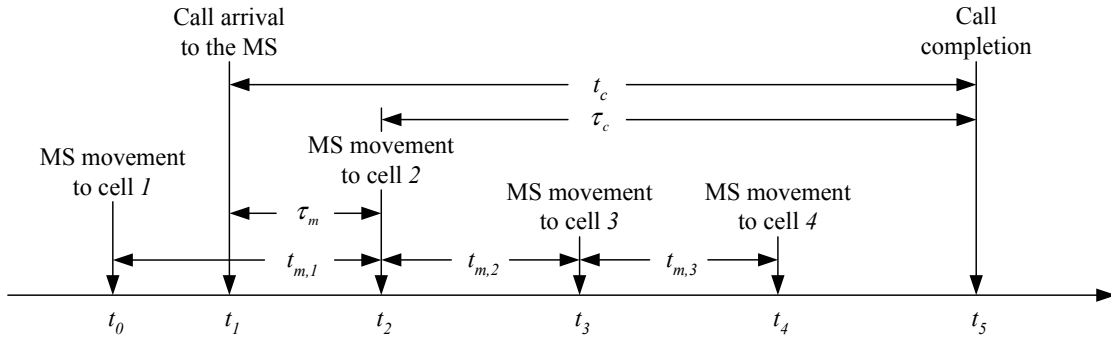
Figure 2.1: The Timing Diagram for MS Movement and Call Arrival (Redraw of Fig. 1.1)

- U: a uniform random variable over the interval $(0, 1)$

- $f_{(U,T)}(u, t)$: the joint density function of U and T

- W: $W = U \times T$

- $f_W(w)$: the density function of W
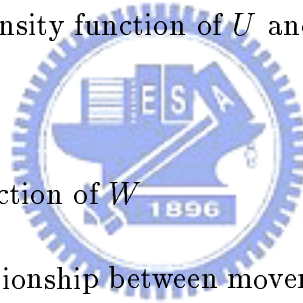
Figure 2.1 illustrates the relationship between movement of an MS and a call session to that MS. The MS moves to cell 1 at time t_0, and then moves to cell i at time t_i for $i > 1$. A call for the MS arrives at time t_1. If the call is not blocked or forced to terminate, it completes at time t_5. At time t_1, if cell 1 does not have enough radio resources to accommodate this call (which can be a plain voice call or a multimedia call), the call is *blocked*. When the MS moves to cell i, the call is handed over from cell $i - 1$ to cell i. If no radio resources are available in cell i, the call is *forced to terminate*. Performance of a cellular telecommunications network is typically evaluated by the *new call blocking probability* (a new call attempt is blocked), *handover force-termination* probability (a handover call is forced to terminate), and the *call incompletion probability* (a call is either blocked or forced to terminate).

13

Many studies [7, 8, 25, 15, 45] have been devoted to evaluate these probabilities for various radio resource allocation strategies exercised in cellular telecommunications networks. Most of them utilized analytic approaches that provide useful insights to cellular network modeling. However, analytic analysis has its limitations. For example, in Figure 2.1, if the *call holding time* $t_c = t_5 - t_1$ is non-exponential (which is probably true for multimedia calls) [6], then it is difficult to derive the remaining call holding time $\tau_c = t_5 - t_i$ after the MS moves into cell i (for $i > 1$). Furthermore, most analytic studies made an approximate assumption that the handover traffic to a cell is a fixed Poisson Process. This assumption is reasonable for large-scale cellular telecommunications networks, but may result in significant inaccuracy for small-scale networks [49, 20]. Also, if the resource allocation policies under consideration are very complicate (which is probably true for wireless data sessions with QoS), it is impossible to find analytic solutions.

An alternative modeling technique to analytic analysis is discrete event simulation. There are two approaches to cellular telecommunications network simulation: the *MS-based* simulation and the *call-based* simulation. In the MS-based simulation, the number of MSs are defined in the simulation, and the MS objects are actually simulated for their movements (even if there are no calls destinated at these MSs). Examples of MS-based simulation can be found in [29]. In the call-based simulation [32, 24], the call arrival rate to the network is considered as the input that drives the simulation progress. In this approach, after a call arrival event is processed, the corresponding MS movement and the call termination events are generated following the timing diagram illustrated in Figure 2.1 (details of the call-based simulation is described in Appendix A). When the number of MSs is small in a cellular telecommunications network, the MS-based simulation will produce more accurate results than the call-based simulation. When the number of MSs is large, both approaches produce results with similar accuracies. On the other hand, the execution time for the MS-based simulation is much longer than that for the call-based simulation (e.g., 100 times longer [29]). Since large MS population is expected in most third generation systems such as UMTS (Universal Mobile Telecommunications System) [1, 30],

the call-based simulation will become more important in advanced cellular telecom-munications studies.

In cellular telecommunications network modeling, several random variables are defined. Two of them are elaborated here; others are described in Appendix A. In Figure 2.1, $t_{m,1} = t_2 - t_0$ is the time interval that the MS resides in cell 1, and $t_{m,i} = t_{i+1} - t_i$ (for $i > 1$) are the time intervals that the MS resides in cell i. These *MS cell residence times* are typically modeled by a random variable with a specific distribution such as gamma and mixed Erlang [32, 24, 10]. The interval $\tau_m = t_2 - t_1$ is the period between when a call arrives and when the MS moves out of the first cell, which is referred to as the *excess life* of the MS cell residence time. In the call-based simulation, it is required to generate the random numbers for the excess life τ_m (see Appendix A). Clearly, the τ_m distribution must be derived from the MS cell residence time distribution. The call arrivals are typically assumed to be random observers of the MS cell residence times. If the MS cell residence times have the exponential distribution, then τ_m also has the same exponential distribution [42]. On the other hand, if the MS cell residence times have an arbitrary distribution, generation of the τ_m random numbers is a non-trivial task. In this chapter, we describe how to generate the τ_m random numbers from the MS cell residence time distribution. For various MS cell residence time distributions, generation of τ_m random numbers need separate treatments. We show how to generate the excess-life random numbers for MS cell residence time random variables with gamma, Pareto, lognormal and Weibull distributions. Our study indicates that the generated random numbers closely match the true excess-life distributions.

2.1   Derivation of Excess Life Distribution

In Figure 2.1, the MS cell residence times $t_{m,i}$ ($i \geq 1$) of an MS are assumed to be i.i.d. random variables. Therefore, we use t_m to represent an arbitrary MS cell residence time with the density function $f_m(t_m)$, the distribution function $F_m(t_m)$ and the mean μ. Let τ_m be the excess life of t_m with the density function $r_m(\tau_m)$

and the distribution function $R_m(\tau_m)$. Since the call arrivals form a Poisson process, a call arrival is a random observer of the MS cell residence times. From the excess life theorem [42], we have

$$r_m(\tau_m) = \frac{1 - F_m(\tau_m)}{\mu} \tag{2.1}$$

It is difficult to generate the random numbers for the excess life of an MS cell residence time random variable using Equation (2.1) because this equation involves the distribution function $F_m(\tau_m)$. To efficiently generate the random numbers τ_m, we shall utilize a variation of $f_m(\tau_m)$. We will prove that $r_m(\tau_m)$ can be derived from the following function

$$f_T(t) = \frac{t f_m(t)}{\mu} \tag{2.2}$$

Since

$$\int_{t=0}^{\infty} \left[\frac{t f_m(t)}{\mu} \right] dt = \left(\frac{1}{\mu} \right) \int_{t=0}^{\infty} t f_m(t) dt = \frac{\mu}{\mu} = 1$$

it is obvious that $f_T(t)$ can be a density function. Let T be a random variable with the density function $f_T(t)$. We have the following Theorem:

Theorem 2.1. Let τ_m be the excess life of t_m. Let random variable U be uniformly distributed over the interval $(0, 1)$. Let T be random variable with the density function $f_T(t) = \dfrac{t f_m(t)}{\mu}$, and U and T are independent. Then the distribution of τ_m is the same as the distribution of $U \times T$.

Proof: The joint density function of U and T is

$$f_{(U,T)}(u,t) = \begin{cases} \dfrac{t f_m(t)}{\mu}, & \text{for } 0 < u < 1 \text{ and } t > 0 \\ 0 & \text{otherwise} \end{cases}$$

Let $W = U \times T$. Then

$$\begin{aligned} \Pr[W \le w] &= \Pr[U \times T \le w] \\ &= \int_{u=0}^{1} \int_{t=0}^{\frac{w}{u}} f_{(U,T)}(u,t) dt du \\ &= \int_{u=0}^{1} \int_{t=0}^{\frac{w}{u}} \frac{t f_m(t)}{\mu} dt du \end{aligned} \tag{2.3}$$

16

From (2.3), the density function $f_W(w)$ of W can be derived as

$$
\begin{aligned}
f_W(w) &= \frac{d \Pr[W \le w]}{dw} \\
&= \int_{u=0}^{1} \left(\frac{w}{u}\right) \left[\frac{f_m(w/u)}{\mu}\right] \times \left(\frac{1}{u}\right) du \\
&= \left(\frac{1}{\mu}\right) \int_{u=0}^{1} \left(\frac{w}{u^2}\right) f_m(w/u) du \qquad (2.4)
\end{aligned}
$$

Let $y = \dfrac{w}{u}$. Then (2.4) can be rewritten as

$$
\begin{aligned}
f_W(w) &= \frac{1}{\mu} \int_{y=w}^{\infty} f_m(y) dy \\
&= \frac{1 - F_m(w)}{\mu} \\
&= r_m(w)
\end{aligned}
$$

which means that $W = U \times T$ has the same distribution as τ_m.

Q.E.D.

Theorem 2.1 allows us to generate a τ_m random number using $f_m(\cdot)$ as follows: We first generate a random number u for the uniform random variable U in (0,1). Then we generate a random number t for the random variable T with the density function $f_T(t)$ (see (2.2)). Then we multiply t by u to obtain the random number for the excess life τ_m. Derivation of $f_T(t)$ is not a trivial task, and some $f_T(t)$ functions cannot not be derived from the corresponding $f_m(t)$ functions. In the next section, we show how to derive $f_T(t)$ for some popular distributions.

2.2 Excess-Life Random Number Generation: Some Examples

This section derives the T distributions for MS cell residence times with distributions such as gamma, Pareto, lognormal and Weibull. Then we show how to generate the excess-life random numbers using Theorem 2.1 and the T distributions.

2.2.1 The Gamma Distribution

Suppose that t_m has a gamma distribution with the shape parameter α and the scale parameter β. Then the mean value is $\mu = \alpha\beta$ and the density function $f_m(t_m)$ is

$$f_m(t_m) = \frac{e^{-\frac{t_m}{\beta}} t_m^{\alpha-1}}{\beta^\alpha \Gamma(\alpha)} \quad \text{for} \quad t_m \geq 0 \tag{2.5}$$

We have the following theorem:

Theorem 2.2. If t_m has a gamma distribution with the parameters (α, β), then T has a gamma distribution with the parameters $(\alpha + 1, \beta)$.

Proof: From (2.2) and (2.5), we have

$$
\begin{aligned}
f_T(t) &= \frac{te^{-\frac{t}{\beta}} t^{\alpha-1}}{\mu\beta^\alpha \Gamma(\alpha)} \quad \text{for} \quad t \geq 0 \\
&= \left[\frac{\beta e^{-\frac{t}{\beta}} t^\alpha}{\mu\beta^{\alpha+1} \Gamma(\alpha+1)} \right] \left[\frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} \right]
\end{aligned}
\tag{2.6}
$$

Since $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$ and $\mu = \alpha\beta$, (2.6) is re-written as

$$f_T(t) = \frac{e^{-\frac{t}{\beta}} t^\alpha}{\beta^{\alpha+1} \Gamma(\alpha+1)} \quad \text{for} \quad t \geq 0 \tag{2.7}$$

From (2.7), it is clear that T has the gamma distribution with parameters $(\alpha + 1, \beta)$.

Q.E.D.

Generation of an excess-life random number for gamma residence time with the parameters (α, β) includes the following steps: We first generate a uniform random number u in $(0, 1)$. Then according to Theorem 2.2, we generate a random number t for the gamma random variable T with the parameters $(\alpha + 1, \beta)$. By multiplying u and t, we obtain a random number for the excess life τ_m. Figure 2.2 plots the $r_m(\tau_m)$ function for gamma excess life. In this figure, the symbols "$\diamond$" and "$\bullet$" represent the values obtained from the random number generation. The solid and dashed curves are directly computed from Equation (2.1). The figure indicates that our random number generation procedure accurately generates the excess-life random numbers for the gamma MS cell residence times.

Figure 2.2: The $r_m(\tau_m)$ Function for Gamma Excess Life

2.2.2 The Pareto Distribution

Suppose that t_m has a Pareto distribution with the parameters (a, b), where a is the shape parameter and b is the scale parameter. Then the mean is

$$\mu = \begin{cases} \dfrac{ab}{a-1}, & \text{if } a > 1 \\ \\ \infty, & \text{if } 0 < a \leq 1 \end{cases} \tag{2.8}$$

and the density function is

$$f_m(t_m) = \frac{ab^a}{t_m^{a+1}}, \quad \text{where} \quad t_m \geq b, \quad a > 0, \quad \text{and} \quad b > 0 \tag{2.9}$$

We have the following theorem.

Theorem 2.3. Suppose that t_m has a Pareto distribution with the parameters (a, b), where $a > 1$. Then T has a Pareto distribution with the parameters $(a - 1, b)$.

Proof: From (2.2), (2.8) and (2.9), we have

$$\begin{aligned} f_T(t) &= \left(\frac{tab^a}{t^{a+1}} \right) \times \left(\frac{a-1}{ab} \right) \\ &= \frac{(a-1)b^{a-1}}{t^{(a-1)+1}} \end{aligned} \tag{2.10}$$

Equation (2.10) is a Pareto density function with the parameters $(a - 1, b)$.

Q.E.D.

By utilizing Theorems 2.1 and 2.3, the τ_m random number generation procedure for Pareto MS cell residence times is similar to that for gamma MS cell residence times. Figure 2.3 plots the $r_m(\tau_m)$ function for Pareto excess life. The figure indicates that our random number generation procedure accurately generates the excess-life random numbers for the Pareto MS cell residence times.
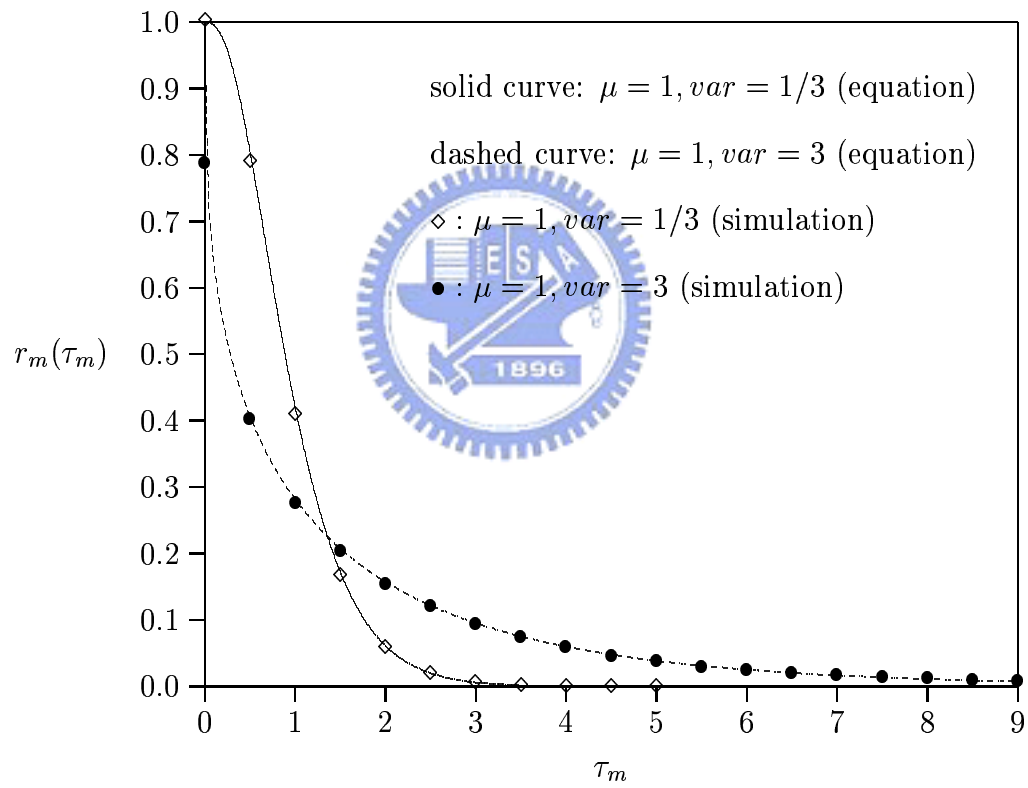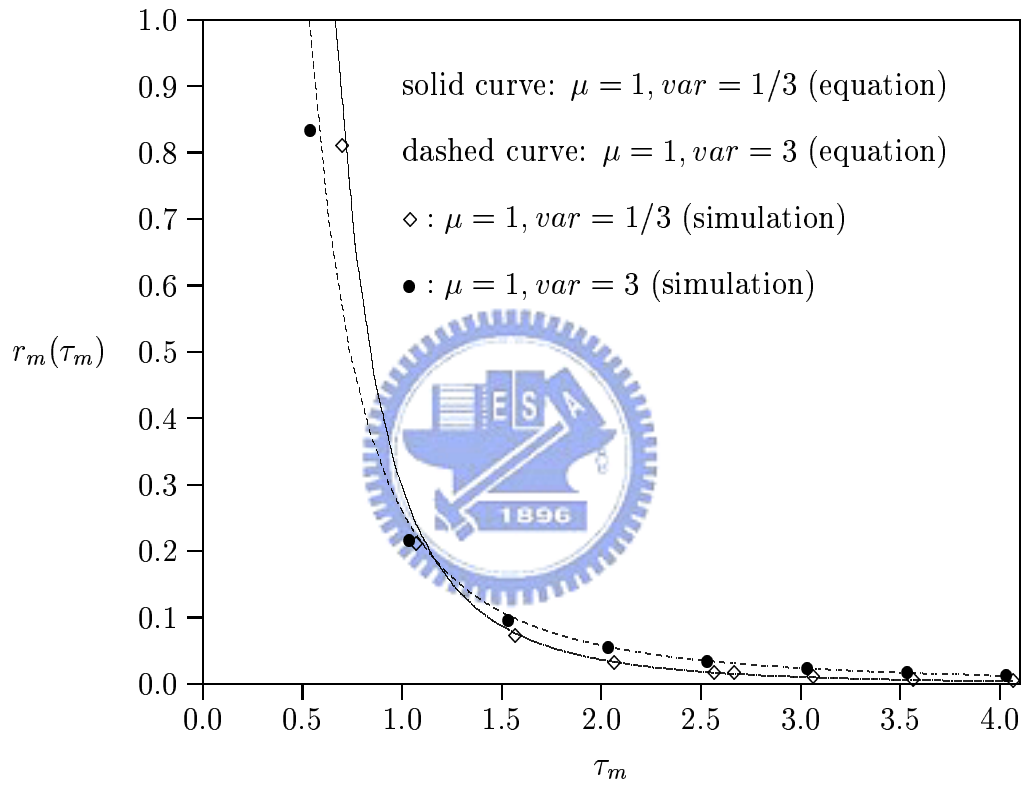
Figure 2.3: The $r_m(\tau_m)$ Function for Pareto Excess Life

2.2.3 The Lognormal Distribution

Suppose that t_m has a lognormal distribution with the parameters (θ, σ). Then the mean value is $\mu = e^{\theta + \frac{\sigma^2}{2}}$ and the density function $f_m(t_m)$ is

$$f_m(t_m) = \left(\frac{1}{\sigma t_m \sqrt{2\pi}} \right) e^{-\frac{(\ln t_m - \theta)^2}{2\sigma^2}} \quad \text{for} \quad t_m \geq 0 \tag{2.11}$$

We have the following theorem.

Theorem 2.4. Suppose that t_m has a lognormal distribution with the parameters (θ, σ). Let $Y = \ln T$. Then Y has a normal distribution with the mean $\mu + \sigma^2$ and the standard deviation σ.

Proof: From (2.2) and (2.11),

$$
\begin{aligned}
f_T(t) &= \left(\frac{1}{\mu \sigma \sqrt{2\pi}} \right) e^{-\frac{(\ln t - \theta)^2}{2\sigma^2}} \quad \text{where} \quad t \geq 0 \\
&= \left(\frac{1}{e^{\theta + \frac{\sigma^2}{2}} \sigma \sqrt{2\pi}} \right) e^{-\frac{(\ln t - \theta)^2}{2\sigma^2}} \tag{2.12}
\end{aligned}
$$

Since $Y = \ln T$, we have $T = e^Y$. According to the Jacobian of the transformation [35], the density function of Y is expressed as

$$f_Y(y) = f_T(e^y) \left| \frac{dt}{dy} \right| = f_T(e^y) \times e^y \tag{2.13}$$

where $-\infty < y < \infty$. Substitute (2.12) into (2.13) to yield

$$
\begin{aligned}
f_Y(y) &= \left(\frac{1}{e^{\theta + \frac{\sigma^2}{2}} \sigma \sqrt{2\pi}} \right) e^{-\frac{(y - \theta)^2}{2\sigma^2}} \times e^y \quad \text{where} - \infty < y < \infty \\
&= \left(\frac{1}{\sigma \sqrt{2\pi}} \right) e^{-\frac{[y - (\theta + \sigma^2)]^2}{2\sigma^2}} \tag{2.14}
\end{aligned}
$$

From (2.14), Y is a normal random variable with the mean $\theta + \sigma^2$ and the standard deviation σ.

Q.E.D.

Generation of an excess-life random number for lognormal MS cell residence time with the parameters (θ, σ) includes the following steps: We first generate a random

number u from the uniform random variable U in $(0,1)$. Then according to Theorem 2.4, we generate a random number y for the normal random variable Y with the mean $\theta + \sigma^2$ and the standard deviation σ. By multiplying u and e^y, we obtain a random number for the excess life τ_m. Details of the lognormal residence time curves will not be presented in this chapter.

2.2.4 The Weibull Distribution

Suppose that t_m has a Weibull distribution with the shape parameter γ and the scale parameter ϕ. Then the mean value is $\mu = \phi^{\frac{1}{\gamma}} \Gamma\left(1 + \frac{1}{\gamma}\right)$ and the density function $f_m(t_m)$ is

$$f_m(t_m) = \begin{cases} \left(\frac{\gamma}{\phi}\right) t_m^{\gamma-1} e^{-\frac{t_m^{\gamma}}{\phi}}, & \text{if } t_m \geq 0, \\ 0, & \text{if } t_m < 0 \end{cases} \tag{2.15}$$

We have the following theorem.

Theorem 2.5. Suppose that t_m has a Weibull distribution with the parameters (γ, ϕ). Let $Y = T^{\gamma}$. Then Y has a gamma distribution with the parameters $\left(1 + \frac{1}{\gamma}, \phi\right)$.

Proof: From (2.2) and (2.15), we have

$$f_T(t) = \left(\frac{\gamma}{\phi}\right) t^{\gamma} e^{-\frac{t^{\gamma}}{\phi}} \left[\phi^{\frac{1}{\gamma}} \Gamma\left(1 + \frac{1}{\gamma}\right)\right]^{-1} \quad \text{where} \quad t \geq 0 \tag{2.16}$$

Let $Y = T^{\gamma}$. Then $T = Y^{\frac{1}{\gamma}}$. According to the Jacobian of the transformation [35], the density function of Y is

$$f_Y(y) = f_T\left(y^{\frac{1}{\gamma}}\right) \left|\frac{dt}{dy}\right| = f_T\left(y^{\frac{1}{\gamma}}\right) \times \left(\frac{y^{\frac{1}{\gamma}-1}}{\gamma}\right) \quad \text{where} \quad y \geq 0 \tag{2.17}$$

Substitute (2.16) into (2.17) to yield

$$f_Y(y) = \frac{y^{\frac{1}{\gamma}} e^{-\frac{y}{\phi}}}{\phi^{1+\frac{1}{\gamma}} \Gamma\left(1 + \frac{1}{\gamma}\right)} \quad \text{where} \quad y \geq 0 \tag{2.18}$$

From (2.18), Y has a gamma distribution with the parameters $\left(1 + \frac{1}{\gamma}, \phi\right)$.

Q.E.D.

Generation of an excess-life random number for Weibull MS cell residence time with the parameters (γ, ϕ) includes the following steps: We first generate a random number u form the uniform random variable U in $(0, 1)$. Then according to Theorem 2.5, we generate a random number y for the gamma random variable Y with the parameters $\left(1 + \frac{1}{\gamma}, \phi\right)$. By multiplying u and $y^{\frac{1}{\gamma}}$, we obtain a random number for the excess life τ_m. Details of the Weibull residence time curves will not be presented in this chapter.

2.3   Conclusions

In performance evaluation of a cellular telecommunications network, it is important to derive the excess life distribution from the MS cell residence times. This distribution determines if a connected call will be handed over to a new cell, and therefore significantly affects the call dropping probability of the network. In cellular telecommunications network simulation, generating the excess-life random numbers is not a trivial task, which has not been addressed in the literature. This chapter showed how to derive the excess life distribution and to generate the random numbers from the excess life distribution. We then developed the excess-life random number generation procedures for cell residence times with gamma, Pareto, lognormal and Weibull distributions. Our study indicates that the generated random numbers closely match the true excess-life distribution (i.e., Equation (2.1)). Compared with the traditional approach [30], the procedures described in this chapter reduce excution time of simulation from several hours to several minutes. Therefore our procedures can be utilized to efficiently generate excess-life random numbers in cellular telecommunications network simulation.

Chapter 3

Dynamic Periodic Location Area Update

Cellular telecommunications networks have been evolved from the second generation (e.g., GSM) to the 2.5 generation (e.g., GPRS), and then to the third generation (e.g., *Universal Mobile Telecommunications System* or UMTS) [30]. In this evolution, the concept of mobility management has remained the same. Consider the circuit-switched domain of UMTS [2, 3]. In order to track the mobile stations (MSs), the cells (the radio coverages of base stations) in UMTS service area are grouped into several location areas (LAs). To deliver services to an MS, the cells in the group covering the MS will page the MS to establish the radio link. To identify the LA of an MS, mobility management is required. In mobility management, the MS informs the network of its location through the LA update procedure. The update procedure is executed in two situations.

Normal location area update (NLAU) is performed when the location of an MS has been changed. Location change of an MS is detected as follows. The cells continuously broadcast their cell identities. The MS periodically listens to the broadcast cell identity, and compares it with the cell identity stored in the MS's buffer. If the comparison indicates that the location area has been changed, then the MS sends the location area update message to the network.

Periodic location area update (PLAU) allows an MS to periodically report its "presence" to the network even if the MS does not move. A *periodic LA update timer* (PLAU timer or T3212 in [3]) is maintained in the MS. Corresponding to the PLAU timer, an *implicit detach* (ID) timer is maintained in the network. When the PLAU timer expires, the MS performs PLAU.

Normal location update has been intensively investigated in the literature (see [5, 30] and the references therein). Most studies on PLAU focused on mobility database failure restoration [12, 26]. However, a major purpose of PLAU is to allow the network to detect if an MS is still attached to the network in the normal network operation situation (i.e., the mobility databases do not fail). To our knowledge, this aspect has not been investigated in the literature. Before we explain an important issue for PLAU, we first introduce the notation used in this chapter as follows.

- μ_i: the incoming (MS terminated) call arrival rate to an MS

- μ_o: the outgoing (MS originated) call arrival rate plus the NLAU rate to an MS

- μ: the net arrival rate to an MS; i.e., $\mu = \mu_o + \mu_i$

- α: the ratio of the incoming (MS terminated) call arrival rate to the net arrival rate; i.e., $\alpha = \dfrac{\mu_i}{\mu} = \dfrac{\mu_i}{\mu_i + \mu_o}$

- m: the network stores the m most recent inter checkpoint event arrival time samples in dynamic PLAU scheme

- N: the number of PLAUs occurring between two checkpoint events (incoming calls, outgoing calls or NLAUs) when the MS is attached

- $N_{d,m}$: the N value for dynamic PLAU

- N_f: the N value for fixed PLAU

- β: the probability that when an MS is abnormally detached, no failure call setup for mobile termination occurs

26

- $\beta_{d,m}$: the β value for dynamic PLAU

- β_f: the β value for fixed PLAU

- $\gamma(t)$: the probability that there is no MS termination occurring in a period t

- $\bar{t}$: the period between when the previous checkpoint event or PLAU occurs and when the abnormal MS detach occurs (in Fig. 3.3)

- t: the period between when the abnormal MS detach occurs and when the next PLAU occurs (in Fig. 3.3)

- $f_{\bar{p},m}(\cdot)$: the density function of both $\bar{t}$ and t

- **A**: the event that no checkpoint event or PLAU occurs in period $\bar{t}$ (refer to Fig. 3.3)

- **B**: the event that no MS call termination occurs in period t (refer to Fig. 3.3)

- t_p: the period for the PLAU/ID timers

- $f_{p,m}(t_p)$: the density function of t_p

- $f_{p,m}^*(s)$: the Laplace transform of the t_p distribution

- c: a selected value which is related to t_p and μ, where $t_p = \dfrac{c}{\mu}$ and $1 \leq c \leq \dfrac{1}{\alpha}$ (see Equation 3.3)

- t_j: the inter checkpoint event arrival time between the jth previous checkpoint event and the $j + 1$st previous checkpoint event, where $1 \leq j \leq m$

- T_n: $T_n = t_{p,1} + t_{p,2} + \ldots + t_{p,n}$, for $n > 0$ (refer to Fig. 3.2)

- k (used in the simulation flow chart): the total number of PLAU events

- $k0$ (used in the simulation flow chart): the total number of CKPNT intervals where no PLAU occurs

- K (used in the simulation flow chart): the total number of CKPNT intervals

27

- n (used in the simulation flow chart): the total number of events that no call setup failure occurs after abnormal detaches

- N (used in the simulation flow chart): the total number of AB_DETACH events

- M: in the simulation experiments, the abnormal MS detach occurs after an attach period exponentially distributed with mean that is M times of an inter checkpoint arrival time interval

An important issue for PLAU is the selection of the period t_p for the PLAU/ID timers. In 3GPP TS 23.012 [2] and TS 24.008 [3], the t_p value is set/changed by the network and broadcasted to every MS in the LA through the L3-RRC SYSTEM INFORMATION BLOCK 1 message on the Broadcast Control Channel (BCCH). In this approach, the t_p value is the same for all MSs in an LA. There are two issues regarding this *fixed PLAU* scheme:

- How is the t_p value determined when an MS first enters an LA?

- Is it appropriate to have a fixed t_p value during the MS's stay in an LA?

This chapter addresses the above two issues. As we will discuss in the following section, the t_p value should be selected based on the call and movement activities of the MS. Therefore we propose a dynamic PLAU scheme in this chapter. We investigate the performance of this scheme and compare it with the fixed PLAU scheme. Our study provides guidelines to select parameters for dynamic PLAU.

3.1 Dynamic PLAU Scheme

Before we describe our solution for PLAU, we first introduces the concept of *attach*. In UMTS, the attach procedure allows an MS to be "known" by the network. For example, after the MS is powered on, the attach procedure must be executed before the MS can obtain access to the UMTS services. In a cellular telecommunications

network, four events can be utilized by the network to detect the presence of an MS attached to the network:

- MS call origination (the MS makes an outgoing call),

- MS call termination (the MS receives an incoming call),

- periodic location area update (PLAU), and

- normal location area update (NLAU).

Note that we consider NLAU as the fourth event although the main purpose of NLAU is to detect the movement of an MS.

When an MS is *detached* (disconnected) from the network due to abnormal reasons (e.g., battery removal, subscriber moving out of the service area, and so on), the MS will not originate a call. The network detects abnormal MS detach in one of the following two cases:

Case 1. The next PLAU occurs before the arrival of the next MS call termination. In this case, the network detects expiration of the ID timer, and considers the MS detached. The next MS call terminations will not be delivered.

Case 2. The next MS call termination occurs before the ID timer expires. In this case, the network attempts to deliver the next incoming call to the MS but fails. After failure of call setup, the network considers that the MS is detached and will disable future MS call terminations and PLAU timer.

In the call termination procedure, the network resources (trunks and so on) are reserved. In Case 2, these network resources are not released until the network detects that call setup fails. In other words, failure call setup wastes network resources, which should be avoided. To reduce the possibility of Case 2, one may shorten the interval t_p of PLAU. On the other hand, short t_p may result in large network signaling overhead. Therefore, t_p should be carefully selected. A perfect PLAU mechanism will satisfy the following criteria:

Criterion 3.1. When the MS is attached, the presence of the MS is detected through call activities (either incoming or outgoing) or NLAUs, and PLAU is never performed.

Criterion 3.2. When the MS is abnormally detached, the network detects the situation through periodic update mechanism (i.e., Case 1 holds), and failure call setup (i.e., Case 2) never occurs.

If Criterion 3.1 is satisfied, the PLAU cost is zero when the MS is attached. If Criterion 3.2 is satisfied, then the network resources will not be wasted due to call setup failure. We show how to select the t_p value with attempt to satisfy both Criteria 1 and 2. Suppose that the outgoing (MS originated) call arrival rate plus the NLAU rate to an MS is μ_o and the incoming (MS terminated) call arrival rate is μ_i. Then the net arrival rate is $\mu = \mu_o + \mu_i$. Let

$$\alpha = \frac{\mu_i}{\mu} = \frac{\mu_i}{\mu_i + \mu_o}$$

Statistics from mobile operators [30] indicate that 40% of the call activities are incoming calls to an MS. Therefore, $\alpha < 0.4$ can be observed in a typical cellular network. When the MS is attached to the network, we expect to see a call (either incoming or outgoing) or an NLAU for every $1/\mu$ interval. If we select $t_p = \frac{c}{\mu}$ (where $c > 1$) and after every call arrival, t_p is reset to $\frac{c}{\mu}$, then there is a good chance that Criterion 3.1 is satisfied. Consider the example in Figure 3.1 (a) where call arrivals at τ_1, τ_2, and τ_3. For the discussion purpose, we define *presence checkpoint* or *checkpoint* as the action to inform the network of the status of an MS (whether the MS is attached or not). We also define *checkpoint event* as an incoming call, an outgoing call or an NLAU. Such an event results in checkpoint action. When the MS is attached, a checkpoint is triggered by an checkpoint event or PLAU. When the MS is abnormally detached, a checkpoint is triggered by expiration of ID timer or failure setup for MS call termination. In Figure 3.1 (a) a checkpoint occurs at τ_1 and the PLAU timer is reset to t_p (i.e., the next PLAU is expected to occur at time $\tau_1 + t_p$). If t_p is sufficiently large so that $\tau_2 \leq \tau_1 + t_p$, then the next checkpoint
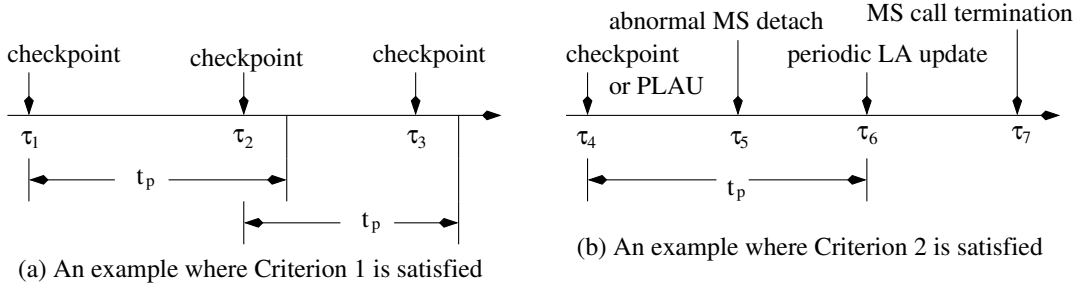
Figure 3.1: Examples where Criteria 1 and 2 are Satisfied

occurs at τ_2 and the PLAU timer is reset to t_p again. In this scenario, the PLAU is never performed and all checkpoints are triggered by call arrivals or NLAUs. If we select t_p so that

$$t_p = \frac{c}{\mu} \geq \frac{1}{\mu} \tag{3.1}$$

then there is good opportunity that Criterion 3.1 is satisfied. However, if t_p is too large, Criterion 3.2 is likely to be violated. Figure 3.1 (b) illustrates a scenario when Criterion 3.2 is satisfied. In this figure, the MS is abnormally detached between two checkpoints. The previous checkpoint occurs at τ_4. The abnormal MS detach occurs at $\tau_5 > \tau_4$. After MS detach, the next periodic LA update occurs at $\tau_6 = \tau_4 + t_p$. The next call termination occurs at τ_7. If $\tau_7 > \tau_6$, then PLAU timer expires before the next call termination arrives, and Criterion 3.2 is satisfied. Since the MS call termination rate is μ_i, to have a good chance to satisfies Criterion 3.2, we suggest that c is selected such that

$$t_p = \frac{c}{\mu} \leq \frac{1}{\mu_i} = \frac{1}{\alpha\mu} \tag{3.2}$$

From (3.1) and (3.2), if $\mu_o \geq 0$, it seems appropriate to select c so that

$$1 \leq c \leq \frac{1}{\alpha} \tag{3.3}$$

Based on the above discussion, we propose a dynamic PLAU scheme that dynamically selects t_p according to the call and NLAU activities of an MS.

Dynamic PLAU Scheme

31

Step 0. Initially a default t_p value is given.

Step 1. When a checkpoint event arrives, the following steps are executed in the network; specifically, the *Visitor Location Register* (VLR) [30]:

>   **Step 1.1.** The interval t_1 between this checkpoint event and the previous checkpoint event is computed and stored in a storage. The network stores the m most recent inter checkpoint event arrival time samples.
>
>   **Step 1.2.** The α statistics is updated.
>
>   **Step 1.3.** Let t_j be the inter checkpoint event arrival time between the jth previous checkpoint event and the $j+1$st previous checkpoint event. The value t_p is computed as
>
> $$t_p = c\left(\frac{t_1 + t_2 + ... + t_m}{m}\right) \tag{3.4}$$
>
>   where c is selected following the guideline (3.3).
>
>   **Step 1.4.** The ID timer in the network is reset with the value t_p. The MS is informed to reset its PLAU timer.

Step 2. When the network receives the PLAU message from the MS, the ID timer is reset with the previously selected t_p.

In the *fixed PLAU* scheme proposed in 3GPP TS 23.012 and TS 24.008, Step 2 is always executed, and Step 1 is never executed. Also note that in the dynamic PLAU, the MS is informed to reset its PLAU timer by the network. In the standard GSM/UMTS procedures, when an MS requests for call origination, NLAU or PLAU, the network always acknowledges the request. The new t_p value is included in GSM/UMTS acknowledgement messages issued by the network. In call termination, the network includes the new t_p value in the call setup message. Therefore, no extra signaling messages are introduced by the dynamic PLAU scheme at the cost that the acknowledgement and call setup messages are slightly modified. Note that in 3GPP TS 24.008, the t_p value is broadcasted to all MSs through the L3-RRC SYSTEM

INFORMATION BLOCK 1 message on the BCCH, which cannot be used in our approach.

In a real GSM/UMTS network, the dynamic PLAU scheme can be implemented in the VLR as a micro-procedure. This implementation can be vendor specific, which does not change any GSM/UMTS message flows. The message flows between the VLR and the MS follow the standard GSM/UMTS procedures.

3.2 Analytic Modeling

This section investigates the performance of dynamic PLAU and compares it with fixed PLAU proposed in 3GPP TS 23.012 and TS 24.008. Two performance measures are considered.

- Let N be the number of PLAUs occurring between two checkpoint events (incoming calls, outgoing calls or NLAUs) when the MS is attached. The smaller the N value, the lower the network signaling overhead caused by the PLAU mechanism. Criterion 3.1 is satisfied when $N = 0$. Let $N_{d,m}$ and N_f be the N values for dynamic PLAU and fixed PLAU, respectively. We will derive the expected values $E[N_{d,m}]$ and $E[N_f]$.

- Let β be the probability that when an MS is abnormally detached, no failure call setup for mobile termination occurs (i.e., Case 1 holds). It is clear that the bigger the β value, the better the PLAU mechanism. Specifically, Criterion 3.2 is satisfied when $\beta = 1$. Let $\beta_{d,m}$ and β_f be the β values for dynamic PLAU and fixed PLAU, respectively.

Telecommunications network operations suggest that incoming and outgoing call arrivals are Poisson streams [30], and the aggregate arrivals of the incoming and outgoing calls together with the NLAUs can be approximated as a Poisson stream. Following the above statement, we assume Poisson checkpoint arrivals as in many other studies [40, 18, 12]. Therefore $t_1, t_2, ..., t_m$ in (3.4) are exponentially distributed, and

previous checkpoint event   next checkpoint event
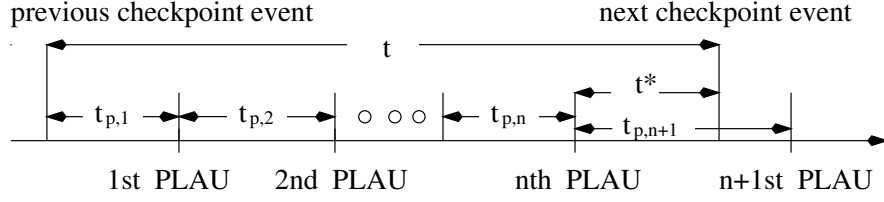
1st PLAU   2nd PLAU   nth PLAU   n+1st PLAU

Figure 3.2: The Number of PLAUs between Two Checkpoint Events When the MS is Attached

t_p has an Erlang-m density function $f_{p,m}(t_p)$ with mean $\frac{c}{\mu}$. That is,

$$f_{p,m}(t_p) = \left(\frac{m\mu}{c}\right)^m \left[\frac{t_p^{m-1}}{(m-1)!}\right] e^{-\left(\frac{m\mu}{c}\right)t_p} \tag{3.5}$$

The Laplace transform of the t_p distribution is

$$f_{p,m}^*(s) = \int_{t_p=0}^{\infty} f_{p,m}(t_p)e^{-st_p}dt_p = \left(\frac{\dfrac{m\mu}{c}}{s+\dfrac{m\mu}{c}}\right)^m = \left(\frac{m\mu}{sc+m\mu}\right)^m \tag{3.6}$$

Figure 3.2 shows the scenario when the MS is attached to the network and there are n PLAUs between two checkpoint events in dynamic PLAU. For $n > 0$, let $T_n = t_{p,1} + t_{p,2} + ... + t_{p,n}$. Then

$$\begin{aligned}
\Pr[N_{d,m} = n] &= \Pr[T_n \le t < T_{n+1}] \\
&= \int_{t_{p,1}=0}^{\infty} \int_{t_{p,2}=0}^{\infty} \cdots \int_{t_{p,n+1}=0}^{\infty} \int_{t=T_n}^{T_{n+1}} \left[\prod_{j=1}^{n+1} f_{p,m}(t_{p,j})\right] \mu e^{-\mu t} dt dt_{p,n+1} \cdots dt_{p,2} dt_{p,1} \\
&= \int_{t_{p,1}=0}^{\infty} \int_{t_{p,2}=0}^{\infty} \cdots \int_{t_{p,n+1}=0}^{\infty} \left[\prod_{j=1}^{n+1} f_{p,m}(t_{p,j})\right] \left[e^{-\mu T_n} - e^{-\mu T_{n+1}}\right] dt_{p,n+1} \cdots dt_{p,2} dt_{p,1} \\
&= \left[f_{p,m}^*(\mu)\right]^n \left[1 - f_{p,m}^*(\mu)\right] \tag{3.7}
\end{aligned}$$

For $n = 0$,

$$\begin{aligned}
\Pr[N_{d,m} = 0] &= \Pr[t < t_{p,1}] = \int_{t_{p,1}=0}^{\infty} \int_{t=0}^{t_{p,1}} f_{p,m}(t_{p,1})\mu e^{-\mu t} dt dt_{p,1} \\
&= 1 - \int_{t_{p,1}=0}^{\infty} f_{p,m}(t_{p,1})e^{-\mu t_{p,1}} dt_{p,1} = 1 - f_{p,m}^*(\mu) \tag{3.8}
\end{aligned}$$

34

From (3.7), the expected number of $N_{d,m}$ is

$$E[N_{d,m}] = \sum_{n=1}^{\infty} n \Pr[N_{d,m} = n] = \left[1 - f_{p,m}^*(\mu)\right] \left\{ \sum_{n=1}^{\infty} n \left[f_{p,m}^*(\mu)\right]^n \right\} = \frac{f_{p,m}^*(\mu)}{1 - f_{p,m}^*(\mu)} \tag{3.9}$$

Substitute (3.6) in (3.9) to yield

$$E[N_{d,m}] = \left(\frac{m}{c+m}\right)^m \left[1 - \left(\frac{m}{c+m}\right)^m\right]^{-1} \tag{3.10}$$

Now we derive $E[N_f]$. Consider Figure 3.2 again. This figure illustrates an example when the MS is attached to the network and there are n PLAUs between two checkpoint events in fixed PLAU. In this example, for $1 \le i \le n+1$, $t_{p,i} = \frac{c}{\mu}$ is a fixed value, and the inter checkpoint event time t is expressed as $t = n\left(\frac{c}{\mu}\right) + t^*$. For $n \ge 1$, the probability that $N_f = n$ is derived as

$$\begin{aligned} \Pr[N_f = n] &= \int_{t^*=0}^{\frac{c}{\mu}} \mu e^{-\mu\left[n\left(\frac{c}{\mu}\right) + t^*\right]} dt^* \\ &= \left(1 - e^{-c}\right) e^{-nc} \end{aligned} \tag{3.11}$$

For $n = 0$,

$$\Pr[N_f = 0] = \Pr[t < t_{p,1}] = \int_{t=0}^{\frac{c}{\mu}} \mu e^{-\mu t} dt = 1 - e^{-c} \tag{3.12}$$

From (3.11), the expected number of N_f is

$$E[N_f] = \sum_{n=1}^{\infty} n \left(1 - e^{-c}\right) e^{-nc} = \frac{e^{-c}}{1 - e^{-c}} \tag{3.13}$$

$E[N_f]$ can also be derived from (3.10). Consider the case when $m \to \infty$ and $s = \mu$. Equation (3.6) is re-written as

$$\lim_{m \to \infty} f_{p,m}^*(\mu) = \lim_{m \to \infty} \left(\frac{m}{m+c}\right)^m = e^{-c} \tag{3.14}$$

From (3.6), (3.8) and (3.12), we have $\Pr[N_f = 0] \ge \Pr[N_{d,m} = 0]$. Also, from (3.8) and (3.14), we have

$$\Pr[N_f = 0] = \lim_{m \to \infty} \Pr[N_{d,m} = 0] = 1 - e^{-c} \tag{3.15}$$

Equation (3.15) is the same as (3.12). From (3.10) and (3.14), we have

$$E[N_f] = \lim_{m \to \infty} E[N_{d,m}] = \frac{e^{-c}}{1 - e^{-c}} \tag{3.16}$$
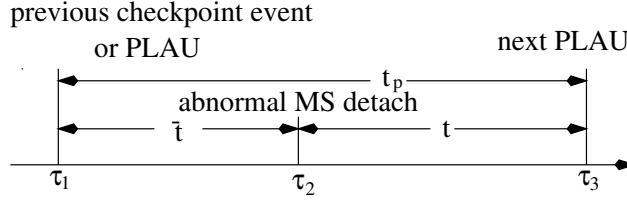
Figure 3.3: Timing Diagram for Deriving $\beta_{d,m}$ and β_f

Equation (3.16) is the same as (3.13).

Probability $\beta_{d,m}$ is derived as follows. Consider Figure 3.3. Suppose that the previous checkpoint or PLAU occurs at time τ_1, the abnormal MS detach occurs at τ_2, and the next PLAU occurs at τ_3. Define two events as follows:

Event A. No checkpoint event or PLAU occurs in period $\bar{t} = \tau_2 - \tau_1$.

Event B. No MS call termination occurs in period $t = \tau_3 - \tau_2$.

It is apparent that Criterion 3.2 is satisfied if and only if event B occurs under the condition that event A occurs. That is $\beta_{d,m} = \Pr[B|A]$. If the occurrence of abnormal MS detach is a random observer, then from residual life theorem and reverse residual life theorem [43], both $\bar{t}$ and t have the same density function $f_{\bar{p},m}$

$$f_{\bar{p},m}(\bar{t}) = f_{\bar{p},m}(t) = \left(\frac{\mu}{c}\right) \int_{t_p=t}^{\infty} f_{p,m}(t_p) dt_p = \left(\frac{\mu}{c}\right) \left\{ \sum_{j=0}^{m-1} \left(\frac{m\mu}{c}\right)^j \left(\frac{t^j}{j!}\right) e^{-\left(\frac{m\mu}{c}\right)t} \right\}$$

(3.17)

Furthermore, events A and B are independent of each other and

$$\beta_{d,m} = \Pr[B|A] = \frac{\Pr[B \cap A]}{\Pr[A]} = \frac{\Pr[B]\Pr[A]}{\Pr[A]} = \Pr[B]$$

(3.18)

Let $\gamma(t)$ be the probability that there is no MS termination occurring in a period t. Since MS termination calls are a Poisson stream with rate $\mu_i = \alpha\mu$, $\gamma(t)$ is expressed as [43]

$$\gamma(t) = e^{-\alpha\mu t}$$

(3.19)

From (3.18), (3.19) and (3.17),

$$\beta_{d,m} = \int_{t=0}^{\infty} f_{\bar{p},m}(t)\gamma(t)dt = \sum_{j=0}^{m-1} \left\{ \left(\frac{\mu}{c}\right) \int_{t=0}^{\infty} \left(\frac{m\mu}{c}\right)^j \left(\frac{t^j}{j!}\right) e^{-\left[\left(\frac{m\mu}{c}\right)+\alpha\mu\right]t} dt \right\}$$

$$= \sum_{j=0}^{m-1} \left(\frac{\mu}{c}\right) \left(\frac{m\mu}{c}\right)^j \left(\frac{m\mu}{c} + \alpha\mu\right)^{-(j+1)} = \left(\frac{1}{\alpha c}\right) \left[1 - \left(\frac{m}{m + \alpha c}\right)^m\right] \quad (3.20)$$

Now we derive β_f. Consider Figure 3.3 again. For fixed PLAU, $t_p = \frac{c}{\mu}$ is a constant. Since occurrence of abnormal MS detach is a random observer, t has a uniform distribution in interval $\left[0, \frac{c}{\mu}\right]$. From (3.19), β_f is derived as

$$\beta_f = \int_{t=0}^{\frac{c}{\mu}} \left(\frac{\mu}{c}\right) \gamma(t) dt = \frac{1 - e^{-\alpha c}}{\alpha c} \quad (3.21)$$

We can also derive β_f from (3.20) and (3.14):

$$\beta_f = \lim_{m \to \infty} \beta_{d,m} = \left(\frac{1}{\alpha c}\right) \left[1 - \lim_{m \to \infty} \left(\frac{m}{m + \alpha c}\right)^m\right] = \frac{1 - e^{-\alpha c}}{\alpha c}$$

The above analytic model is validated against the simulation experiments. We use a C program to implement the simulation model that consists of three types of events: (1) Checkpoint (CKPNT); (2) PLAU; and (3) MS abnormal detach (AB_DETACH). The next CKPNT and AB_DETACH event arrival times are generated by the exponential random number generator, and all events are processed according to their timestamps. The simulation flowchart is shown in Figure 3.4. For a CKPNT event, steps 6-9 are executed. For a PLAU event, steps 10-11 are executed. For an AB_DETACH event, steps 12-16 are executed. In the simulation experiments, the abnormal MS detach occurs after an attach period exponentially distributed with mean that is M times of an inter checkpoint arrival time interval. For $M > 100$, the simulation results are not sensitive to the M values. In our simulation experiments, the confidence intervals of the 99% confidence levels are within 3% of the mean values in most cases. Figure 3.5 shows that analytic analysis and simulation experiments are consistent for the β values. The comparison results for other performance measures are similar and will not be presented in this chapter.

3.3 Numerical Examples

Based on the analysis in the previous section, we use numerical examples to investigate the performance of dynamic PLAU and compare it with fixed PLAU.

- ❖ **K: total number of CKPNT intervals;**
- ❖ **k: total number of PLAU events;**
- ❖ **k0: total number of CKPNT intervals where no PLAU occurs;**
- ❖ **N: total number of AB_DETACH events;**
- ❖ **n: total number of events that no call setup failure occurs after abnormal detaches;**
- ❖ **$\beta_{d,m}$=n/N;**
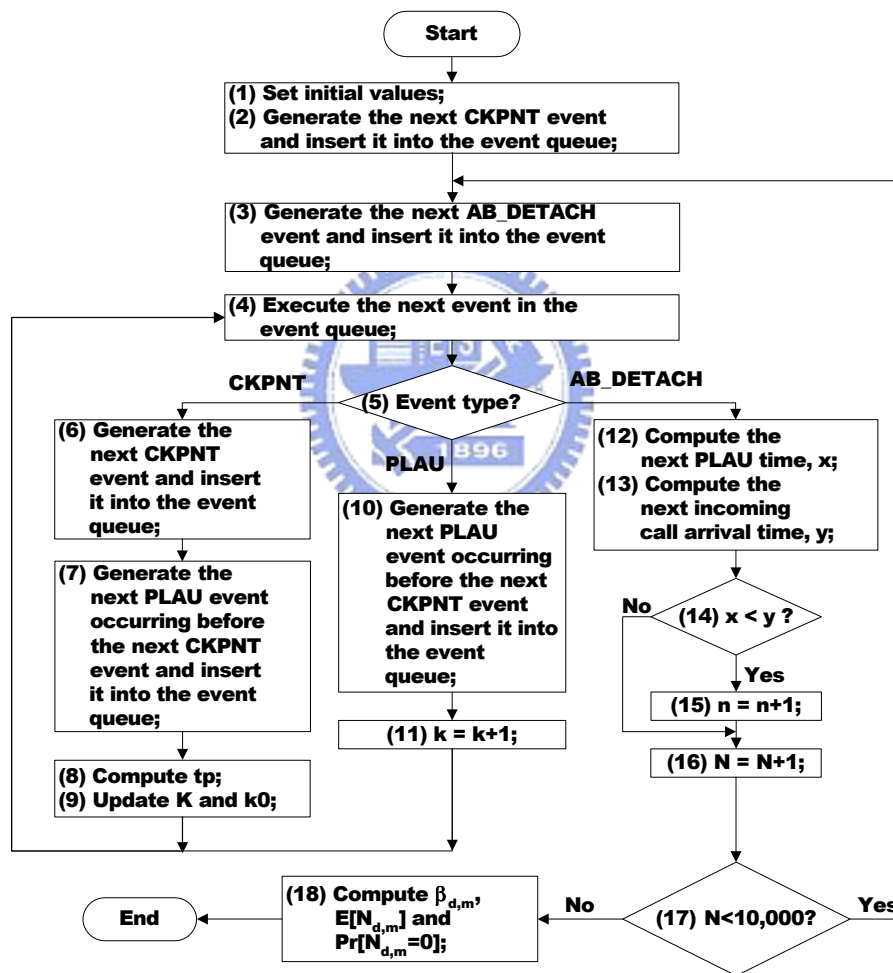- ❖ **$E[N_{d,m}]$=k/K;**
- ❖ **$Pr[N_{d,m}=0]$=k0/K;**

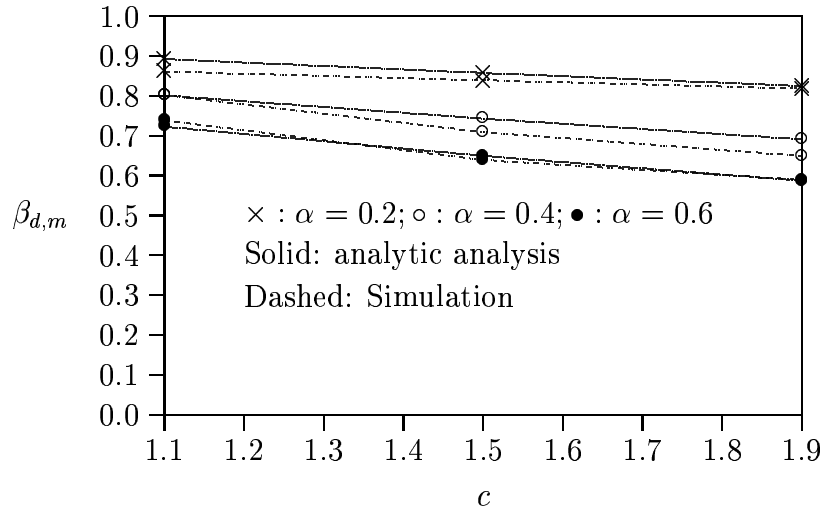Figure 3.4: The Simulation Flowchart

38

Figure 3.5: Comparing Analytic Analysis with Simulation Experiments ($m = 20$)
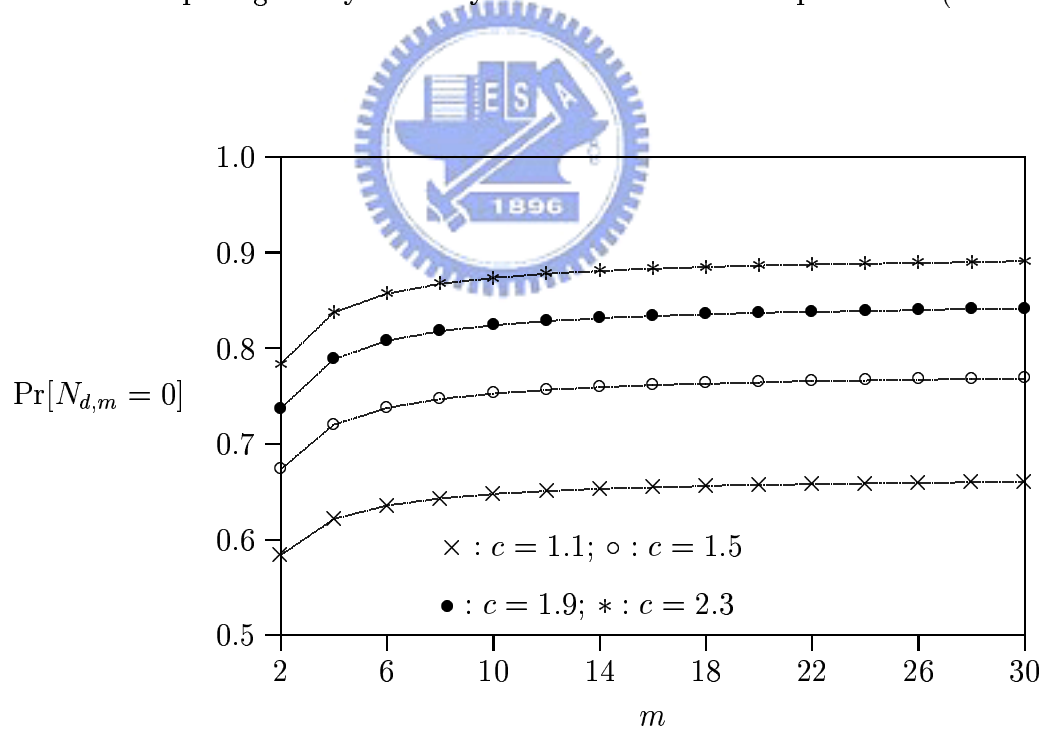
Figure 3.6: The $\Pr[N_{d,m} = 0]$ Performance
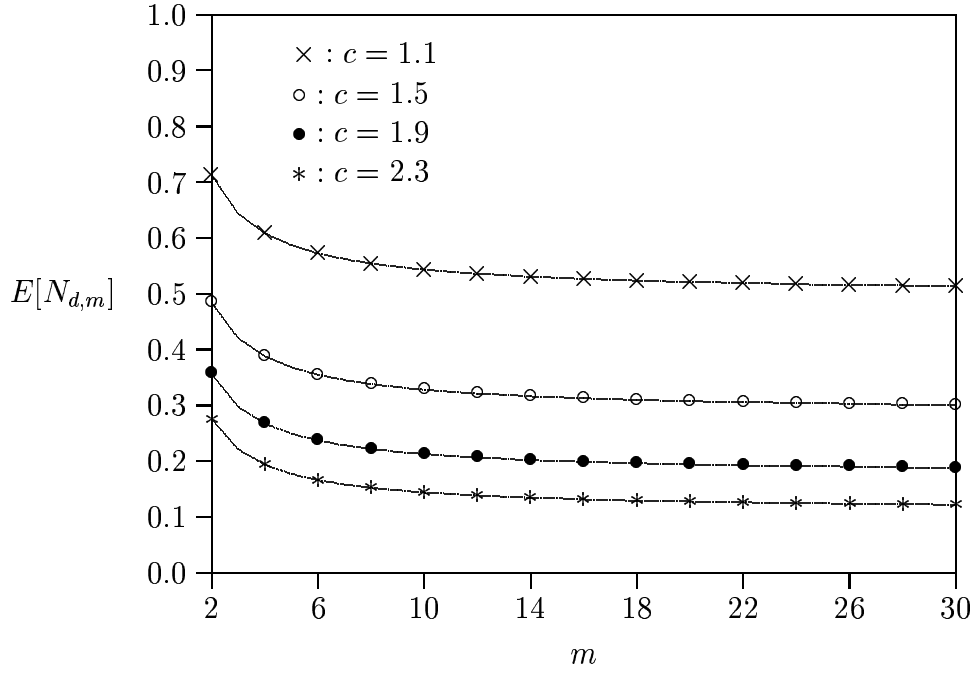
Figure 3.7: The $E[N_{d,m}]$ Performance

By using (3.8), Figure 3.6 plots $\Pr[N_{d,m} = 0]$ as a function of m. The figure indicates that even we choose a small c (e.g., $c = 1.1$ or 1.5), Criterion 3.1 can be satisfied with probability higher than 0.5. In this figure, the $\Pr[N_f = 0]$ values are 0.66713, 0.77687, 0.85043 and 0.89974 for $c =$1.1, 1.5, 1.9 and 2.3 respectively.

Based on (3.10), Figure 3.7 plots $E[N_{d,m}]$ against m, where $1.1 \leq c \leq 2.3$. The figure indicates that $E[N_{d,m}]$ is a decreasing function of m. When m is small (i.e., $m < 6$), if dynamic PLAU measures one more inter checkpoint arrival time sample (i.e,. m is incremented by one), the $E[N_{d,m}]$ performance is significantly improved. On the other hand, when m is large ($m > 20$), measuring more inter checkpoint arrival time samples will not improve the $E[N_{d,m}]$ performance. In Figure 3.7, the $E[N_f]$ values are 0.49896, 0.28722, 0.17588 and 0.11142 for $c =$1.1, 1.5, 1.9 and 2.3 respectively.

Based on (3.20), Figure 3.8 plots $\beta_{d,m}$ against m, where $\alpha = 0.2, 0.4$ and 0.6, and $c = 1.1, 1.5$, and 1.9. For $m > 10$, $\beta_{d,m}$ is not sensitive to the change of m. Suppose

Figure 3.8: The $\beta_{d,m}$ Performance

that $m = 20$ is selected. When $\alpha = 0.2$, $0.82 \leq \beta_{d,m} \leq 0.9$ for $1.1 \leq c \leq 1.9$. When $\alpha = 0.4$, $0.6 \leq \beta_{d,m} \leq 0.8$. When $\alpha = 0.6$, $0.58 \leq \beta_{d,m} \leq 0.72$. Therefore, the $\beta_{d,m}$ performance is significantly affected by α (i.e., the frequency of incoming calls to the MS). 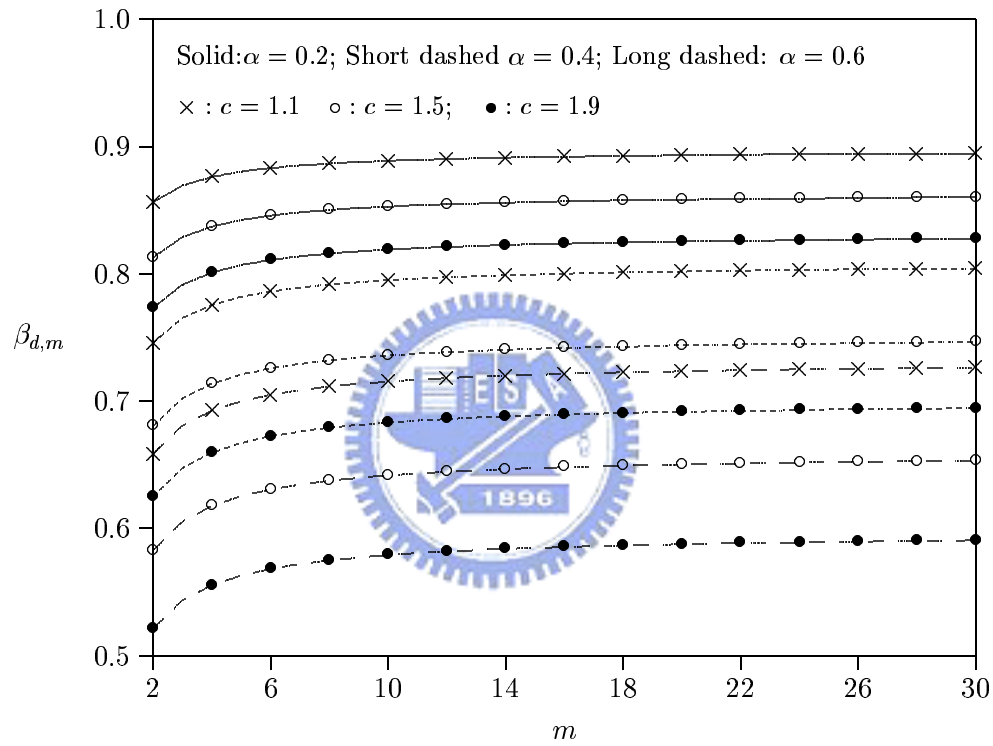As we mentioned before, $\alpha = 0.4$ is observed in cellular network operations, and good $\beta_{d,m}$ performance can be expected. In this figure, the β_f values are shown below. For $\alpha = 0.2$, the β_f values are 0.89764, 0.86394 and 0.83194 for $c = 1.1$, 1.5 and 1.9 repectively; for $\alpha = 0.4$, the β_f values are 0.80901, 0.75198 and 0.70044 for $c = 1.1$, 1.5 and 1.9 repectively; for $\alpha = 0.6$, the β_f values are 0.73204, 0.65937 and 0.59665 for $c = 1.1$, 1.5 and 1.9 repectively.

Note that the $\Pr[N_f = 0]$, $E[N_f]$, and β_f values corresponding to Figures 5, 6, 7 are "optimal". That is, such good performance can only be achieved when the "optimal fixed t_p values" are found. In reality, it is very difficult (if not possible) to guess such "optimal" values in advance. In Figures 5, 6, 7, we demonstrate that by the adaptive mechanism, the dynamic PLAU scheme can achieve good performance close to the optimal fixed PLAU scheme.

The results in Figures 3.7 and 3.8 indicate that $E[N_{d,m}]$ and $\beta_{d,m}$ have conflicting goals. In other words, it is impossible to choose the c values that minimize $E[N_{d,m}]$ and maximize $\beta_{d,m}$ at the same time. However, by choosing appropriate c values, dynamic PLAU can satisfy both $E[N_{d,m}]$ and $\beta_{d,m}$ restrictions, if such solutions do exist. For example, consider $\alpha = 0.4$. If the system requires that $E[N_{d,m}] < 0.6$ and $\beta_{d,m} > 0.7$, then a c value such as 1.1 satisfies the requirement (with $m = 20$).

Note that if the frequency of checkpoint events (call activities and NLAUs) changes from time to time, dynamic PLAU can automatically adapt to the change. Consider the scenario where $\mu = \mu_1$ for a long time (the first period), and then μ changes to $10\mu_1$ (the second period). Assume that the intervals for both the first and the second periods are the same. For dynamic PLAU, the t_p period is adjusted as μ changes so that the c value is kept as a constant (except for the short period for transition from μ_1 to $10\mu_1$). On the other hand, the period t_p is fixed in fixed PLAU. Therefore, $c = t_p\mu_1$ in the first period and $c = 10t_p\mu_1$ in the second period. In other words, the c value in the second period is 10 times that in the first period. After the
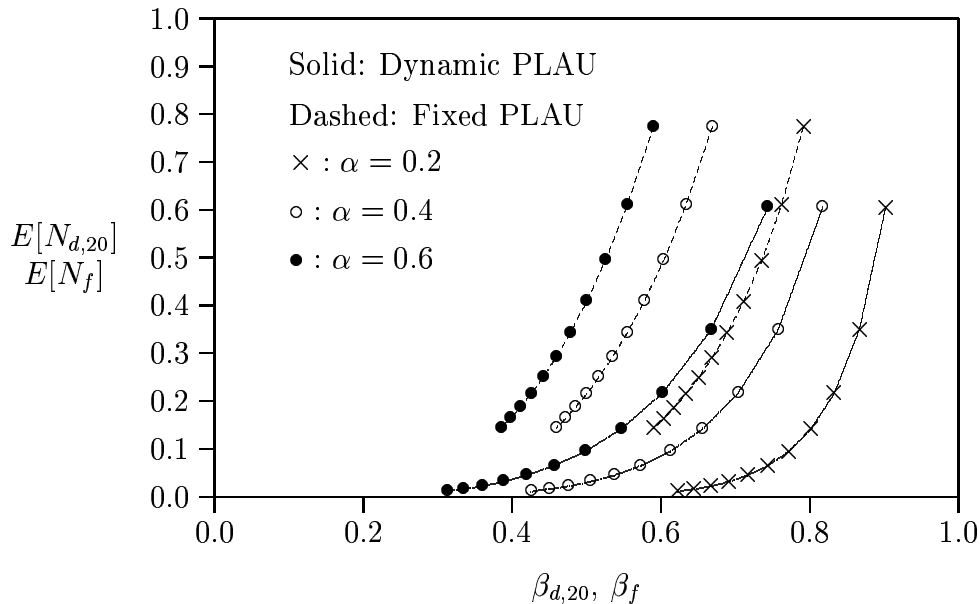
Figure 3.9: Relationship between $E[N]$ and β $(m = 20)$

checkpoint rate changes, the fixed PLAU may only slightly improve the N_f performance at the cost of significantly degrading the β_f performance. The consequence is that the average performance of fixed PLAU for the above mixed checkpoint traffic is worse than that for dynamic PLAU. Figure 3.9 plots $E[N]$ against β for dynamic PLAU and fixed PLAU, where $m = 20$, and $\alpha = 0.2, 0.4$, and 0.6. Note that β is a function of m, α, and c, while $E[N]$ is a function of m and c. Therefore, when we choose a specific set of (m, α, c), the corresponding β and $E[N]$ values can be computed. Then we use these computed $(\beta, E[N])$ sets to plot this figure. The figure indicates that to achieve the same β performance for the mixed checkpoint traffic patterns, much less network signaling overhead for LA updates is expected in dynamic PLAU as compared with that in fixed PLAU.

3.4 Conclusions

In cellular telecommunications networks, periodic location area update (PLAU) is utilized to detect presence of a mobile station (MS). In 3GPP Technical Specifica-

43

tions 23.012 and 24.008, a fixed PLAU scheme was proposed for UMTS where the interval between two PLAUs is of fixed length. We observe that MS presence can also be detected through call and movement activities. Therefore we proposed a dynamic PLAU scheme where the PLAU interval is dynamically adjusted based on the call and NLAU traffic. An analytic model was developed to validate against the simulation model with the excess life modeling technique. Then we investigated the performance of dynamic and fixed PLAU schemes. Our study indicated that compared with fixed PLAU, dynamic PLAU significantly reduces the network signaling traffic caused by periodic location area update.

As a final remark, in dynamic PLAU, storage and the mechanism maintaining $m = 20$ or 30 inter checkpoint arrival time samples for an MS can be practically implemented in the UMTS network (specifically, in the VLR). The value t_p can be efficiently computed using the *window averaging* technique [30].

Chapter 4

Modeling Channel Assignment of Small-Scale Cellular Networks

Emerging *cellular telecommunications network* technologies have attracted considerable attention in academic research as well as commercial deployment. A cellular network supports telephony services when users are in movement [30]. The cellular phone service area is populated with base stations (BSs). The radio coverage of a BS is referred to as a *cell*. Customers within a cell can connect to the corresponding BS via *mobile stations* (MSs) or mobile phones. When a call for a customer occurs, one radio channel of the BS is used for connecting the MS and the BS. If all radio channels are in use when a new call is attempted, the call will be blocked and cleared from the system. If the call is accepted, a radio channel will be occupied until the call is completed, or until the MS moves out of the cell. When a communicating MS moves from one cell to another, the occupied channel in the old cell is released, and an idle channel is acquired in the new cell. During this handover procedure, if no channel is available in the new cell, the call is forced to terminate before its completion. When the call is connected, the call may be completed after several successful handovers, or may be forced to terminate due to a failed handover. The duration of a call connection (if the call is completed) is referred to as the *call holding time.*

For billing and network planning purposes, the handover behavior and the prob-

ability of call completion need to be analyzed. Several analytic studies have contributed to cellular network performance evaluation [34, 10, 7, 48, 19, 13, 36, 39]. Most studies assume that the handover traffic to a cell is a fixed-rate Poisson process. This assumption is reasonable for large-scale cellular networks, or when the networks experience light load traffic [9]. In reality, the handover traffic to a cell depends on the workloads of the neighboring cells. This fact has significant impact on modeling of small-scale cellular networks. Before we investigate the impact, we first introduce the notation used in this chapter as follows.

- n: the number of cells in the cellular system under consideration

- c: the number of channels in each cell in the cellular system under consideration

- S_i: the index set of cell i's neighbors, for $1 \leq i \leq n$

- $|S_i|$: the number of cell i's neighbors, for $1 \leq i \leq n$

- s: we assume that $|S_i| = |S_j| = s$ for $1 \leq i, j \leq n$

- λ: the new call arrival rate to a cell; the new call arrivals are a Poisson stream, and the new call arrivals to each cell are independent

- $1/\mu$: the mean call holding time; the call holding times have an exponential distribution

- $1/\eta$: the mean MS cell residence time (i.e., η is the MS mobility rate); the MS cell residence times are independent and identically distributed (i.i.d.)

- N_i: a random variable for the number of busy channels in cell i, for $1 \leq i \leq n$

- p_b: the new call blocking probability; i.e., the number of new call blockings divided by the number of new calls

- p_f: the handover force-termination probability; i.e., the number of forced terminations divided by the number of handovers

- p_{nc}: the call incompletion probability (the probability that a call is either blocked or forced to terminate); i.e., the sum of the numbers of new call blockings and forced terminations divided by the number of new calls (note that $p_{nc} \neq p_b + p_f$)

- A_i: the event that a call is handed over into cell i, where $1 \leq i \leq n$

- B_k: the event that a call is handed over out of a cell with k busy channels, where $1 \leq k \leq c$

- C: the event that a handover occurs in the cellular system under consideration

- D_m: the event that a call is handed over out of cell m, where $1 \leq m \leq n$

- $E_{j,k}$: the event that a call is handed over out of a specific cell j with k busy channels, where $1 \leq j \leq n$ and $1 \leq k \leq c$

- $f_m^*(s)$: the Laplace transform of the MS cell residence time distribution

- ϵ: a pre-defined value for the threshold in our algorithm which approximates the general MS cell residence time distribution by an exponential distribution to compute the probabilities p_b, p_f, and p_{nc} ($\epsilon = 0.001$ in our example)

- t_1: the call holding time in Figure 4.1 (note that the definition of t_1 in Fig. 4.1 is different from that in Fig. 1.1 and Fig. 2.1)

- t_2: the MS cell residual time at cell 1 (i.e., the interval between when the call arrives and when the MS moves out of cell 1) in Figure 4.1 (note that the definition of t_2 in Fig. 4.1 is different from that in Fig. 1.1 and Fig. 2.1)

- $t_{3,i}$: the remaining call holding time if the call is successfully handed over to cell i for $i = 2, 3, \ldots$ in Figure 4.1

- $t_{4,i}$: the MS cell residence time at cell i (i.e., the interval between when the MS enters cell i and when it moves out of cell i) for $i = 2, 3, \ldots$ in Figure 4.1

47
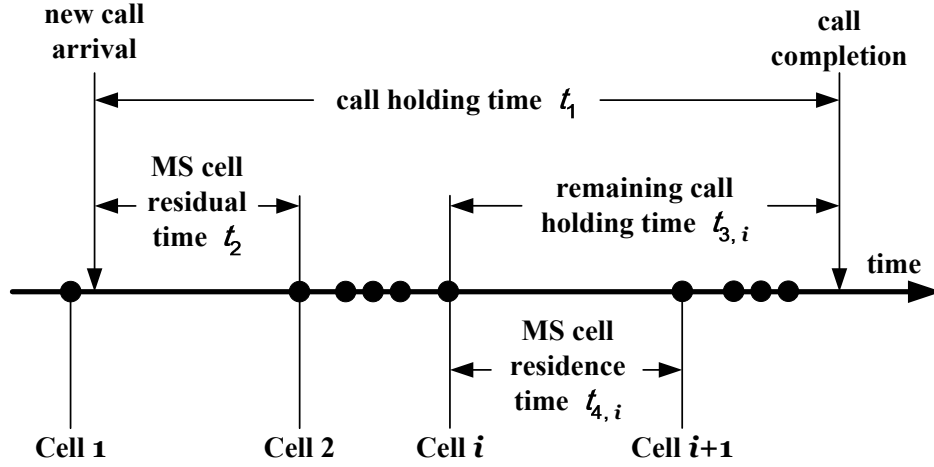
Note: The dots (●) in the time line represent MS movements.

Figure 4.1: The Timing Diagram (Note That the Notation Used in This Figure May Have Different Definition from That in Fig. 1.1 and Fig. 2.1)

- α_1: $\alpha_1 = \Pr[t_1 > t_2]$, which is the probability that a new call is not completed before the MS moves out of the first cell

- α_2: $\alpha_2 = \Pr[t_{3,i} > t_{4,i}]$, which is the probability that a handover is not completed before the MS moves out of cell i, where $i = 2, 3, \ldots$

- X: a random variable for the number of call completions for a connected call; note that the value of X is either one or zero

- Y: a random variable for the number of handovers for a connected call

- θ: $\theta = \dfrac{E[X]}{E[X+Y]}$

- η^*: the mobility rate for the approximate exponential distribution which approximates the general MS cell residence time distribution

- $\mathbf{Q}$: the Markov process state transition rate matrix in the approximate algorithm for the general MS cell residence time distribution (refer to Appendix B for more details)

48

- k_i: the number of busy channels in cell i for $1 \leq i \leq n$ and $0 \leq k_i \leq c$ (refer to Appendix B for more details)

- $(k_1, k_2, \cdots, k_n)$: a state in the Markov process (refer to Appendix B for more details)

- $\pi_{k_1, k_2, \cdots, k_n}$: the steady state probability for state $(k_1, k_2, \cdots, k_n)$ (refer to Appendix B for more details)

- Π: the stationary probability distribution vector; i.e., $\Pi = [\pi_{k_1, k_2, \cdots, k_n}]_{1 \times (c+1)^n}$ (refer to Appendix B for more details)

Figure 4.2 plots the call incompletion probability p_{nc} against the user mobility η and the call arrival rate λ where the number of radio channels in a cell is 9. The "•" curve is generated from a previous analytic model that assumes fixed-rate handover traffic [34]. The "⋄" curve is generated from simulation of a 64-cell mesh configuration, and the "∗" curve is generated from simulation of a 3-cell configuration (illustrated in Fig. 1.3). The simulation model [31] actually simulates the MS movement in mesh or hexagonal networks of cells. The figure indicates that the fixed-rate assumption is acceptable when the number of cells is reasonably large, but is inaccurate for small-scale cellular networks. In this chapter, we derive the exact equation for the handover force-termination probability when the MS cell residence times are exponentially distributed. Then we propose an approximate model with general MS cell residence times. We use the analytic model to validate against the simulation model with the excess life modeling technique. Then the analytic results are compared with the previously proposed model [34], where the simulation results are used as the baseline. Our comparison study indicates that the new model can capture the handover behavior much better than the old one for small-scale cellular networks.
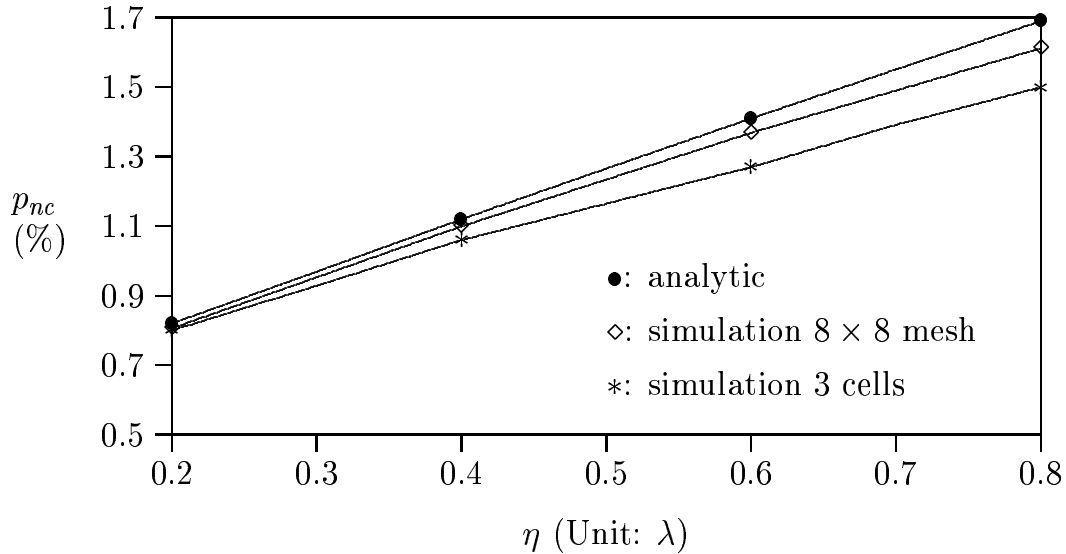
Figure 4.2: Effect of Network Size on the Call Incompletion Probability ($\mu = 0.3\lambda$)

4.1 Exact Analytic Model for Exponential MS Cell Residence Times

This section describes an exact analytic solution for exponential MS cell residence times. Consider a cellular system with n cells. For $1 \leq i \leq n$, let S_i be the index set of cell i's neighbors. That is, cell j is a neighbor of cell i if $j \in S_i$. Let $|S_i|$ be the number of cell i's neighbors. For the illustration purpose, we consider a homogeneous system conforming to the following requirements:

Capacity: Each cell has c channels.

Movement: The routing probabilities to the neighboring cells are the same. That is, for an MS at cell i, it moves to each of cell i's neighbors with probability $1/|S_i|$. We further assume that $|S_i| = |S_j| = s$ for $1 \leq i, j \leq n$. The MS cell residence time distributions are the same for all cells.

Call traffic: The new call arrival rates to all cells are the same. The call holding time distributions are the same for all calls.

The following input parameters are considered in our model.

- λ: the new call arrival rate to a cell. The new call arrivals are a Poisson stream [22], and the new call arrivals to each cell are independent.

- $1/\mu$: the mean call holding time. The call holding times have an exponential distribution.

- $1/\eta$: the mean MS cell residence time. The MS cell residence times are independent and identically distributed (i.i.d.). This section assumes exponential MS cell residence times. In the next section, we will consider general MS cell residence time distributions.

We assume that the call holding time and the MS cell residence time are independent of each other. Let random variable N_i be the number of busy channels in cell i ($1 \leq i \leq n$). The following output measures are evaluated in our study.

- p_b (the new call blocking probability): The number of new call blockings divided by the number of new calls. Since the system is homogeneous, p_b for all cells are the same and for a cell i, p_b can be expressed as

$$
\begin{aligned}
p_b &= \Pr[\text{A new call is blocked} \mid \text{this new call occurs at cell } i] \\
&= \Pr[N_i = c].
\end{aligned}
\tag{4.1}
$$

- p_f (the handover force-termination probability): The number of forced terminations divided by the number of handovers.

- p_{nc} (the call incompletion probability; i.e., the probability that a call is either blocked or forced to terminate): The sum of the numbers of new call blockings and forced terminations divided by the number of new calls. Note that $p_{nc} \neq p_b + p_f$.

To derive p_f, we first define five events:

Event A_i. A call is handed over into cell i. For $1 \leq i \leq n$, A_i are mutually exclusive events.

Event B_k. A call is handed over out of a cell with k busy channels. For $1 \leq k \leq c$, B_k are mutually exclusive events.
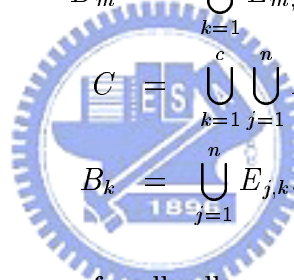
Event C. A handover occurs in the cellular system. The relationship among C, A_i and B_k is

$$C = \bigcup_{i=1}^{n} A_i = \bigcup_{k=1}^{c} B_k. \tag{4.2}$$

Event D_m. A call is handed over out of cell m. For $1 \leq m \leq n$, D_m are mutually exclusive events, and

$$C = \bigcup_{m=1}^{n} D_m. \tag{4.3}$$

Event $E_{j,k}$. A call is handed over out of a specific cell j with k busy channels. For $1 \leq j \leq n$ and $1 \leq k \leq c$, $E_{j,k}$ are mutually exclusive events. Note that

$$\begin{aligned}
D_m &= \bigcup_{k=1}^{c} E_{m,k} \\
C &= \bigcup_{k=1}^{c} \bigcup_{j=1}^{n} E_{j,k} \\
B_k &= \bigcup_{j=1}^{n} E_{j,k}.
\end{aligned} \tag{4.4}$$

Since the system is homogeneous, p_f for all cells are the same, which can be expressed as

$$p_f = \Pr\left[\{N_i = c\} \cap A_i | A_i\right]. \tag{4.5}$$

Since $A_i = A_i \cap \left(\bigcup_{k=1}^{c} B_k\right)$, (4.5) can be re-written as

$$p_f = \Pr\left[\{N_i = c\} \cap A_i \cap \left(\bigcup_{k=1}^{c} B_k\right) \middle| A_i\right]. \tag{4.6}$$

Equation (4.6) says that to compute p_f, we need to consider how the flow-in handover traffic behaves. For example, there is no flow-in handover traffic into cell i if $k = 0$ for all cell i's neighbors (i.e., there is no busy channel in any of cell i's neighbors). Since $B_1, B_2, \cdots, B_c$ are mutually exclusive events, from [41], (4.6) can be expressed

as

$$p_f = \sum_{k=1}^{c} \Pr\left[\{N_i = c\} \cap A_i \cap B_k | A_i\right]$$

$$= \sum_{k=1}^{c} \frac{\Pr\left[\{N_i = c\} \cap B_k \cap A_i\right]}{\Pr[A_i]}$$

$$= \sum_{k=1}^{c} \frac{\Pr\left[\{N_i = c\} \cap B_k \cap A_i\right]}{\Pr[B_k \cap A_i]} \times \frac{\Pr[B_k \cap A_i]}{\Pr[A_i]}$$

$$= \sum_{k=1}^{c} \Pr[\{N_i = c\} \cap B_k \cap A_i | B_k \cap A_i] \times \Pr[B_k | A_i]. \qquad (4.7)$$

In (4.7), let

$$p_{1,k} = \Pr[\{N_i = c\} \cap B_k \cap A_i | B_k \cap A_i], \qquad (4.8)$$

which is the probability that all channels in cell i are busy given that a call is handed over from a cell with k busy channels to cell i. From (4.4), (4.8) can be re-written as

$$p_{1,k} = \Pr\left[\{N_i = c\} \cap B_k \cap A_i \cap \left(\bigcup_{l=1}^{n} E_{l,k}\right) \middle| B_k \cap A_i\right]. \qquad (4.9)$$

Since $E_{1,k}, E_{2,k}, \cdots, E_{n,k}$ are mutually exclusive events, from [41], (4.9) can be expressed as

$$p_{1,k} = \sum_{l=1}^{n} \Pr[\{N_i = c\} \cap B_k \cap A_i \cap E_{l,k} | B_k \cap A_i]$$

$$= \sum_{l=1}^{n} \frac{\Pr[\{N_i = c\} \cap E_{l,k} \cap B_k \cap A_i]}{\Pr[B_k \cap A_i]}$$

$$= \sum_{l=1}^{n} \frac{\Pr[\{N_i = c\} \cap E_{l,k} \cap B_k \cap A_i]}{\Pr[E_{l,k} \cap B_k \cap A_i]} \times \frac{\Pr[E_{l,k} \cap B_k \cap A_i]}{\Pr[B_k \cap A_i]}$$

$$= \sum_{l=1}^{n} \Pr[N_i = c | E_{l,k} \cap B_k \cap A_i] \times \Pr[E_{l,k} \cap B_k \cap A_i | B_k \cap A_i]. \quad (4.10)$$

From (4.4), we have $E_{l,k} \cap B_k = E_{l,k}$ and (4.10) is re-written as

$$p_{1,k} = \sum_{l=1}^{n} \Pr[N_i = c | E_{l,k} \cap A_i] \times \Pr[E_{l,k} \cap A_i | B_k \cap A_i]. \qquad (4.11)$$

In (4.11), $\Pr[E_{l,k} \cap A_i | B_k \cap A_i]$ represents the routing probability that a call is handed over from cell l with k busy channels to cell i, given that the call is handed over from a cell with k busy channels to cell i. Since we consider homogeneous topology and

53

routing pattern, the MS can only move into cell i from any one of cell i's neighbors with probability $1/s$. Therefore,

$$\Pr[E_{l,k} \cap A_i | B_k \cap A_i] = \begin{cases} 1/s, & \text{if } l \in S_i, \\ 0, & \text{otherwise.} \end{cases} \quad (4.12)$$

Substitute (4.12) into (4.11) to yield

$$p_{1,k} = \sum_{l \in S_i} \Pr[N_i = c | E_{l,k} \cap A_i] \times \frac{1}{s}. \quad (4.13)$$

Due to homogeneous assumptions, (4.13) can be re-written as

$$\begin{aligned} p_{1,k} &= \Pr[N_i = c | E_{l,k} \cap A_i] \\ &= \Pr[N_i = c | \{\text{A call is handed over from cell } l \text{ to cell } i\} \\ &\quad \cap \{N_l = k\}], \end{aligned} \quad (4.14)$$

where $1 \le k \le c$. It is clear that Event $\{N_i = c\}$ and Event $\{$A call is handed over from cell l to cell $i\}$ are conditionally independent given Event $\{N_l = k\}$, where $1 \le k \le c$. Therefore, from [41], (4.14) can be expressed as

$$\begin{aligned} p_{1,k} &= \Pr[N_i = c | N_l = k] \\ &= \frac{\Pr[\{N_i = c\} \cap \{N_l = k\}]}{\Pr[N_l = k]} \\ &= \frac{\Pr[N_l = k | N_i = c] \times \Pr[N_i = c]}{\Pr[N_l = k]}, \end{aligned} \quad (4.15)$$

where $1 \le k \le c$. In (4.7), let

$$p_{2,k} = \Pr[B_k | A_i]. \quad (4.16)$$

Given that a call is handed over into cell i, $p_{2,k}$ is the probability that the call is handed over out of a cell with k busy channels. From (4.3) and $B_k = B_k \cap (\bigcup_{l=1}^{n} D_l)$, (4.16) can be re-written as

$$p_{2,k} = \Pr\left[B_k \cap \left(\bigcup_{l=1}^{n} D_l \right) \middle| A_i \right]. \quad (4.17)$$

54

Since $D_1, D_2, \cdots, D_n$ are mutually exclusive events, from [41], (4.17) can be expressed as

$$
\begin{aligned}
p_{2,k} &= \sum_{l=1}^{n} \Pr[B_k \cap D_l | A_i] \\
&= \sum_{l=1}^{n} \frac{\Pr[B_k \cap D_l \cap A_i]}{\Pr[A_i]} \\
&= \sum_{l=1}^{n} \frac{\Pr[B_k \cap D_l \cap A_i]}{\Pr[D_l \cap A_i]} \times \frac{\Pr[D_l \cap A_i]}{\Pr[A_i]} \\
&= \sum_{l=1}^{n} \Pr[B_k | D_l \cap A_i] \times \Pr[D_l | A_i].
\end{aligned}
\tag{4.18}
$$

In (4.18), $\Pr[D_l | A_i]$ is the probability that under the condition that a call is handed over into cell i, it came from cell l. Because of homogeneous topology and routing pattern, the handover call is from any of cell i's neighbors with the same probability. Therefore,

$$
\Pr[D_l | A_i] = \begin{cases} 1/s, & \text{if } l \in S_i, \\ 0, & \text{otherwise.} \end{cases}
\tag{4.19}
$$

Substitute (4.19) into (4.18) to yield

$$
\begin{aligned}
p_{2,k} &= \sum_{l \in S_i} \Pr[B_k | D_l \cap A_i] \times \frac{1}{s} \\
&= \sum_{l \in S_i} \Pr[N_l = k | D_l \cap A_i] \times \frac{1}{s},
\end{aligned}
\tag{4.20}
$$

where $1 \le k \le c$. Due to network homogeneity, $\sum_{l \in S_i} \Pr[N_l = k | D_l \cap A_i] = s \times \Pr[N_l = k | D_l \cap A_i]$. Thus, (4.20) is re-written as

$$
\begin{aligned}
p_{2,k} &= s \times \Pr[N_l = k | D_l \cap A_i] \times \frac{1}{s} \\
&= \Pr[N_l = k | D_l \cap A_i] \\
&= \Pr[\{N_l = k\} \cap D_l \cap A_i | D_l \cap A_i],
\end{aligned}
\tag{4.21}
$$

where $1 \le k \le c$. In (4.21), Event $D_l \cap A_i$ represents the event that a call is handed over from cell l to cell i, and Event $\{N_l = k\} \cap D_l \cap A_i$ means that a call is handed over from cell l with k busy channels to cell i. Note that the mobility rate of an

55

MS is η. Because of homogeneous topology and routing pattern, if a busy channel in cell l will be released due to MS movement to cell i, then the channel is released at rate η/s, and (4.21) can be re-written as

$$p_{2,k} = \frac{\Pr[N_l = k] \times k \times \left(\frac{\eta}{s}\right)}{\displaystyle\sum_{k=1}^{c} \Pr[N_l = k] \times k \times \left(\frac{\eta}{s}\right)} = \frac{k \Pr[N_l = k]}{E[N_l]}. \tag{4.22}$$

Substituting (4.15) and (4.22) into (4.7), we have

$$\begin{aligned}
p_f &= \sum_{k=1}^{c} p_{1,k} \times p_{2,k} \\
&= \sum_{k=1}^{c} \frac{\Pr[N_l = k | N_i = c] \times \Pr[N_i = c]}{\Pr[N_l = k]} \times \frac{k \Pr[N_l = k]}{E[N_l]} \\
&= \left\{ \frac{\Pr[N_i = c]}{E[N_l]} \right\} \left\{ \sum_{k=1}^{c} k \Pr[N_l = k | N_i = c] \right\} \\
&= p_b \times \frac{E[N_l | N_i = c]}{E[N_l]}. \tag{4.23}
\end{aligned}$$

The probability p_{nc} was derived in [34], which is expressed as follows.

$$p_{nc} = p_b + \frac{\eta (1 - p_b) [1 - f_m^*(\mu)] p_f}{\mu [1 - (1 - p_f) f_m^*(\mu)]} \tag{4.24}$$

where $f_m^*(s)$ is the Laplace transform of the MS cell residence time distribution. For exponential MS cell residence time distributions, $f_m^*(\mu) = \dfrac{\eta}{\mu + \eta}$, and (4.24) can be expressed as

$$\begin{aligned}
p_{nc} &= p_b + \frac{\eta (1 - p_b) \left(\dfrac{\mu}{\mu + \eta}\right) p_f}{\mu \left[1 - (1 - p_f) \left(\dfrac{\eta}{\mu + \eta}\right)\right]} \\
&= \frac{\mu p_b + \eta p_f}{\mu + \eta p_f}. \tag{4.25}
\end{aligned}$$

$\Pr[N_i = c]$, $E[N_l]$ and $E[N_l | N_i = c]$ in (4.1) and (4.23) are derived in Appendix B by solving an n-dimensional Markov process, which are then used to compute p_{nc}. We note that our approach is similar to the one proposed in [37], where the model can be extended for heterogeneous network modeling (e.g., the numbers of channels in cells are different).

4.2 Approximate Model for General MS Cell Residence Times

This section proposes an approximate solution for modeling general MS cell residence times. The idea is to adjust the exact analytic solution developed in the previous section. Specifically, we approximate the general MS cell residence time by an exponential distribution with the adjusted rate η^*.

Consider the timing diagram in Fig. 4.1. In this figure, a call arrives when the MS resides in cell 1. The call holding time is t_1. The MS cell residual time at cell 1 (i.e., the interval between when the call arrives and when the MS moves out of cell 1) is t_2. For $i = 2, 3, \ldots$, if the call is successfully handed over to cell i, then the remaining call holding time is $t_{3,i}$. For $i = 2, 3, \ldots$, the MS cell residence time at cell i (i.e., the interval between when the MS enters cell i and when it moves out of cell i) is $t_{4,i}$. Since the call holding times are exponentially distributed, t_1 and $t_{3,i}$ have the same exponential distribution. Let random variable X be the number of call completions for a connected call. Note that the value of X is either one or zero, depending on whether the call is eventually completed or forced to terminate. Let random variable Y be the number of handovers for a connected call. We derive $E[X]$ and $E[Y]$ as follows. In Fig. 4.1, let $\alpha_1 = \Pr[t_1 > t_2]$ be the probability that a new call is not completed before the MS moves out of the first cell, and $\alpha_2 = \Pr[t_{3,i} > t_{4,i}]$ be the probability that a handover call is not completed before the MS moves out of cell i, where $i = 2, 3, \ldots$ From [27], we have

$$\alpha_1 = \left(\frac{\eta}{\mu}\right)\left[1 - f_m^*(\mu)\right] \quad \text{and} \quad \alpha_2 = f_m^*(\mu). \tag{4.26}$$

With (4.26), $E[X]$ is derived as follows.

$$
\begin{aligned}
E[X] &= 1 \times \Pr[\text{A connected call is completed}] \\
&\quad + 0 \times \Pr[\text{A connected call is forced to terminate}] \\
&= (1 - \alpha_1) + \sum_{k=1}^{\infty} \left\{ \alpha_1 \left[(1 - p_f)\alpha_2\right]^{k-1} (1 - p_f)(1 - \alpha_2) \right\}. \tag{4.27}
\end{aligned}
$$

In the right hand side of (4.27), $(1 - \alpha_1)$ is the probability that a new call is completed

before the MS moves out of the cell, and $\left\{\alpha_1\left[(1-p_f)\alpha_2\right]^{k-1}(1-p_f)(1-\alpha_2)\right\}$ is the probability that a call is successfully handed over for k times and is completed at the $k+1$-st cell, where $k \geq 1$. From (4.27), we have

$$E[X] = (1-\alpha_1) + \frac{\alpha_1(1-p_f)(1-\alpha_2)}{1-(1-p_f)\alpha_2}. \tag{4.28}$$

Substitute (4.26) into (4.28) to yield

$$E[X] = 1 - \frac{\eta\left[1-f_m^*(\mu)\right]p_f}{\mu\left[1-(1-p_f)f_m^*(\mu)\right]}. \tag{4.29}$$

Similarly, $E[Y]$ is derived as

$$E[Y] = \sum_{k=1}^{\infty} k\left\{\alpha_1\left[(1-p_f)\alpha_2\right]^{k-1}\left[(1-p_f)(1-\alpha_2)+p_f\right]\right\} \tag{4.30}$$

$$= \frac{\alpha_1}{1-(1-p_f)\alpha_2}. \tag{4.31}$$

In the right hand side of (4.30), $\left\{\alpha_1\left[(1-p_f)\alpha_2\right]^{k-1}\left[(1-p_f)(1-\alpha_2)+p_f\right]\right\}$ is the probability that a call is successfully handed over for k times before it is completed, or is successfully handed over for $k-1$ times and forced to terminate at the k-th handover. Substitute (4.26) into (4.31) to yield

$$E[Y] = \frac{\eta[1-f_m^*(\mu)]}{\mu[1-(1-p_f)f_m^*(\mu)]}. \tag{4.32}$$

Define θ as

$$\theta = \frac{E[X]}{E[X+Y]}. \tag{4.33}$$

Although X and Y are dependent random variables, we have [14]

$$E[X+Y] = E[X] + E[Y]. \tag{4.34}$$

From (4.34), (4.33) is re-written as

$$\theta = \frac{E[X]}{E[X]+E[Y]}. \tag{4.35}$$

From (4.29) and (4.32), (4.35) can be expressed as

$$\theta = 1 - \frac{\eta[1-f_m^*(\mu)]}{\mu\left\{1-(1-p_f)f_m^*(\mu)+\left(\dfrac{\eta}{\mu}\right)(1-p_f)\left[1-f_m^*(\mu)\right]\right\}}. \tag{4.36}$$

For the exponential MS cell residence time distribution, $f_m^*(\mu) = \dfrac{\eta}{\mu + \eta}$, and (4.36) can be simplified as

$$\theta = \frac{\mu}{\mu + \eta}. \tag{4.37}$$

To approximate the general MS cell residence time distribution by an exponential distribution, we define η^* as the mobility rate for the approximate exponential distribution. From (4.37), the approximate mobility rate is

$$\eta^* = \frac{\mu}{\theta} - \mu. \tag{4.38}$$

By substituting (4.36) into (4.38), we have

$$\eta^* = \frac{\eta[1 - f_m^*(\mu)]}{1 - (1 - p_f)f_m^*(\mu) - \left(\dfrac{\eta}{\mu}\right)\left[1 - f_m^*(\mu)\right]p_f}. \tag{4.39}$$

Therefore, for MS cell residence time distribution with Laplace transform $f_m^*(s)$, this distribution can be approximated by an exponential distribution with mobility rate η^* (expressed in (4.39)) for the channel assignment model. The probabilities p_b, p_f, and p_{nc} are computed as follows (refer to Appendix B for more details).

Input parameters: λ, μ, η, c, and $f_m^*(\mu)$

Output measures: θ, p_b, p_f, and p_{nc}

Step 1. Select an initial value for p_f.

Step 2. $p_{f,old} \leftarrow p_f$. Compute η^* by using (4.39).

Step 3. Construct the state transition rate matrix $\mathbf{Q}$, where η in Fig. B.1 is replaced by η^*. Then the Π vector is solved from $\Pi\mathbf{Q} = \mathbf{0}$ and $\displaystyle\sum_{k_1=0}^{c}\sum_{k_2=0}^{c}\cdots\sum_{k_n=0}^{c}\pi_{k_1,k_2,\cdots,k_n} = 1$. Finally, the Π vector is used to compute p_b and p_f by using (4.1), (4.23), (B.1), (B.2), and (B.3).

Step 4. Let ϵ be a pre-defined value ($\epsilon = 0.001$ in our example). If $|p_f - p_{f,old}| > \epsilon$, then go to Step 2. Otherwise, go to Step 5.

Step 5. The values for θ, p_b, and p_f converge. Compute p_{nc} by using (4.24).

4.3　Numerical Examples

This section uses numerical examples to compare the analytic model in Sections 4.1 and 4.2 (called the "new model") with the model we proposed in [34] (called the "old model"). For the demonstration purpose, we consider a three-cell cellular system (i.e., $n = 3$; see Fig. 1.3), where each cell has nine channels (i.e., $c = 9$) and two neighbors (i.e., $s = 2$). Such systems have been manufactured and deployed in Asia [47]. For other small-scale cell configurations, similar results are observed, which will not be presented in this chapter.

Fig. 4.3 plots the blocking probabilities against the mobility rate η for the exponential MS cell residence time model, where the call holding times are exponentially distributed with rate $\mu = 0.3\lambda$. The figure indicates that both p_b and p_f decrease as η increases. Since the number of handovers increases as η increases, p_{nc} increases as η increases. Same phenomena were found in [7, 28], and the reader is referred to these pervious studies for more details. We observe that the new analytic results almost match the simulation results, while the errors between the old analytic model and the simulation experiments can be up to 18%. The figure suggests that the new analytic model is more accurate than the old one. Fig. 4.3 also indicates that the higher the mobility, the more the inconsistency between the old analytic and the simulation results. Therefore, the advantage of the new analytic model becomes significant when the mobility is high. The old analytic model is not as accurate as the new one because it assumes $p_f = p_b$, while in the new analytic model, we have proven that $p_f = p_b \times \dfrac{E[N_l | N_i = c]}{E[N_l]} \neq p_b$ (see (4.23)). When the mobility is low, the value of $E[N_l | N_i = c]$ is close to $E[N_l]$. Consequently, $p_f \approx p_b$ and the old analytic model works well.

For a general MS cell residence time distribution, its variance ν may have significant impact on the output measures. For Gamma MS cell residence times, Table 4.1 shows the call incompletion probability p_{nc} values and their errors between analytic and simulation models, where the MS cell residence time variances are $\nu = 0.01/\eta^2$,
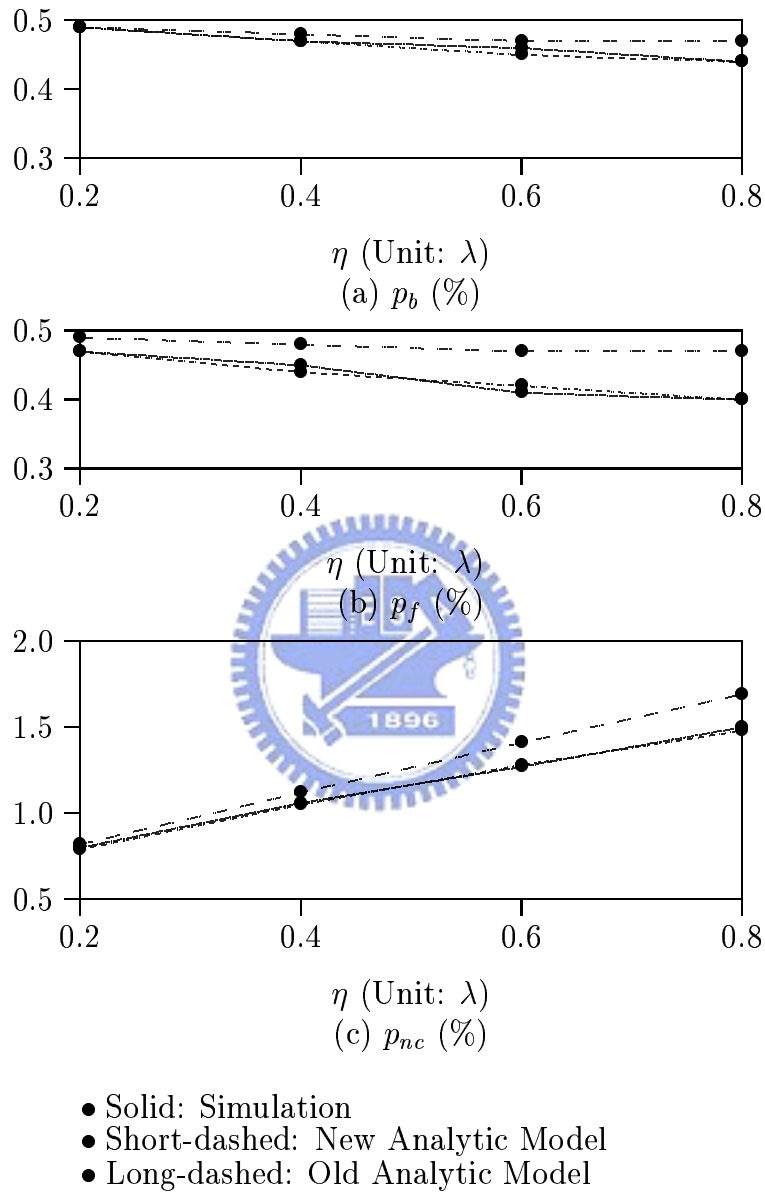
Figure 4.3: Results for Exponential MS Cell Residence Time Model ($\mu = 0.3\lambda$)

Table 4.1: The p_{nc} Values and Errors: Analytic Models vs. Simulation ($\eta = 0.5\lambda$)

ν (Unit: $1/\eta^2$)	0.01	0.1	1
$\mu = 0.1\lambda$			
Simulation	39.429%	39.5817%	39.1805%
Old Analytic Model	40.915%	40.7124%	40.4659%
New Analytic Model	39.4335%	39.4122%	39.2026%
$\mu = 0.2\lambda$			
Simulation	8.3426%	8.4134%	8.2144%
Old Analytic Model	9.2013%	9.1963%	9.1491%
New Analytic Model	8.2463%	8.2428%	8.2131%
$\mu = 0.3\lambda$			
Simulation	1.2036%	1.1719%	1.1532%
Old Analytic Model	1.2674%	1.2673%	1.2659%
New Analytic Model	1.1665%	1.1665%	1.1655%

(a) p_{nc}

ν (Unit: $1/\eta^2$)	0.01	0.1	1
$\mu = 0.1\lambda$			
Old Analytic Model	3.77%	2.86%	3.28%
New Analytic Model	0.01%	0.43%	0.06%
$\mu = 0.2\lambda$			
Old Analytic Model	10.29%	9.31%	11.38%
New Analytic Model	1.15%	2.03%	0.02%
$\mu = 0.3\lambda$			
Old Analytic Model	5.3%	8.14%	9.77%
New Analytic Model	3.08%	0.46%	1.07%

(b) Errors of p_{nc}

$0.1/\eta^2$, and $1/\eta^2$. In this table, the mobility rate is $\eta = 0.5\lambda$ and the call completion rates are $\mu = 0.1\lambda$, 0.2λ, and 0.3λ respectively. The results suggest that for various ν, the p_{nc} values of the new analytic model are much closer to the simulation results than that of the old analytic model.

4.4 Conclusions

Most analytic modeling studies for cellular networks assume that the handover traffic to a cell is a fixed-rate Poisson process. This assumption may introduce significant inaccuracy for modeling small-scale cellular networks. This chapter showed that the handover traffic to a cell depends on the workloads of the neighboring cells. We derived the exact equation for the handover force-termination probability when the MS cell residence times are exponentially distributed. Then we proposed an approximate model for general MS cell residence time distributions. We used the analytic model to validate against the simulation model with the excess life modeling technique. Then the analytic results were compared with a previously proposed model, which indicated that the new model can capture the handover behavior much better than the old one for small-scale cellular networks.

Chapter 5

Conclusions and Future Work

This dissertation emphasizes the issues related to the mobile user mobility, especially the MS cell residence times and the excess life of a mobile user in a cell or an LA. We explored excess-life random number generation, fixed and dynamic PLAU schemes, and performance modeling for small-scale cellular telecommunications networks. This chapter summarizes our studies, and briefly discusses directions of the future work.
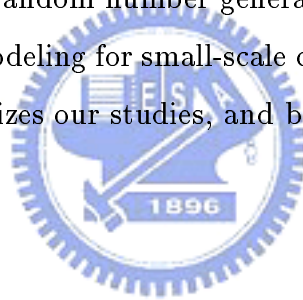
5.1 Summary

In performance evaluation of a cellular telecommunications network, it is important to derive the excess life distribution from the MS cell residence times. This distribution determines if a connected call will be handed over to a new cell, and therefore significantly affects the handover force-termination probability of the network. Since we use simulation results to validate against the analytical results, we need to generate the excess-life random numbers in the simulation model. However, generating the excess-life random numbers is not a trivial task, which has not been addressed in the literature. Chapter 2 showed how to derive the excess life distribution and to generate the random numbers from the excess life distribution. We then developed the excess-life random number generation procedures for MS cell residence times with gamma, Pareto, lognormal and Weibull distributions. Our

64

study indicated that the generated random numbers closely match the true excess-life distribution (i.e., Equation (2.1)). Therefore, our procedures can be utilized to efficiently generate excess-life random numbers in cellular telecommunications network simulations. These procedures were used in the subsequent research topics addressed in this dissertation, as described below.

The first example was shown in Chapter 3. In cellular communications networks, PLAU is utilized to detect the presence of an MS. In 3GPP Technical Specifications 23.012 and 24.008, a fixed PLAU scheme was proposed for UMTS where the interval between two PLAUs is of fixed length. We observed that MS presence can also be detected through call and movement activities. Therefore we proposed a dynamic PLAU scheme where the PLAU interval is dynamically adjusted based on the call and NLAU traffic. An analytic model was developed to validate against the simulation model with the excess life modeling technique. Then we investigated the performance of dynamic and fixed PLAU schemes. Our study indicated that compared with fixed PLAU, dynamic PLAU significantly reduces the network signaling traffic caused by PLAU. As a final remark, in dynamic PLAU, storage and the mechanism maintaining $m = 20$ or 30 inter checkpoint arrival time samples for an MS can be practically implemented in the UMTS network (specifically, in the VLR). The PLAU/ID timer value can be efficiently computed using the *window averaging* technique [30].

The second example was shown in Chapter 4. Most analytic modeling studies for cellular telecommunications networks assume that the handover traffic to a cell is a fixed-rate Poisson process. This assumption may introduce significant inaccuracy for modeling small-scale cellular telecommunications networks. This chapter showed that the handover traffic to a cell depends on the workloads of the neighboring cells. We derived the exact equation for the handover force-termination probability when the MS cell residence times were exponentially distributed. Then we proposed an approximate model for general MS cell residence time distributions. We used the analytic model to validate against the simulation model with the excess life modeling technique. Then the analytic results were compared with a previously proposed

model, which indicated that the new model can capture the handover behavior much better than the old one for small-scale cellular telecommunications networks.

5.2 Future Work

In performance modeling, Laplace transforms play an important role in analytic models (e.g., equations (3.7), (3.8) and (3.9) in Chapter 3, and equations (4.24) and (4.39) in Chapter 4). However, while some distributions have closed-form Laplace transforms (e.g., exponential distribution, Erlang distribution, gamma distribution, etc.), some other heavy-tailed distributions (e.g., Pareto distribution, lognormal distribution, and Weibull distribution with the value of the shape parameter being within the range $(0, 1)$) do not have closed-form Laplace transforms, which introduce significant complexity to the derivation of the analytical formulas. We will investigate a special Laplace transform approximation technique called *Transform Approximation Method* (TAM) as our future work (see Appendix C for more details).

Appendix A

The Call-based Simulation

This appendix describes the basic call-based discrete event simulation for mobile telecommunications network. Several random variables are defined: the *inter call arrival time* (the call arrivals are typically modeled as a Poisson process), the *call holding time*, the *MS cell residence time* and the *excess life* of the MS cell residence time. Three basic event types are considered: the **arrival** event (a call arrival), the **move** event (an MS movement), and the **complete** event (a call completion). Every event is associated with a timestamp representing the time when the event occurs. All unprocessed events are inserted in an *event list* and are processed in the non-decreasing timestamp order. The notation used in this appendix is listed below.

- t_c: the call holding time

- τ_c: the remaining call holding time

- t_m: the MS cell residence time

- τ_m: the excess life of the MS cell residence time

Details of the call-based simulation are described in the following steps:

Step 1 (Initialization): Generate the first **arrival** event and insert it in the event list.

Step 2. Remove the next event from the event list. If the event type is **arrival** then go to Step 3. If the type is **move** then go to Step 5. If the type is **complete** then go to Step 6.

Step 3 (arrival). Check if the cell can accommodate this call based on some wireless resource allocation policy. If not, reject the call, update the call statistics, and go to Step 4. Otherwise, generate the random numbers for the excess life τ_m of the MS cell residence time and the call holding time t_c.

> **Step 3.1.** If $\tau_m > t_c$, generate a **complete** event with timestamp "current time+t_c".
>
> **Step 3.2.** If $\tau_m < t_c$, generate a **move** event with timestamp "current time+τ_m". Note that when the **move** event occurs, the remaining call holding time is $\tau_c = t_c - \tau_m$.

Insert the generated event into the event list.

Step 4. Generate the next **arrival** event according to the Poisson process and insert it into the event list. Go to Step 2.

Step 5 (move). The MS moves from the old cell to the new cell. Check if the new cell can accommodate this handover call. If not, drop the call, update the call statistics, and go to Step 2. Otherwise, generate the MS cell residence time t_m. The remaining call holding time is τ_c.

> **Step 5.1.** If $t_m > \tau_c$, generate a **complete** event with timestamp "current time+τ_c".
>
> **Step 5.2.** If $t_m < \tau_c$, generate the next **move** event with timestamp "current time+t_m". Note that when the next **move** event occurs the remaining call holding time is $\tau_c \leftarrow \tau_c - t_m$.

Insert the generated event into the event list. Go to Step 2.

Step 6 (complete). Reclaim the resources used by this call. Update the call statistics, and go to Step 2.

The simulation can be terminated based on various criteria. For example, at Step 3, we may check if some terminating conditions are satisfied (e.g., 1000,000 call arrivals have been simulated). If so, the simulation terminates.

Appendix B

Solving the n-Dimensional Markov Process

This appendix solves the n-dimensional Markov Process for $\Pr[N_i = c]$, $E[N_l]$ and $E[N_l | N_i = c]$ in (4.1) and (4.23) in Chapter 4. The notation used in this appendix is listed below.

- N_i: a random variable for the number of busy channels in cell i, for $1 \leq i \leq n$

- n: the number of cells in the cellular system under consideration

- c: the number of channels in each cell in the cellular system under consideration

- S_i: the index set of cell i's neighbors, for $1 \leq i \leq n$

- s: we assume that $|S_i| = |S_j| = s$ for $1 \leq i, j \leq n$, where $|S_i|$ is the number of cell i's neighbors for $1 \leq i \leq n$ and $|S_j|$ is the number of cell j's neighbors for $1 \leq j \leq n$

- δ_j: for $1 \leq j \leq n$,

$$\delta_j = \begin{cases} 0, & \text{if } k_j < c, \\ 1, & \text{if } k_j = c. \end{cases}$$

- k_i: the number of busy channels in cell i for $1 \leq i \leq n$ and $0 \leq k_i \leq c$

70

- $(k_1, k_2, \cdots, k_n)$: a state in the Markov process

- λ: the transition rate from state $(k_1, \cdots, k_i, \cdots, k_l, \cdots, k_n)$ to state $(k_1, \cdots, k_i + 1, \cdots, k_l, \cdots, k_n)$, where $k_i < c$ and $1 \le i \ne j \le n$ (note that λ is also the new call arrival rate to a cell mentioned in Chapter 4)

- μ: the call completion rate (note that $1/\mu$ is the mean call holding time mentioned in Chapter 4)

- η: the MS mobility rate (note that $1/\eta$ is the mean MS cell residence time mentioned in Chapter 4)

- $\pi_{k_1, k_2, \cdots, k_n}$: the steady state probability for state $(k_1, k_2, \cdots, k_n)$

- Π: the stationary probability distribution vector; i.e., $\Pi = [\pi_{k_1, k_2, \cdots, k_n}]_{1 \times (c+1)^n}$

- Q: the state transition rate matrix for the Markov process

- p_b: the new call blocking probability; i.e., the number of new call blockings divided by the number of new calls

- p_f: the handover force-termination probability; i.e., the number of forced terminations divided by the number of handovers

- p_{nc}: the call incompletion probability (the probability that a call is either blocked or forced to terminate); i.e., the sum of the numbers of new call blockings and forced terminations divided by the number of new calls (note that $p_{nc} \ne p_b + p_f$)

We assume that the call holding times and the MS cell residence times have exponential distributions. The n-cell cellular system can be modeled by an n-dimensional Markov process. A state in the Markov process is of the form $(k_1, k_2, \cdots, k_n)$, where k_i is the number of busy channels in cell i for $1 \le i \le n$ and $0 \le k_i \le c$. The total number of states for this Markov process is $(c + 1)^n$. Let $\pi_{k_1, k_2, \cdots, k_n}$ denote the steady state probability for state $(k_1, k_2, \cdots, k_n)$, and $\Pi = [\pi_{k_1, k_2, \cdots, k_n}]_{1 \times (c+1)^n}$
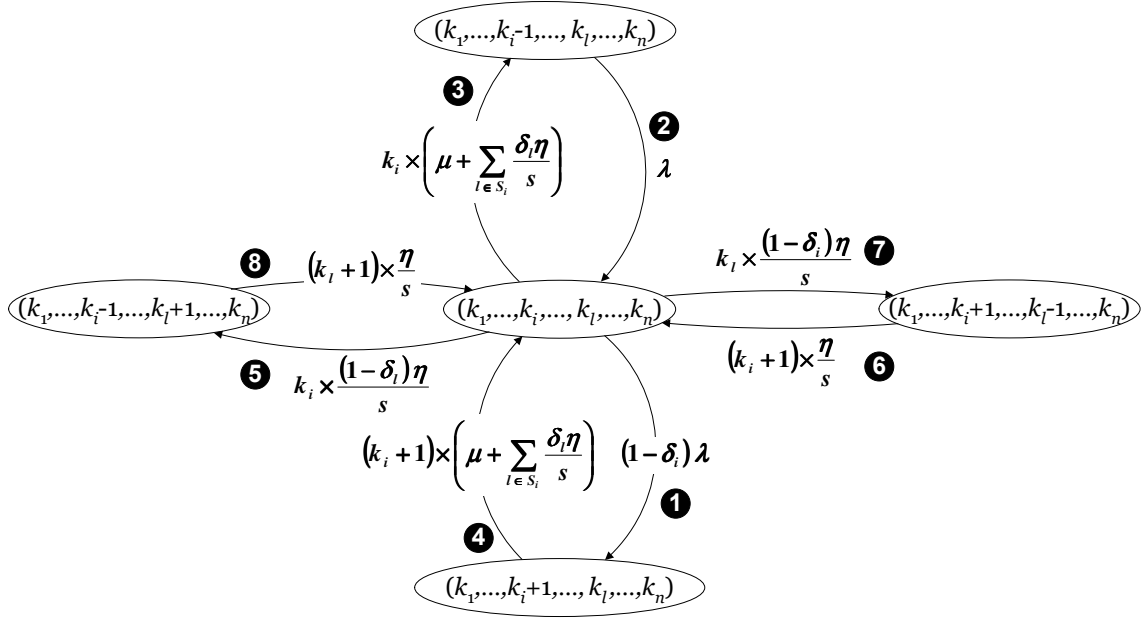
Figure B.1: The Transitions for State $(k_1, \cdots, k_i, \cdots, k_l, \cdots, k_n)$

denote the stationary probability distribution vector. Fig. B.1 illustrates the state transition diagram for state $(k_1, \cdots, k_i, \cdots, k_l, \cdots, k_n)$ in this Markov process, where $1 \leq i \leq n$ and $l \in S_i$. In this figure, δ_j (for $1 \leq j \leq n$) is defined as

$$\delta_j = \begin{cases} 0, & \text{if } k_j < c, \\ 1, & \text{if } k_j = c. \end{cases}$$

The transitions for state $(k_1, \cdots, k_i, \cdots, k_l, \cdots, k_n)$ are described as follows.

Transition 1 (New Call Arrival): When a new call arrives at cell i, the process moves from state $(k_1, \cdots, k_i, \cdots, k_l, \cdots, k_n)$ to state $(k_1, \cdots, k_i+1, \cdots, k_l, \cdots, k_n)$. Since the transition rate is λ for $k_i < c$, and 0 for $k_i = c$, the net transition rate is $(1 - \delta_i)\lambda$. Transition 2 is similar to Transition 1, and the details are omitted.

Transition 3 (Channel Release due to Call Completion or Handover to

a Fully Occupied Cell): The Markov process moves from state $(k_1, \cdots, k_i, \cdots, k_l, \cdots, k_n)$ to state $(k_1, \cdots, k_i - 1, \cdots, k_l, \cdots, k_n)$ in two cases:

72

1. A call is completed at cell i with rate μ.

2. The MS moves from cell i to a neighboring cell l where $k_l = c$. Clearly, this call is forced to terminate. The MS moves to a neighboring cell l with rate $\dfrac{\eta}{s}$ and the handover fails if $\delta_l = 1$. Therefore, the total transition rate is $\displaystyle\sum_{l \in S_i} \dfrac{\delta_l \eta}{s}$.

Since there are k_i busy channels (i.e., k_i ongoing calls) at cell i, the rate for Transition 3 is

$$k_i \times \left(\mu + \sum_{l \in S_i} \frac{\delta_l \eta}{s} \right).$$

Transition 4 is similar to Transition 3, and the details are omitted.

Transition 5 (Successful Handover to a Non-fully Occupied Cell): When a call is handed over from cell i to cell l where $k_l < c$, the process moves from state $(k_1, \cdots, k_i, \cdots, k_l, \cdots, k_n)$ to state $(k_1, \cdots, k_i - 1, \cdots, k_l + 1, \cdots, k_n)$. Since there are k_i ongoing calls at cell i, the transition rate is $k_i \times \dfrac{(1 - \delta_l)\,\eta}{s}$. Transitions 6, 7, and 8 are similar to Transition 5, and the details are omitted.

Let $\mathbf{Q}$ denote the state transition rate matrix for the Markov process. We solve Π from the equilibrium equations $\Pi \mathbf{Q} = \mathbf{0}$ and $\displaystyle\sum_{k_1=0}^{c} \sum_{k_2=0}^{c} \cdots \sum_{k_n=0}^{c} \pi_{k_1,k_2,\cdots,k_n} = 1$. Then we derive $\Pr[N_i = c]$, $E[N_l]$ and $E[N_l | N_i = c]$ in (4.1) and (4.23) as follows. For $1 \leq i \leq n$,

$$\Pr[N_i = c] = \begin{cases} \displaystyle\sum_{k_2=0}^{c} \sum_{k_3=0}^{c} \cdots \sum_{k_n=0}^{c} \pi_{k_1=c,k_2,\cdots,k_n} & \text{for } i = 1, \\[2ex] \displaystyle\sum_{k_1=0}^{c} \cdots \sum_{k_{i-1}=0}^{c} \sum_{k_{i+1}=0}^{c} \cdots \sum_{k_n=0}^{c} \pi_{k_1,\cdots,k_i=c,\cdots,k_n} & \text{for } 1 < i < n, \\[2ex] \displaystyle\sum_{k_1=0}^{c} \sum_{k_2=0}^{c} \cdots \sum_{k_{n-1}=0}^{c} \pi_{k_1,k_2,\cdots,k_n=c} & \text{for } i = n. \end{cases} \quad (\text{B.1})$$

For $1 \leq l \leq n$,

$$\begin{aligned} E[N_l] &= \sum_{k_l=0}^{c} k_l \Pr[N_l = k_l] \\ &= \sum_{k_1=0}^{c} \cdots \sum_{k_l=0}^{c} \cdots \sum_{k_n=0}^{c} k_l \pi_{k_1,\cdots,k_l,\cdots,k_n}. \end{aligned} \quad (\text{B.2})$$

73

For $1 \le i \le n$ and $l \in S_i$,

$$E[N_l | N_i = c]$$

$$= \sum_{k_l=0}^{c} k_l \Pr[N_l = k_l | N_i = c]$$

$$= \sum_{k_l=0}^{c} k_l \times \frac{\Pr[\{N_l = k_l\} \cap \{N_i = c\}]}{\Pr[N_i = c]}$$

$$= \begin{cases} \dfrac{\sum\limits_{k_2=0}^{c} \cdots \sum\limits_{k_l=0}^{c} \cdots \sum\limits_{k_n=0}^{c} k_l \pi_{k_1=c,\cdots,k_l,\cdots,k_n}}{\sum\limits_{k_2=0}^{c} \sum\limits_{k_3=0}^{c} \cdots \sum\limits_{k_n=0}^{c} \pi_{k_1=c,k_2,\cdots,k_n}} & \text{for } i = 1, 1 < l < n, \\[4mm] \dfrac{\sum\limits_{k_2=0}^{c} \sum\limits_{k_3=0}^{c} \cdots \sum\limits_{k_n=0}^{c} k_n \pi_{k_1=c,k_2,\cdots,k_n}}{\sum\limits_{k_2=0}^{c} \sum\limits_{k_3=0}^{c} \cdots \sum\limits_{k_n=0}^{c} \pi_{k_1=c,k_2,\cdots,k_n}} & \text{for } i = 1, l = n, \\[4mm] \dfrac{\sum\limits_{k_1=0}^{c} \cdots \sum\limits_{k_l=0}^{c} \cdots \sum\limits_{k_{n-1}=0}^{c} k_l \pi_{k_1,\cdots,k_l,\cdots,k_n=c}}{\sum\limits_{k_1=0}^{c} \sum\limits_{k_2=0}^{c} \cdots \sum\limits_{k_{n-1}=0}^{c} \pi_{k_1,k_2,\cdots,k_n=c}} & \text{for } i = n, 1 < l < n, \\[4mm] \dfrac{\sum\limits_{k_1=0}^{c} \sum\limits_{k_2=0}^{c} \cdots \sum\limits_{k_{n-1}=0}^{c} k_1 \pi_{k_1,k_2,\cdots,k_n=c}}{\sum\limits_{k_1=0}^{c} \sum\limits_{k_2=0}^{c} \cdots \sum\limits_{k_{n-1}=0}^{c} \pi_{k_1,k_2,\cdots,k_n=c}} & \text{for } i = n, l = 1, \text{(B.3)} \\[4mm] \dfrac{\sum\limits_{k_1=0}^{c} \cdots \sum\limits_{k_{i-1}=0}^{c} \sum\limits_{k_{i+1}=0}^{c} \cdots \sum\limits_{k_n=0}^{c} k_1 \pi_{k_1,\cdots,k_i=c,\cdots,k_n}}{\sum\limits_{k_1=0}^{c} \cdots \sum\limits_{k_{i-1}=0}^{c} \sum\limits_{k_{i+1}=0}^{c} \cdots \sum\limits_{k_n=0}^{c} \pi_{k_1,\cdots,k_i=c,\cdots,k_n}} & \text{for } 1 < i < n, l = 1, \\[4mm] \dfrac{\sum\limits_{k_1=0}^{c} \cdots \sum\limits_{k_{i-1}=0}^{c} \sum\limits_{k_{i+1}=0}^{c} \cdots \sum\limits_{k_l=0}^{c} \cdots \sum\limits_{k_n=0}^{c} k_l \pi_{k_1,\cdots,k_i=c,\cdots,k_l,\cdots,k_n}}{\sum\limits_{k_1=0}^{c} \cdots \sum\limits_{k_{i-1}=0}^{c} \sum\limits_{k_{i+1}=0}^{c} \cdots \sum\limits_{k_n=0}^{c} \pi_{k_1,\cdots,k_i=c,\cdots,k_n}} & \text{for } 1 < i \ne l < n, \\[4mm] \dfrac{\sum\limits_{k_1=0}^{c} \cdots \sum\limits_{k_{i-1}=0}^{c} \sum\limits_{k_{i+1}=0}^{c} \cdots \sum\limits_{k_n=0}^{c} k_n \pi_{k_1,\cdots,k_i=c,\cdots,k_n}}{\sum\limits_{k_1=0}^{c} \cdots \sum\limits_{k_{i-1}=0}^{c} \sum\limits_{k_{i+1}=0}^{c} \cdots \sum\limits_{k_n=0}^{c} \pi_{k_1,\cdots,k_i=c,\cdots,k_n}} & \text{for } 1 < i < n, l = n \end{cases}$$

Substituting (B.1), (B.2) and (B.3) into (4.1) and (4.23), we obtain p_b and p_f respectively. Substituting these p_b and p_f expressions into (4.25), we obtain p_{nc}.

Appendix C

Transform Approximation Methods (TAM)

In performance modeling, analytic solutions often contain the Laplace transforms in their formulas [23, 4, 11, 38, 21, 33]. Some distributions have closed-form Laplace transforms, such as exponential distribution, Erlang distribution, gamma distribution, etc. However, heavy-tailed distributions, e.g., Pareto distribution, lognormal distribution, and Weibull distribution with shape parameter in the range $(0, 1)$, often do not have closed-form Laplace transforms, which introduce significant complexity to the derivation of the analytical formulas. To avoid the complexity, one feasible way is to use simulations instead of analytic solutions, which however could make the modeling results less persuasive due to the lack of analytical analysis. To resolve the problem, approximation methods for Laplace transforms are taken into account in the literature. In this appendix, we focus on a previously proposed approximation technique called *Transform Approximation Method* (TAM) [16, 17, 44]. We discuss both the previously proposed TAM techniques [16, 17, 44] and the modified TAM techniques we investigated. The notation used in this appendix is listed below.

- α: the shape parameter of a Pareto distribution

- β: the scale parameter of a Pareto distribution

- a: a value used in M2 (one of the modified TAM techniques that we proposed), where the value is suggested to be derived by the criterion that the "TAM distribution mean" matches the "true distribution mean"

- b: a randomly chosen value used in M3 and M4

- $f(x)$: the density function for an arbitrary distribution

- $F(x)$: the CDF for an arbitrary distribution

- $F^{-1}(x)$: the inverse function of $F(x)$

- $f^*(s)$: the Laplace transform for an arbitrary distribution

- $\hat{f}^*(s)$: the approximate Laplace transform for $f^*(s)$

- n: the number of partitions in TAM techniques

- γ_i: the partition interval length in the SFGM TAM and M1

- $\max\limits_{1 \leq i \leq n} \gamma_i$: the norm of the partition in the SFGM TAM

- ν: a value used in the SFGM TAM, where the value is suggested to be derived by the criterion that the "TAM distribution mean" matches the "true distribution mean"

The TAM technique was first presented in [16, 17] (called the "HMBF TAM"), and then generalized in [44] (called the "SFGM TAM"). For a distribution with CDF $F(x)$, HMBF TAM approximates the Laplace transform

$$f^*(s) = \int_{x=0}^{\infty} e^{-sx} f(x) dx = \int_{F(x)=0}^{1} e^{-sx} dF(x), \tag{C.1}$$

by a Riemann sum

$$\hat{f}^*(s) \equiv \frac{1}{n} \left(\sum_{i=1}^{n} e^{-sx_i} \right), \tag{C.2}$$

where

$$x_i = F^{-1} \left(\frac{i - 0.5}{n} \right), \text{ for } i = 1, 2, \cdots, n. \tag{C.3}$$
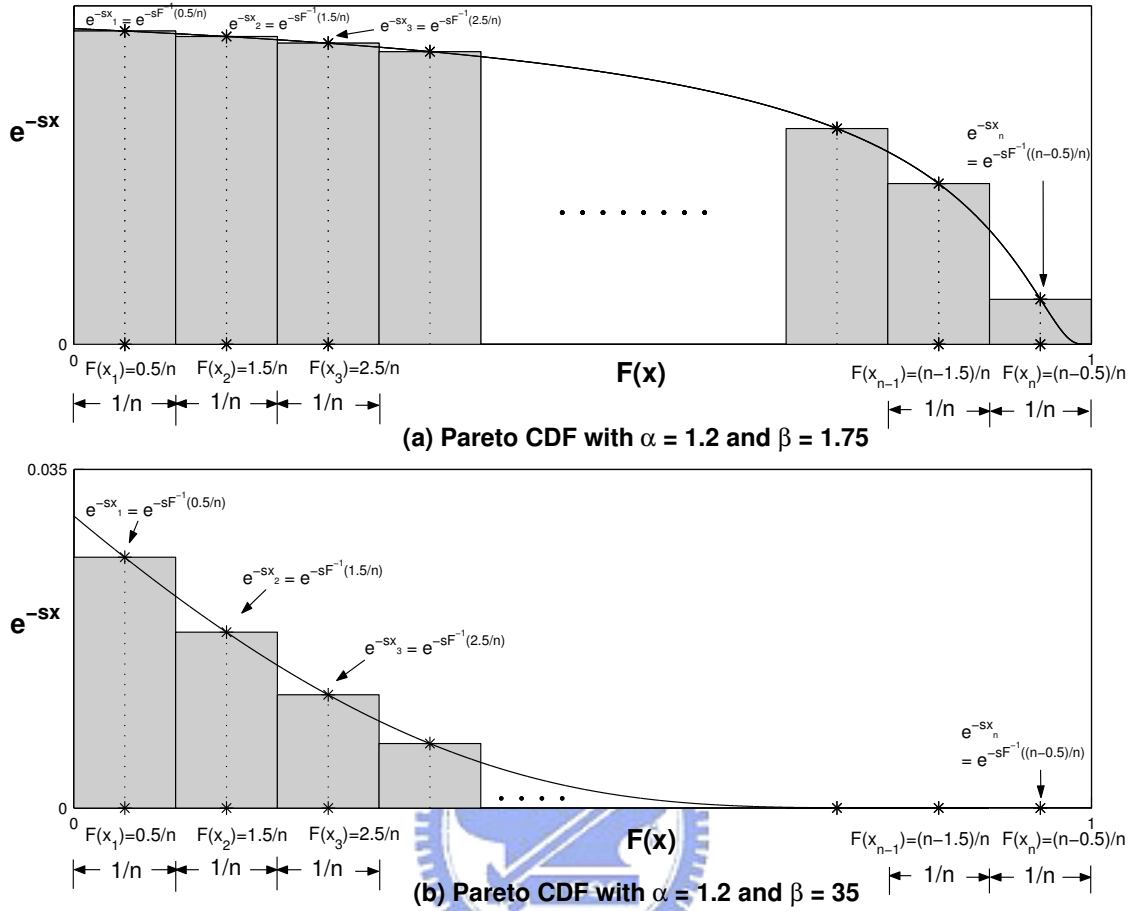
Figure C.1: Integral Area Slicing Based on HMBF TAM $(s = 0.1)$

Figure C.1 demonstrates how the integral area is sliced into smaller slots based on (C.2) for Pareto distribution with shape parameter $\alpha = 1.2$ and scale parameters $\beta = 1.75$ (Figure C.1(a)) and $\beta = 35$ (Figure C.1(b)) respectively (note that Figures C.1(a) and C.1(b) demonstrate different slope tendencies of the e^{-sx} curves respectively). In this figure, the x-axis represents $F(x)$, which is the variable of the integral in (C.1); and the y-axis represents the integrand e^{-sx} in (C.1). Note that both e^{-sx} and $F(x)$ are functions of x. From (C.3), we have $F(x_i) = (i - 0.5)/n$ for $i = 1, 2, \cdots, n$ in this figure. From (C.2), the partition intervals are equally sliced with length $1/n$. In Figure C.1(a), SFGM TAM approximates the curve well when $F(x)$ is small, but incurs significant errors when $F(x)$ is large. In Figure C.1(b), we

77

see the reverse results.

In [44], SFGM TAM approximates the Laplace transform $f^*(s)$ as

$$\hat{f}^*(s) \equiv \sum_{i=1}^{n} \gamma_i e^{-sx_i}, \tag{C.4}$$

where

$$x_i = F^{-1}(1 - \nu^i), \text{ for } i = 1, 2, \cdots, n, \text{ and } 0 < \nu < 1, \tag{C.5}$$

and the partition intervals are

$$\gamma_i = \begin{cases} \dfrac{F(x_1) + F(x_2)}{2}, & i = 1, \\ \dfrac{F(x_{i+1}) - F(x_{i-1})}{2}, & i = 2, 3, \cdots, n-1, \\ 1 - \dfrac{F(x_{n-1}) + F(x_n)}{2}, & i = n. \end{cases} \tag{C.6}$$

Based on (C.6), Figure C.2 demonstrates how the integral area is sliced into smaller slots for Pareto distribution with shape parameter $\alpha = 1.2$ and scale parameters $\beta = 1.75$ (Figure C.2(a)) and $\beta = 35$ (Figure C.2(b)) respectively. From (C.5), we have

$$F(x_i) = 1 - \nu^i, \text{ for } i = 1, 2, \cdots, n, \text{ and } 0 < \nu < 1. \tag{C.7}$$

By substituting (C.7) into (C.6), we have

$$\gamma_i = \begin{cases} 1 - \dfrac{\nu + \nu^2}{2}, & i = 1, \\ \dfrac{\nu^{i-1}(1 - \nu^2)}{2}, & i = 2, 3, \cdots, n-1, \\ \dfrac{\nu^{n-1} + \nu^n}{2}, & i = n. \end{cases} \tag{C.8}$$

From (C.7) and (C.8) (ignoring the boundary exceptions "$i = 1$" and "$i = n$" in (C.8)), it is clear that the larger the $F(x_i)$ value, the smaller the partition interval (the γ_i value; as illustrated in Figure C.2). The SFGM TAM [44] suggests to derive ν by the following criterion that the "TAM distribution mean" matches the "true distribution mean", i.e.,

$$-\left.\frac{d\hat{f}^*(s)}{ds}\right|_{s=0} = \int_{x=0}^{\infty} x f(x) dx. \tag{C.9}$$
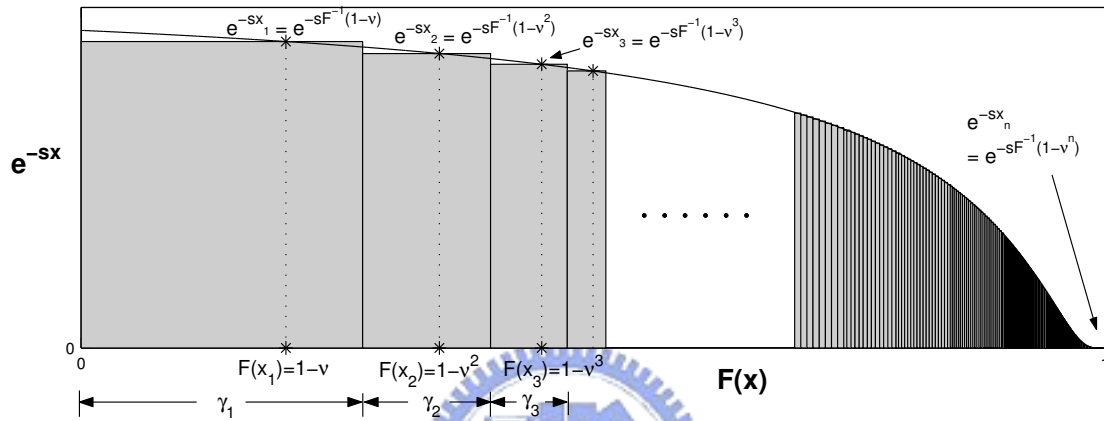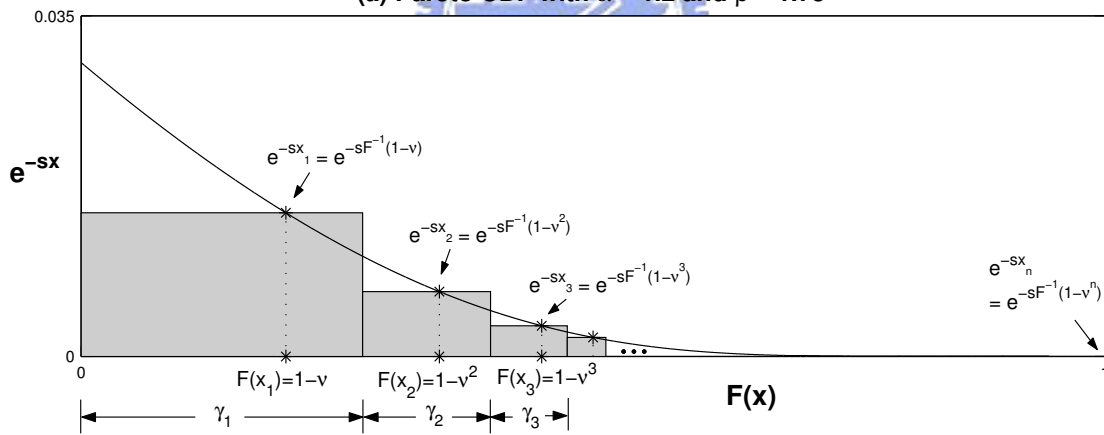
Figure C.2: Integral Area Slicing Based on SFGM TAM ($s = 0.1$)

Also note that $\max_{1 \le i \le n} \gamma_i$ is the norm of the partition [46]. From the definition of the definite integral as a limit of Riemann sum, we have the following theorem [46].

Theorem A.1. If $\lim_{n \to \infty} \max_{1 \le i \le n} \gamma_i = 0$, then the Riemann sum $\hat{f}^*(s)$ converges to the integral $f^*(s)$ for all s as $n \to \infty$.

From Theorem A.1, it is desirable to derive ν such that $\lim_{n \to \infty} \max_{1 \le i \le n} \gamma_i = 0$. In the following theorem, we show how to judge whether a derived ν can guarantee $\lim_{n \to \infty} \max_{1 \le i \le n} \gamma_i = 0$.

Theorem A.2. $\lim_{n \to \infty} \max_{1 \le i \le n} \gamma_i = 0$ if and only if the following two requirements are satisfied.

    Requirement A.1. $\lim_{n \to \infty} \nu = 1$ (i.e., as $n \to \infty$, $\nu \to 1$).

    Requirement A.2. $\lim_{n \to \infty} \nu^n = 0$ (i.e., as $n \to \infty$, $\nu^n \to 0$).

Proof :

$\Leftarrow$) We first prove the hypothesis that if both Requirements A.1 and A.2 are satisfied, then $\lim_{n \to \infty} \max_{1 \le i \le n} \gamma_i = 0$. For $2 \le i \le n - 1$ in (C.8), we have

$$\max_{2 \le i \le n-1} \gamma_i = \gamma_2 = \frac{\nu(1 - \nu^2)}{2}. \tag{C.10}$$

From Requirement A.1 and (C.10), we have

$$\lim_{n \to \infty} \max_{2 \le i \le n-1} \gamma_i = \lim_{n \to \infty} \frac{\nu(1 - \nu^2)}{2} = 0. \tag{C.11}$$

From Requirement A.1 and (C.8), we have

$$\lim_{n \to \infty} \gamma_1 = \lim_{n \to \infty} \left(1 - \frac{\nu + \nu^2}{2} \right) = 0. \tag{C.12}$$

From Requirement A.2 and (C.8), we have

$$\lim_{n \to \infty} \gamma_n = \lim_{n \to \infty} \left(\frac{\nu^{n-1} + \nu^n}{2} \right) = 0. \tag{C.13}$$

From (C.11), (C.12) and (C.13), the hypothesis holds.

80

$\Rightarrow$) Now we prove the hypothesis that if $\lim\limits_{n\to\infty}\max\limits_{1\leq i\leq n}\gamma_i = 0$, then both Requirements A.1 and A.2 are satisfied. For $i = 1$ in (C.8), we have the following inequality:

$$\gamma_1 = 1 - \frac{\nu + \nu^2}{2} \geq 1 - \frac{\nu + \nu}{2} = 1 - \nu. \tag{C.14}$$

For $i = n$ in (C.8), we have the following inequality:

$$\gamma_n = \frac{\nu^{n-1} + \nu^n}{2} \geq \frac{\nu^n + \nu^n}{2} = \nu^n. \tag{C.15}$$

Since $\lim\limits_{n\to\infty}\max\limits_{1\leq i\leq n}\gamma_i = 0$, we have

$$\lim\limits_{n\to\infty}\gamma_1 = 0 \text{ and } \lim\limits_{n\to\infty}\gamma_n = 0. \tag{C.16}$$

From (C.14), (C.16) and the fact that $\lim\limits_{n\to\infty}\gamma_1 = 0$, we have $\lim\limits_{n\to\infty}(1 - \nu) = 0$, which implies that $\lim\limits_{n\to\infty}\nu = 1$ (Requirement A.1 is satisfied). From (C.15), (C.16) and because $\lim\limits_{n\to\infty}\gamma_n = 0$, we have $\lim\limits_{n\to\infty}\nu^n = 0$ (Requirement A.2 is satisfied).

Q.E.D.

According to Theorem A.2, we show that criterion (C.9) does not necessarily yield appropriate ν that guarantees $\lim\limits_{n\to\infty}\max\limits_{1\leq i\leq n}\gamma_i = 0$:

Theorem A.3. The ν value derived from criterion (C.9) is affected by the value of n (see (C.4)). Therefore, ν may not converge to 1 as $n \to \infty$, which violates Requirement A.1.

Proof: Consider the example where $\nu = \left(\dfrac{n}{n+1}\right)^n$. From [46],

$$\lim\limits_{n\to\infty}\nu = \lim\limits_{n\to\infty}\left(\frac{n}{n+1}\right)^n = \lim\limits_{n\to\infty}\left(1 - \frac{1}{n+1}\right)^n = e^{-1}. \tag{C.17}$$

From (C.17), $\lim\limits_{n\to\infty}\nu = e^{-1} \neq 1$ and Requirement A.1 is violated.

Q.E.D.

Theorem A.4. Because the ν value derived from criterion (C.9) is affected by the value of n (see (C.4)), the ν^n value is not assured to converge to 0 as $n \to \infty$, which violates Requirement A.2.

Proof: If $\nu = \dfrac{n}{n+1}$, then

$$\lim_{n \to \infty} \nu^n = \lim_{n \to \infty} \left(\frac{n}{n+1}\right)^n = \lim_{n \to \infty} \left(1 - \frac{1}{n+1}\right)^n = e^{-1}. \qquad (C.18)$$

From (C.18), $\lim_{n \to \infty} \nu^n = e^{-1} \neq 0$ and Requirement A.2 is violated.

Q.E.D.

From Theorems A.2-A.4, criterion (C.9) may not yield appropriate ν that guarantees $\lim_{n \to \infty} \max_{1 \le i \le n} \gamma_i = 0$. Therefore, SFGM TAM may not produce accurate result (see Theorem A.1). A "tail behavior" problem caused by the lack of "$\lim_{n \to \infty} \max_{1 \le i \le n} \gamma_i = 0$" is specifically pointed out as follows. Shortle *et al.* [44] claimed that SFGM TAM with approximation formula (C.4) captures the tail behavior of heavy-tailed distributions, since $F(x_i) = 1 - \nu^i$ (see (C.7)) geometrically approaches 1 when i becomes large. However, since criterion(C.9) does not guarantee that the derived ν satisfies Requirement A.2, $F(x_i) = 1 - \nu^i$ may not approach 1 when i becomes large. Therefore, SFGM TAM may not capture the tail behavior of heavy-tailed distributions. Even though $F(x_i) = 1 - \nu^i$ happens to approach 1 geometrically when i becomes large, the same goal (to capture the tail behavior) can also be achieved if we choose x_i such that x_i approaches infinity at a suitable rate (which implies that $F(x_i)$ approaches 1 at a suitable rate). Note that we say "at a suitable rate" in the sense that $F(x_i)$ need not "geometrically" approach 1. For example, in HMBF TAM, the partition intervals are equally sliced with length $\dfrac{1}{n}$ (see (C.2)), and we say that $F(x_i)$ approaches 1 "at a suitable rate." Another example (to be elaborated later) is to let $x_i = F^{-1}\left(\dfrac{i}{n} + a\right)$ for $i = 0, 1, \cdots, n-1$ and $0 < a < \dfrac{1}{n}$. In this example, the partition intervals are also equally sliced with length $\dfrac{1}{n}$, and we say that $F(x_i)$ approaches 1 "at a suitable rate." Note that ν does not exist in the x_i expression in this example.

In addition, SFGM TAM implicitly made an inappropriate assumption on the slope of the e^{-sx} curve against $F(x)$. In general, when using the Riemann sum $\hat{f}^*(s)$ to approximate the integral $f^*(s)$, the integral area should be sliced into smaller slots when the slope is steepened in order to capture the integral area more accurately. On the other hand, the area can be sliced into larger slots without introducing significant inaccuracy when the slope is less steepened. From (C.7) and (C.8) (ignoring the boundary exceptions in (C.8)), we observe that when i is large, γ_i becomes small while $F(x_i) = 1 - \nu^i$ becomes large (see the points $F(x_i)$ and the intervals γ_i in Figure C.2), which implies that SFGM TAM assumed that the slope of the e^{-sx} curve against $F(x)$ becomes more and more steepened when $F(x_i)$ becomes large (see Figure C.2(a)).However, the e^{-sx} curve against $F(x)$ may behave differently. The curve in Figure C.2(b) gives an example where SFGM TAM may result in inaccurate approximation of $f^*(s)$ due to the inappropriate assumption on the slope.

To summarize, two requirements for accurate approximation of $f^*(s)$ are listed below:

Requirement A.A. $\lim\limits_{n \to \infty} \max\limits_{1 \leq i \leq n} \gamma_i = 0.$

Requirement A.B. The integral area must be sliced appropriately according to the slope of the e^{-sx} curve against $F(x)$.

Now we elaborate the proposed modified TAM techniques in **M1**, **M2**, **M3** and **M4** as follows. Note that all of them use the same approximation formula (C.4) with alternative x_i expressions.

M1. We propose that

$$x_i = F^{-1}(\nu^i), \text{ for } i = 1, 2, \cdots, n, \text{ and } 0 < \nu < 1, \qquad (\text{C.19})$$

where we use criterion (C.9) to derive ν (called the "Modified TAM 1"). Note that from (C.19), we have

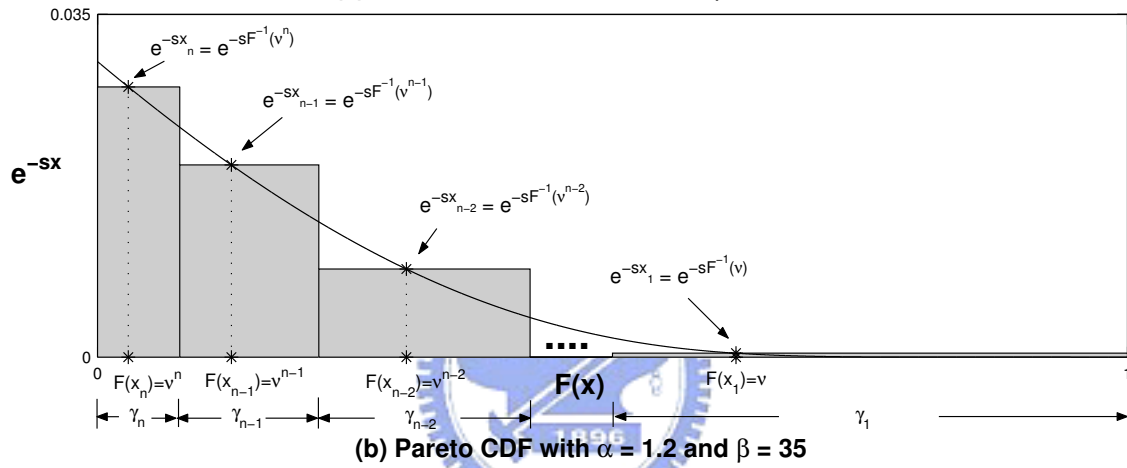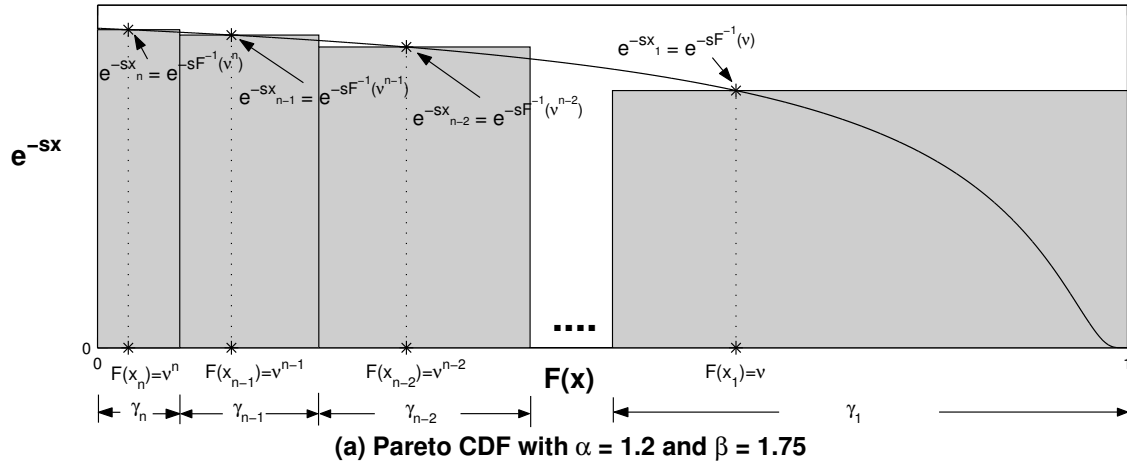$$F(x_i) = \nu^i, \text{ for } i = 1, 2, \cdots, n, \text{ and } 0 < \nu < 1. \qquad (\text{C.20})$$

Figure C.3: Integral Area Slicing for Modified TAM 1 $(s = 0.1)$

In this modification, the γ_i is expressed as follows.

$$\gamma_i = \begin{cases} 1 - \dfrac{F(x_2) + F(x_1)}{2}, & i = 1, \\ \dfrac{F(x_{i-1}) - F(x_{i+1})}{2}, & i = 2, 3, \cdots, n-1, \\ \dfrac{F(x_n) + F(x_{n-1})}{2}, & i = n. \end{cases} \tag{C.21}$$

A demonstration of integral area slicing for this modification (based on (C.21)) is shown in Figure C.3 (note that Figures C.3(a) and C.3(b) demonstrate different slope tendencies of the e^{-sx} curves respectively). By substituting

(C.20) into (C.21), we have

$$
\gamma_i = \begin{cases}
1 - \dfrac{\nu + \nu^2}{2}, & i = 1, \\[2ex]
\dfrac{\nu^{i-1}(1 - \nu^2)}{2}, & i = 2, 3, \cdots, n - 1, \\[2ex]
\dfrac{\nu^{n-1} + \nu^n}{2}, & i = n.
\end{cases}
\tag{C.22}
$$

It is obvious that (C.22) has exactly the same form of (C.8). Therefore, Theorem A.2 is applicable for this modification. We observe that this modification has the same problem as SFGM TAM, and Requirements A.1 and A.2 may not be satisfied.

- Criterion (C.9) does not guarantee that $F(x_1) = \nu \to 1$ as $n \to \infty$ (Requirement A.1 may not be satisfied).

- Neither does criterion (C.9) guarantee that $F(x_n) = \nu^n \to 0$ as $n \to \infty$ (Requirement A.2 may not be satisfied).

From Theorem A.2, we conclude that Requirement A.A may not be satisfied by this modification. However, we can prove that this modification satisfies Requirement A.B for the curves behaving as shown in Figure C.3(b) in the following lemma.

Lemma A.1. Modified TAM 1 satisfies Requirement A.B for the e^{-sx} curves behaving as shown in Figure C.3(b).

Proof: From (C.20) and (C.22) (ignoring the boundary exceptions in (C.22)), we observe that when i is large, γ_i becomes small while $F(x_i) = \nu^i$ becomes small too (see the points $F(x_i)$ and the intervals γ_i in Figure C.3). This characteristic assures that for the e^{-sx} curves behaving as shown in Figure C.3(b), the integral area can be sliced appropriately according to the slope of the e^{-sx} curve against $F(x)$. In other words, Modified TAM 1 satisfies Requirement A.B for the e^{-sx} curves behaving as shown in this figure.

Q.E.D.

M2. We propose that

$$x_i = F^{-1}\left(\frac{i}{n} + a\right), \text{ for } i = 0, 1, \cdots, n-1, \text{ and } 0 < a < \frac{1}{n}. \qquad (C.23)$$

In (C.23), we use criterion (C.9) to derive the a value (called the "Modified TAM 2"). In this modification, $\hat{f}^*(s)$ is slightly modified as follows.

$$\hat{f}^*(s) = \sum_{i=0}^{n-1} \gamma_i e^{-sx_i}, \qquad (C.24)$$

where $\gamma_i = \dfrac{1}{n}$ for $i = 0, 1, \cdots, n-1$. A demonstration of integral area slicing for this modification (based on (C.24)) is shown in Figure C.4 (note that Figures C.4(a) and C.4(b) demonstrate different slope tendencies of the e^{-sx} curves respectively). In the following lemma, we prove that this modification satisfies Requirement A.A.

Lemma A.2. Modified TAM 2 satisfies Requirement A.A.

Proof: Since $\gamma_i = \dfrac{1}{n}$ for $i = 0, 1, \cdots, n-1$, we have $\lim\limits_{n \to \infty} \max\limits_{0 \le i \le n-1} \gamma_i = \lim\limits_{n \to \infty} \dfrac{1}{n} = 0$, and Requirement A.A is satisfied.

Q.E.D.

Note that although Requirement A.B is not satisfied by this modification, the area can still be captured well as $n \to \infty$. Moreover, although the form of (C.24) with $\gamma_i = 1/n$ is similar to that of HMBF TAM (see (C.2)), the a value in (C.23) is supposed to make this modification outperform HMBF TAM, because the a value can be adjusted according to different distributions or different distribution parameters based on criterion (C.9).
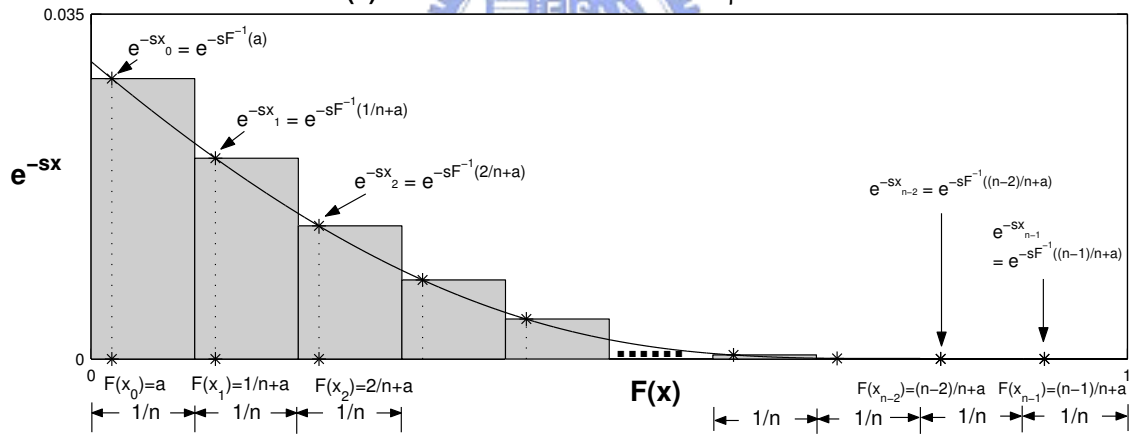
M3. We propose that

$$x_i = F^{-1}(1 - \nu^i), \text{ for } i = 1, 2, \cdots, n \text{ and } 0 < \nu < 1. \qquad (C.25)$$

In (C.25), we set

$$\nu = \left(\frac{n}{n+1}\right)^{n^b}, \text{ where } 0 < b < 1 \qquad (C.26)$$
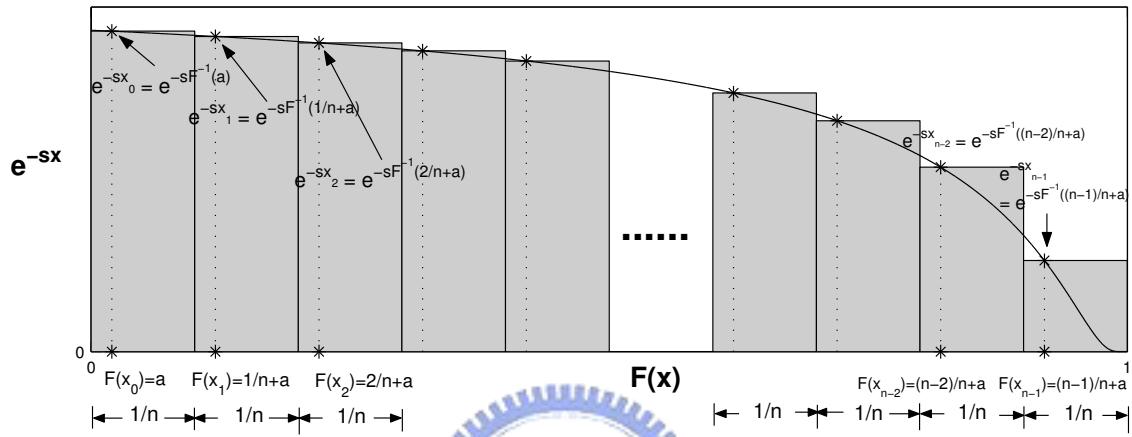
Figure C.4: Integral Area Slicing for Modified TAM 2 ($s = 0.1$)

(called the "Modified TAM 3"). In this modification, γ_i's are the same as (C.8) because this modification and SFGM TAM have the same x_i expression (see (C.5) and (C.25); note that how ν is derived makes no difference). Therefore, Theorem A.2 is applicable for this modification. Also, the demonstration of integral area slicing for this modification is similar to Figure C.2. We observe that this modification has the same problem as SFGM TAM: the area under the steepened part of e^{-sx} may not be captured well when encountering the e^{-sx} curves as shown in Figure C.2(b); that is, Requirement A.B may not be satisfied in the situation such as shown in this figure. Despite of this remaining problem, we show that this modification satisfies Requirement A.A in the following lemma.

Lemma A.3. Modified TAM 3 satisfies Requirement A.A.

Proof:

(A) We first prove the hypothesis that Modified TAM 3 satisfies Requirement A.1 (i.e., $\lim\limits_{n \to \infty} \nu = 1$). From (C.26), as $n \to \infty$, we have

$$\lim_{n \to \infty} \nu = \lim_{n \to \infty} \left(\frac{n}{n+1} \right)^{n^b} = \lim_{n \to \infty} \left[\left(\frac{n}{n+1} \right)^n \right]^{n^{b-1}} = \lim_{n \to \infty} \left[\left(1 - \frac{1}{n+1} \right)^n \right]^{n^{b-1}} \quad \text{(C.27)}$$

By substituting (C.17) into (C.27), (C.27) can be re-written as

$$\lim_{n \to \infty} \nu = \lim_{n \to \infty} e^{-n^{b-1}} = 1, \quad \text{(C.28)}$$

and the hypothesis holds.

(B) Then we prove the hypothesis that Modified TAM 3 satisfies Requirement A.2 (i.e., $\lim\limits_{n \to \infty} \nu^n = 0$). From (C.26), as $n \to \infty$, we have

$$\lim_{n \to \infty} \nu^n = \lim_{n \to \infty} \left[\left(\frac{n}{n+1} \right)^{n^b} \right]^n = \lim_{n \to \infty} \left[\left(\frac{n}{n+1} \right)^n \right]^{n^b} = \lim_{n \to \infty} \left[\left(1 - \frac{1}{n+1} \right)^n \right]^{n^b} \quad \text{(C.29)}$$

By substituting (C.17) into (C.29), (C.29) can be expressed as

$$\lim_{n \to \infty} \nu^n = \lim_{n \to \infty} e^{-n^b} = 0, \quad \text{(C.30)}$$

and the hypothesis holds.

From Theorem A.2, since both Requirements A.1 and A.2 are satisfied, Requirement A.A is satisfied as well.

Q.E.D.

It should be noted that the b value is not derived from criterion (C.9), but merely randomly chosen. Otherwise, the b value is affected by the value of n (see (C.4)), which could invalidate the proof of Lemma A.3 (see (C.28) and (C.30)). For example, when $b = \dfrac{\ln \ln 2}{\ln n} + 1$ (b is affected by n), (C.28) can be expressed as

$$\lim_{n \to \infty} \nu = \lim_{n \to \infty} e^{-n^{b-1}} = \lim_{n \to \infty} e^{-n^{\frac{\ln \ln 2}{\ln n}}} \tag{C.31}$$

$$= \lim_{n \to \infty} e^{-\left(n^{\frac{1}{\ln n}}\right)^{\ln \ln 2}}. \tag{C.32}$$

From the property of the logarithm, we have $n^{\frac{1}{\ln n}} = e$, and (C.32) can be re-written as

$$\lim_{n \to \infty} \nu = \lim_{n \to \infty} e^{-e^{\ln \ln 2}} = \lim_{n \to \infty} \left(e^{e^{\ln \ln 2}}\right)^{-1} = \lim_{n \to \infty} 2^{-1} = \lim_{n \to \infty} \frac{1}{2} = \frac{1}{2} \neq 1. \tag{C.33}$$

Another example is when $b = \dfrac{\ln \ln 2}{\ln n}$ (b is affected by n), (C.30) can be expressed as

$$\lim_{n \to \infty} \nu^n = \lim_{n \to \infty} e^{-n^b} = \lim_{n \to \infty} e^{-n^{\frac{\ln \ln 2}{\ln n}}}. \tag{C.34}$$

It is clear that (C.34) is the same as (C.31). Therefore, (C.34) can be derived to the same result as (C.33), and we have $\lim_{n \to \infty} \nu^n = \dfrac{1}{2} \neq 0$. As a consequence, Requirement A.A may not be satisfied if the b value is derived from criterion (C.9). Therefore, the b value is randomly chosen instead of being derived from criterion (C.9) (in this paper, we use "$b = 0.5$" as an example).

M4. We propose that

$$x_i = F^{-1}(\nu^i), \text{ for } i = 1, 2, \cdots, n \text{ and } 0 < \nu < 1, \tag{C.35}$$

where the ν expression is the same as (C.26) (called the "Modified TAM 4"). In this modification, the γ_i expression is the same as (C.22) because this

modification and Modified TAM 1 have the same x_i expression (see (C.19) and (C.35); note that how ν is derived makes no difference). Therefore, Theorem A.2 is applicable for this modification. Also, the demonstration of integral area slicing for this modification is similar to Figure C.3. In the following lemma, we prove that this modification satisfies Requirement A.A. This modification also satisfies Requirement A.B if the e^{-sx} curve against $F(x)$ behaves as shown in Figure C.3(b).

Lemma A.4. Modified TAM 4 satisfies Requirement A.A. If the e^{-sx} curve against $F(x)$ behaves as shown in Figure C.3(b), Modified TAM 4 also satisfies Requirement A.B.

Proof:

(A) As we have shown in Lemma A.3, since Modified TAM 4 has the same ν expression as that of Modified TAM 3 (see (C.26)), it satisfies both Requirements A.1 and A.2. Therefore, from Theorem A.2, Requirement A.A is also satisfied by this modification.

(B) As we have shown in Lemma A.1, since $F(x_i) = \nu^i$ (see (C.35)), Requirement A.B is satisfied if the e^{-sx} curve against $F(x)$ behaves as shown in Figure C.3(b).

Q.E.D.

Note that with the same reason as stated in **M3**, the b value is randomly chosen instead of being derived from criterion (C.9) (in this paper, we use "$b = 0.5$" as an example).

Bibliography

[1] 3GPP. 3rd Generation Partnership Project; Technical Specification Group Services and Systems Aspects; General Packet Radio Service (GPRS); Service Descripton; Stage 2. Technical Specification 3G TS 23.060 version 3.6.0 (2001-01), 2000.

[2] 3GPP. 3rd Generation Partnership Project; Technical Specification Group Core Network; Location Management Procedures (Release 4). 3GPP TS 23.012 V4.0.0, 2001.

[3] 3GPP. 3rd Generation Partnership Project; Technical Specification Group Core Network; Mobile radio interface layer 3 specification; Core Network Protocols - Stage 3 (Release 5). 3GPP TS 24.008 V5.0.0, 2001.

[4] Akyildiz, I.F., and Wang, W. A Dynamic Location Management Scheme for Next-Generation Multitier PCS Systems. *IEEE Trans. on Wireless Communications*, 1(1):178–189, 2002.

[5] Akyildiz, I.F., McNair, J., Ho, J.S.M., Uzunalioglu, H., and Wang W. Mobility Management in Next Generation Wireless Systems. *IEEE Proceedings*, 87(8):1347–1385, 1999.

[6] Bolotin, A.V. Modeling Call Holding Time Distributions for CCS Network Design And Performance Analysis. *IEEE JSAC*, 12(3):433–438, 1994.

[7] Chlamtac, I., Fang, Y., and Zeng, H. Call Blocking Analysis for PCS Networks under General Cell Residence Time. *IEEE WCNC, New Orleans*, September 1999.

[8] Chlamtac, I., Liu, T., and Carruthers, J. Location Management for Efficient Bandwidth Allocation and Call Admission Control. *IEEE WCNC, New Orleans*, September 1999.

[9] Chlebus, E., and Ludwin, W. Is Handoff Traffic Really Poissonian? *Proc. IEEE ICUPC*, pages 348–353, 1995.

[10] Fang, Y., and Chlamtac, I. Teletraffic Analysis and Mobility Modeling for PCS Networks. *IEEE Transactions on Communications*, 47(7):1062–1072, July 1999.

[11] Fang, Y., and Chlamtac, I. Analytical Generalized Results for Handoff Probability in Wireless Networks. *IEEE Trans. on Communications*, 50(3):396–399, 2002.

[12] Fang, Y., Chlamtac, I., and Fei, H. Analytical Results for Optimal Choice of Location Update Interval for Mobility Database Failure Restoration in PCS networks. *IEEE Transactions on Parallel and Distributed Systems*, 2000.

[13] Fantacci, R. Performance Evaluation of Prioritized Handoff Schemes in Mobile Cellular Networks. *IEEE Trans. on Vehicular Technology*, 49(2):485–493, March 2000.

[14] Gallager, R. G. *Discrete Stochastic Processes*. Kluwer Academic Publishers, 1996.

[15] Giancristofaro, D., Ruggieri, M., And Santucci, F. Analysis of queue-Based handover procedures for mobile communications. *Proc. IEEE ICUPC*, pages 168–172, 1993.

[16] Harris, C.M., and Marchal, W.G. Distribution Estimation using Laplace Transforms. *INFORMS Journal on Computing*, 10(4):448–458, 1998.

[17] Harris, C.M., Brill, P.H., and Fischer, M.J. Internet-Type Queues with Power-Tailed Interarrival Times and Computational Methods for Their Analysis. *INFORMS Journal on Computing*, 12(4):261–271, 2000.

[18] Hong, D. And Rappaport, S.S. Traffic Model And Performance Analysis for Cellular Mobile Radio Telephone Systems with Prioritized And Non-protection Handoff Procedure. *IEEE Trans. Veh. Techol.*, VT-35(3):77–92, August 1986.

[19] Hong, D., and Rappaport, S.S. Priority Oriented Channel Access for Cellular Systems Serving Vehicular and Portable Radio Telephones. *IEE PROCEEDINGS-I*, 136(5):339–346, 1989.

[20] Hung, H.-N., Lee, P.-C., Lin, Y.-B., and Peng, N.-F. Modeling Channel Assignment of Small-Scale Cellular Networks. *IEEE Trans. on Wireless Communications*, 4(2):646–652, 2005.

[21] Hung, H.-N., Lin, Y.-B., Peng, N.-F., and Yang, S.-R. Resolving Mobile Database Overflow with Most Idle Replacement. *IEEE Journal on Selected Areas in Communications*, 19(10):1953–1961, 2001.

[22] Jedrzycki, C., and Leung, V.C.M. Probability Distribution of Channel Holding Time in Cellular Telephony Systems. *IEEE VTC*, 1996.

[23] Kleinrock, L. *Queueing Systems: Volume I – Theory*. New York: Wiley, 1976.

[24] Lin, P. and Lin, Y.-B. Channel Allocation for GPRS. *IEEE Trans. on Vehicular Technology*, 50(2), 2000.

[25] Lin, P., Lin, Y.-B. and Jeng, J.-Y. Improving GSM Call Completion by Call Re-Establishment. *IEEE Journal on Selected Areas in Communications*, 1999.

[26] Lin, Y.-B. Failure Restoration of Mobility Databases for Personal Communication Networks. *ACM-Baltzer Journal of Wireless Networks*, 1:365–372, 1995.

[27] Lin, Y.-B. Impact of PCS Handoff Response Time. *IEEE Communications Letters*, 1(6):160–162, 1997.

94

[28] Lin, Y.-B. Performance Modeling for Mobile Telephone Networks. *IEEE Network Magazine*, 11(6):63–68, November/December 1997.

[29] Lin, Y.-B. And Chen, W. Impact of Busy Lines And Mobility on Call Blocking in A PCS Network. *Intl. J. of Commun. Systems*, 9:35–45, 1996.

[30] Lin, Y.-B., and Chlamtac, I. *Wireless and Mobile Network Architectures*. John Wiley & Sons, 2001.

[31] Lin, Y.-B., And Mak, V.W. Eliminating The Boundary Effect of A Large-scale Personal Communication Service Network Simulation. *ACM Tran. on Modeling And Computer Simulation*, 4(2), 1994.

[32] Lin, Y.-B., Lai, W.R., and Chen, R.J. Performance Analysis for Dual Band PCS Networks. *IEEE Transactions on Computers*, 49(2):148–159, 2000.

[33] Lin, Y.-B., Lee, P.-C., and Chlamtac, I. Dynamic Periodic Location Area Update in Mobile Networks. *IEEE Trans. on Vehicular Technology*, 51(6):1494–1501, 2002.

[34] Lin, Y.-B., Mohan, S., And Noerpel, A. Queueing Priority Channel Assignment Strategies for PCS Hand-Off And Initial Access. *IEEE Trans. Veh. Technol.*, 43(3):704–712, 1994.

[35] Meyer, P.L. *Introductory Probability and Statistical Applications*. Addison-Wesley, 1968.

[36] Orlik, P. V. and Rappaport, S. S. On the Handoff Arrival Process in Cellular Communications. *ACM/Baltzer Wireless Networks*, 7:147–157, March/April 2001.

[37] Orlik, P.V., and Rappaport, S.S. Traffic Performance and Mobility Modeling of Cellular Communications with Mixed Platforms and Highly Variable Mobilities. *Proceedings of the IEEE*, 86(7):1464 –1479, 1998.

[38] Pang, A.-C., Lin, P., and Lin, Y.-B. Modeling Mis-Routing Calls Due to User Mobility in Wireless VoIP. *IEEE Communications Letters*, 4(12):394–397, 2000.

[39] Peha, J. M. and Sutivong, A. Admission Control Algorithms for Cellular Systems. *ACM/Baltzer Wireless Networks*, 7:117–125, March/April 2001.

[40] Rappaport, S.S. Blocking, hand-off And traffic performance for Cellular communication Systems with mixed platforms. *IEE PROCEEDINGS-I*, 140(5):389–401, 1993.

[41] Ross, S. *A First Course in Probability*. Prentice-Hall International, Inc., 1998.

[42] Ross, S.M. *Introduction to Probability Models*. Academic Press, Inc., 1985.

[43] Ross, S.M. *Stochastic Processes*. John Wiley & Sons, 1996.

[44] Shortle, J.F., Fischer, M.J., Gross, D., and Masi, D.M.B. Using the Transform Approximation Method to Analyze Queues with Heavy-Tailed Service. *Journal of Probability and Statistical Science*, 1(1):17–30, 2003.

[45] Tekinary, S. And Jabbari, B. A Measurement Based Prioritization Scheme for Handovers in Cellular and Microcellular Networks. *IEEE J. Select. Areas Commun.*, pages 1343–1350, October 1992.

[46] Thomas, G.B., Jr., and Finney, R.L. *Calculus and Analytic Geometry*. Addison-Wesley Publishing Company, Inc., 1998.

[47] Widax. http://www.widax.com.tw/.

[48] Xhafa, A. and Tonguz, O. K. Dynamic Priority Queueing of Handoff Requests in PCS. *IEEE International Conference on Communications*, 2:341–345, 2001.

[49] Zeng, H., and Chlamtac, I. Handoff Traffic Distribution in Cellular Networks. *IEEE WCNC, New Orleans*, September 1999.

Curriculum Vita

Lee, Pei-Chun was born in Taoyuan, Taiwan, R.O.C., in 1976. She received her B.S. and M.S. degrees in Computer Science & Information Engineering (CSIE) from National Chiao Tung University (NCTU) in 1998 and 2000, respectively. She is currently a Ph.D. candidate in Computer Science (CS), NCTU. Her research interest includes Performance Modeling.

Publication List

- **Journal Publication**

  1. Lin, Y.-B., Lee, P.-C. and Chlamtac, I., Dynamic Periodic Location Area Update in Mobile Networks. IEEE Transactions on Vehicular Technology, 51(6): 1494-1501, November 2002.

  2. Hung, H.-N., Lee, P.-C., Lin, Y.-B., and Peng, N.-F., Modeling Channel Assignment of Small-Scale Cellular Networks. IEEE Transactions on Wireless Communications, 4(2): 646-652, March 2005.

  3. Hung, H.-N., Lee, P.-C., and Lin, Y.-B., Random Number Generation for Excess Life of Mobile User Residence Time. Accepted and to appear in IEEE Transactions on Vehicular Technology.

- **Conference Papers**

  1. Hung, H.-N., Lee, P.-C., Lin, Y.-B., and Peng, N.-F., Modeling Small-Scale Cellular Networks. Proceedings of the 2003 International Symposium on Communications, December 2003.

  2. Hung, H.-N., Lee, P.-C., and Lin, Y.-B., Random Number Generation for Residual Life of Mobile Phone Movement. Proceedings of the 2004 IEEE International Conference on Networking, Sensing and Control, March 2004.