

A Generic Framework for Mobility Modeling and Performance Analysis in Next-generation Heterogeneous Wireless Networks

Ahmed H. Zahran and Ben Liang

Department of Electrical and Computer Engineering

University of Toronto

Email: [zahran,liang]@comm.utoronto.ca

Abstract

The integration of different wireless radio cellular technologies is emerging as an effective approach to accommodate the increasing demand of next-generation multi-media based applications. In such systems, user roaming among different technologies, commonly known as vertical handoff, will significantly affect different aspects of network design and planning due to the characteristically wide-ranging diversity in access technologies and supported applications. Hence, the development of new mobility models that accurately depict vertical mobility is crucial for studying different design problems in these heterogeneous systems. This article presents a generic framework for mobility modeling and performance analysis of integrated heterogeneous networks using phase-type distributions. This framework realizes all modeling requirements in next-generation user mobility including accuracy, analytical tractability, and accommodating the correlation between different residence times within different access technologies. Additionally, we present general guidelines to evaluate application performance based on the new mobility models introduced in this article. We show the accuracy of our modeling approach through simulation and analysis given different applications.

I. INTRODUCTION

The Service convergence of different radio access technologies (RAT) is foreseen as an effective approach to accommodate the demand increase in next-generation heterogeneous wireless networks (NGHWN). These technologies have great characteristic diversity in terms of coverage, resources, and cost. Hence, their service convergence brings several advantages including achieving better system resource utilization and improving the user perceived quality-of-service (QoS). On the other hand, such convergence mandates the development of new operational mechanisms and system designs to accommodate the vertical handoff (VHO) impact on different network performance metrics such as service cost, signaling, QoS, and resource utilization. In particular, the development of accurate mobility and traffic models has become a crucial requirement for the assessment of these operational mechanisms and system designs.

Generally, mobility models can be categorized into analytical and simulation models. *Analytical models* usually limit user mobility to a specific region of residence within the network. In contrast, *simulation models* provide a larger amount of details about the user's location by periodically tracking the user in small time steps. Random trip [1] and Gauss-Markov [2] models are examples of common generic mobility models in wireless cellular system simulations. Clearly, the simulation models presented in the literature can be also used for NGHWN simulation after introducing heterogeneous network overlays into the simulation setup. However, such models are generally intractable when the network coverage topology is under consideration. Therefore, the goal of this work is to develop new analytical models to accommodate the vertical roaming impact on the resource utilization in heterogeneous overlay cells.

In homogeneous wireless systems, e.g., traditional cellular networks, the user mobility is usually represented by its cell residence time (CRT), defined as the duration spent by the user within a cell of a specific technology. This level of details completely defines the resources utilized by the user and is sufficient to derive different performance metrics and important conclusions regarding critical network dimensioning problems. In the literature, several distributions are used to model the CRT to evaluate the performance of these networks. The key idea is to propose a good mobility model that is flexible enough

to fit field data and at the same time analytically tractable to accurately estimate the required performance metrics. Hence, several works [3], [4] adopt phase-type (PH) distributions [5] as a versatile dense class of distributions that can approximate any non-negative statistics.

In NGHWN, the integration of overlapping technologies such as wide area RATs, e.g. third-generation (3G) cellular networks, metropolitan area network, e.g. WiMax, and local area RATs, e.g. IEEE 802.11, will create zones with different technological coverage combinations. Noting that short-range technologies (e.g., IEEE 802.11) usually provide more resources and charge lower prices, users are expected to vertically roam between different technologies as they move from one zone to another. Hence, the representation of these zones is a crucial requirement in NGHWN mobility models for proper performance evaluation. More importantly, the overlapping of different technological zones and 3G cells in NGHWN originates a new type of correlation between the zone residence times and the CRT. Hence, accommodating this correlation evolves as a crucial requirement in NGHWN mobility models. Note that ignoring this correlation may lead to significant errors in performance evaluation as shown in [6].

In Section II, we present a generic mobility modeling framework that realizes all the aforementioned requirements. This framework employ PH modeling using two different approaches: a *fictitious* approach to model the zone residence times and a *physical* approach to represent the overlay CRT as a stochastic sum of these zone residence times using different PH structures to accommodate the aforementioned correlation. Section III presents an analytical framework that considers the technological nature of different phases of the developed CRT models and provides guidelines for estimating several performance metrics including network utilization times, session cost, signaling load, and user perceived QoS assuming PH session holding times. Most of these metrics are estimated using Markov reward models [7] and PH distribution properties. The numerical results presented in Section IV illustrate the accuracy of the presented mobility modeling and performance evaluation frameworks.

II. MOBILITY MODELING IN NGHWN

In this section, we present a new mobility modeling approach for NGHWN. The models developed using this approach employ mobility traces that record the durations spent by the MT in different technological zones within the borders of 3G cells. Additionally, they record the type of the first visited technological zone in each cell. Using these traces, different technological zones in the system are determined and their residence times are modeled as PH distributions by direct fitting to their corresponding measurements in the mobility traces. The CRT structure is then defined and its parameters are calculated to accommodate the aforementioned correlation. This modeling approach is applied in our extended-Coxian model (ECM) for two-tier systems and our novel n-tier generic zone residence time model (ZRTM).

A. PH Distributions

PH distributions are highly versatile stochastic models that can be used to approximate the distribution of any non-negative random variable (RV). Each distribution defines the absorption time of an underlying Markov chain to a single absorbing state. A PH RV is denoted as $PH(\alpha, \mathbf{T})$, where \mathbf{T} corresponds to the Markov chain transient dynamics and α represents the initial probabilities of the transient states. It is worth mentioning that all the statistical characteristics of a PH distribution can be expressed in terms of α and T . For example, Figure 1 shows the structure and representation of a two-phase hypo-exponential (also termed generalized Erlang) distribution and a two-phase hyper-exponential distribution.

PH distributions enjoy an interesting closure property by which several operations on PH distributions produce another PH distribution. This property is algorithmically useful as several operations that may require integration are replaced by simple matrix operations. The sum of two RVs is a typical example to show the power of this property. The distribution of the resultant sum is usually associated with tedious integrals. However, if the original distributions are PH, the sum will be another PH RV with easily computable parameters. This addition operation is graphically represented in Figure 1. Intuitively, when two PH RVs are added together, the first is absorbed into the second and the resultant distribution represents the absorption time of the combined process. The initial probabilities of the resultant PH RV is given completely by the initial distribution of the first RV since the process starts in one of its Markov states, while the transient dynamics of the resultant PH RV depends on the dynamics of both original RVs. This addition procedure plays a key role in modeling CRT distribution from its corresponding zone residence times.

B. Extended-Coxian Model (ECM)

ECM [6] is mainly proposed for two-tier systems and inherits its structure from the Coxian PH RV [8] shown in Figure 2. The Coxian structure is originally proposed for queuing systems in which customers probabilistically visit different numbers of servers. After visiting any server i , the customer may exit the system with probability b_i or visit the following server with probability a_i except for the last server at which the customer exits with probability 1. In ECM, the overlay network cell is analogous to the queuing system, where the duration spent in different zones are analogous to the time spent at different servers and the cell exit probabilities are analogous to the system exit probabilities. Consequently, the total duration spent by the user inside the cell is modeled as a probabilistic sum of the durations spent in different zones.

The term *extended* is used since the duration spent in each stage is modeled as a generic PH distribution as opposed to the exponential stages in the original Coxian distribution¹. This extension provides a flexible means for fitting the residence times in each stage and produces better estimates for time-based performance metrics. In Figure 2, the dynamics of each zone z is modeled as a PH distribution, denoted as $PH(\alpha_z, \mathbf{T}_z)$, whose parameters are fitted to the experimentally collected residence times of the corresponding unique and double coverage zones. The cell exit probabilities b_i , where $i = 1, 2, \dots, k - 1$, are calculated as the ratio between the number of cells with exactly i technology visits to the number of cells with i or more technology visits. Noting that the inter-stage transition probabilities, a_i , complement the probability space of b_i , i.e. $a_i + b_i = 1$, the extended-Coxian model parameters are completely defined.

ECM has been shown to provide a great accuracy in estimating different performance metrics of a two-tier NGHWN [6]. However, its complex structure limits its applicability to multi-tier systems. These challenges are alleviated by ZRTM, which can be used to model any n-tier system using a simpler structure.

C. Zone Residence Time Model (ZRTM)

As new technologies become pervasive, the integrated network will extend to include more technologies. Hence, developing new mobility models for general n-tier systems will evolve as a crucial requirement for their design. In any n-tier system, the wide area network may overlay one or more technologies creating zones with different technological coverage combinations. For any n-tier system, there may be up to 2^{n-1} different coverage combinations assuming that the 3G network has universal coverage. For example, assuming a three-tier system integrating a wide area network technology, a metropolitan area technology, and a local area technology, the user may traverse four types of zones with different technological coverage combinations including

- zones uniquely covered by a wide area network,
- zones covered by both wide and metropolitan area technologies,
- zones covered by both wide and local area technologies, or
- zones covered by three technologies.

These zones represent the basic building blocks of the CRT structure. Each zone is modeled as a PH distribution using the collected measurements from mobility traces.

The second step is to choose an appropriate structure to model the CRT as a probabilistic sum of the residence times of these zones. ZRTM uses one PH stage per zone and employs one cell exit probability for each stage. Hence, at the end of any ZRT, the user may exit the cell with probability b_i or remain in the cell with probability a_i . In the latter case, the user moves from zone i to zone j , where $i \neq j$, with a zone transition probability of p_{ij} . The values of zone transition probabilities p_{ij} are determined, from the experimentally collected traces, as the ratio between the number of transitions from zone i to zone j to the total number of transitions from zone i to all other zones. The correlation between the CRT and ZRTs is accommodated by estimating the exit probabilities b_i such that the average duration spent in zone i equals the corresponding average value obtained from the collected traces. Hence, the total duration spent in different stages (zones) will match the CRT.

Figure 3 shows a two-tier ZRTM in which one stage represents the unique coverage zone and the second stage represents the double coverage zone. Clearly, the ZRTM structure is simpler compared to the ECM structure. Hence, it involves simpler

¹Note that the exponential distribution is a special case of the PH distribution.

queuing models too. Another advantage is its fixed CRT model size, which is independent of user mobility pattern. This is not the case for ECM, whose size is proportional to mobility pattern randomness. Furthermore, it will be shown later that both ZRTM and ECM have a similar level of accuracy in estimating different performance metrics. However, these advantages of ZRTM are achieved at the cost of non-trivial, though not complex, estimation procedure for the cell exit probabilities. Note that in case of ECM, the exit probabilities are directly estimated from the collected traces.

It is worth noting that the granularity of zone definition is flexible and can be easily restructured according to the problem of interest. For example, we may define only two zones if we are only interested in unique and multi-coverage zones irrespective of multi-coverage details, such as the number of access technologies available in a multi-coverage zone. Additionally, the model can accommodate different levels of trace collection details. For example, assume the collected traces for the underlay network in a two-tier system are downgraded to the base station level rather than the technology level; we can then define two identical sub-zones, each sub-zone representing the duration spent by the user attached to one base station, and model the zone residence time in the same fashion the CRT is modeled using zone residence times. Note that this approach will enable us to capture the horizontal handoff details within the underlay technology at the cost of extended trace collection.

III. NGHWN PERFORMANCE ANALYSIS

In general, the performance modeling and analysis of cellular systems can be conducted at two levels [9]. The first-level modeling uses the number of radio channels available in the cell as an input parameter to determine the blocking and forced termination probabilities of new and handoff calls respectively. The second-level modeling uses the new call-blocking and the forced termination probabilities as input parameters to study the call-completion probability, expected effective call hold times, and expected number of handoffs. The integration of both approaches to accommodate the interaction between user behavior and network behavior can be realized by extending the iterative algorithm proposed by Hong and Rappaport [10] for homogeneous networks. The main extensions are introducing vertical handoff rates and redefining the channel occupation time to consider the VHO possibility within cell borders. Note that this article only deals with the second-level modeling and performance analysis.

The analytical framework of NGHWN described in this article assumes PH session holding time. For simplicity, we will initially assume that applications have exponentially distributed holding time and will relax this condition at the end of this section. First, a cell-based session model is developed for each mobility model by combining the application and mobility details. The guidelines for developing such session models are presented in Subsection III-A and are applicable to both ECM and ZRTM. Second, we present a generic framework for estimating cell-level performance metrics using Markov reward models in Subsection III-B. In Subsection III-C, we present an alternate session model that considers the user activity during the entire session, i.e. not limited to cell residence, to estimate several session-level performance metrics. We would like to point out that the presented framework is directly applicable to ZRTM, but requires some more steps to apply it to ECM, in which two mobility sub-models are required to differentiate between the two possible initial-phase technologies. Interested readers are referred to [6].

A. Session Modeling

The main conceptual difference between any mobility model and its corresponding session model is the consideration of user application activity in addition to user mobility. Without loss of generality, we assume that users have a predefined static network preference that controls the VHO decision in different technological zones, e.g. in a two-tier system, the user always prefers the underlay technology to the 3G network. Active users request resources from the base station of the RAT that they are attached to. If the base station has enough resources, the user is granted the requested resources; otherwise, the user request is denied. In the latter case, the session will be dropped if the base station is the only available point of attachment to the network. Generally, we assume that short-range technologies always have sufficient resources and a user will never be blocked as it performs vertical handoff from an overlay network to an underlay network. On the other hand, the topmost overlay network has limited resources and an active session may be dropped as the user moves from an underlay network to an overlay network with a specific vertical blocking probability P_{vb} . Similarly, the session may be dropped as the user performs horizontal handoff from one cell to another with a specific horizontal blocking probability P_{hb} .

The applications are assumed to have exponentially distributed holding times t_{ih} with parameter λ_{ih} , where i represents a technology index. Note that the values of λ_{ih} depends on the application behavior in different technologies. From this perspective, applications can be categorized as *symmetric* and *asymmetric*. The former preserves its session holding time independent of the available network resources, while the latter may have technology dependent session holding time. Conversational applications, such as voice over IP (VoIP) and video conference (V-conf), are examples of symmetric applications, while streaming applications with buffering capabilities, such as video on demand (VoD) and radio on demand (RoD), are examples of asymmetric applications.

The session model accounts for the interaction between user mobility, application dynamics, and network reaction to user requests by defining four different absorption states, TERM, SHH, HHFT, and VHFT, which respectively correspond to normal session termination, successful handoff to a neighbor overlay cell, session dropping during horizontal handoff, and session dropping during vertical handoff. The absorption rate from any phase to a TERM state equals its session holding time parameter λ_{ih} . The absorption rate from any phase to the remaining absorbing states is determined by its corresponding mobility event and the network reaction to this event. For example, assuming that the cell exit rate from a specific phase is λ_e in the mobility model, then the absorption rates from this phase to SHH and HHFT in the session model respectively equal $(1 - P_{hb})\lambda_e$ and $P_{hb}\lambda_e$ for overlay phases and λ_e and zero for underlay phases. Note that if an active user is exiting the cell while it is using the underlay technology, it does not request any resources from the cellular network. Similar to horizontal handoff, the transitions from any underlay technology phase to any overlay technology phase may get absorbed to VHFT with probability P_{vb} . Figure 4 shows a ZRT session model assuming exponentially distributed zone residence. To this end, it is worth mentioning that the resultant session model has a PH structure provided that the absorbing states are aggregated into one state.

B. Cell-Level Analysis

In this subsection, we present a generic framework to obtain cell-level metrics using Markov reward models (MRM) and PH distribution properties. MRMs are generally extended from Markov chains by associating a specific reward/cost structure to the states and transitions. Rewards associated with states and transitions are called rate rewards and impulse rewards respectively [7]. Rate rewards are accumulated during state residence, while impulse rewards are collected as lump sums on transitions. Generally, the assigned reward can be time dependent; however, stationary rewards are usually sufficient to estimate most performance metrics. Using MRMs, the session models presented in this article can be used to estimate a large variety of transient and steady state performance metrics by varying the reward structure assigned to different states. In this article, we focus on a specific type of Markov chains with absorbing state(s), for which steady state metrics are estimated until absorption. These metrics are calculated by considering the transient part of the infinitesimal Markov chain generator matrix, i.e. the \mathbf{T} matrix of corresponding PH distributions.

In our analytical framework, we use MRMs to estimate time-based performance metrics, such as network utilization and session cost, and transition-based metrics such as horizontal handoff and vertical handoff rates. These metrics can be estimated by assigning different rate reward structures to the session model transient states. A performance metric m is obtained as an average accumulated reward of a specific reward structure and is expressed as

$$m = \sum_i r_i l_i, \quad (1)$$

where l_i represents the duration spent by the chain in state i until absorption, and r_i represents the reward rate associated with state i . In order to estimate time-based metrics, constant reward values are assigned to different states. For example, if binary rewards, i.e. either 1 or 0, are assigned to the transient states such that a specific set of states S , e.g. phases belonging to a specific zone, are assigned 1 and other phases are assigned 0, the estimated metric is the time spent in this zone until absorption. In order to estimate transition-based metrics, phases from which these transitions may be fired are assigned a reward equivalent to the transition rate. For example, to estimate vertical handoff rate from the overlay technology to the underlay technology in a two-tier system, each overlay technology phase is assigned a reward equivalent to the total of transition rates from this phase to all phases of the underlay technology.

Clearly, different reward structures can be defined to estimate different performance metrics using MRM provided that we can estimate the duration spent in every state of the Markov chain until absorption l_i . One way to obtain these values is by using [11, Theorem 2.4.3]. This theorem states that $(-T^{-1})_{rs}$ is the expected total time spent in phase s until absorption given that the initial phase is r . Hence, the expected time spent in different phases until absorption, $\mathbf{L} = [l_i]$, can be expressed as

$$\mathbf{L} = -\alpha \mathbf{T}_s^{-1}, \quad (2)$$

where \mathbf{T}_s is the session model matrix. To this end, the presented framework enables the stochastic analysis of mobility models developed in Section II. In particular, this can be used for handoff sessions as they involve only the complete CRT of a user.

However, new sessions usually arrive after the user has spent some time inside the cell. In this case, cell residence is determined by the user's remaining time within the cell, known as the residual time. This residual time can be shown to be $PH(\beta, \mathbf{T}_m)$ [11, Theorem 3.3.1], where $\beta = (\alpha \mathbf{T}_m^{-1} \mathbf{e})^{-1} \alpha \mathbf{T}_m^{-1}$ in which \mathbf{T}_m equals the transition matrix in the mobility model, i.e. \mathbf{T}_m^{EC} or \mathbf{T}_m^{ZRT} and \mathbf{e} is an all-one column vector. Note that both the residual and residence time have identical user mobility dynamics, determined by the transition matrix, while the main difference between them is the initial distribution. Intuitively, the user behavior inside the cell is independent of the session arrival, and the main stochastic factor in the residual time is the user location inside the cell when the session arrives. This location is determined by the initial distribution β , which intuitively represents the ratio between the expected duration spent in each phase of the mobility model PH distribution, $\alpha \mathbf{T}_m^{-1}$, and the expected total duration spent by the user in the cell, $\alpha \mathbf{T}_m^{-1} \mathbf{e}$.

In general, the estimated performance metrics for both new and handoff sessions will be different. Hence, the overall performance metric can be estimated as a weighted sum of the estimated new session and handoff session metrics, where the weights represent the probabilities of a specific session type. The probability of a session being of either type, termed session probability, can be estimated as a function of the number of visited cells per session. For example, assuming that a session visits n cells during its activity, then the new and handoff session probabilities are $1/n$ and $(n-1)/n$ respectively.

C. Session-Level Analysis

Session-level analysis estimates performance metrics for the entire session. For example, a system designer may be interested in the number of visited cells per session. These metrics are estimated using an extended ZRT session model that combines two identical cell models from the previous subsection. Figure 5 shows an example for a two-tier ZRT session model with exponentially distributed zone residence. In this model, SHH state is replaced by another transient dynamics structure representing a handoff cell. In this model, the session alternates between two identical transient parts until it terminates. The resultant model again has a PH structure provided that the absorption states are aggregated into one state. The initial distribution of the extended session model can be expressed as $[\beta \mathbf{0}]$ since the initial phase of the session will be determined by the residual time in the first transient part.

Using the guidelines presented in the previous subsection, different rewards can be defined to estimate different performance metrics including the VHO and horizontal handoff signaling load, the number of visited cells per session, and the network utilization in different access technologies. Additionally, the session termination probabilities can be estimated as the absorption probabilities to the HHFT, VHFT, and TERM states. These absorption probabilities can be estimated from the embedded Markov chain [12] of the extended session model.

D. PH Session Holding Time

Although the exponential session holding time assumption is common in wireless networks due to their pricing strategies and limited power resources, the introduction of new services to the system such as media streaming invalidate this assumption. Several studies [13] have shown that files transmitted on the Internet feature a large variance of their sizes. Hence, the sessions holding time will have similar statistical characteristics, which is particularly well captured by the hyper-exponential distribution. Technically, the hyper-exponential distribution is a special case of the PH distribution and represents a probabilistic mixture of exponential distributions. Hence, the corresponding *hyper-metric* can be similarly estimated as a probabilistic sum of the metrics estimated for the corresponding exponential phases in the hyper-exponential session holding time distribution.

IV. NUMERICAL RESULTS

We present simulation results of a two-tier system that integrates a 3G cellular network and a wireless local area network (WLAN). We used square cells for simplicity of illustration. Each cell is composed of N square subdivisions, where WLANs are randomly located, so that each subdivision represents a potential WLAN site given the WLAN coverage percentage P_{ow} . When a user is handed off to another cell, it experiences a new random WLAN topology. In order to emulate practical operation, a handoff area of five seconds is assumed between overlay 3G cells. This delay corresponds to the hysteresis introduced in the handoff algorithms to decrease the ping-ponging impact during horizontal handoff. Additionally, the VHO to WLAN is delayed with three seconds as a typical delay required for WLAN discovery and handoff signaling [14]. For mobility simulation, we adopt a two-dimensional Gauss-Markov movement model from [2], as it can be easily tuned to represent a wide range of user mobility patterns between the two extreme cases of random-walk and constant velocity fluid-flow. The collected mobility traces are generated by using the aforementioned network setup since publicly available traces do not have the required level of details in the user mobility.

We simulate VoIP and V-conf as examples of symmetric applications with average session holding times of 3 and 30 minutes respectively (for both access technologies). Additionally we simulate RoD and VoD as examples of asymmetric applications with mean cellular holding times of 60 and 90 minutes and mean WLAN holding times of 10 and 15 minutes respectively. We assume that the WLANs are overlapping with 30% of the 3G coverage. For other coverage percentages, the results are similar and are omitted for formatting limitations. Additionally, we simulate the complete mobility randomness spectrum between random-walk and fluid-flow mobility patterns. This is realized by tuning the memory factor of the Gauss-Markov model, α_v , between zero and one, which respectively correspond to random-walk and fluid-flow mobility patterns. The parameters of the zone residence PH distributions are estimated using the EMpht package [15]. According to the coefficient of variation, θ_x^2 , the measurements of the ZRT are fitted to hyper-exponential, hypo-exponential, or exponential if $\theta_x > 1$, $\theta_x < 1$, or $\theta_x = 1$ respectively [8]. In the shown results, we compare the simulation results with the analytical results of ZRTM, ECM, and independence modeling (IM) approaches. The latter approach is a naive extension of PH traditional modeling in which the correlation between different time parameters is ignored [6].

Figure 6 plots the session cellular utilization and session VHO rate versus user mobility pattern randomness in the 3G-WLAN integrated system. The figures show excellent match between the simulation and analysis results of both ECM and ZRTM. The mismatch between the simulation and analysis results is less than 8% for all derived metrics even for highly random mobility patterns. Additionally, Figure 6 shows that ignoring the correlation between different the CRT and other zone residence times leads to significant errors in the estimated performance metrics as shown for IM. Hence, these results verify the accuracy of the presented modeling and performance analysis frameworks and emphasize their importance in estimating different performance metrics. Figure 6(a) shows that the cellular utilization of a symmetric application is independent of user mobility pattern and it is directly proportional to the mobility randomness for asymmetric applications. Additionally, Figure 6(b) shows that VHO signaling is inversely proportional to the mobility randomness. These results are logical consequences of the inverse relation between motion randomness and the user locality. Fluid-flow travelers preserve their direction and speed while random walkers continuously change both their direction and their velocity. Hence, a fluid-flow traveler has a shorter zone residence time compared to a random-walker, but the former visits more zones during an active session. Consequently, fluid-flow travelers tend to complete their asymmetric session faster due to frequent WLAN visits at the cost of larger signaling load.

V. CONCLUSION

In this article, we have presented a generic framework for mobility modeling and performance evaluation of NGHWN. The modeling framework realizes all the mobility model requirements including flexibility, analytical tractability, and accounting for the correlation between different model parameters. Additionally, we present general guidelines for evaluating the performance of NGHWN using the proposed mobility and session models. These guidelines provide a means to accurately estimate time-based and transition-based performance metrics for a wide range of mobility patterns.

² $\theta_x = \frac{\sigma_x}{\mu_x}$, where σ_x and μ_x represent the standard deviation and the mean of the corresponding measurements respectively.

REFERENCES

- [1] J.-Y. Le Boudec and M. Vojnovic, "Perfect simulation and stationarity of a class of mobility models," in *INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE*, vol. 4, 13-17 March 2005, pp. 2743–2754vol.4.
- [2] B. Liang and Z. J. Haas, "Predictive distance-based mobility management for multi-dimensional PCS networks," *IEEE/ACM Trans. on Netw.*, vol. 11, no. 5, pp. 718–732, October 2003.
- [3] P. Orlik and S. Rappaport, "A model for teletraffic performance and channel holding time characterization in wireless cellular communications with general session and dwell time distributions," *IEEE J. on Select. Areas in Commun.*, vol. 16, no. 5, pp. 788–803, June 1998.
- [4] Y. Fang, "Hyper-erlang distribution and its applications in wireless and mobile networks," *Wireless Networks (WINET)*, vol. 7, no. 3, pp. 211–219, 2001.
- [5] M. F. Neuts, *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*. The Johns Hopkins University Press, 1981.
- [6] A. H. Zahran, B. Liang, and A. Saleh, "Modeling and performance analysis for beyond 3G integrated wireless networks," in *Proc. IEEE International Conference on Communications (ICC)*, June 2006.
- [7] G. Bolch, S. Greiner, H. de Meer, and K. S. Trivedi, *Queuing networks and Markov Chains: Modeling and Performance Evaluation with Computer Science Applications*, 2nd ed. Wiley, August 1998.
- [8] D. R. Cox, *Renewal Theory*. Methuen and Co., Ltd., 1962.
- [9] Y. Fang, I. Chlamtac, and Y.-B. Lin, "Call performance for a PCS network." *IEEE J. on Select. Areas in Commun.*, vol. 15, no. 8, pp. 1568–1581, 1997.
- [10] D. Hong and S. S. Rappaport, "Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and nonprioritized handoff procedures," *IEEE Trans. Veh. Technol.*, vol. 35, no. 3, pp. 77 – 92, Aug. 1986.
- [11] G. Latouche and V. Ramaswami, *Introduction to Matrix analytic Methods in Stochastic Modeling*, ser. ASA-SIAM series on Statistics and Applied Probability. SIAM, 1999.
- [12] A. Papoulis and S. Pillai, *Probability, Random Variables and Stochastic Processes*, 4th ed. McGraw-Hill, 2002.
- [13] M. Li, M. Claypool, R. Kinicki, and J. Nichols, "Characteristics of streaming media stored on the web," *ACM Trans. Inter. Tech.*, vol. 5, no. 4, pp. 601–626, 2005.
- [14] A. Zahran, B. Liang, and A. Saleh, "Application signal threshold adaptation for vertical handoff in heterogeneous wireless networks," *ACM/Spring Mobile Networks and Applications (MONET), Special Issue on Soft Radio Enabled Heterogeneous Networks*, vol. 11, no. 4, pp. 625 – 640, Aug 2006.
- [15] S. Asmussen, O. Nerman, and M. Olsson, "Fitting phase-type distribution via the EM algorithm," *Scand. J. Statist.*, vol. 23, pp. 419 – 441, 1996.

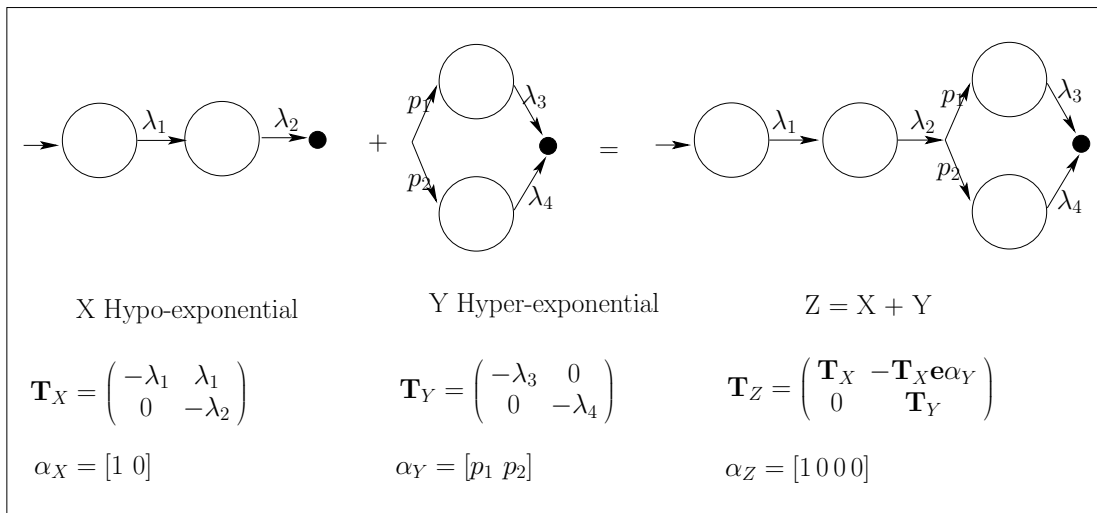


Fig. 1. PH distribution closure under addition

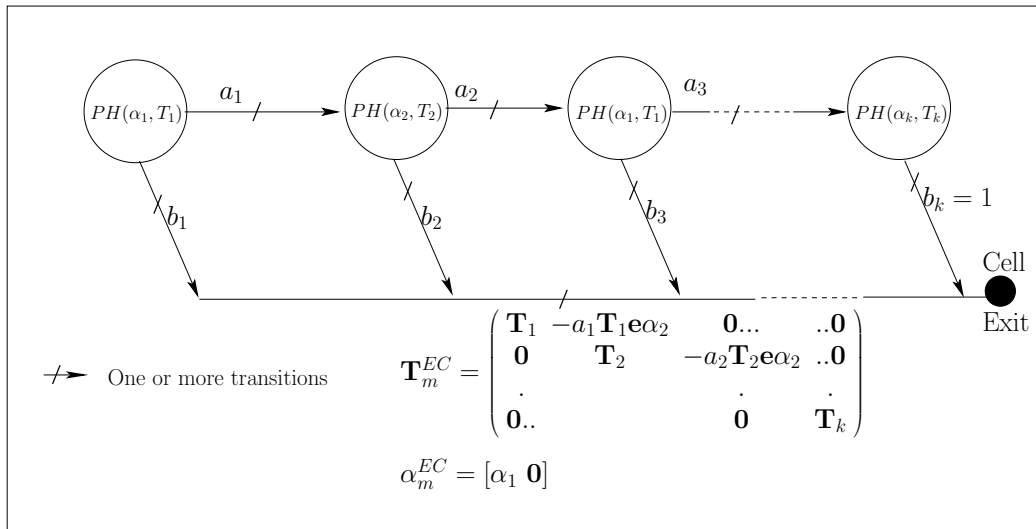


Fig. 2. Coxian and extended-Coxian mobility models

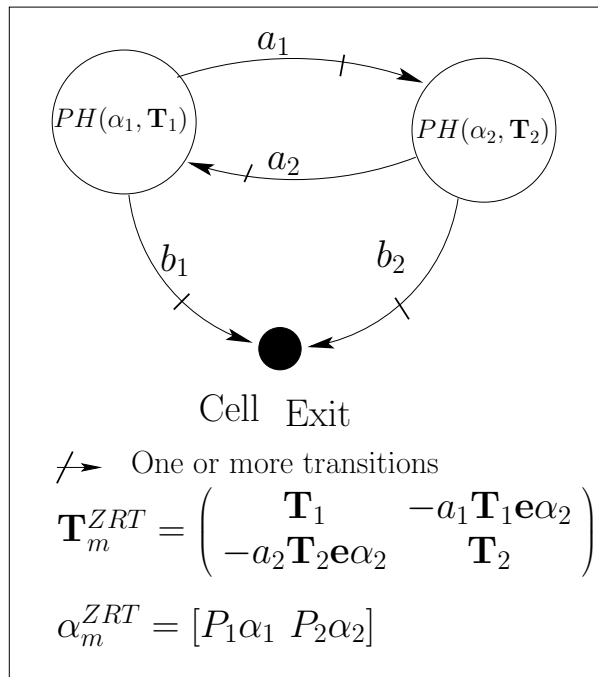


Fig. 3. Two-tier ZRT mobility model

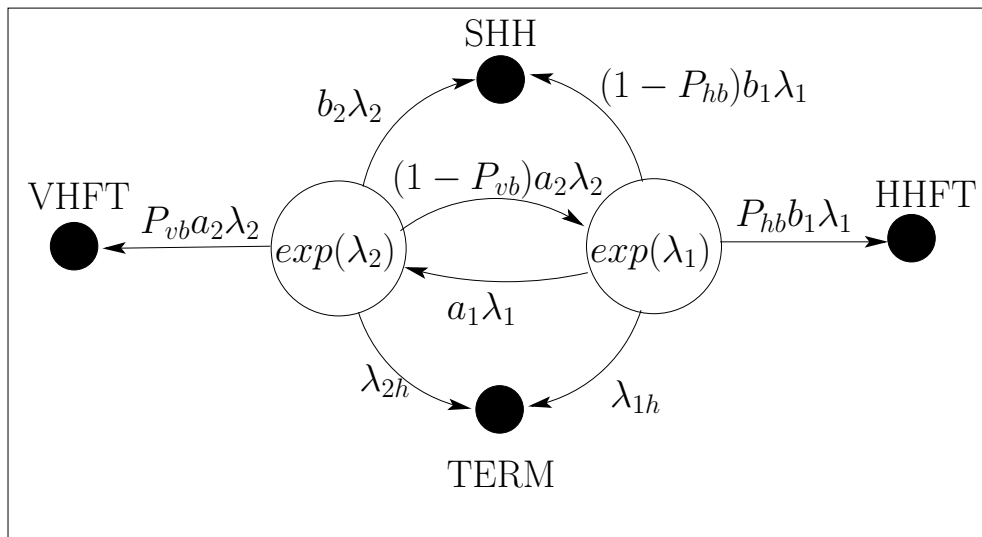


Fig. 4. ZRT session model

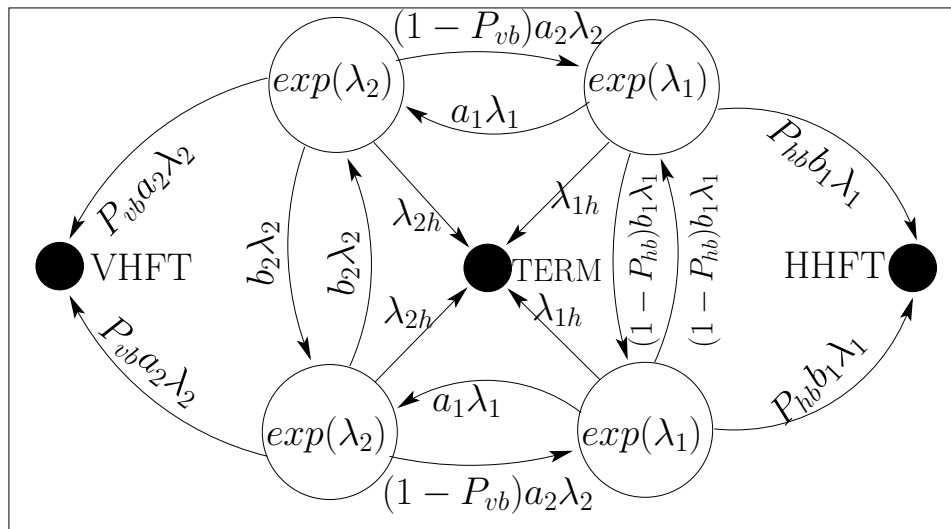


Fig. 5. ZRTM session Analysis

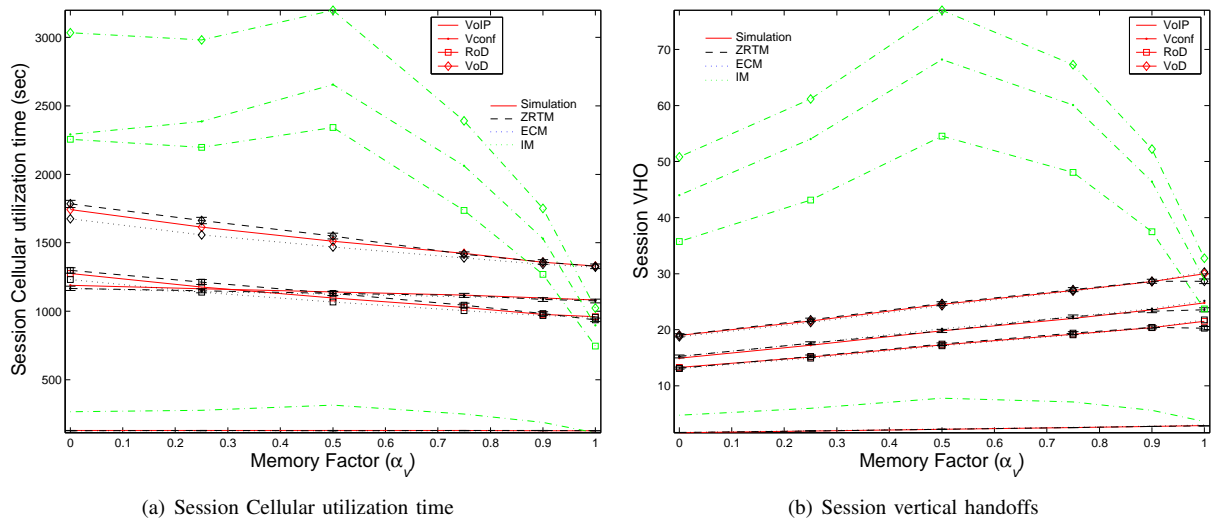


Fig. 6. Performance metrics vs. mobility pattern randomness