A Novel Framework for Handoff Analysis Under Generalized Session and Mobility Statistics

Wolfgang Bziuk
Technical University of Braunschweig
Hans-Sommer-Straße 66
D-38106 Braunschweig, Germany
bziuk@ida.ing.tu-bs.de

Said Zaghloul
Technical University of Braunschweig
Hans-Sommer-Straße 66
D-38106 Braunschweig, Germany
zaghoul@ida.ing.tu-bs.de

Admela Jukan
Technical University of Braunschweig
Hans-Sommer-Straße 66
D-38106 Braunschweig, Germany
jukan@ida.ing.tu-bs.de

ABSTRACT

Recently, cellular networks have witnessed major developments pertaining to user mobility, rich multimedia service offering and a melange of network access options including 3G/4G, WiFi, WiMAX, etc. As a result, sessions are expected to last longer and users are more likely to roam between access technologies and to other networks. As architectural design is advancing in all-IP cellular systems such as the Long Term Evolution (LTE), the question we propose to address is whether the established results on handoff and roaming statistics used in cellular theory still apply. To this end, we revisit the theory for handoff statistics and take up the challenge on extending the current model under general assumptions for session distributions, user mobility, network coverage and access technology. We show that the derived model yields estimates of handoff frequency and roaming statistics which cannot be obtained otherwise. The key strength of the proposed analysis is offering closed form results, which are easy to use and can lead to more accurate conclusions about the signaling load and the observed QoS, as a direct function of handoff statistics.

Categories and Subject Descriptors

C.4 [**Performance of System**]: Performances attributes

General Terms

Theory, Performance

1. INTRODUCTION

Recently, cellular networks have witnessed major developments pertaining to user mobility, rich multimedia service offering and a melange of network access options including 3G/4G, WiFi, WiMAX, etc. As a result, sessions are expected to last longer and users are expected to generate more usage anytime and on the move. In such systems, mobility between cells, groups of cells, gateways managing different cells, or even cells belonging to different radio technologies can trigger various sets of control plane signaling procedures at the radio and IP levels over numerous protocol interfaces [1, 2]. Although numerous mobility aware protocols are standardized within such systems for handoff management, policy and QoS authorization, the handoff and roaming statistics remain as key performance indicators as they generally reflect the amount of IP and higher layer signaling as well as the likelihood of session dropping. Thus, the fundamental question we pose in this article is whether the established results on handoff statistics used in the current cellular research still cover emerging cellular deployment scenarios.

Most of the current results on handoff statistics have obtained the mean number of handoffs during a call by the ratio of the mean session duration to the cellular residence time, under the assumption of infinite network size and identically distributed cell residence times. This was achieved by using complex analysis as in [3–5], probability generating function derivatives as in [6], or by assuming Gamma fits for the residence time as in [7]. Such homogeneous residence time assumptions were used extensively in the evaluation and the optimization of latency performance of mobility protocols such as Mobile IP as in [8, 9]. However, recent studies [10] have shown by simulations that irregular cell boundaries, which are common in practice, affect the residence time in a cell, and hence impact the estimated handoff rate in the system. Irregularities are indeed an issue when considering different radio access technologies such as 3G, 4G, and WiFi in which cell sizes and shapes vary considerably. This aspect motivated the researchers in [20] to address the handoff rate in heterogenous 3G-WLAN integrated systems. They studied one 3G cell including some WLAN hot spots and used a phase type distribution, where the different phases are used to model different exponentially distributed residence times. Their generalized Coxian phase model allows to approximate more general distribution with the disadvantage of a possible state explosion. The rigorous estimation of the handoff statistics in the context of new cellular architectures requires a more generic approach beyond

current methods as they entail unsuitable assumptions of infinite network size [5] or single cell [20] arrangements which nullify the possibility of roaming between networks and ignore the geographic user distribution.

In this article, we derive, for the first time, closed form expressions for the mean number of handoffs and roaming likelihood under generic assumptions of session and residence times, network coverage, and mobility. This is a non-trivial extension as it involves not only the consideration of temporal factors of session and residence times, but also spatial aspects of area/gateway arrangements and mobility patterns between them and to other technologies and networks. Our framework incorporates both aspects using a two dimensional Markovian mobility model and offers closed form results using complex analytic techniques. The closed form solution and the generality of our assumptions makes it easy to use the model by plugging in all needed parameters from accounting records which are readily available at the authentication, authorization, and accounting (AAA) systems or at the downstream billing systems. In our results section, we rigorously prove and show that modeling the number of handoffs as the ratio of the mean session to residence time can be inaccurate for a wide range of scenarios, as it is in fact a non-linearly increasing function of this ratio. The derived expression can be used to evaluate number of handoffs belonging to different access technologies as well as deployment scenarios that consider non-uniform user concentrations, which have not been addressed to date. The rest of the paper is organized as follows. In Section 2, we present a case study of an emerging network architecture and define the mobility model as well as the related residence times in terms of the spatial arrangement of the network. In Section 3, we generalize the commonly known handoff probability to consider non-identically distributed residence times. In Section 4, we incorporate the area layout and mobility patterns using a transient Markov chain and derive closed form solutions. Section V presents the numerical results.

2. OVERVIEW AND MODELING

2.1 Example of all-IP Cellular Networks

To illustrate the complexity of the control plane, we consider the scenario in Figure 1 which illustrates a simplified all-IP cellular network [1, 2] of two operators, one supporting two different access technologies (e.g., 3GPP LTE and non-3GPP such as WiMAX or EVDO), and the other as the roaming partner. The operator's network under consideration consists of the radio access network which includes base stations and a myriad of core network components. Core network components include serving gateways (S-GW) which route the data traffic between the radio access and the packet data network gateway (P-GW). The Mobility Management Entity (MME) handles users' mobility between S-GWs and within location or paging areas consisting of a few cells served by the S-GW. In addition, the S-GW and the P-GW interact with the Policy and Charging Rules Function (PCRF) to authorize QoS and obtain charging rules for services as users initiate sessions or move between cells, location areas, or S-GW regions. In non-3GPP networks, Access Gateways (AGW) are used to achieve similar functionality to the S-GW. When users roam to other operators, more complex signaling is needed to handle mobility and authorize QoS. In such architecture, the handoff event has a

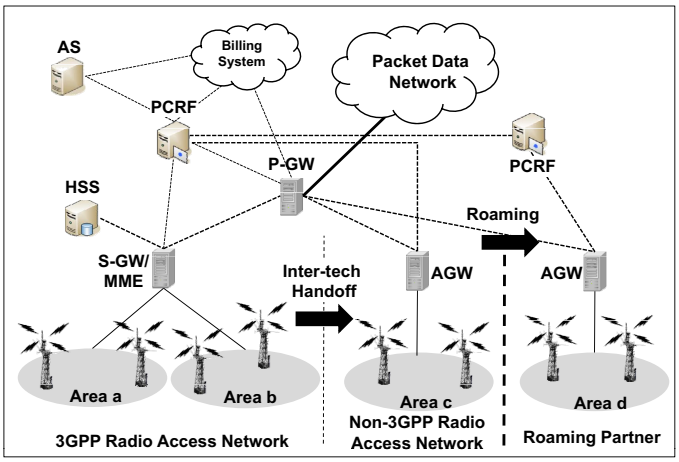


Figure 1: Simplified All-IP cellular network architecture. [AS: Application Server, PCRF: Policy and Charging Rules Function, HSS: Home Subscriber Subsystem, S-GW: Serving Gateway, MME: Mobility Management Entity, P-GW: Packet data network Gateway, AGW: Access Gateway].

broader meaning beyond movement between cells and layer 2 signaling to cover mobility events of movements between location areas or even gateways serving different radio access technologies. The handoff event in its broader meaning impacts more components that manage mobility, policy, and user credit. For instance, suppose that we have a location based service where users are charged differently depending on their location areas (see [12]). In this case, the policy system (i.e., PCRF) is contacted to generate new charging rules whenever the users move between location areas *during the sessions*. When the user is prepaid, the billing system is also contacted for service rating purposes. From this example, it is clear that the residence time within cells and thus within location areas can not be assumed identical due to coverage (e.g., 4G cells are usually smaller than 3G cells) and usage behavior which will have a significant impact on the observed number of handoffs between areas during the session. For simplicity, let us use the term "area" to generally refer to a cell or a cellular region like a location area.

2.2 Solution Methodology

In a nutshell, the derivation steps for the handoff and roaming statistics solution are structured as follows: First, we formally define the mobility model as well as the related residence times in terms of the spatial arrangement of the network areas. Then, we generalize the commonly known handoff probability [4] to consider non-identically distributed residence times. Afterwards, we incorporate the area layout and mobility patterns between gateways using a transient Markov chain. Finally, using complex analysis theory, we derive an easy-to-use closed form solution for the mean number of handoffs and the roaming likelihood.

In our analysis, the user mobility inside an area is modeled by the residence time. Since handoff signaling is relevant to active sessions, we use in-session residence time measurements (see [19, 25]) which can be easily obtained

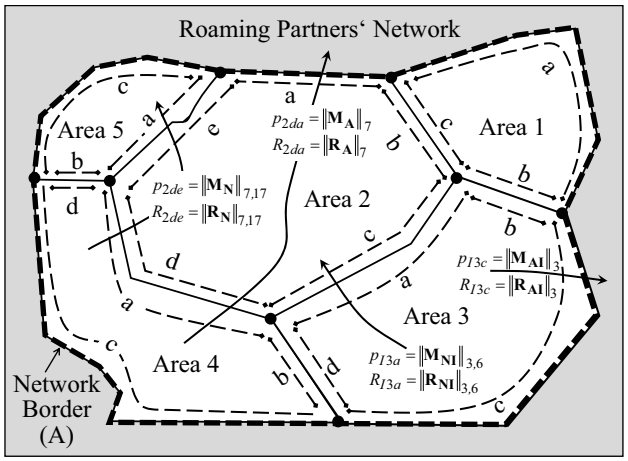


Figure 2: Sample network topology with ($n = 5$) areas. Borders are marked $a, b, c, ...$ with mobility transition probabilities and residence times for a session starting in Area 3 and another already established session entering Area 2 through its border d.

from network components for all users. In fact our method parameters can be obtained from accounting records at the AAA system which includes complete record of the session movement events and usage counters as it moves in the network including information such as the serving base station, the serving gateway, the session duration, the time duration between handoff events (see [16] for instance). This information was also used in [23] for designing an online mechanism which optimizes the reliability of accounting traffic. Our model can also accommodate theoretical models for residence time for cells and for areas composed of different cells as in our work in [14]. Unlike current work, we allow non-identically distributed residence times and relate them to any arbitrary mobility pattern between areas. This is achieved by considering the residence time depending on the entry and exit borders of the serving area according to the mobility pattern (e.g. residence time R_{2de} denotes the time needed to enter from border d and exit from border e in Area 2, see Fig.2).

To characterize mobility patterns between areas, we use a Markovian 2D chain as its parameters can be obtained from the AAA accounting records - see [21] for an overview on mobility modeling. In our Markovian model, when a mobile node enters an area its future movement is described by a set of transition probabilities, which are denoted as p_{jxy}, where j denotes the area, x and y define the entering and the exiting borders respectively (see Fig.2 for instance). As such, the model is able to characterize correlated movement pattern. The advantage of our model is that it directly applies to the pixel based mobility models developed for realistic network planning and simulation [13], as well as to the Pedestrian and Manhattan-like Urban Deployment model and the Vehicular Environment Deployment model standardized by ETSI [22]. Notice that since accounting records pertaining to a session show the sequence of areas traversed by a mobile, the probabilities, p_{jxy}, can be obtained accordingly.

2.3 Definitions and Model parameters

Before we start, let us list our assumptions,

- The session duration, S, has density $f_S(t)$, a rational Laplace transform $f_S^*(s)$ and a mean of E_s.
- The (in-session) residence time R is generally distributed with an existing Laplace transform $f_R^*(s)$. Subsequent residence times are independent.
- For simplicity, we assume that sessions are always resumed after handoffs. The inclusion of blocking is straightforward and can be carried out as in [3].

Let us define $B_I = \{1, 2, ..., n\}$ as set of all areas and the lexograhically ordered set of entry borders of all areas as $B = \{1a, 1b, 1c, ..., na, nb, ...\}$. Then the transition probabilities can be arranged into the one step transition matrix $\mathbf{M_N}$ and vector $\mathbf{M_A}$. The columns and rows of $\mathbf{M_N}$ are numbered according to the set B. The matrix, $\mathbf{M_N}$, describes the movement of a session between areas using all probabilities, p_{jxy}. For example, a session entering Area 2 through its border d as shown in Fig.2 correspondents to the row ($2d$) of the matrix $\mathbf{M_N}$ given below, where the state numbering is added on top and to the right side for clarity. A session leaving Area 2 through border e enters Area 5 at border a with probability p_{2de} (see the example below and Fig.2).

$$\mathbf{M_N} = \begin{array}{c} \begin{array}{ccccccc} (1b) & (1c) & \cdots & (3a) & \cdots & (5a) & \cdots \end{array} \\ \left(\begin{array}{ccccccc} \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & p_{2db} & \cdots & p_{2dc} & \cdots & p_{2de} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{array}\right) \end{array} \begin{array}{l} \vdots \\ (2d) \\ \vdots \end{array}, \quad \mathbf{M_A} = \begin{array}{c} (A) \\ \left(\begin{array}{c} \vdots \\ p_{2da} \\ \vdots \end{array}\right) \end{array}$$

To simplify the notation and without loss of generality, we do not differentiate movements to different roaming partners, and hence consider one network entry border denoted as A. Thus for an area at the network boundary there is only one "roaming" border and all transitions to the roaming partners are listed in the $1 \times n$ column vector $\mathbf{M_A}$ (e.g. a session entering Area 2 through border d and leaving it at border a, will leave the network with probability p_{2da}, see Fig.2).

Similarly, new sessions starting in Area j and leaving at border y are described by the transition probabilities p_{Ijy} (e.g. see Fig.2), which are arranged in the matrix $\mathbf{M_{NI}}$. $\mathbf{M_{NI}}$ has $m_{NI(i,j)}, i \in B_I, j \in B$ elements. Again, all transitions to roaming partners are combined in the column vector $\mathbf{M_{AI}}$. In summary, the mobility transition probabilities are given as,

$$\mathbf{M_N}\ ,\ \mathbf{M_A}\ ,\ \mathbf{M_{NI}}\ ,\ \mathbf{M_{AI}} \tag{1}$$

When structures are regular (e.g., a network build by rectangularly arranged areas), the matrices can be formulated by simple rules as has been shown in [13]. Finally let us denote the row vector of the session initialization probabilities for a session starting in area $i \in B_I$ as

$$\mathbf{P_I} = \{p_I(1), ..., p_I(i), ..., p_I(n)\} \tag{2}$$

Using the same notation, we define the residence time R_{kxy} for a session entering Area k through its border x and leaving it through border y. Thus, they are related to each possible movement direction as shown in Fig.2, and are summarized

in matrices $\mathbf{R_N}$ and $\mathbf{R_A}$. Since a new session starts inside an area, the corresponding residence times in the initial areas are different from the subsequent ones[1]. The corresponding residence times are represented by the matrix $\mathbf{R_{NI}}$ and the vector $\mathbf{R_{AI}}$. In our analysis, we frequently use the Laplace transform of the densities of the residence times, denoted as,

$$\mathbf{R^*_N(s)}\ ,\ \mathbf{R^*_A(s)}\ ,\ \mathbf{R^*_{NI}(s)}\ ,\ \mathbf{R^*_{AI}(s)} \tag{3}$$

For roaming sessions coming from other operators, the model still applies, by replacing the session duration by its remaining lifetime and adjusting the session initial probabilities to reflect only entering from network borders (see [15]).

To further simplify the notation, we order the set B such that each state representing an area entry border is mapped to the set of integers. For example, for the network in Fig.2 we have $B = \{1a, 1b, 1c, 2a, 2b, 2c, 2d, 2e, 3a, ...\} \Leftrightarrow \{1, 2, 3, 4, 5, 6, 7, 8, 9, ...\}$. This allows us to write the mobility probabilities (1) and the residence times in the generic form,

$$x_{i,j} = \begin{cases} x_{N(i,j)} & = \|\mathbf{X_N}\|_{i,j} \quad , \quad (i,j) \in B \\ x_{NI(i,j)} & = \|\mathbf{X_{NI}}\|_{i,j} \quad , \quad i \in B_I, j \in B \\ x_{A(i)} & = \|\mathbf{X_A}\|_{i} \quad , \quad i \in B, j = A \\ x_{AI(i)} & = \|\mathbf{X_{AI}}\|_{i} \quad , \quad i \in B_I, j = A \end{cases} \tag{4}$$

where we can set $\mathbf{X} \equiv \mathbf{M}$, or $\mathbf{X} \equiv \mathbf{R}$ respectively. As an example refer to Fig.2, where we have $m_{N(7,17)} = \| M_N \|_{7,17} = p_{2de}$ and $m_{A(7)} = \| M_A \|_7 = p_{2da}$. The Laplace transform of the residence time density is similarly given as $R^*_{N(7,17)}(s) = \| R^*_N(s) \|_{7,17} = R^*_{2de}(s)$.

3. HANDOFF PROBABILITIES FOR NON-IDENTICAL RESIDENCE TIMES

In this section, we derive the probability, P_H, that a session makes one more handoff given that it made k handoffs. Our challenge is to generalize current results for P_H ([5]) because we consider non-identically distributed residence times. We derive P_H, first over a single movement path and then over all possible paths.

3.1 A Single Path Analysis

All movement patterns related to a handoff can be represented by a simple Trellis diagram as in Fig.3. We use the sets B_I and B to define the states related to a movement inside the network. On the other hand, we use the state A (shown twice for clarity) as destination for a session roaming outside the network. Each state transition is associated with two variables, the mobility transition probabilities and the corresponding residence times according to the notation in (4).

Using the Trellis in Fig.3, we define a set W of ordered pairs $w = (o, d)$. $o \in B_I$ indicates the area in which a session originates, and $d \in B$ defines the entry border of the current area in which the session is served. The set of all possible paths $\Upsilon^{(o,d)}(k)$ is hence the sequence of areas, a session may have visited, after k handoffs and is associated with each

[1]In the absence of measurements the initial residence time can be evaluated by $R_{Iky} = \sum_{x \in B_k} p_{kxy} \tilde{R}_{kxy}$, where $\tilde{R}$ denotes the residual of the residence time, which has density $f_{\tilde{R}}(t) = \frac{1 - f_R(t)}{E_R}$ [3]

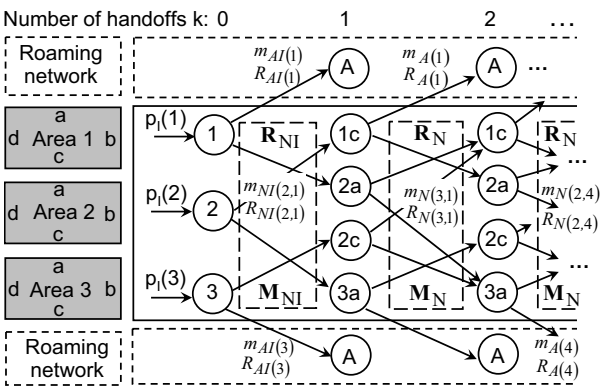


Figure 3: Trellis diagram with mobility transition probabilities ($\mathbf{M_{NI}}$, $\mathbf{M_N}$, $\mathbf{M_{AI}}$, $\mathbf{M_A}$) and residence times ($\mathbf{R_{NI}}$, $\mathbf{R_N}$, $\mathbf{R_{AI}}$ and $\mathbf{R_A}$) for a linear arrangement of 3 areas (state A is shown twice to keep the diagram simple).

pair $w \in W$. Let us define $\nu(k) \in \Upsilon^{(o,d)}(k)$ as an ordered set of elements, $\nu(k) = \{(o,i),(i,l),...,(j,d)\}$, describing a specific path of length $k \geq 1$ from state o to state d, and $\nu(0) = \emptyset$. Let us denote the i^{th} ordered pair in the path as $\nu(k,i), i = 1, ..., k$. Using these definitions, the probability that a session will take a specific path after k handoffs, is determined by the mobility transition probability ($\mathbf{X} \equiv \mathbf{M}$ in (4)) as,

$$\pi_{\nu(k)} = \prod_{(i,j) \in \nu(k)} m_{i,j} = \prod_{l=1}^{k} m_{\nu(k,l)} \tag{5}$$

Similarly, the sum of residence times incurred over a specific path $\nu(k) \in \Upsilon^{(o,d)}(k)$ since the session start until the k^{th} handoff event is given as,

$$R_{\nu(k)} = \sum_{(i,j) \in \nu(k)} R_{i,j} = \sum_{l=1}^{k} R_{\nu(k,l)} \tag{6}$$

and has density $f_{R\nu(k)}(t)$. We simply refer to (6) as the *path residence time*. Since the residence times are assumed to be independent, the Laplace transform of $f_{R\nu(k)}(t)$ is given as,

$$f^*_{R\nu(k)}(s) = \prod_{(i,j) \in \nu(k)} f^*_{R_{i,j}}(s), \ k \geq 1. \tag{7}$$

Let us define the handoff probability, $P_{H(\nu(k),(d,n))}$, as the probability that a session, which has already incurred k handoffs over the path $\nu(k)$ and has entered the current area from border d, is long enough to include at least one more handoff (the $(k+1)^{th}$) to the next area's border $n \in B$ as,

$$\begin{aligned} P_{H(\nu(k),(d,n))}(k+1) &= P\left\{R_{\nu(k)} + R_{d,n} \leq S \,\middle|\, R_{\nu(k)} \leq S\right\} \\ &= \frac{P\left\{R_{\nu(k)} + R_{d,n} \leq S\right\}}{P\left\{R_{\nu(k)} \leq S\right\}}, \ k \geq 1 \end{aligned} \tag{8}$$

For $k = 0$, (8) simplifies to $P_{H(\nu(0),(d,n))}(1) = P\{R_{d,n} \leq S\}$. Using complex analysis as in [4,5] and the Laplace transform (7), the probability $P\{R_{\nu(k)} \leq S\}$ is given as the solution of

the complex contour integral as,

$$P\{R_{\nu(k)} \leq S\} = \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} f^*_{R\nu(k)}(s)\frac{f^*_S(-s)}{s}ds \quad (9)$$

$$= -\sum_{s_P \in \Xi_{S-}} \underset{s=s_p}{Res}\, f^*_{R\nu(k)}(s)\,\frac{f^*_S(-s)}{s} \quad (10)$$

where it is assumed that the set of poles Ξ_{S-} of $f^*_S(-s)$ only contains poles s_p in the right half side of the complex plane. If $f^*_S(s)$ is a rational function (e.g. Hyper Erlang distribution), (10) has a simple closed form solution [17].

3.2 Generalized Handoff Probability

Up to now we have focused on the properties of a single path in the Trellis diagram. Let us now consider the possible aggregation of the mobility behavior from all possible paths a session can take. Let us assume that a session has entered an area through its entry border $j \in B$ at the k^{th} handoff event. Using (2) and (5), the probability to be in state $j \in B$, taking into account all possible paths of length k, is given as,

$$\pi_{Dj}(k) = \|\mathbf{P_D(k)}\|_j = \sum_{i\in B_I}\sum_{\nu(k)\in\Upsilon^{(i,j)}(k)} p_I(i)\pi_{\nu(k)}, \quad (11)$$

where the probability row vector, $\mathbf{P_D(k)}$, can be written as,

$$\mathbf{P_D(k)} = \mathbf{P_I M_{NI} M_N}^{k-1}. \quad (12)$$

The matrix multiplication starting with $\mathbf{M_{NI}}$ for the initial movement, takes into account all possible paths from the initial set of states B_I to the final destination states $j \in B$. Using (6) and (11), the weighted path residence time over all possible paths of length k to the destination state $j \in B$ is given as,

$$R_{Dj}(k) = \sum_{i\in B_I}\sum_{\nu(k)\in\Upsilon^{(i,j)}(k)} \frac{p_I(i)\pi_{\nu(k)}}{\pi_{Dj}(k)} R_{\nu(k)} \quad (13)$$

PROPOSITION 3.1. *Let us assume that the paths are chosen from the set $\cup_i \Upsilon^{(i,j)}(k)$ of all paths from the initial states to state $j \in B$ are independently selected. Then, the probability of making k or more handoffs over all possible paths is,*

$$P\{R_{Dj}(k)\leq S\} = \sum_{i\in B_I}\sum_{\nu(k)\in\Upsilon^{(i,j)}(k)} \frac{p_S(i)\pi_{\nu(k)}}{\pi_{Dj}(k)} P\{R_{\nu(k)}\leq S\} \quad (14)$$

The proof is obviously using (13).

Eq. (14) gives us two alternatives to derive the probability $P\{R_{Dj}(k) \leq S\}$. In the first method, we calculate the weighted path residence time and then compare it to the session time. Alternatively, we can take the weight over the individual path survival probabilities $P\{R_{\nu(k)} \leq S\}$.

Next we consider the probability for a session to arrive at a specific state over a single path, depending on both mobility and session time. As shown in Fig.4, for a session starting ($k = 0$) in Area i, the transition probability to state $l \in B$ is given by the product of $m_{NI(i,l)}$ for the initial mobility movement and the initial handoff probability, $P_{H(\nu(0),(i,l))}(1) = P\{R_{NI(i,l)} \leq S\}$, because session

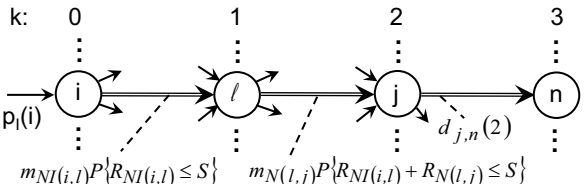


Figure 4: Partial Trellis diagram for the movement of a single session

statistics and mobility are assumed to be independent. For $k = 1$, the transition probability from state l to state j is then given by $m_{N(l,j)}P_{H(\nu(1),(l,j))}(2) = m_{N(l,j)}P\{R_{NI(i,l)} + R_{N(l,j)} \leq S\}/P\{R_{NI(i,l)} \leq S\}$ and depends on the whole residence time history. Finally the path probability follows from the multiplication of all one step transition probabilities included in that path.

Now, we derive the probability for a session to be in state j after making k handoffs, $p_j(k)$, which is obtained by weighting the path probabilities over all possible paths $\nu(k) \in \cup_i \Upsilon^{(i,j)}(k)$ to the destination state as $p_j(k) = \sum_{i\in B_I} p_I(i) \sum_{\nu(k)\in\Upsilon^{(i,j)}(k)} \pi_{\nu(k)} \prod_{l=1}^{k} P_{H(\nu(l-1),\nu(l,l))}(l)$. By extending the path residence time $R_{\nu(l-1)} + R_{\nu(l,l)} = R_{\nu(l)}$, we get $\prod_{l=1}^{k} P_{H(\nu(l-1),\nu(l,l))}(l) = P\{R_{\nu(k)} \leq S\}$, because the denominator cancels out in (8). Using (11) and (14) we have

$$p_j(k) = \sum_{i\in B_I}\sum_{\nu(k)\in\Upsilon^{(i,j)}(k)} p_I(i)\pi_{\nu(k)}P\{R_{\nu(k)} \leq S\}$$

$$= \pi_{Dj}(k)P\{R_{Dj}(k) \leq S\} \quad (15)$$

Notice that $p_j(k)$ is given by the product of the probability $\pi_{Dj}(k)$ which only considers spatial mobility aspects and the weighted path delay $P\{R_{Dj}(k) \leq S\}$ which only includes temporal aspects. Now, we consider the case where the movement path is extended for one more handoff (i.e., from state j to state n as in Fig.4) in the following proposition,

PROPOSITION 3.2. *Assuming that the session is in state j after making k handoffs, then the transition probability to state n at the $(k+1)^{th}$ handoff instant is given as,*

$$d_{j,n}(k) = m_{j,n}P_{H(j,n)}(k+1) \quad (16)$$

where $P_{H(j,n)}(k+1)$ is the handoff probability for a session which made k handoffs and is currently in state j, to make the next handoff to the neighboring area's entry border $n \in B$,

$$P_{H(j,n)}(k+1) = \frac{P\{R_{Dj}(k) + R_{j,n} \leq S\}}{P\{R_{Dj}(k) \leq S\}} \quad (17)$$

The proof is given in Appendix A.

Note that for a session which incurred k handoffs, the transition probability $d_{j,n}(k)$ only depends on the the current state j. Although the handoff probability of an individual session must consider the complete movement history, the handoff probability (17) only depends on the current state, because the sum of the residence time of all individual paths can be aggregated into the weighted path residence time

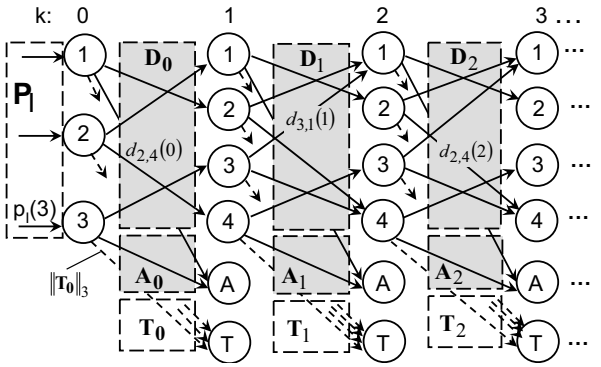


Figure 5: Diagram of the Markov chain with selected transition probabilities, absorbing states A and T (all transient states can reach state T) and state mapping $S = \{1c, 2a, 2c, 3a\} \Leftrightarrow \{1, 2, 3, 4\}$.

$R_{Dj}(k)$. Thus, our model behaves like a discrete time non homogenous Markov chain.

4. GENERALIZED HANDOFF STATISTICS

4.1 Transient Markov Chain Formulation

In this subsection, the handoff statistics is derived using the transient Markov chain shown in Fig.5. Compared to the Trellis diagram of Fig.3 the additional absorbing state T is added to model the session termination. Handoffs inside the network are transitions between transient states (set of states B_I and B) and summarized in the matrices $\mathbf{D_k}$, where k gives the number of handoffs a session has already done to reach the current state. $\mathbf{D_0}$ takes into account the initial transitions. Using (16) - (17), the matrix elements are given by

$$\| \mathbf{D_k} \|_{j,n} = d_{j,n}(k) = m_{j,n} P_{H(j,n)}(k+1) \ , \ k \geq 0 \qquad (18)$$

Eq.(15) gives the probability, $p_j(k)$, that a session reaches state $j \in B$ with the k^{th} handoff. Let us define the corresponding row vector $\mathbf{P}(k)$. Than, from the properties of a Markov chain and with the transition probabilities defined in (18), it is obvious that this vector is alternatively given by

$$\mathbf{P}(k) = \mathbf{P_I} \prod_{j=0}^{k-1} \mathbf{D_j} \ , \ k \geq 1 \ , \mathbf{P}(0) = \mathbf{P_I} \qquad (19)$$

The analysis done for (16) - (17) can be extended to derive the row vector $\mathbf{A_k}$ of the transition probabilities to absorbing state $\{A\}$. Using the same notation its elements are given by

$$\| \mathbf{A_k} \|_j = m_{j,A} P_{H(j,A)}(k+1) \ , \ k \geq 0 \qquad (20)$$

Based on the result (19) and from Fig.5 it is easy to verify that the probability for being absorbed into state A with the k^{th} handoff, which means that there are only $k-1$ handoffs inside the network (the last is the roaming one), is given as

$$P_A(1) = \mathbf{P_I A_0} \, , \ P_A(k) = \mathbf{P_I} \prod_{j=0}^{k-2} \mathbf{D_j A_{k-1}} \, , \ k \geq 2 \qquad (21)$$

Finally the transition probability to the absorbing state T results from the normalization. The row vector of transition probabilities, $\mathbf{T_k}$, has elements $\| \mathbf{T_k} \|_i = 1 - \sum_{j \in B} \| \mathbf{D_k} \|_{i,j} - \| \mathbf{A_k} \|_i$. Similar to (21), we define the probability of session termination after the $(k-1)^{th}$ handoff, $P_T(k)$, as,

$$P_T(1) = \mathbf{P_I T_0} \, , \ P_T(k) = \mathbf{P_I} \prod_{j=0}^{k-2} \mathbf{D_j T_{k-1}} \, , \ k \geq 2 \qquad (22)$$

Equations (19) - (22) completely describe the statistics of the handoff process.

4.2 The Mean Number of Handoffs

The mean number of handoffs (MNH) a session does until it leaves the network or terminates is made up of two different parts: the mean number of handoffs done inside the network, $E\{N\}$, and the probability for the last handoff (the chain goes to absorbing state A), which is the probability of roaming, β_A.

Let $\mathbf{e} = \{1, 1, ..., 1\}$ be a vector with a suitable dimension, where $\mathbf{e}^T$ denotes the transpose, then using (19) - (22), it is easy to show that the mean number of handoffs made inside the network (excluding the roaming one) can be written as,

$$E\{N\} = \sum_{k=0}^{\infty} k \cdot [P_A(k+1) + P_T(k+1)] = \sum_{k=1}^{\infty} \mathbf{P}(k) \mathbf{e}^T \qquad (23)$$

For space limitations, we omit the derivation of (23). The elements of the row vector $\sum_{k=1}^{\infty} \mathbf{P}(k)$ can be interpreted as the mean number of handoffs through area border j,

$$\| \sum_{k=1}^{\infty} \mathbf{P}(k) \|_j = E\{N_j\} \, , \quad j \in B \qquad (24)$$

Thus, the mean number of handoffs in the area under consideration, is given by summing $E\{N_j\}$ in (24) over all its corresponding borders.

From Fig.5 it is obvious that the probability of roaming is given by the summation over all the probabilities of being absorbed into state A. Using (21) we have,

$$\beta_A = \sum_{k=1}^{\infty} P_A(k) = \sum_{k=0}^{\infty} \mathbf{P}(k) \mathbf{A_k} \qquad (25)$$

Although (23) and (25) gives the solution for the MNH, it implies an infinite summation. Thus in a practical implementation the summation has to be truncated to a reasonable size, which on the other side may result in a lower accuracy. In the following subsection we derive a closed formulation which avoids the infinite summation.

4.3 Closed Form Solution

Let us define the element wise multiplication

$$\mathbf{X} = \mathbf{Y} \odot \mathbf{Z} \ \Leftrightarrow \ x_{i,j} = y_{i,j} z_{i,j} \qquad (26)$$

This allows us to formulate the following Proposition.

PROPOSITION 4.1. *Using* (26)*, let us combine the matrices of the Laplace transforms of the residence times* (3) *and*

the mobility matrices (1) *as*

$$\mathbf{Q^*_{NI}(s)} = \mathbf{M_{NI}} \odot \mathbf{R^*_{NI}(s)} \ , \ \mathbf{Q^*_N(s)} = \mathbf{M_N} \odot \mathbf{R^*_N(s)} \quad (27)$$

Than, the probability to be in a transient state $j \in S$ with the k^{th} handoff is given by the row vector

$$\mathbf{P}(k) = \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} \mathbf{P_I Q^*_{NI}(s)} \left[\mathbf{Q^*_N(s)}\right]^{k-1} \frac{f^*_S(-s)}{s} ds \quad (28)$$

The proof is given in Appendix B.

Now with (28) and using (23), the mean number of handoffs inside the network can be derived as

$$E\{N\} = \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} \mathbf{P_I Q^*_{NI}(s) M^*_R}(s)\, \mathbf{e}^T \frac{f^*_S(-s)}{s} ds \quad (29)$$

where the infinite geometric sum results into the matrix

$$\mathbf{M^*_R}(s) = \left(\mathbf{I} - \mathbf{Q^*_N(s)}\right)^{-1} = \left(\mathbf{I} - \mathbf{M_N} \odot \mathbf{R^*_N(s)}\right)^{-1} \quad (30)$$

and $\mathbf{I}$ is the identity matrix. Eq.(29) and (30) highlight the relationship between the spatial effects of mobility and the related residence times on one side and the session duration on the other side. The direction of movement and the observed residence time is correlated, which results in the strong relationship given by the matrix operation $\mathbf{M_N} \odot \mathbf{R^*_N(s)}$. On the other hand, the sequence of residence times a session incurs during its lifetime is assumed to be independent. This reflects on a product form of the individual residence times of a sample path in the Laplace domain. Both effects imply the elementwise multiplication defined for this matrix operation.

Let us apply the same analysis to the roaming probability. From Fig.5 we notice that the absorbing state A is reached by expanding the path from every transient state $j \in B$ to state A. Thus the analysis derived for (28) can be applied to $P_A(k)$ if the matrices (27) relevant for the last transient movement are replaced by the column vectors $\mathbf{Q^*_{AI}(s)}$ if $k = 0$, and $\mathbf{Q^*_A(s)}$ if $k \geq 1$ respectively. This yields

$$\mathbf{Q^*_{AI}(s)} = \mathbf{M_{AI}} \odot \mathbf{R^*_{AI}(s)} \ , \ \mathbf{Q^*_A(s)} = \mathbf{M_A} \odot \mathbf{R^*_A(s)} \quad (31)$$

Using (21), the probability for being absorbed into state A after the session has done k handoffs can be written as,

$$P_A(1) = \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} \mathbf{P_I Q^*_{AI}(s)} \frac{f^*_S(-s)}{s} ds \quad (32)$$

$$P_A(k) = \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} \mathbf{P_I Q^*_{NI}(s)} \left[\mathbf{Q^*_N(s)}\right]^{k-2} \mathbf{Q^*_A(s)} \frac{f^*_S(-s)}{s} ds$$

Now with (32) and (25) the final result for β_A is given as

$$\beta_A = P_A(1) + \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} \mathbf{P_I Q^*_{NI}(s) M^*_R}(s)\ \mathbf{Q^*_A(s)} \frac{f^*_S(-s)}{s} ds \quad (33)$$

4.4 Analytical Example

Let us assume that the session time is hyper-Erlang distributed, because it is able to approximate any continuous density function with vanishing error (Theorem 1 out of [24]). Its Laplace transform is given by

$$f^*_S(s) = \sum_{j=1}^{J} \alpha_j \left(\frac{\mu_j}{s+\mu_j}\right)^{m_j}. \quad (34)$$

The poles of $f^*_S(-s)$ are located at $s_{Pj} = \mu_j > 0$ and we can use (10) to solve each vector component of the complex integral (29), which yields

$$E\{N_j\} = -\sum_{j=1}^{J} \frac{\alpha_j (-\mu_j)^{m_j}}{(m_j-1)!} \mathbf{P_I} \lim_{s\to\mu_j} \frac{d^{m_j-1}}{ds^{m_j-1}} \mathbf{M_V}(s)$$

where we have set $\mathbf{M_V}(s) = \mathbf{Q^*_{NI}(s) M^*_R}(s)/s$. For the practically interesting case of $m_j = 2$, where first and second moment matching is straightforward to calculate [24], the required derivatives can be easily obtained as,

$$\left.\frac{d\mathbf{M^*_V}(s)}{ds}\right|_{\mu_j} = \frac{1}{\mu_j}\left[\mathbf{M_{NI}} \odot \left.\frac{d\mathbf{R^*_{NI}(s)}}{ds}\right|_{\mu_j}\right] \mathbf{M^*_R}(\mu_j) + \frac{1}{\mu_j}\left[\mathbf{M_{NI}} \odot \mathbf{R^*_{NI}(\mu_j)}\right]\left[\left.\frac{d\mathbf{M^*_R}(s)}{ds}\right|_{\mu_j} - \frac{\mathbf{M^*_R}(\mu_j)}{\mu_j}\right]$$

$$\left.\frac{d\mathbf{M^*_R}(s)}{ds}\right|_{\mu_j} = \mathbf{M^*_R}(\mu_j)\left[\mathbf{M_N} \odot \left.\frac{d\mathbf{R^*_N(s)}}{ds}\right|_{\mu_j}\right] \mathbf{M^*_R}(\mu_j)$$

The matrix $\mathbf{R^*_N(s)}$ summarizes the Laplace transform of all residence times, which are assumed to be Gamma distributed with mean $E\{R_{i,j}\} = \alpha_{i,j}\theta_{i,j}$, coefficient of variation $c_{R_{i,j}} = 1/\sqrt{\alpha_{i,j}}$ and transform

$$\|\mathbf{R^*_N(s)}\|_{i,j} = f^*_{R_{i,j}}(s) = \frac{1}{(1+\theta_{i,j}s)^{\alpha_{i,j}}} \quad (35)$$

For $f^*_{R_{i,j}}(s)$ it is straight forward to derive the residual residence time required for $\mathbf{R^*_{NI}(s)}$ and the required derivatives as well. The main numerically expensive part is the determination of the matrix $\mathbf{M^*_R}(\mu_j)$, because it requires matrix inversion. The matrix has size $|B| \times |B|$ where $|B|$ is proportional to the number of considered areas times the mean number of borders per area. It should be noted that the numerical complexity of our model is reasonable and comparable to the recent model in [20] which assumes exponentially distributed residence times and a single cell layout.

5. NUMERICAL RESULTS

Recall that our model relaxes two major assumptions of what we refer to as the *homogenous model*: the identically distributed residence times in all areas and the infinite network size without roaming. In this section, we validate our *general model* in (29), (33) by simulations by considering an exemplary region of 16 areas and compare our results to the homogeneous model. In our C++ event based simulations, we follow the guidelines in [13, 18] for session generation and mobility behavior. The number of handoffs per area is collected and confidence intervals are determined using the mean batch method (5% significance levels). In our scenario, the central four areas (e.g., downtown area) belong to a 4G technology (e.g., LTE) and the others belong to a 3G technology (e.g., EVDO). The residence time follows a Gamma distribution as it is known to match the lognormal distribution which is observed in measurement studies [25]. Unless specified otherwise, the residence time has a coefficient of variation of $c_R = 1.2$ and a mean of $R_{\text{LTE}} = 20$ mins in the LTE zone and $R_{\text{EVDO}} = 80$ mins in the EVDO zone to reflect the coverage differences between technologies. The average residence time is defined as $E_{R_{avg}} = 0.75R_{\text{EVDO}} + 0.25R_{\text{LTE}} = 65$ mins. We consider 3

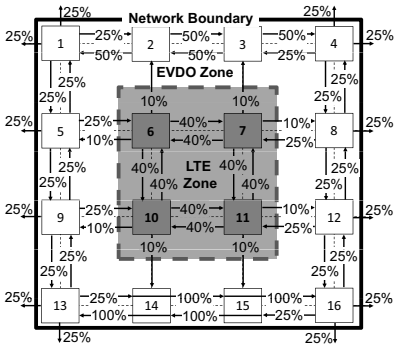


Figure 6: EVDO and LTE zones with arbitrary mobility pattern

mobility patterns: random movement (4×4) areas, directed mobility in a linear arrangement of the areas, and arbitrary pattern (see Fig.6). In random mobility all directions of movement are equally likely. In the directed movement, once a direction is picked at the session start, it is maintained until termination. For the arbitrary mobility pattern in Fig.6,we have assumed mixtures of 2-way and 4-way random mobility as well as directed movement. The session duration is assumed to be Erlang distributed with coefficient of variation of $\sqrt{0.5}$ and a variable mean of E_S to reflect different mobility ratios (i.e., $E_S/E_{R_{avg}}$ $[3, 6]$). The session initialization probability vector P_I is used to represent users' concentrations within areas, because P_I can be assumed to be proportional to the session arrival rate times the service penetration rate times the user density of each area.

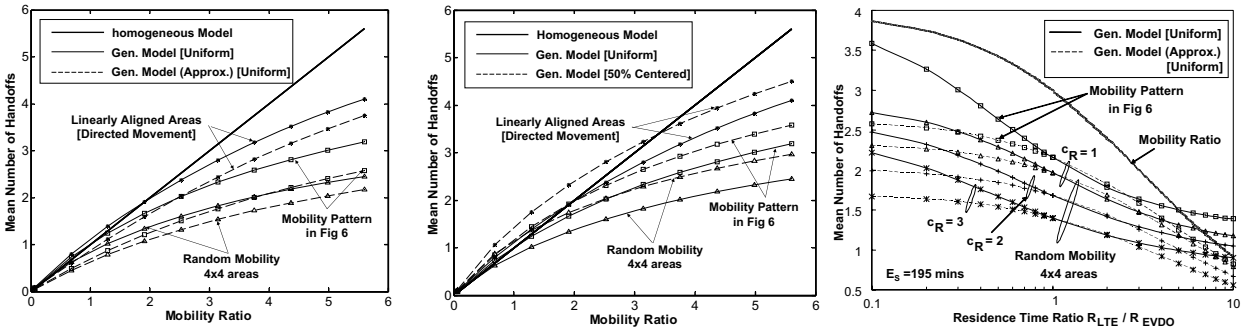
Let us first study the impact of the mobility ratio by varying the mean session duration. We consider the three mobility patterns above and compare them to our general model using average residence time $E_{R_{avg}}$ in all areas, referred as *gen. model (approximation)* and to the homogenous model. As shown in Fig.7(a), we see that the mobility pattern highly impacts the observed mean number of handoffs (MNH) during the session. For instance, the random mobility results in the lowest number of handoffs due to its localized movement effect while the directed movement results in the highest number of handoffs. Notice the large error in the homogenous model (thick solid line) as sessions become longer (i.e., higher mobility ratio). We also see that using the average residence time in all areas can lead to inaccuracies of the order of 20% (see dashed and solid lines).

The effect of user concentrations using the initial session starting probabilities is shown in Fig.7(b). In addition to a uniform concentration we consider the case of 50% user concentration in the LTE zone. If we neglected the different user concentrations, as in the homogenous model, we may underestimate the number of handoffs observed as in the case of the directed movement pattern (by approx. 30%) when traffic is concentrated in the LTE zone and also largely overestimate handoffs otherwise. We notice that ignoring the users concentration can lead to around 20% estimation error in our example. For completeness, we also compared areas with irregular shapes as in Fig.2 with the 4x4 topology. The results (not shown for space) revealed similar trends as Fig.7(a)-7(b), as expected.

In Fig.7(c) we consider the effect of the residence time ratio $R_{\text{LTE}}/R_{\text{EVDO}}$ on the MNH, where R_{EVDO} is fixed at 65 mins. Varying R_{LTE} changes the mean residence time and thus the mobility ratio, which is shown in Fig.7(c) for completeness. For both the random mobility and the arbitrary pattern and for different values of c_R we observe a quite similar behavior of the MNH if we compare the exact general model with the approximate one (using $E_{R_{avg}}$). For the arbitrary pattern in Fig.6, we observe that the results from the exact and approximate model match well in the range of ratios of $0.8 < R_{\text{LTE}}/R_{\text{EVDO}} < 3$. On the other hand, the error increases as the ratio deviates from one. For the range $R_{LTE} << R_{EVDO}$, the MNH estimates of the general model strongly increases over those of the approximate model because more handoffs are incurred in the LTE zone. For the range $R_{\text{LTE}} >> R_{\text{EVDO}}$, we observe a similar behavior, but in this case due to the low residence time in the EVDO zone. For random mobility this effect is not as pronounced because a random movement averages out the differences in the residence times, however with error in the order of $20-30\%$. In our example, we also see that in all cases the MNH decreases as the coefficient of variation of the residence time increases (i.e. $c_R = 1$ establishes an upper bound).

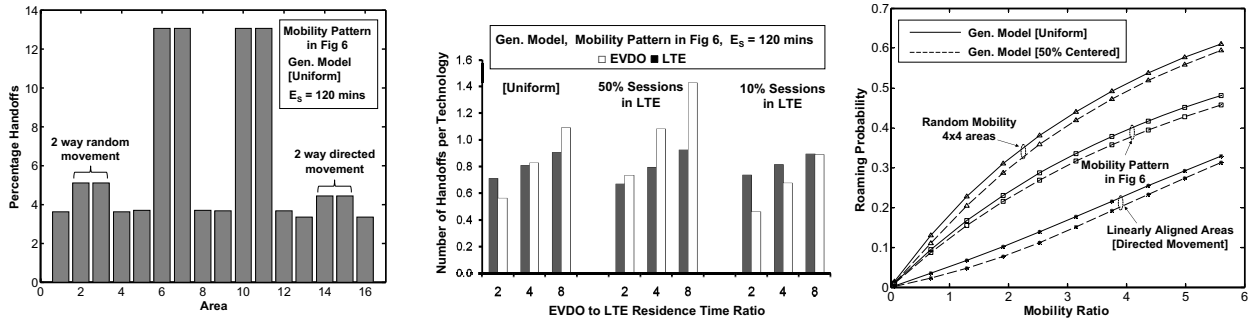
We proceed by estimating the number of handoffs within each area and for each technology using (24). This feature is very useful as it allows estimating the signaling load to the core network from each area separately. In Fig.8(a) we show the portion of the mean number of handoffs in all areas for the arbitrary mobility pattern under uniform user distribution. Clearly, the LTE zones are likely to incur more handoffs due to their relatively short residence time (i.e., R_{LTE}= 20 min) while the EVDO network (i.e., R_{EVDO}= 80 min) exhibits relatively uniform behavior among all areas. The little deviation in the EVDO network in areas 2, 3, 14, 15 is due to their non-homogeneous mobility pattern (see Fig.6). We proceed in Fig.8(b) by comparing the number of handoffs per technology as function of the incurred residence times as well as the projected user concentrations in their regions. The latter is interesting as it can reflect the deployment phase of a given technology. In our example, we consider uniform user concentration as a baseline, and also evaluate initial LTE deployment with 10% user concentration, as well as the case of 50% concentration which reflects users adoption of the technology down the road. We set the EVDO area residence time to 80 mins and vary its ratio to LTE area from 2 to 8. We see the large impact of the deployment profile on the observed mean number of handoffs in the network. As in the previous cases, we also see that the residence time ratio between LTE and EVDO highly affects the number of handoffs.

Finally, in Fig.8(c) we investigate the roaming probabilities to other operators as function of the mobility ratio reflecting short and long session durations, mobility patterns, and user concentrations. This is useful for estimating the inter-domain signaling rate to the roaming partners [15]. As expected, we observe that the roaming likelihood is higher for longer sessions (i.e., large mobility ratios). We see that the user concentration has less impact on the roaming proba-



(a) Mobility Pattern Impact (b) User Concentration Impact (c) Residence Time Ratio Impact

Figure 7: Study Case on the Mean Number of Handoffs for non homogenous residence times. [Uniform: uniform user concentration, Centered: 50% user concentration in areas 6, 7, 10 and 11. 95% confidence levels of simulation results are within marker sizes.]



(a) MNH per Area (b) MNH per Technology (c) Roaming Probability

Figure 8: Mean Number of Handoffs (MNH) per area and technology, and Roaming Likelihoods.

bility than the number of handoffs in Fig.7(b). In addition, we see that the random mobility assumption, can lead to relatively high roaming likelihoods as it ignores the user behavior and attraction points (e.g., a central LTE zone in a crowded region) while the linear arrangements exhibit the least roaming likelihood as users are only allowed to leave from the edge areas 1 and 16.

6. CONCLUSIONS

In this paper, we revisited the theory for handoffs and roaming statistics under generic assumptions of mobility patterns, spatial arrangement of gateways, as well as generic session times and non-identically distributed residence times. Using transient Markov chains and complex analysis techniques, we derived a closed form solution which not only addresses temporal aspects of session and residence times but also spatial aspects of mobility and user concentrations. The results show that significant estimation errors for the number of handoffs, using current models, can be reduced when applying our general approach, especially in scenarios involving multiple access technologies. Future work includes applying our framework to evaluate performance measures relevant to signaling protocols (e.g., Proxy MobileIP and Diameter) such as the signaling load in the core network and the associated signaling latency.

7. REFERENCES

[1] Ali, I., et al.. Network Based Mobility Management in the Evolved 3GPP Core Network. *IEEE Commun. Magazine*, vol. 47, no. 2, pp. 58-66, 2009.

[2] Balbas, J., et al.. Policy and Charging Control in the Evolved Packet System. *IEEE Communications Magazine*, vol. 47, no. 2, pp. 68-74, 2009.

[3] Fang, Y. and Chlamtac, I.. Call performance of a PCS network. *IEEE J. Selec. Areas Commun*, vol. 15, no. 8, pp. 1568-1581, 1997.

[4] Fang, Y. et al.. Analytical Generalized Results for Handoff Probabilities in Wireless Networks. *IEEE Trans. Comm.*, vol. 50, no. 3, 2002.

[5] Fang, Y.. Modeling and Performance Analysis for Wireless Mobile Networks: A New Analytical Approach. *IEEE Trans. Network.*, vol. 13, no. 5, pp. 989-1002, 2005.

[6] Rodriguez-Dagnino et al.. Counting Handovers in a Cellular Mobile Communication Network: Equilibrium Renewal Process Approach. *Performance Evaluation*, vol. 52, no. 2, pp. 153-174, 2003.

[7] Alwakeel, M.. Approximating handoff number and handoff rate in cellular networks with arbitrary call holding time and cell residence time distributions,' 2nd IEEE International Conference on Computer Science and Information Technology, pp. 22-26, 2009.

[8] Cheng, S., et al.. Performance Modeling on Handover Latency in Mobile IP Regional Registration. *Proc. 19th*

IEEE Internat. Symposium on Personal, Indoor and Mobile Radio Communications PIMRC'08, 2008.

[9] Ma, W. and Fang, Y.. Dynamic hierarchical mobility management strategy for mobile IP networks. *IEEE Jour. Selec. Areas in Commun.*, 22(4), pp. 664-676, 2004.

[10] Khan, A. N., Sha, X., Zhang, X.. Determining rate of handoff for irregularly bounded cellular environments in future mobile networks. *Information Technology Journal*, vol. 7, no. 2, pp. 344-349, 2008.

[11] Gowrishankar, R., Raju, G. and Satyanarayana, P.. Performability Model of Vertical Handoff in Wireless Data Network. *4th International Conference on Wireless and Mobile Communications*, pp. 309-314, 2008.

[12] 3GPP TS 23.203, Policy and charging control architecture (Release 9), V9.3.0. Dec 2009.

[13] Tuerke, U. et al.. Evaluation of the influence of user mobility on average traffic load, D2.5 Models and Simulations for Network Planning and Control of UMTS, (MOMENTUM). 2003.

[14] Zaghloul, S., Bziuk, W., Jukan, A.. Relating the Number of Access Gateway Handoffs to Mobility Management: A Fundamental Approach. *International Teletraffic Congress ITC21*, 2009.

[15] Zaghloul, S., Bziuk, W., Jukan, A.. Signaling and Handoff Rates at the Policy Control Function (PCF) in IP Multimedia Subsystem (IMS). *IEEE Commun. Letters Journal*, vol. 12, no. 7, pp. 526-528, 2008.

[16] 3GPP2 X.S0011-005-C, 'Accounting Services and 3GPP2 RADIUS VSAs,' vol. 1.0, Aug. 2003.

[17] Bziuk, W., Zaghloul, S., Jukan, A.. A New Framework for Characterizing the Number of Handoffs in Cellular Networks. *European Trans. on Telecommunications ETT*, no. 20, pp. 689-700, 2009.

[18] Hung, H., Lee, P., Lin, Y.. Random number generation for excess life of mobile user residence time. *IEEE Trans, on Vehicular Technol.*, vol. 55(3), pp. 1045-1050, 2006.

[19] Sricharan, M. S., Vaidehi, V.. A Pragmatic Analysis of User Mobility Patterns in Macrocellular Wireless Networks. *Pervasive and Mobile Computing*, no. 4, pp. 616-632, 2008.

[20] Zahran, A. H. , Liang, B., Saleh, A.. Mobility Modeling and Performance Evaluation of Heterogenous Wireless Networks. *IEEE Trans. Mobile Computing* vol. 7, no. 8, pp. 1041-1056, 2008.

[21] Ilyas, M., Mahgoub, I.. Mobile Computing Handbook, Auerbach Publications, 1028 pages, 2005.

[22] ETSI TR 101 112 V3.1.0, Universal Mobile Telecommunications System (UMTS); Selection procedures for the choice of radio transmission technologies of the UMTS, 1997.

[23] Zaghloul, S., Jukan, A.. Optimal Accounting Policies for AAA Systems in Mobile Telecommunications Networks. *IEEE Transactions on Mobile Computing*, vol. 9, no.6, pp.865-880, June 2010.

[24] Thuemmler, A., Buchholz, P., Telek, M. A Novel Approach for Phase-Type Fitting with the EM Algorithm. I*EEE Trans. Dependable a. Secure Computing*, vol. 3, no. 3, pp. 245-258, 2006.

[25] Yavuz, E. A., Leung, V. C. M. Modeling Channel Occupancy Times for Voice Traffic in Cellular Networks. *IEEE ICC 2007*, pp. 332-337, 2007.

APPENDIX

A. PROOF OF PROPOSITION 3.2

The joint probability to reach state $j \in B$ with the $(k)^{th}$ handoff and state $n \in B$ with the $(k+1)^{th}$ handoff can be derived from (15), if all paths terminating at state j are extended to state n. Using Proposition 3.1 this will give

$$p_{j,n}(k+1) = \sum_{i \in B_I} p_I(i) \sum_{\nu(k) \in \Upsilon^{(i,j)}(k)} \pi_{\nu(k)} m_{j,n} P\{R_{\nu(k)} + R_{j,n} \leq S\}$$

$$= \pi_{Dj}(k) m_{j,n} P\{\sum_{i \in B_I} \sum_{\nu(k) \in \Upsilon^{(i,j)}(k)} \frac{p_I(i)\pi_{\nu(k)}(R_{\nu(k)} + R_{j,n})}{\pi_{Dj}(k)} \leq S\}$$

Using (11), (13), (15) and the fact that $\pi_{Dj}(k) R_{j,n} = \sum_{i \in S_I} \sum_{\nu(k) \in \Upsilon^{(i,j)}(k)} p_I(i) \pi_{\nu(k)} R_{j,n}$, we have

$$\begin{aligned} p_{j,n}(k+1) &= \pi_{Dj}(k) m_{j,n} P\{R_{Dj}(k) + R_{j,n} \leq S\} \\ &= p_j(k) m_{j,n} \frac{P\{R_{Dj}(k) + R_{j,n} \leq S\}}{P\{R_{Dj}(k) \leq S\}} \\ &= p_j(k) m_{j,n} P_{H(j,n)}(k+1) \end{aligned} \tag{36}$$

The transition probability to reach state n with the $(k+1)^{th}$ handoff given that we are in state j after the $(k)^{th}$ handoff follows to be $d_{j,n}(k) = \frac{p_{j,n}(k+1)}{p_j(k)} = m_{j,n} P_{H(j,n)}(k+1)$.

B. PROOF OF PROPOSITION 4.1

Using the complex integral (9) for $P\{R_{\nu(k)} \leq S\}$, the probability for a path (5) and the Laplace of its density (7), the probability (15) to be in transient state j is given by

$$\begin{aligned} p_j(k) &= \sum_{i \in B_I} \sum_{\nu(k) \in \Upsilon^{(i,j)}(k)} p_I(i) \prod_{(h,l) \in \nu(k)} m_{h,l} \times \\ &\times \frac{1}{2\pi j} \int_{\sigma - j\infty}^{\sigma + j\infty} \prod_{(h,l) \in \nu(k)} f^*_{R_{h,l}}(s) \frac{f^*_S(-s)}{s} ds \\ &= \frac{1}{2\pi} \int_{\sigma - j\infty}^{\sigma + j\infty} \sum_{i \in B_I} \sum_{\nu(k) \in \Upsilon^{(i,j)}(k)} p_I(i) \prod_{(h,l) \in \nu(k)} m_{h,l} f^*_{R_{h,l}}(s) \frac{f^*_S(-s)}{s} ds \end{aligned}$$

The element wise multiplication for $\mathbf{Q^*_{NI}(s)}$ and $\mathbf{Q^*_N(s)}$ has been defined in (27) and the summation over all paths in (11). Then with (1) - (4) and adapting the matrix formulation of (12), we get the elements of the row vector (28) given as,

$$p_j(k) = \frac{1}{2\pi} \int_{\sigma - j\infty}^{\sigma + j\infty} \left\| \mathbf{P_I Q^*_{NI}(s)} \left[\mathbf{Q^*_N(s)} \right]^{k-1} \right\|_j \frac{f^*_S(-s)}{s} ds.$$