

# Modeling Spatial and Temporal Dependencies of User Mobility in Wireless Mobile Networks

Wei-Jen Hsu, *Member, IEEE*, Thrasyvoulos Spyropoulos, *Member, IEEE*,  
Konstantinos Psounis, *Senior Member, IEEE*, and Ahmed Helmy, *Member, IEEE*

**Abstract**—Realistic mobility models are fundamental to evaluate the performance of protocols in mobile ad hoc networks. Unfortunately, there are no mobility models that capture the non-homogeneous behaviors in both space and time commonly found in reality, while at the same time being easy to use and analyze. Motivated by this, we propose a time-variant community mobility model, referred to as the TVC model, which realistically captures spatial and temporal correlations. We devise the communities that lead to skewed location visiting preferences, and time periods that allow us to model time dependent behaviors and periodic reappearances of nodes at specific locations.

To demonstrate the power and flexibility of the TVC model, we use it to generate synthetic traces that match the characteristics of a number of qualitatively different mobility traces, including wireless LAN traces, vehicular mobility traces, and human encounter traces. More importantly, we show that, despite the high level of realism achieved, our TVC model is still theoretically tractable. To establish this, we derive a number of important quantities related to protocol performance, such as the average node degree, the hitting time, and the meeting time, and provide examples of how to utilize this theory to guide design decisions in routing protocols.

**Index Terms**—Mobility modeling, mobility trace-based model, spatial and temporal correlation in mobility.

## I. INTRODUCTION

MOBILE ad hoc networks (MANETs) are self-organized, infrastructureless networks that could potentially support many applications, such as vehicular networking (VANET) [4], wildlife tracking [19], and Internet provision to rural areas [16], to name a few. Mobility also enables message delivery in sparsely connected networks, generally known as delay tolerant networks (DTNs). As the devices are easily portable and the scenarios of deployment are inherently dynamic, *mobility* becomes one of the key characteristics in most of these networks. It has been shown that mobility impacts MANETs in multiple ways, such as network capacity [9], routing performance [1], and cluster maintenance [24]. In short, the evaluation of protocols and services for MANETs seems to be inseparable from the underlying mobility models. It is, thus, of crucial importance to

have suitable mobility models as the foundation for the study of ad hoc networks.

Ideally, a good mobility model should achieve a number of goals: (i) capturing *realistic* mobility patterns of scenarios in which one wants to eventually operate the network; (ii) being *mathematically tractable*; this is very important to allow researchers to derive performance bounds and understand the limitations of various protocols under the given scenario, as in [6], [9], [30], [31]; (iii) being *flexible* enough to provide qualitatively and quantitatively different mobility characteristics by changing some parameters of the model, yet in a repeatable and scalable manner.

Most existing mobility models excel in one or, less often, two aspects of the above requirements, but none satisfies all of them at the same time. Our goal in this paper is, on one hand, to improve the existing *random mobility models* (e.g., random walk, random direction, etc.) and *synthetic mobility models* (e.g., [11], [12], [17]) on the front of *realism*, by considering empirically observed mobility characteristics from the traces [14]. On the other hand, the construction of the model should be simple enough to allow in-depth *theoretical analysis*, and *flexible* enough to have wider applicability than the *mobility traces* (which provide only a single snapshot of the underlying mobility process) and current *trace-based mobility models* [22], [23], [33] which focus mainly on matching mobility characteristics with a specific class of traces.

The main contribution of this paper is the proposal of a *time-variant community mobility model*, referred to as the TVC model. One salient characteristic in the TVC model is *location preference*. Another important characteristic is the *time-dependent, periodical behavior* of nodes. To our best knowledge, this is the first *synthetic mobility model* that captures non-homogeneous behavior in both *space* and *time*.

To establish the realism and flexibility of our TVC model, we show that one can match its two prominent properties, *location visiting preferences* and *periodical reappearance*, with *multiple WLAN traces* collected from various environments. More interestingly, although we motivate the TVC model with the observations made on WLAN traces, our model is generic enough to have wider applicability. We validate this claim by examples of matching our TVC model with two additional mobility traces: a vehicle mobility trace [36] and a human encounter trace [6]. In the latter case, we are even able to match our TVC model with some other mobility characteristics not explicitly incorporated in our model by its construction, namely the inter meeting time and encounter duration between different users/devices.

Finally, in addition to the improved realism, the TVC model can be mathematically treated to derive analytical expressions for important quantities of interest, such as the *average node degree*, the *hitting time* and the *meeting time*. These quantities are

Manuscript received November 25, 2007; revised August 05, 2008 and November 21, 2008; approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor B. Levine. First published June 19, 2009; current version published October 14, 2009. This work was supported by the National Science Foundation under NSF Grants 0724658, 0134650, CNS-0520017, 0435505, and 0832043, and by funding by Cisco.

W.-J. Hsu and A. Helmy are with the Department of Computer and Information Science and Engineering, University of Florida, Gainesville, FL 32611 USA (e-mail: wjhsu@ufl.edu; helmy@ufl.edu).

T. Spyropoulos is with the Computer Engineering and Networks Laboratory, ETH Zurich, 8092 Zurich, Switzerland (e-mail: spyropoulos@tik.ee.ethz.ch; tsyropou@gmail.com).

K. Psounis is with the Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089 USA (e-mail: kpsounis@usc.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TNET.2008.2011128

often fundamental to theoretically study issues such as routing performance, capacity, connectivity, etc. We show that our theoretical derivations are accurate through simulation cases with a wide range of parameter sets, and additionally provide examples of how our theory could be utilized in actual protocol design. To our best knowledge, this is the first synthetic mobility model proposed that matches with traces from multiple scenarios, and has also been theoretically treated to the extent presented in this paper. We make the code of the TVC model available at [40].

The rest of this paper is organized as follows. In Section II we discuss related work. Our *TVC model* is then introduced in Section III. In Section IV, we show how to generate realistic mobility scenarios matched with various traces. Then, in Section V, we present our theoretical framework and derive generic expressions of various quantities. Simulation validates the accuracy of these expressions in Section VI. Additionally, in Section VII, we motivate our theoretical framework further, by applying our analysis to performance predictions in protocol design. Finally, we conclude the paper in Section VIII.

## II. RELATED WORK

Mobility models have been long recognized as one of the fundamental components that impacts the performance of wireless ad hoc networks. A wide variety of mobility models are available in the research community (see [5] for a good survey). Among all mobility models, the popularity of *random mobility models* (e.g., random walk, random direction, and random waypoint) roots in its simplicity and mathematical tractability. A number of important properties for these models have been studied, such as the stationary nodal distribution [3], the hitting and meeting times [29], and the meeting duration [18]. These quantities in turn enable routing protocol analysis to produce performance bounds [30], [31]. However, *random mobility models* are based on over-simplified assumptions, and as has been shown recently and we will also show in the paper, the resulting mobility characteristics are very different from real-life scenarios. Hence, it is debatable whether the findings under these models will directly translate into performance in real-world implementations of MANETs.

More recently, an array of *synthetic mobility models* are proposed to improve the realism of the simple *random mobility models*. More complex rules are introduced to make the nodes follow a popularity distribution when selecting the next destination [12], stay on designated paths for movements [17], or move as a group [11]. These rules enrich the scenarios covered by the *synthetic mobility models*, but at the same time make theoretical treatment of these models difficult. In addition, most *synthetic mobility models* are still limited to i.i.d. models, and the mobility decisions are also independent of the current location of nodes and time of simulation.

A different approach to mobility modeling is by *empirical mobility trace collection*. Along this line, researchers have exploited existing wireless network infrastructure, such as wireless LANs (e.g., [2], [10], [25]) or cellular phone networks (e.g., [7]), to track user mobility by monitoring their locations. Such traces can be replayed as input mobility patterns for simulations of network protocols [13]. More recently, DTN-specific testbeds [4], [6], [19] aim at collecting encounter events between mobile nodes instead of the mobility patterns. Some initial efforts to mathematically analyze these traces can be found in [6], [20]. Yet, the size of the traces and the environments in which the

experiments are performed can not be adjusted at will by the researchers. To improve the flexibility of traces, the approach of *trace-based mobility models* have also been proposed [22], [23], [33]. These models discover the underlying mobility rules that lead to the observed properties (such as the duration of stay at locations, the arrival patterns, etc.) in the traces. Statistical analysis is then used to determine proper parameters of the model to match it with the particular trace.

The goal of this work is to combine the strengths of various approaches to mobility modeling and propose a *realistic, flexible, and mathematically tractable synthetic mobility model*. Our work is partly motivated by several prominent, common properties in multiple WLAN traces (e.g., traces available from [37], [38]) we observed in [14], based on which we construct the TVC model. This model extends the concept of communities proposed by us in [29] and also introduces time-dependent behavior. A preliminary version of the model has been presented in [15]. In this work we highlight the *flexibility* of the TVC model by matching the synthetic traces with two additional, qualitatively different traces to WLAN traces (i.e., vehicular and human encounter traces, in Section IV). We also present more generic theoretical results under the scenario with multiple communities (Section V), and display its applications on protocol performance prediction (Section VII).

We differentiate our work from other trace-based models [22], [23], [33] in several aspects. First, among all efforts of providing realistic mobility models, to our best knowledge, this is the first work to explicitly capture time-variant mobility characteristics. Although capturing time-dependent behavior is suggested in [22], it has not been incorporated in the particular paper. Second, while previous works emphasize the capability to truthfully recreate the mobility characteristics observed from the traces, we also strive to ensure at the same time the mathematical tractability of the model. Our motivation is to facilitate the application of our model for performance prediction of various communication protocols. Finally, most of the other trace-based models have not been shown as capable to match mobility characteristics of a diverse set of traces, since their focus is mostly on one particular trace or at most a single class of traces (e.g., WLAN trace). We go beyond that and reproduce matching mobility characteristics of several *qualitatively different* traces, including WLAN, vehicle, and human encounter traces.

As a final note, in [26], the authors assume the attraction of a community (i.e., a geographical area) to a mobile node is derived from the number of friends of this node currently residing in the community. In our paper we assume that the nodes make movement decisions independently of the others (nonetheless, node sharing the same community will exhibit mobility correlation, capturing the social feature indirectly). Mobility models with inter-node dependency require a solid understanding of the social network structure, which is an important area under development. We plan to work further in this direction in the future.

## III. TIME-VARIANT COMMUNITY MOBILITY MODEL

### A. Mobility Characteristics Observed in WLAN Traces

The main objective of this paper is to propose a mobility model that captures the important mobility characteristics observed in daily life. To better understand this mobility, we have conducted extensive analysis of a number of wireless LAN

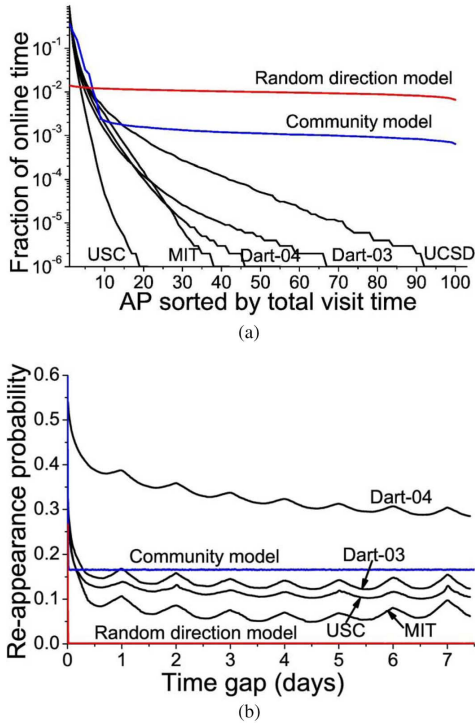


Fig. 1. Two important mobility features observed from WLAN traces. Labels of traces used: MIT: trace from [2], dart: trace from [10], UCSD: trace from [25], USC: trace from [14]. (a) Skewed location visiting preferences. (b) Periodical reappearance at the same location.

traces collected by several research groups (e.g., traces available at [38] or [37]). The reason for this choice is that WLAN traces log information regarding large numbers of nodes, and thus are reliable for statistical analysis. After analyzing a large number of traces, we have observed two common properties in all of them: (a) *skewed location visiting preferences* and (b) *time-dependent mobility behavior* [14].

More specifically, the *location visiting preference* refers to the percentage of time a node spends at a given access point (AP). We refer to the coverage area of an access point as a *location*. In Fig. 1(a), we draw the probability density function of the percentages of online time an average user spends at each location, ranking the locations from the most favorite place to the least for various traces. The distribution appears highly *skewed*; more than 95% of user's online time is spent at only top five APs. The *time-dependent mobility behavior* refers to the observation that nodes visit different locations, depending the time of the day. In Fig. 1(b) we plot the probability of a node reappearing at the same location at some time in the future, as a function of the elapsed time. It is clear that this probability displays some amount of periodicity, as the mobile nodes have stronger tendency to reappear at a previously visited location after a time gap of integer multiples of days. A slightly higher peak on the 7th day, suggesting a stronger weekly correlation in location visiting preferences, could also be observed in some curves (e.g., MIT).

Unfortunately, these two prominent realistic mobility characteristics are not captured by commonly used simple random models, as they do not possess any space or time dependent features. This is demonstrated in Fig. 1 by a straight line (uniform distribution) for the Random Direction model (the red curves).

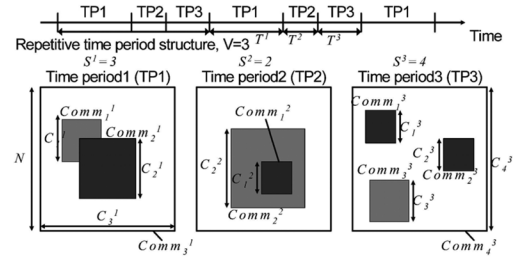


Fig. 2. Illustration of a generic scenario of time-variant mobility model, with three time periods and different numbers of communities in each time period.

The same could be obtained from Random Waypoint, Random walk, etc., or even more sophisticated models without spatial-temporal preferences (e.g., [11], [17]). There are some more recent models (e.g., [12], [23], [29], [33]) that aim at capturing spatial preference explicitly. As shown in Fig. 1(a) using the simple community model [29] (the blue curves), with appropriately assigned parameters this model is able to capture the *skewed location visiting preference*, to some extent. However, time-dependent behavior is not captured, and thus the *periodical reappearance* property cannot be reproduced, as shown by the flat curve in Fig. 1(b).

It is our goal to design a mobility model that successfully captures the *skewed location preference* and *time-dependency* mobility properties observed in the traces *in an analytically tractable fashion*. We believe that although the above observations are made based on WLAN traces, the two properties in question are indeed prevalent in real-life mobility. This belief is supported by typical daily activities of humans: most of us tend to spend most time at a handful of frequently visited locations, and a recurrent daily or weekly schedule is an inseparable part of our lives. It is essential to design a model that captures such spatial-temporal preferences of human mobility in many contexts.

## B. Construction of the Time-Variant Community Model

In this section, we present the design of our *time-variant community (TVC) mobility model*. We illustrate the model with an example in Fig. 2 and use this example to introduce the notations we use in the rest of the paper (see Table I).

First, to induce *skewed location visiting preferences*, we define some *communities* (or heavily-visited geographic areas). Take time period 1 (TP1) in Fig. 2 as an example, the communities are denoted as  $Comm_j^1$  and each of them is a square geographical area with edge length  $C_j^1$ .<sup>1</sup> A node visits these communities with different *probabilities* (details are given later) to capture its spatial preference in mobility. In the TVC model, the mobility process of a node consists of *epochs* in these communities. When the node chooses to have an *epoch* in community  $j$  (we say that *the node is in state  $j$*  during this epoch), it starts from the end point of the previous epoch within  $Comm_j^1$  and the epoch length (movement distance) is drawn from an exponential distribution with average  $\bar{L}_j$ , in the same order of the community edge length. The node then picks a random speed uniformly in  $[v_{min}, v_{max}]$ , and a direction (angle) uniformly in  $[0, 2\pi]$ , and performs a random direction movement within the

<sup>1</sup>For all parameters used in the paper, we follow the convention that the subscript of a quantity represents its community index, and the superscript represents the time period index.

TABLE I  
 PARAMETERS OF THE TIME-VARIANT COMMUNITY MOBILITY MODEL<sup>1</sup>

$N$	Edge length of simulation area
$V$	Number of time periods
$T^t$	Duration of $t$ -th time period
$S^t$	Number of communities in time period $t$
$C_j^t$	Edge length of community $j$ in time period $t^1$
$Comm_j^t$	The $j$ -th community during time period $t^1$
$p_{i,j}^t$	The probability to choose community $j$ when the previous community is $i$ , during time period $t^1$
$\pi_j^t$	Stationary probability of an epoch in community $j$ during time period $t^1$
$v_{min}, v_{max}, \bar{v}$	Minimum, maximum, and average speed
$D_{max,j}, \bar{D}_j$	Maximum and average pause time after each epoch
$L_j$	Average epoch length for community $j$
$P_{move,j}^t   P_{pause,j}^t$	Probability that a node is moving   pausing when being in community $j$ during period $t^1$
$P_j^t$	Fraction of time the node is in state $j$ ( $P_j^t = P_{move,j}^t + P_{pause,j}^t$ ) <sup>1</sup>
$K$	Transmission range of nodes
$A(a_j^t, b_k^t)$	The overlapped area between $Comm_j^t$ of node $a$ and $Comm_k^t$ of node $b$
$w^t$	A specific relationship between a target coordinate and the communities in time period $t$
$\Omega^t$	The set of all possible relationships between a target coordinate and the communities in time period $t$
$P_h(w^t)$	Unit-time hitting probability under the specific scenario $w^t$
$P_H(w^t)$	Hitting probability for a time period $t$ under specific scenario $w^t$
$P_m^t$	Unit-time meeting probability in time period $t$
$P_M^t$	Meeting probability for a time period $t$

chosen community with the chosen epoch length.<sup>2</sup> The first difference between the TVC model and the standard Random Direction model is hence the spatial preference and location-dependent behavior. Note that, a node can still roam around the whole simulation area during some epochs, by assigning an additional community that corresponds to the whole simulation field (e.g.  $Comm_1^t$ ). We refer to such epochs as *roaming epochs*.

We next explain how a node selects the next community for a sequence of epochs. At the completion of an epoch, the node remains stationary for a pause time uniformly chosen in  $[0, D_{max,j}]$ . Then, depending on its current state  $i$  and time period  $t$ , the node chooses the next epoch to be in community  $j$  with probability  $p_{i,j}^t$ . This community selection process is essentially a time-variant Markov chain that captures the spatial and temporal dependencies in nodal mobility and thus makes the community selection process in the TVC model non-i.i.d., an important feature absent in many synthetic mobility models even if they consider non-uniform mobility features. Now, if the end point of the previous epoch is in  $Comm_j^t$  (this can be the case when the node has two consecutive epochs in  $Comm_j^t$ , or  $Comm_j^t$  contains  $Comm_i^t$ ), the node starts the next epoch directly. If, on the other hand, the node is currently not in  $Comm_j^t$ , a *transitional epoch* is inserted to bridge the two epochs in disjoint communities. The node selects a random coordinate point in the next community, moves directly towards this point on the shortest straight path with a random speed drawn from  $[v_{min}, v_{max}]$ , and then continues with an epoch in the next community. Hence the movement trajectory of a node is always continuous in space.

<sup>2</sup>To avoid boundary effects, if the node hits the community boundary it is reinserted from the other end of the area (i.e., ‘torus’ boundaries). Note that we could also choose random waypoint or random walk models for the type of movement during each epoch.

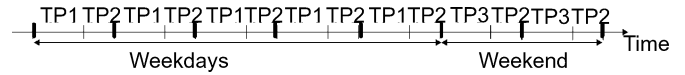


Fig. 3. An illustration of a simple weekly schedule, where we use time period 1 (TP1) to capture weekday working hour, TP2 to capture night time, and TP3 to capture weekend day time.

We next introduce the structure in time. To capture time-dependent behavior, one creates multiple *time periods* with different community and parameter settings. As an example, there are  $V = 3$  time periods with duration  $T^1$ ,  $T^2$ , and  $T^3$  in Fig. 2. These time periods follow a *periodic structure* (e.g., a simple recurrent structure in Fig. 2 or the weekly schedule in Fig. 3). This setup naturally captures the *temporal preferences* (e.g., go to work during the days and home during the nights) and *periodicity* in human mobility. On the time boundaries between time periods, each node continues with its ongoing epoch, and decides the next epoch according to the new parameter settings in the new time period when it finishes the current epoch.

As a final note, we choose to construct the TVC model with simple building blocks introduced above due to its amenability to theoretical analysis [29] and flexibility. To further explain the flexibility of our TVC model, we note that the number of communities in each time period (denoted as  $S^t$ ) can be different, and the communities can overlap (as in TP1 in Fig. 2) or contain each other (as in TP2 in Fig. 2). Finally, the time period structure, communities, and all other parameters could be assigned differently for *each node* to capture node-dependent mobility (e.g., people following different schedules, with different working places, etc.), while nodes can share some communities (i.e., the popular locations) as well. This construction allows for maximum flexibility when setting up the simulations for nodes with heterogeneous behaviors.<sup>3</sup>

The benefit of using simple building blocks will become evident in Section V. At the same time, we will show next that these choices do not compromise our model’s ability to accurately capture real life mobility scenarios.

#### IV. GENERATION OF MOBILITY SCENARIOS

The TVC model described in the previous section provides a general framework to model a wide range of mobility scenarios. In this section, our aim is to demonstrate the model’s flexibility and validate its realism by generating various synthetic traces from the model, with matching mobility characteristics to well-known, publicly-available traces (e.g., WLANs, VANET, and human encounter traces). However, it is important to note that the use of such a model is not merely to match it with any specific trace instance available; this is only done for validation and calibration purposes. Rather, the goal is to be able to reproduce a much larger range of realistic mobility instances than a single trace can provide.<sup>4</sup>

We first outline a general 3-step systematic process to construct specific mobility scenarios. Then, we demonstrate our success to generate matching mobility characteristics with three

<sup>3</sup>When necessary, we use a pair of parentheses to include the node ID for a particular parameter, e.g.,  $C_j^t(i)$  denotes the edge length of the  $j$ th community during time period  $t$  for node  $i$ .

<sup>4</sup>We have made our mobility trace generator available at [40]. The tool provides mobility traces in both ns-2 [39] compatible format and time-location (i.e.,  $(t, x, y)$ ) format.

qualitatively different traces. All the parameter values we use in this section are also available in [40].<sup>5</sup>

### Step 1) Determine the Structure in Space and Time

- (1.1 Number of communities) Each community in the TVC model corresponds to a location visited frequently by nodes (i.e., the most visited location in Fig. 1(a) corresponds to the most popular community in the model, and so on). The number of communities needed is thus determined by how closely one wants the mobility characteristics to match with the curves in Fig. 1(a). Due to the nature of skewed location visiting preference, in our experience, only two or three communities are needed to capture up to 85% of the user online time spent at the most popular locations. Such a simple spatial structure yields simple theoretical expressions. However, if one wants the model to capture more details (e.g., for detailed simulation), the user can instantiate as many communities as needed to explicitly represent the less visited locations.
- (1.2 Location of communities) If the map of the target environment is available, one should observe the map and identify the points of attraction in the given environment to assign the communities accordingly. The methods described in [22] could be applied to help choosing the “hot spots” on campus, by adding up the time users spend at each location on a 2-D map and identifying the peaks. Alternatively, if the map is not available, one can instantiate communities at *random locations*.<sup>6</sup> One way to do so is to simply divide the simulation area into equal-sized grid cells, and assign randomly chosen cells as communities.
- (1.3 Time period structure) From the curves in Fig. 1(b), one observes the *reappearance* periodicity and decides on the time period structure accordingly. Typically, human activities are bounded by daily and weekly schedules so a time period structure shown in Fig. 3 would suffice for most applications. If capturing finer behavior based on time-of-day is necessary, one could additionally split the day into time periods with different mobile node behavior. We illustrate this in our third case study, the human encounter trace.

### Step 2) Assign Other Parameters

After the space/time structure is determined, one has to determine the remaining parameters for each community and time period. This includes  $\pi_j^t$ ,  $\overline{D}_j^t$ , and  $\overline{L}_j^t$ , which represent the stationary probability (using simple Markov chain theory once we select proper  $p_{i,j}^t$ 's based on the transition probabilities between locations), the average pause time, and the average epoch length, respectively, at community  $j$  during time period  $t$ . These parameters can be determined by referring to the curves in Fig. 1. We give some general rules of how the parameters change the curves in Fig. 1. The detailed adjustments we make for each specific case studies will be discussed later.

- The average epoch length in each community,  $\overline{L}_j^t$ , should be at least in the same order as the edge length of the community,  $C_j^t$ . This is to ensure that the end point of the epoch becomes almost independent of its starting point, since the

<sup>5</sup>Due to space limitations, we cannot list all parameters in this paper.

<sup>6</sup>Concerning matching with the two mobility properties shown in Fig. 1, the actual locations of the communities do not make a difference.

mixing time of the corresponding process becomes quite small. (The motivation for this requirement is to keep the theoretical analysis tractable.)

- The average duration the node stays in community  $j$  is given by  $\pi_j^t(\overline{D}_j^t + \overline{L}_j^t/\bar{v})$ . The ratio between the durations the node stays in each community shapes the location visiting preference curve in Fig. 1(a).
- The highest peak of the reappearance probability curve (on the 7th day under the weekly schedule) in Fig. 1(b) is determined by the weighted average probability of the node appearing in the same community during the same type of time period. This value is  $\sum_{t=1}^V (T^t / \sum_{k=1}^V T^k) \sum_{j=1}^{S^t} (P_j^t)^2$ , where  $P_j^t$  denotes the fraction of time the node spends in community  $j$ .

### Step 3) Adjust User On-off Pattern (Optional)

The mobility trace generated by the TVC model is an “always-on” mobility trajectory (i.e., the mobile nodes are always present somewhere in the simulation field). However, in some situations some nodes might be absent occasionally. For example, in a WLAN setting, nodes (e.g. laptops) are often turned off when travelling from one location to another and the “off” time is often not negligible [14]. Thus one may need to make *optional* adjustments to turn nodes off in the generated trace, depending on the actual environment to match with. To address this we assign a probability  $P_{on,j}$  as the probability for the node to be “on” in community  $j$ . In two of the case studies we present (WLAN and vehicular trace), we utilize this feature as the nodes are not always-on in the actual traces.

Note that it is possible to automate part of the above community and parameter selection. This can be done by feeding the curves in Fig. 1 and the desired level of matching to a program that executes the above steps. Automatic generation of proper synthetic traces is a direction of our future work.

Next, we look into three specific case studies and apply the aforementioned procedure in each case, to display that the TVC model successfully produces synthetic mobility traces with matching characteristics observed in the real traces.

#### A. WLAN Traces

In the first example, we show that the TVC model can recreate the *location preferences* and *reappearance probability* curves observed in WLANs. We use the MIT WLAN trace (first presented in [2]) as the main example here.<sup>7</sup> We split the MIT trace into two halves and generate a matching synthetic trace with observed mobility characteristics from the first half (the training data set). We then compare our synthetic trace with the mobility characteristics of the second half (the validation data set). Note that, the mobility characteristics are similar across the two halves (shown by the two very close thick black curves in Fig. 4). We generate two synthetic traces with the TVC model, a *simplified* one and a *complex* one, to display its flexibility to have different levels of matching to the WLAN trace.

The simplified model (shown by thin black curves) uses only one community and two time periods (for the day time and night time), with parameters listed as *Model-1* in Table II. The simple

<sup>7</sup>We also achieve good matching with the USC [14] or the Dartmouth [10] traces, but do not show it here due to space limitations.

TABLE II  
 PARAMETERS FOR THE SCENARIOS IN THE SIMULATION

Model name	Description	$N$	$C_l^1$	$C_l^2$	$D_{max,l}$	$D_{max,r}$	$L_l$	$L_r$	$\pi_l^1$	$\pi_r^1$	$\pi_l^2$	$\pi_r^2$	$T^1$	$T^2$
Model 1	Match with the MIT trace	1000	100	100	100	50	80	520	0.714	0.286	0.8	0.2	5760	2880
Model 2	Highly attractive communities	1000	200	50	100	200	52	520	0.667	0.333	0.889	0.111	3000	2000
Model 3	Not attractive communities	1000	100	100	50	200	80	800	0.5	0.5	0.667	0.333	2000	1000
Model 4	Large-size communities	1000	200	250	50	100	200	800	0.7	0.3	0.889	0.111	2000	1000

We use the same movement speed for all node:  $v_{max} = 15$  and  $v_{min} = 5$  in all scenarios. In all cases we use two time periods and they are named as time period 1 and 2 for consistency. We only list the parameters for the simple models (*Model 1–4*) here. Please refer to [40] for the details of the generic models (*Model 5–7*).

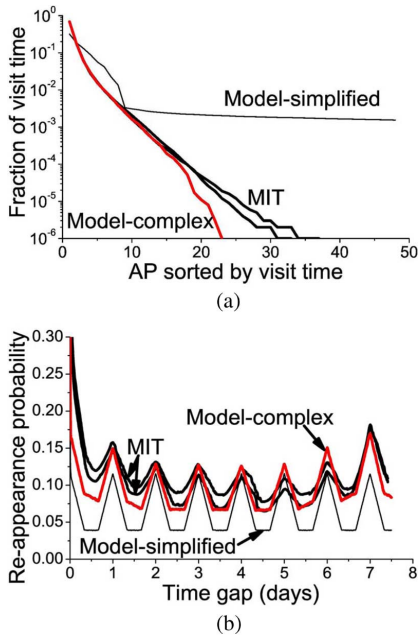


Fig. 4. Matching mobility characteristics of the synthetic traces to the MIT WLAN trace. (a) Skewed location visiting preferences. (b) Periodical reappearance at the same location.

model captures the major trends but still shows several noticeable differences: (a) the tail in the *model-simplified* curve in Fig. 4(a) is “flat” as opposed to the exponentially diminishing tail of the *MIT* curve. (b) the peaks in the *model-simplified* curve in Fig. 4(b) are of equal heights.

We can improve the matching between the synthetic trace and the real trace by adding complexity in both space and time, with the following detailed procedure. (STEP1): We divide the simulation area into 10-by-10 grid cells. Since we want to have a close match with the curve in Fig. 4(a), we assign randomly 15 of the cells as communities to each node (Intuitively, this number corresponds to the number of distinct access points that a person may connect to on a university campus over a period of one month.). For the time period structure we use the simple weekly structure shown in Fig. 3, allocating 8 hours for day time (TP1, TP3) and 16 hours for night time (TP2), as this trace is collected from a corporate environment. (STEP2 and STEP3): In the actual WLAN trace the nodes are “on” only for a low percentage of time. We capture this phenomenon with an additional parameter,  $P_{on,j}^t$ , the probability the node is “on” in state  $j$ . In WLAN, the nodes are typically “on” (i.e., appear at the current location) when they are not moving. Under this on-off pattern,  $P_{on,j}^t = \bar{D}_j^t / (\bar{D}_j^t + \bar{L}_j^t / \bar{v})$ . We then consider the on-off pattern and parameter assignment jointly.

(1) We first assign the same  $\bar{D}_j^t, \bar{L}_j^t, P_{on,j}^t$  to all communities, then assign  $\pi_j^t$  with a value equal to the fraction of time spent at the  $j$ th location in Fig. 1(a).<sup>8</sup> This assignment strategy makes the node “on” for the same amount of time in each community during each visit, and the total time in each community (and hence the observed location visiting preference curve) is therefore determined by the value of  $\pi_j^t$ . (2) Due to the on-off pattern, the peak value in the reappearance probability curve becomes  $\sum_{t=1}^V (T^t / \sum_{k=1}^V T^k) \sum_{j=1}^{S_t} (P_j^t)^2 (P_{on,j}^t)^2$ . To shape the reappearance probabilities, we adjust the  $\bar{D}_j^t$  values, which, in turn, adjust the values of  $P_{on,j}^t$  and set the reappearance probabilities to the desirable values to match with the curve in Fig. 1(b). Note that by adjusting the  $\bar{D}_j^t$  values in a consistent manner among all communities we do not change the location visiting probability curve that has already been matched in the previous step.

As it is evident from the red curves with labels *Model-complex* in Fig. 4, this model yields synthetic traces whose characteristics match very closely with those of the MIT trace.

## B. Vehicle Mobility Traces

In this example we display that *skewed location visiting preferences* and *periodical reappearance* are also prominent mobility properties in vehicle mobility traces. We obtain a vehicle movement trace from [36], a website that tracks participating taxis in the greater San Francisco area. We process a 40-day trace obtained between Sep. 22, 2006 and Nov. 1, 2006 for 549 taxis to obtain their mobility characteristics. The results are shown in Fig. 5 with the label *Vehicle-trace*. It is interesting that the trend of vehicular movements is very similar to that of WLAN users in terms of these two properties.

We use 30 communities and the weekly time schedule in (STEP1). We need more communities for this trace as the taxis are more mobile and visit more places than people on university campuses. From the actual trace, we discover that the taxis are offline (i.e., not reporting their locations) when not in operation. Hence we assume that the nodes are “on” only when they are moving. The pause times between epochs are considered as breaks in taxi operation. Therefore in (STEP3),  $P_{on,j}^t = (\bar{L}_j^t / \bar{v}) / (\bar{D}_j^t + \bar{L}_j^t / \bar{v})$ , and we adjust the parameters in a similar way as described in the previous section. The curves in Fig. 5 with label *Model* match with the curves with *Vehicle-trace* label well. As a final note, although vehicular movements are generally constrained by streets and our TVC model does not capture such microscopic behaviors, designated paths and other constraints could still be added in the model’s map (for vehicular

<sup>8</sup>One simple way to achieve this is to assign  $p_{i,j}^t = \pi_j^t, \forall i$ . Of course, one could make proper adjustments to  $p_{i,j}^t$  based on the observed transition probabilities in the trace if this information is available.

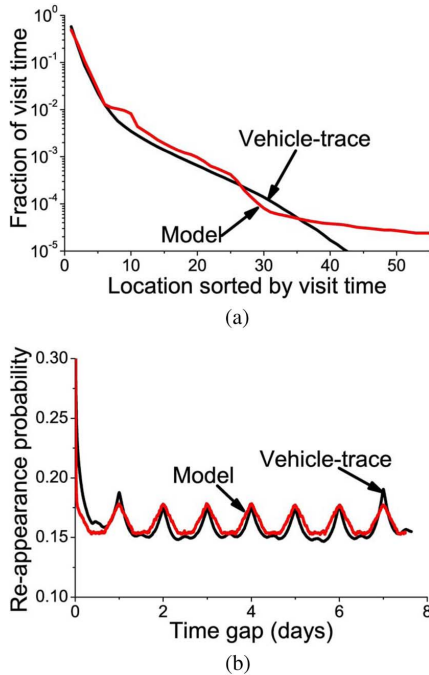


Fig. 5. Matching mobility characteristics of the synthetic trace to the vehicle mobility trace. (a) Location visiting preferences. (b) Periodical reappearance.

or human mobility) without losing its basic properties. We defer this for future work.

### C. Human Encounter Traces

In this example, we show that the TVC model is generic enough to mimic the encounter properties of mobile human networks observed in an experiment performed at INFOCOM 2005 [6]. In this experiment, wireless devices were distributed to 41 participants at the conference to log encounters between nodes (i.e., coming within Bluetooth communication range) as they moved around the premises of the conference area. The inter-meeting time and the encounter duration distributions of all 820 pairs of users obtained from this trace are shown in Fig. 6 with label *Cambridge-INFOCOM-trace*.

To mimic such behaviors using our TVC model, we observe the conference schedule at INFOCOM, and set up a daily recurrent schedule with five different types of time periods (STEP1): technical sessions, coffee breaks, breakfast/lunch time, evening, and late night (see [40] for the detailed parameters). For each time period we set up communities as the conference rooms, the dining room, etc. We also generate a community that is far away from the rest of the communities for each node and make the node sometimes isolated in this community to capture the behavior of patrons skipping part of the conference. In (STEP2), we use the theory presented in Section V to adjust the parameters and shape the inter-meeting time and encounter duration curves. For example, a stronger tendency for nodes to choose roaming epochs (setting larger  $\pi_r^t$ ) would increase the meeting probability (see, e.g., (18)), hence reducing inter-meeting times. Finally, since the devices used to collect the encounter traces are always-on, we do not apply any changes to the synthetic trace (STEP3). We randomly generate 820 pairs of users and show their corresponding distributions of the inter-meeting time and

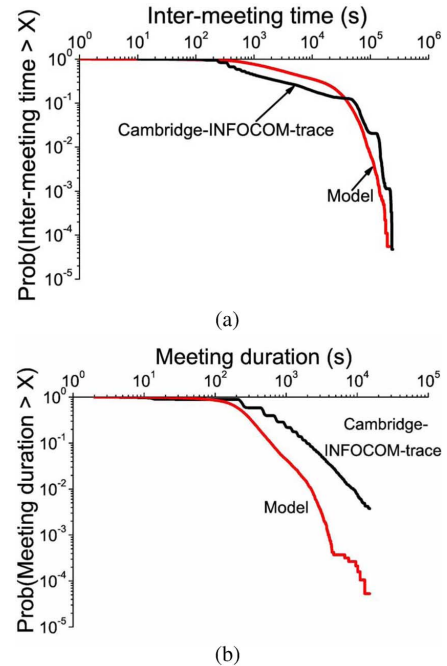


Fig. 6. Matching *inter-meeting time* and *encounter duration* distributions with the encounter trace. (a) Inter-meeting time. (b) Encounter duration.

the encounter duration in Fig. 6 with label *Model*. It is clear that our TVC model has the capability to reproduce the observed distributions, even if it is not constructed explicitly to do so. This displays its success in capturing the decisive factors of typical human mobility.

It is clear from the cases studied here that the TVC model is flexible to capture mobility characteristics from various environments well. In addition, with the respective configuration, it is possible to generate synthetic traces with much larger scale (i.e., more nodes) than the empirical ones while maintaining the same mobility characteristics. It is also possible to generate multiple instances of the synthetic traces with the same mobility characteristics to complement the original, empirically collected trace.

## V. THEORETICAL ANALYSIS OF THE TVC MODEL

So far, we have established the flexibility of the TVC model in terms of its ability to reproduce the properties observed in qualitatively different mobility traces. Yet, one of the biggest advantages of our model is that, in addition to the realism, it is also analytically tractable with respect to some important quantities which determine protocol performance. In the rest of this paper, we focus on demonstrating this last point.

We start here by deriving the theoretic expressions of various properties of the proposed mobility model assuming the nodes are always “on”. The properties of interest are defined below.

- The *average node degree* is the average number of nodes residing within the communication range of a given node. This quantity has major implication on the success rate of various tasks (e.g. geographic routing [28]) in mobile ad hoc networks.
- The *hitting time* is the time it takes a node, starting from the stationary distribution, to move within transmission range of a fixed, randomly chosen target coordinate in the field.
- The *meeting time* is the time until two mobile nodes, both starting from the stationary distribution, move into the

transmission range of each other. The *hitting and meeting times* are of interest due to their close relationship to the performance of DTN routing protocols.

We note that a preliminary version of the theoretical derivations appear under a special case of our TVC model (with one community and two time periods only) in [15]. Here, we generalize all derivations for any community and time-period structure. We start the theory derivations with a useful lemma that calculates the probability of a node to reside in a particular state.

*Lemma 5.1:* The probability that a node moves, pauses (after the completion of an epoch) in state  $j$ , or performs a transitional epoch at any given time instant during time period  $t$ , respectively, is

$$P_{move,j}^t = \pi_j^t \left( \overline{L}_j^t / \overline{v}_j^t \right) / \Psi, \quad (1)$$

$$P_{pause,j}^t = \pi_j^t \overline{D}_j^t / \Psi, \quad (2)$$

$$P_{tr}^t = \sum_{k=1}^{S^t} \pi_k^t \sum_{\forall n} p_{k,n}^t \overline{L}_{tr(k,n)} / \overline{v}_k^t / \Psi, \quad (3)$$

where  $\Psi = \sum_{k=1}^{S^t} \pi_k^t (\overline{L}_k^t / \overline{v}_k^t + \overline{D}_k^t + \sum_{\forall n} p_{k,n}^t \overline{L}_{tr(k,n)} / \overline{v}_k^t)$  and  $\overline{L}_{tr(k,n)}$  is the average length of a transitional epoch from community  $k$  to community  $n$ .

*Proof:* The probability for a node to be in state  $j$  ( $\pi_j^t$ ) can be easily derived with Markov chain theory from the state transition probabilities ( $p_{i,j}^t$ ). The above result follows from the ratio of the average durations of the moving part ( $\overline{L}_j^t / \overline{v}_j^t$ ) and the pause part ( $\overline{D}_j^t$ ) of regular epochs, and the transitional epochs ( $\overline{L}_{tr(k,n)} / \overline{v}_k^t$ ), weighted by the probabilities of the states. The expected length of the transitional epochs,  $\overline{L}_{tr(k,n)}$ , can be calculated as follows. Note that if community  $n$  contains community  $k$ , no transitional epoch is needed (i.e.,  $L_{tr(k,n)} = 0$ ). The transitional epoch is thus needed for a roaming node to go back to a smaller community, and as the previous roaming epoch ends at a random location in the whole simulation field, by symmetry, the expected length of the transitional epoch is the average length to move to the center of the simulation field from a random point in the field. Numerical analysis concludes  $\overline{L}_{tr} = 0.3826N$  in this case. ■

Note that the above stationary probabilities can be calculated for each time period and node separately. We use  $P_j^t(i)$  to denote the probability that node  $i$  is in state  $j$  during time period  $t$  (i.e.,  $P_j^t(i) = P_{move,j}^t(i) + P_{pause,j}^t(i)$ ).

#### A. Derivation of the Average Node Degree

The average node degree of a node is defined as the expected number of nodes falling within its communication range. Each node contributes to the average node degree independently, as nodes make independent movement decisions.

*Lemma 5.2:* Consider a pair of nodes,  $a$  and  $b$ . Assume further that, in time period  $t$ , community  $j$  of node  $a$  and community  $k$  of node  $b$  overlap with each other for an area  $A(a_j^t, b_k^t)$ . Then, the contribution of node  $b$  to the average node degree of node  $a$ , when  $a$  resides in its  $j$ th community and  $b$  resides in its  $k$ th community, is given by

$$\frac{\pi K^2}{C_j^{t^2}(a)} \frac{A(a_j^t, b_k^t)}{C_k^{t^2}(b)} \quad (4)$$

where  $K$  is the communication range of the nodes.

*Proof:* Since nodes follow random direction movement in each epoch, they appear at any point within the community equally likely. The probability for node  $b$  to fall in the  $j$ th community of node  $a$  is simply the ratio of the overlapped area over the size of the  $k$ th community of node  $b$ . Node  $a$  covers any given point in its community equal-likely, hence given node  $b$  is in the overlapped area, it is within the communication range of node  $a$  with probability  $\pi K^2 / C_j^{t^2}(a)$ . ■

Following the same principle in Lemma 5.2, we include all community pairs and arrive at the following Theorem.

*Theorem 5.3:* The average node degree of a given node  $a$  is

$$\sum_{\forall \text{Comm}_j^t(a)} P_j^t(a) \sum_{\forall b} \sum_{\forall \text{Comm}_k^t(b)} P_j^t(b) \frac{\pi K^2}{C_j^{t^2}(a)} \frac{A(a_j^t, b_k^t)}{C_k^{t^2}(b)}. \quad (5)$$

*Proof:* Eq. (5) is simply a weighted average of the node degree of node  $a$  conditioning on its states. For each state with probability  $P_j^t(a)$ , the expected node degree is a sum over all other nodes' probability of being within the communication range of node  $a$ , again conditioning on all possible states. Transitional epochs are treated the same way as roaming epochs here. That is, when considering a node in the transitional state with probability  $P_{tr}^t$ , it has equivalent contribution to the node degree as when it is in the roaming state (i.e., the node appears uniformly in the simulation field during transitional epochs since the it moves from anywhere in the simulation field back to the local community.). Hence, with probability  $P_{tr}^t + P_{roam}^t$ ,<sup>9</sup> the node has an effective community size of the simulation field,  $N$ . ■

*Corollary 5.4:* In the special case when all nodes choose their communities uniformly at random among the simulation field, (5) degenerates to  $\sum_{\forall b} (\pi K^2 / N^2)$ .

*Proof:* This result follows from the fact that a randomly chosen community is anywhere in the field equally likely. ■

#### B. Derivation of the Hitting Time

In the calculation of the average node degree, the dependence between consecutive epochs did not affect the derivation. In fact, only the stationary occupancy probabilities  $\pi_j^t$  and  $P_j^t$  (i.e. the probability of being found in community  $j$ ) are needed, since we were looking only at a random *snapshot* of the model. In the case of hitting and meeting times, we are interested in counting the number of epochs until a given target coordinate is found ("hit"). Our approach is to try to calculate the "hit probability" for a given epoch, and then count the number of such epochs needed on average until the destination point is hit. If these probabilities were independent, then one could use a simple geometric distribution to derive the result. However, 1) consecutive epochs are strongly related, as the ending point of one epoch is, naturally, the beginning point of the next. This introduces a *seeming* dependency between the hit probabilities of consecutive epochs. What is more, 2) the *transition* between communities (and epochs performed in each) are governed by the TVC model's Markov chain and the respective community transition probabilities  $p_{i,j}^t$ . Thus, looking only at the stationary probabilities for "choosing" the next community  $j$  (as in the previous section) no longer suffices. Finally, 3) the *transitional epochs* themselves introduce further complications, as they cannot, in this case, be handled as regular in-community or even roaming epochs.

<sup>9</sup>If the node has no roaming state in this time period, then we consider only  $P_{tr}^t$ .



The above three observations introduce dependencies that, at first glance, complicate our task. In the following, we will show how these dependencies can be “washed out” under a (minimally restricting) set of assumptions, and that stationary probabilities still suffice to derive a simple formula for the respective hitting time that holds *in the limit*. The basis of our argument is found in the proof of Lemma 5.7, upon which the rest of results in this section depend. (In a nutshell, the fast mixing of the mobility process takes care of (1), the large number of epochs required to hit a target takes care of (2) in the limit, and the dominance of local and roaming epochs over transitional epochs takes care of (3).) In Section VI, we show that the accuracy of our theory is not compromised by these assumptions in most practical scenarios considered.

The sketch of the derivation of the hitting time is as follows: (i) We first condition on the relative location of the target coordinate with respect to a node’s communities (Lemma 5.5). We identify all possible sub-cases (i.e. whether the target is inside or outside one or more of the node’s communities). Using simple geometric arguments, we calculate the probability of each of these sub-cases (Lemma 5.6) and take the weighted average of the respective hitting times (to be calculated per sub-case). (ii) For a given sub-case, we derive the expected number of epochs (and the expected number of time units) until the target is found (Lemma 5.7). (iii) Finally, we introduce the time-period factor, and account for the total number of time periods needed to hit the target (Theorem 5.9).

The most influential factor for the hitting time is whether the target coordinate is chosen inside the node’s communities. We denote the possible relationships between the target location and the set up of communities during time period  $t$  as the set  $\Omega^t$ . Note that the cardinality of set  $\Omega^t$  is at most  $2^{S^t}$  (i.e. for each of the  $S^t$  communities, the target coordinate is either in or out of it).

*Lemma 5.5:* By the law of total probability, the average hitting time can be written as

$$HT = \sum_{w^1 \in \Omega^1, \dots, w^V \in \Omega^V} P(w^1, \dots, w^V) HT(w^1, \dots, w^V) \quad (6)$$

where  $w^1, w^2, \dots, w^V$  denote one particular relationship (i.e. a combination of  $\{out, in\}^{S^t}$ ) between the target coordinate and the community set up during time period  $1, 2, \dots, V$ , respectively. Functions  $P(\cdot)$  and  $HT(\cdot)$  denote the corresponding probability for this scenario and the conditional hitting time under this scenario, respectively. Note that each sub-case  $\{w^1, w^2, \dots, w^V\}$  is disjoint from all other sub-cases.

To evaluate (6), we need to calculate  $P(w^1, \dots, w^V)$  and  $HT(w^1, \dots, w^V)$  for each possible sub-case  $(w^1, \dots, w^V)$ .

*Lemma 5.6:* If the target coordinate is chosen independent of the communities and the communities in each time period are chosen independently from other periods, then

$$P(w^1, \dots, w^V) = \prod_{t=1}^V P(w^t) \quad (7)$$

where  $P(w^t) = A(w^t)/N^2$ , i.e., the probability of a sub-case  $w^t$  is proportional to the area  $A(w^t)$  that corresponds to the specific scenario  $w^t$ , which is a series of conditions of the following type:  $(\{target \in comm_1^t\}, \{target \notin comm_2^t\}, \dots, \{target \in comm_S^t\})$ .

*Proof:* The result follows from simple geometric arguments. ■

The first step for calculating  $HT(w^1, \dots, w^V)$  is to derive the unit-time hitting probability in time period  $t$  under target coordinate-community relationship  $w^t$ , denoted as  $P_h^t(w^t)$ .

*Lemma 5.7:* For a given time period  $t$  and a specific scenario  $w^t$ ,

$$P_h^t(w^t) = \sum_{j=1}^{S^t} I(\text{target} \in \text{Comm}_j^t | w^t) P_{move,j}^t 2K \bar{v}_j^t / C_j^{t^2}, \quad (8)$$

where  $I(\cdot)$  is the indicator function.

*Proof:* To calculate the expected hitting time, we first count the total number of epochs needed. Assume that  $N_e$  epochs are needed in total, and let us denote as epoch  $E_{k_m}(m)$  the  $m$ th epoch in sequence (that is occurring in community  $k_m$ ). Let further,  $P(k_1, k_2, \dots, k_m)$  denote the probability of the specific sequence of epochs occurring. Then, the probability that the target has not been found after  $n$  epochs is

$$P(N_e > n) = \sum_{k_1, k_2, \dots, k_n} P(k_1, k_2, \dots, k_n) \cdot P(E_{k_1}(1) = \text{miss}, E_{k_2}(2) = \text{miss}, \dots, E_{k_n}(n) = \text{miss}). \quad (9)$$

In order to simplify the above equation, we need to deal with the inherent dependencies introduced by the transition of epochs. First, since node movement is continuous, the end of one epoch  $E_j(m)$ , performed in community  $j$ , is the beginning of the next,  $E_j(m+1)$ , if performed in the same community.<sup>10</sup> Nevertheless, as explained in Section III-B, the expected “length” of an epoch  $\bar{L}_j^t$  performed in community  $j$  is in the order of the square root of the community size  $C_j^t$ . This is sufficient for the node to “mix” in the community after just one epoch [29]. Consequently, we can write

$$P(N_e > n) = \sum_{k_1, k_2, \dots, k_n} P(k_1, k_2, \dots, k_n) \cdot \prod_{i=1}^n P(E_{k_i} = \text{miss}). \quad (10)$$

An additional dependency arises from the transitions between communities and the calculation of term  $P(k_1, k_2, \dots, k_n)$ . If epoch  $E_{k_m}(m)$  is performed in community  $k_m$  the next epoch  $E_{k_{m+1}}(m+1)$  will be performed in community  $k_{m+1}$  with probability  $p_{k_m, k_{m+1}}$  (the transition probability in the Markov Chain governing the community transitions in the TVC model). Let us assume that out of  $N_e$  epochs needed, there are  $l_j$  epochs of type  $j$  (i.e. performed in community  $j$ ) in the mix. When  $N_e \rightarrow \infty$ , then  $l_i \rightarrow \pi_i N_e$ , that is, the total number of epochs in community  $i$  depends *only* on the stationary probability of community  $i$ ,  $\pi_i$ . Thus,

$$P(k_1, k_2, \dots, k_n) = \pi_{k_1} \cdot \pi_{k_2} \dots \pi_{k_n}. \quad (11)$$

Consequently, (10) becomes

$$P(N_e > n) = \prod_{i=1}^n \pi_{k_i} \cdot P(E_{k_i} = \text{miss}). \quad (12)$$

This implies that, *in the limit*,<sup>11</sup> the total number of epochs needed to hit the target can be approximated by a geometric distribution, where the “average” epoch has a hit probability of

$$\sum_i \pi_i \cdot P(E_i = \text{hit}). \quad (13)$$

<sup>10</sup>For the moment, we will ignore transitional epochs, and assume that all epochs are performed inside some community; we deal with transitional epochs later.

<sup>11</sup>In practice, the requirement is that a large number of epochs is needed on average until the target is hit. In the sparse networks we’re interested in, this is a reasonable assumption, and as we shall show in Section VI the achieved accuracy is indeed high.

As the final step, we need to calculate the probability of a given epoch in community  $j$  to hit the target. Instead of using this per epoch hit probability, we revert now to what we call the *unit-step* hit probabilities,  $P_h$ . The unit-step probability is the probability of encountering the target exactly within the next time-unit (rather than within the duration of a whole epoch). This discrete approximation provides an equivalent formulation to the above continuous case (see [29]), however it is more convenient to manipulate in the case of time-period boundaries and meeting times calculated later. (Note that this approximation is again only possible when the average epoch length is in the order of the respective community size, ensuring mixing after one epoch.)

Note that the hitting event can only occur when the node is physically moving inside the community where the target is located. Whether the target is located inside community  $j$  is denoted using the indicator function  $I(\text{target} \in \text{Comm}_j^t | w^t)$ . If the target is outside the community, then this probability of hit is zero.<sup>12</sup> If the target lies within community  $j$ , then when a node moves with average speed  $\bar{v}_j$ , on average it covers a new area of  $2K\bar{v}_j$  in unit time. Since a node following random direction movements visits the area it moves about with equal probability, and the target coordinate is chosen at random, it falls in this newly covered area with probability  $2K\bar{v}_j/C_j^t$  [29]. Hence the contribution to the unit-time hitting probability by movements made in state  $j$  is  $P_{\text{move},j}^t 2K\bar{v}_j/C_j^t$ . Thus, in (13),  $\pi_j$  is replaced by  $I(\text{target} \in \text{Comm}_j^t | w^t) P_{\text{move},j}^t$  in the unit-step case, and  $P(E_i = \text{hit})$  by  $2K\bar{v}_j/C_j^t$ .

As a final remark, the contribution of transitional epochs to the unit-time hitting probability is not equivalent to other epochs (due to the dependency of end-points on local communities, which introduces bias after communities have been chosen). Nevertheless, in a normal mobility scenario, we expect a node to spend the majority of its time within one of the communities rather than in transitional epochs. Specifically, we will assume that community transition probabilities exhibit a strong *positive correlation*, that is, if a node resides in community  $j$ , it has a higher probability of staying within this community for the next epoch, rather than leaving. In this case, the total contribution of transitional epochs is small, and can be safely ignored in order to not complicate our analysis. The above is a reasonable assumption for many target scenarios we can imagine; simulation results show further that the time a node resides in transitional state is indeed less than 10% in the scenarios considered, not significantly affecting the accuracy of the above expression. ■

Given the aforementioned assumptions about unit-step hitting probabilities, the corollary below follows.

*Corollary 5.8:* The probability for at least one hitting event to occur during time period  $t$  under scenario  $w^t$  is

$$P_H^t(w^t) = 1 - (1 - P_h^t(w^t))^{T^t}. \quad (14)$$

Finally, using the law of total probability, we derive the conditional hitting time under a specific target-community relationship,  $HT(w^1, \dots, w^V)$ .

*Theorem 5.9:*

$$HT(w^1, \dots, w^V) = \sum_{t=1}^V HT(w^1, \dots, w^V | \text{first hit in period } t) \cdot P(w^1, \dots, w^V, \text{first hit in period } t), \quad (15)$$

<sup>12</sup>We neglect the small probability that the target is chosen out of the community but close to it, and make the contributions from epochs in state  $j$  zero if the chosen target coordinate is not in community  $j$ .

where the probability for the first hitting event to happen in time period  $t$  is

$$P(w^1, \dots, w^V, \text{first hit in period } t) = \frac{\prod_{i=1}^{t-1} (1 - P_H^i(w^i)) \cdot P_H^t(w^t)}{P}, \quad (16)$$

and the hitting time under this specific condition is

$$HT(w^1, \dots, w^V | \text{first hit in period } t) = \sum_{i=1}^V T^i \cdot \left( \frac{1}{P} - 1 \right) + \sum_{i=1}^{t-1} T^i + \frac{1}{P_h^t(w^t)}, \quad (17)$$

where  $P = 1 - \prod_{t=1}^V (1 - P_H^t(w^t))$  is the hitting probability for one full cycle of time periods.

*Proof:* Eq. (16) holds as each cycle of time periods follows the same repetitive structure, and for the first hitting event to occur in time period  $t$  it must not occur in time period  $1, \dots, (t-1)$ . The first term in (17) corresponds to the expected duration of full time period cycles until the hitting event occurs. Since for each cycle the success probability of hitting the target is  $P$ , in expectation it takes  $1/P$  cycles to hit the target, and there are  $1/P - 1$  full cycles. The second term in (17) is the sum of duration of time periods before the time period  $t$  in which the hitting event occurs in the last cycle. Finally, the third term is the fraction of the last time period before the hitting event occurs. Note that the last part is an approximation which holds if the time periods we consider are much longer than unit-time. ■

### C. Derivation of the Meeting Time

The procedures of the derivation of the meeting time is similar to that of the hitting time detailed in the last section. In short, we derive the unit-step (or unit-time) meeting probability,  $P_m$ , and the meeting probability for each type of time period,  $P_M$ , and put them together to get the overall meeting time in a similar fashion as in Theorem 5.9.

Similar to Lemma 5.7, we add up the contributions to the meeting probability from all community pairs from node  $a$  and  $b$  in the following Lemma.

*Lemma 5.10:* Let community  $j$  of node  $a$  and community  $k$  of node  $b$  overlap with each other for an area  $A(a_j^t, b_k^t)$  in time period  $t$ . Then, the conditional unit-time meeting probability in time period  $t$  when node  $a$  and  $b$  are in its community  $j$  and  $k$ , respectively, is

$$P_m^t(a_j^t, b_k^t) = P_{\text{move},j}^t(a) P_{\text{move},k}^t(b) \hat{v} \frac{2K\bar{v}}{A(a_j^t, b_k^t)} \frac{A(a_j^t, b_k^t)}{C_j^{t^2}(a)} \frac{A(a_j^t, b_k^t)}{C_k^{t^2}(b)} + P_{\text{move},j}^t(a) P_{\text{stop},k}^t(b) \frac{2K\bar{v}}{A(a_j^t, b_k^t)} \frac{A(a_j^t, b_k^t)}{C_j^{t^2}(a)} \frac{A(a_j^t, b_k^t)}{C_k^{t^2}(b)} + P_{\text{stop},j}^t(a) P_{\text{move},k}^t(b) \frac{2K\bar{v}}{A(a_j^t, b_k^t)} \frac{A(a_j^t, b_k^t)}{C_j^{t^2}(a)} \frac{A(a_j^t, b_k^t)}{C_k^{t^2}(b)}. \quad (18)$$

*Proof:* Equation (18) consists of two parts:

(I) Both of the nodes are moving within the overlapped area. This adds the first term in (18) to the meeting probability. The two ratios,  $A(a_j^t, b_k^t)/C_j^{t^2}(a)$  and  $A(a_j^t, b_k^t)/C_k^{t^2}(b)$ , capture the probabilities that the nodes are in the overlapped area of the communities. The contribution to the unit-time meeting probability is the product of probabilities of both nodes moving within

the overlapped area and the term  $2K\bar{v}/A(a_j^t, b_k^t)$ , which reflects the covered area in unit time. We use the fact that when both nodes move according to the random direction model, one can calculate the effective (extra) area covered by assuming that one node is static, and the other is moving with a (higher) *relative speed*. This is captured with the multiplicative factor  $\hat{v}$  [29].

(II) One node is moving in the overlapped area, and the other one pauses within the area. This adds the remaining two terms in (18) to the unit-time meeting probability. These terms follow similar rationale as the previous one, but now only one node is moving. The second term corresponds to the case when node  $a$  moves (and  $b$  is static), and the third term corresponds to the case when node  $b$  moves (and  $a$  is static).

The derivation of the unit-time meeting probability between nodes  $a$  and  $b$  for time period  $t$  includes all possible scenarios of community overlap. If node  $a$  has  $S^t(a)$  communities and node  $b$  has  $S^t(b)$  communities, there can be at most  $S^t(a)S^t(b)$  community-overlapping scenarios in time period  $t$ . For similar reason detailed in the proof of Lemma 5.7, we neglect the contribution of transitional epochs to the unit-time meeting probability. ■

Note that (18) is the general form of Equation (13) and (14) in [15]. If we assume perfect overlap and a single community from both nodes, we arrive at (14). If we assume no overlap, we result in (13). Also note in the general expressions presented in this paper, the whole simulation area is also considered as a community. Therefore we do not have to include a separate term to capture the roaming epochs.

*Corollary 5.11:* The probability for at least one meeting event to occur during time period  $t$  is

$$P_M^t = 1 - \sum_{\forall(j,k)} \left\{ P_{ov}(a_j^t, b_k^t) \cdot (1 - P_M^t(a_j^t, b_k^t))^{T^t} \right\} \quad (19)$$

where  $P_{ov}(a_j^t, b_k^t)$  is the probability that the community  $j$  of node  $a$  overlaps with community  $k$  of node  $b$ . This quantity is simply 1 if the communities have fixed assignments and  $A(a_j^t, b_k^t) \neq 0$ . If the communities are chosen randomly, this probability can be derived by Lemma 4.5 in [15]. Due to space constraint, the Lemma is not reproduced here.

Finally, similarly to Theorem 5.9, the expected meeting time can be calculated using the Lemmas in this section.

*Theorem 5.12:* The expected meeting time is

$$MT = \sum_{t=1}^V MT(\text{meet in period } t)P(\text{meet in period } t). \quad (20)$$

The quantities in the above equation are calculated by

$$P(\text{meet in period } t) = \frac{\prod_{i=1}^{t-1} (1 - P_M^i) \cdot P_M^t}{Q}, \quad (21)$$

$$MT(\text{meet in period } t) = \sum_{i=1}^V T^i \cdot \left( \frac{1}{Q} - 1 \right) + \sum_{i=1}^{t-1} T^i + \frac{1}{P_M^t}, \quad (22)$$

where  $Q = 1 - \prod_{i=1}^V (1 - P_M^i)$  is the meeting probability for one full cycle of time periods.

*Proof:* The proof is parallel to that of Theorem 5.9 and is omitted due to space limitations (see [40] for details). ■

As a final note, we can easily modify the above theory to account for potential “off” periods (e.g. by introducing a per step or per epoch “off” probability, and a respective multiplicative factor). Due to space limitations, we do not include here these modifications.

## VI. VALIDATION OF THE THEORY WITH SIMULATIONS

In this section, we compare the theoretical derivations of the previous section against the corresponding simulation results, for various scenarios and parameter settings to establish the accuracy of the theoretical framework. Due to space limitations, we can only show some examples of the simulation results we have. More complete results can be found at [40].

We summarize the parameters for the tested scenarios in Table II. We use two different setup of the TVC model for the simulation cases. The parameters listed in Table II are for the simple models (*Models 1–4*), where we have two time periods with two communities in each time period (one of the communities is the whole simulation field). We also construct more generic setup of the TVC model (*Models 5–7*, refer to [40] for their parameters). For the generic models, we use three communities and experiment with two ways of community placement: in a tiered fashion (as drawn in TP2 in Fig. 2), or in a random fashion. Our discrete-time simulator is written in C++. More details about the simulator, as well as the source code, can be found at [40].

### A. The Average Node Degree

For the average node degree, we create simulation scenarios with 50 nodes in the simulation area, and calculate the average node degree of each node by taking the time average across snapshots taken every second during the simulation, and then average across all nodes. All the simulation runs last for 60000 seconds in this subsection.

As we show in Corollary 5.4, when the communities are randomly chosen, the average node degree turns out to be the average number of nodes falling in the communication range of a given node, *as if all nodes are uniformly distributed*. Hence the average node degree does not depend on the exact choices of community setup (i.e. single, multiple, or multi-tier communities) or other parameters. In Fig. 7(a), we see the simulation curves follow the prediction of the theory well.

To make the scenario a bit more realistic, we simulate some more scenarios when the communities are fixed. Among the 50 nodes, we make 25 of them pick the community centered at (300, 300) and the other 25 pick the community centered at (700, 700). We simulate scenarios for all seven sets of parameters, and show some example results in Fig. 7(b). In the simulations, when the communication ranges are small as compared to the edge of the communities, the relative errors are low, indicating a good match between the theory and the simulation. However, as the communication range increases, the area covered by the communication disk becomes comparable to the size of the community and (4) is no longer accurate since the communication disk extends out of the overlapped area in most cases. That is the reason for the discrepancies between the theory and simulation. Besides *Model-3*, we observe at most 20% of error when the communication disk is less than 20% the size of the inner-most community, indicating that our theory is valid when the communication range is relatively small.

### B. The Hitting Time and the Meeting Time

We perform simulations for the hitting and the meeting times for 50,000 independent iterations for each scenario, and compare the average results with the theoretical values derived from

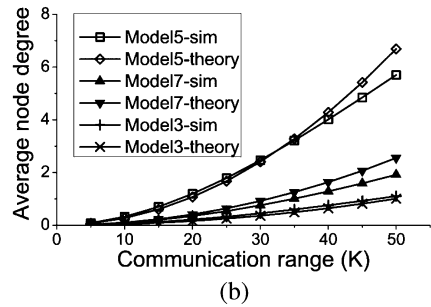
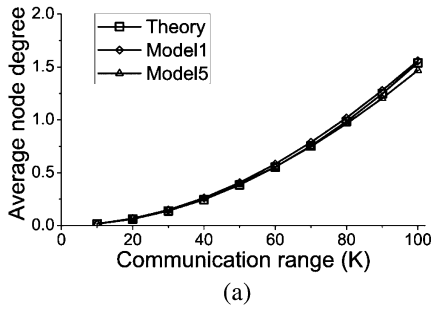


Fig. 7. Examples of simulation results (the average node degree). (a) Randomly placed community. (b) Fixed communities.

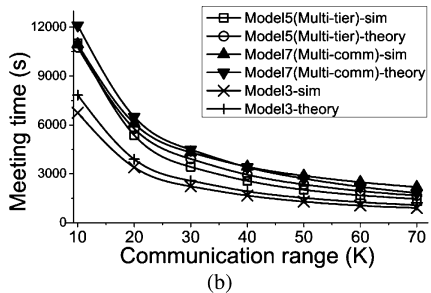
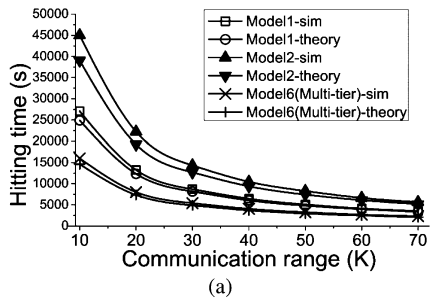


Fig. 8. Examples of simulation results. (a) Hitting time. (b) Meeting time.

the corresponding equations (i.e. (6) and (20)). To find out the hitting or the meeting time, we move the nodes in the simulator indefinitely until they hit the target or meet with each other, respectively.

Again we show some example results in Fig. 8. For all the scenarios (including the ones not listed here), the relative errors are within acceptable range. The absolute values for the error are within 15% for the hitting time and within 20% for the meeting time. For more than 70% of the tested scenarios, the error is below 10% (refer to [40] for other figures). These results display the accuracy of our theory under a wide range of parameter settings. The errors between the theoretical and

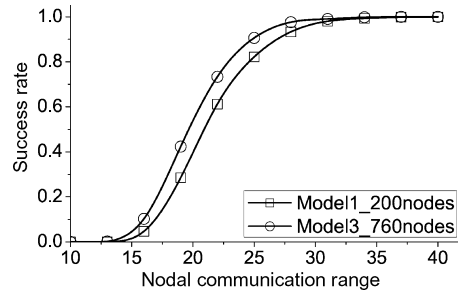


Fig. 9. Geographic routing success rate under different mobility parameter sets and node numbers.

simulation results are mainly due to some of the approximations we made in the various derivations, for example, the approximation of the hitting and meeting processes with discrete, unit-time Bernoulli trials. Furthermore, there exist some border effects—when a node is close to the border of a community, it could also “see” some other nodes outside of the community if its transmission range is large enough. However, we have chosen to ignore such occurrences to keep our analysis simpler. Nevertheless, as shown in the figures, the errors are always within acceptable ranges, justifying our simplifying assumptions.

### VII. USING THEORY FOR PERFORMANCE PREDICTION

Although the various theoretical quantities derived for the TVC model in Section V are interesting in their own merit, they are particularly useful in predicting protocol performance, which in turn can guide the decisions of system operation. We illustrate this point with two examples in this section.

#### A. Estimation of the Number of Nodes Needed for Geographic Routing

It has been shown in geographic routing that the average node degree determines the success rate of messages delivered [28]. Thus, using the results of Section V-A we can estimate the number of nodes (as a function of the average node degree) needed to achieve a target performance for geographic routing, for a given scenario.

We consider the same setup as in Section VI-A, where half of the nodes are assigned to a community centered at (300, 300) and the other half are assigned to another community centered at (700, 700). We are interested in routing messages across one of the communities, from coordinate (250, 250) to coordinate (350, 350) with simple geographic routing (i.e., greedy forwarding only, without face routing [21]). Using simulations we obtain the success rate of geographic routing under various communication ranges when 200 nodes move according to the mobility parameters of *Model-1* (Table II). Results are shown in Fig. 9 (each point is the percentage of success out of 2000 trials). If we assume the mobility model is different, say *Model-3*, we would like to know how many nodes we need to achieve similar performance. Using (5) we find that 760 nodes are needed to create a similar average node degree for *Model-3*. To validate this, we also simulate geographic routing for a scenario where 760 nodes follow *Model-3*. From Fig. 9, we see that similar success rates are achieved under the same transmission range, which confirms the accuracy of our analysis.

### B. Predicting Message Delivery Delay With Epidemic Routing

Epidemic routing is a simple and popular protocol that has been proposed for networks where nodal connectivity is intermittent (i.e., in Delay Tolerant Networks) [34]. It has been shown that message propagation under epidemic routing can be modeled with sufficient accuracy using a simple fluid-based SI (Susceptible-Infected) model [35]. Note that its performance has also been analyzed using Markov Chain [8] and Random Walk [30] models. The gist of the SI model is that the rate by which the number of “infected” nodes increases (“infected” nodes here are nodes who have received a copy of the message) can be approximated by the product of three quantities: the number of already infected nodes, the number of susceptible (not yet infected) nodes, and the pair-wise contact rate,  $\beta$  (assuming nodes meet independently—this contact rate is equivalent to the unit-step meeting probabilities calculated in (18)). Yet, in the TVC model (and often in real life) there are multiple groups of nodes with different communities, and thus different pair-wise contact rates that depend on the community setup. For this reason, we extend the basic SI model to a more general scenario.

For a case study, we use *Model-3* (Table II) for the mobility parameters. A total of  $M = 50$  nodes are divided into two groups of 25 nodes each. One group has its community centered at (300, 300) and the other at (700, 700). One packet starts from a randomly picked source node and spreads to all other nodes in the network. The propagation of the message can be described by the following equations:

$$\begin{cases} \frac{dI_1(t)}{dt} = \beta_{ov}I_1(t)S_1(t) + \beta_{no\_ov}I_2(t)S_1(t) \\ \frac{dI_2(t)}{dt} = \beta_{ov}I_2(t)S_2(t) + \beta_{no\_ov}I_1(t)S_2(t) \\ S_1(t) + I_1(t) = M/2 \\ S_2(t) + I_2(t) = M/2 \end{cases} \quad (23)$$

where  $S_x(t)$  and  $I_x(t)$  denote the number of susceptible and infected nodes at time  $t$  in group  $x$ , respectively. Parameters  $\beta_{ov}$  and  $\beta_{no\_ov}$  represent the pair-wise unit-time meeting probability when the communities are overlapped (i.e., for nodes in the same group) and not overlapped (i.e., nodes in different groups), respectively. We use (18) to obtain these quantities. This model is an extension from the standard SI model [35] and similar extensions can be made for more than two groups [32]. The first equation governs the change of infected nodes in the first group. Notice that the infection to susceptible nodes in the group ( $S_1(t)$ ) can come from the infected nodes in the same group ( $I_1(t)$ ) or the other group ( $I_2(t)$ ). We can solve the system of equations in (23) to get the evolution of the total infected nodes. In Fig. 10, the *theory* curve closely follows the *simulation* curve. This indicates that scenarios generated by our mobility model are still amenable to fluid model based mathematical analysis, despite the increased complexity introduced by the concept of communities. It also shows that results produced can be used by a system designer to predict how fast messages propagate in a given network environment. This might, for example, determine if extra nodes are needed in a wireless content distribution network to speed up message dissemination.

As a final note, in addition to epidemic routing, the theoretical results for the hitting and meeting times could be applied to predict the delay of various other DTN routing protocols (see e.g. [29], [30], [35]), for a large range of mobility scenarios that can be captured by the TVC model.

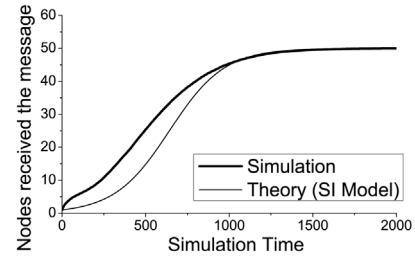


Fig. 10. Packet propagation with epidemic routing with *two* node groups with different communities.

### VIII. CONCLUSION AND FUTURE WORK

We have proposed a *time-variant community mobility model* for wireless mobile networks. Our model preserves common mobility characteristics, namely *skewed location visiting preferences* and *periodical reappearance* observed in empirical mobility traces. We have tuned the TVC model to match with the mobility characteristics of various traces (WLAN traces, a vehicle mobility trace, and an encounter trace of moving human beings), displaying its flexibility and generality. A mobility trace generator of our model is available at [40]. In addition to providing realistic mobility patterns, the TVC model can be mathematically analyzed to derive several quantities of interest: the *average node degree*, the *hitting time* and the *meeting time*. Through extensive simulations, we have verified the accuracy of our theory.

In the future we would like to further analyze the performance of various routing protocols (e.g., [30], [31]) under the time-variant community mobility model. We also would like to construct a systematic way to automatically generate the configuration files, such that the model parameters are set to capture the inter-node encounter properties we observe in various traces (for example, the Small World encounter patterns observed in WLAN traces [13]).

### ACKNOWLEDGMENT

The authors would like to thank the reviewers for their valuable comments, and M. Piorkowski for sharing archived data.

### REFERENCES

- [1] F. Bai, N. Sadagopan, and A. Helmy, “The IMPORTANT framework for analyzing the impact of mobility on performance of routing for ad hoc networks,” *AdHoc Netw. J.*, vol. 1, no. 4, pp. 383–403, Nov. 2003.
- [2] M. Balazinska and P. Castro, “Characterizing mobility and network usage in a corporate wireless local-area network,” in *Proc. MOBISYS 2003*, May 2003.
- [3] C. Bettstetter, G. Resta, and P. Santi, “The node distribution of the random waypoint mobility model for wireless ad hoc networks,” *IEEE Trans. Mobile Computing*, vol. 2, no. 3, pp. 257–269, Jul. 2003.
- [4] J. Burgess, B. Gallagher, D. Jensen, and B. Levine, “MaxProp: Routing for vehicle-based disruption-tolerant networks,” in *Proc. IEEE INFOCOM*, Apr. 2006.
- [5] T. Camp, J. Boleng, and V. Davies, “A survey of mobility models for ad hoc network research,” *Wireless Communication & Mobile Computing (WCMC): Special Issue on Mobile Ad Hoc Networking: Research, Trends and Applications*, vol. 2, no. 5, pp. 483–502, 2002.
- [6] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott, “Impact of human mobility on the design of opportunistic forwarding algorithms,” in *Proc. IEEE INFOCOM*, Apr. 2006.
- [7] N. Eagle and A. Pentland, “Reality mining: Sensing complex social systems,” *J. Pers. Ubiquitous Comput.*, vol. 10, no. 4, May 2006.
- [8] R. Groenevelt, P. Nain, and G. Kooile, “The message delay in mobile ad hoc networks,” in *Proc. PERFORMANCE*, Oct. 2005.

- [9] M. Grossglauser and D. Tse, "Mobility increases the capacity of ad hoc wireless networks," *IEEE/ACM Trans. Networking*, vol. 10, no. 4, pp. 477–486, Aug. 2002.
- [10] T. Henderson, D. Kotz, and I. Akyuz, "The changing usage of a mature campus-wide wireless network," in *Proc. ACM MOBICOM*, Sep. 2004.
- [11] X. Hong, M. Gerla, G. Pei, and C. Chiang, "A group mobility model for ad hoc wireless networks," in *Proc. 2nd ACM Int. Workshop on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, Aug. 1999.
- [12] W. Hsu, K. Merchant, H. Shu, C. Hsu, and A. Helmy, "Weighted waypoint mobility model and its impact on ad hoc networks—Mobicom 2004 poster abstract," *Mobile Comput. Commun. Rev.*, Jan. 2005.
- [13] W. Hsu and A. Helmy, "On nodal encounter patterns in wireless LAN traces," in *Proc. 2nd Int. Workshop on Wireless Network Measurement (WinMee 2006)*, Apr. 2006.
- [14] W. Hsu and A. Helmy, "On important aspects of modeling user associations in wireless LAN traces," in *Proc. 2nd Int. Workshop on Wireless Network Measurement*, Apr. 2006.
- [15] W. Hsu, T. Spyropoulos, K. Psounis, and A. Helmy, "Modeling time-variant user mobility in wireless mobile networks," in *Proc. IEEE INFOCOM*, May 2007.
- [16] S. Jain, K. Fall, and R. Patra, "Routing in a delay tolerant network," in *Proc. ACM SIGCOMM*, Aug. 2004.
- [17] A. Jardosh, E. Belding-Royer, K. Almeroth, and S. Suri, "Towards realistic mobility models for mobile ad hoc networks," in *Proc. ACM MobiCom*, Sep. 2003.
- [18] A. Jindal and K. Psounis, "Fundamental mobility properties for realistic performance analysis of intermittently connected mobile networks," in *Proc. IEEE PerCom Workshop on Intermittently Connected Mobile Ad Hoc Networks (ICMAN)*, 2007.
- [19] P. Juang, H. Oki, Y. Wang, M. Martonosi, L.-S. Peh, and D. Rubenstein, "Energy-efficient computing for wildlife tracking: Design tradeoffs and early experience with ZebraNet," in *Proc. ASPLOS-X Conf.*, Oct. 2002.
- [20] T. Karagiannis, J.-Y. Le Boudec, and M. Vojnovic, "Power law and exponential decay of inter contact times between mobile devices," in *Proc. ACM MobiCom*, Sep. 2007.
- [21] B. Karp and H. Kung, "GPSR: Greedy perimeter stateless routing for wireless networks," in *Proc. ACM MobiCom*, Aug. 2000.
- [22] M. Kim, D. Kotz, and S. Kim, "Extracting a mobility model from real user traces," in *Proc. IEEE INFOCOM*, Apr. 2006.
- [23] D. Lelescu, U. C. Kozat, R. Jain, and M. Balakrishnan, "Model T++: An empirical joint space-time registration model," in *Proc. ACM MOBIHOC*, May 2006.
- [24] A. B. McDonald and T. F. Znati, "A mobility-based framework for adaptive clustering in wireless ad hoc networks," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 8, Aug. 1999.
- [25] M. McNett and G. Voelker, "Access and mobility of wireless PDA users," *ACM SIGMOBILE Mobile Comput. Commun. Rev.*, vol. 7, no. 4, Oct. 2003.
- [26] M. Musolesi and C. Mascolo, "A community based mobility model for ad hoc network research," in *Proc. 2nd Int. Workshop on Multi-Hop Ad Hoc Networks (REALMAN)*, May 2006.
- [27] S. Ross, *Introduction to Probability Models*, 8th ed. Boston, MA: Academic, Dec. 2002.
- [28] K. Seada, M. Zuniga, A. Helmy, and B. Krishnamachari, "Energy-efficient forwarding strategies for geographic routing in lossy wireless sensor networks," in *Proc. ACM Sensys*, Nov. 2004.
- [29] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, "Performance analysis of mobility-assisted routing," in *Proc. ACM MOBIHOC*, 2006.
- [30] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, "Efficient routing in intermittently connected mobile networks: The single-copy case," *IEEE/ACM Trans. Networking*, vol. 16, no. 2, Feb. 2008.
- [31] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, "Efficient routing in intermittently connected mobile networks: The multi-copy case," *IEEE/ACM Trans. Networking*, vol. 16, no. 2, Feb. 2008.
- [32] S. Tanachaiwiwat and A. Helmy, "Computer worm ecology in encounter-based networks (invited paper)," in *Proc. IEEE Broadnets*, Sep. 2007.
- [33] C. Tudeuce and T. Gross, "A mobility model based on WLAN traces and its validation," in *Proc. IEEE INFOCOM*, Mar. 2005.
- [34] A. Vahdat and D. Becker, "Epidemic routing for partially connected ad hoc networks," *Duke Univ., Durham, NC, Tech. Rep. CS-200006*, Apr. 2000.
- [35] X. Zhang, G. Neglia, J. Kurose, and D. Towsley, "Performance modeling of epidemic routing," in *Proc. IFIP Networking 2006*.

- [36] Cab Spotting: A Project That Tracks Taxi Mobility in the San Francisco Bay Area. Trace. [Online]. Available: <http://cabspotting.org/api>
- [37] CRAWDAD: A Community Resource for Archiving Wireless Data at Dartmouth. [Online]. Available: <http://crawdadd.cs.dartmouth.edu/index.php>
- [38] MobiLib: Community-Wide Library of Mobility and Wireless Networks Measurements. [Online]. Available: <http://nile.usc.edu/MobiLib>
- [39] L. Breslau, D. Estrin, K. Fall, S. Floyd, J. Heidemann, A. Helmy, P. Huang, S. McCanne, K. Varadhan, Y. Xu, and H. Yu, "Advances in network simulation," *IEEE Computer*, vol. 33, pp. 59–67, May 2000.
- [40] Simulation codes used in this work and their detailed description. [Online]. Available: [http://nile.cise.ufl.edu/~weijenh/TVC\\_model](http://nile.cise.ufl.edu/~weijenh/TVC_model)



**Wei-Jen Hsu** (M'08) was born in Taipei, Taiwan, in March 1977. He received the B.S. degree in electrical engineering and the M.S. degree in communication engineering from National Taiwan University in 1999 and 2001, respectively. He received the Engineer degree in electrical engineering from the University of Southern California, Los Angeles, in 2006, and the Ph.D. degree in computer science from the University of Florida, Gainesville, in 2008.

His main research interest involves the utilization of realistic measurement data in various tasks in computer networks, including user modeling and behavior-aware protocol design.



**Thrasylvoulos Spyropoulos** (M'07) was born in Athens, Greece, in July 1976. He received the Diploma in electrical and computer engineering from the National Technical University of Athens, Greece, in 2000, and the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, in 2006.

During 2006–2007, he was a Postdoctoral Researcher at INRIA, Sophia Antipolis, France. He is currently a Senior Researcher with the Swiss Federal Institute of Technology (ETH), Zurich.



**Konstantinos Psounis** (SM'08) received his first degree from the National Technical University of Athens, Greece, in 1997, and the M.S. and Ph.D. degrees from Stanford University, Stanford, CA, in 1999 and 2002, respectively.

He is an Associate Professor of electrical engineering and computer science at the University of Southern California, Los Angeles. He models and analyzes the performance of a variety of networks, and designs methods and algorithms to solve problems related to such systems. He is the author of

more than 60 research papers, has received faculty awards from NSF, CISCO Systems, the METRANS transportation center, and the Zumberge foundation, and has been a Stanford graduate fellow throughout his graduate studies.



**Ahmed Helmy** (M'07) received the Ph.D. degree in computer science in 1999 from the University of Southern California (USC), Los Angeles, the M.S. degree in electrical engineering in 1995 from USC, the M.S. degree in engineering and mathematics in 1994, and the B.S. degree in electrical engineering in 1992 from Cairo University, Cairo, Egypt.

He was a key researcher in the NS-2 and PIM projects at USC/ISI during 1995–1999. In 1999, he joined the faculty of the Electrical Engineering Department at USC. In Fall 2006, he became an

Associate Professor and the Director of the Wireless Networking Laboratory in the CISE Department at the University of Florida, Gainesville. His research interests include design, analysis and measurement of wireless ad hoc, sensor and mobile social networks, mobility modeling, multicast protocols, IP mobility and network simulation.

Dr. Helmy received the NSF CAREER Award in 2002, the 2000 Zumberge Research Award, and the Best Paper Award at IEEE/IFIP MMNS'02. In 2004 and 2005, he received the best merit ranking in the USC Electrical Engineering faculty. In 2008 and 2007, he was a finalist and winner in the ACM MobiCom SRC.