

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

Wei-jen Hsu*, Thrasyvoulos Spyropoulos[†], Konstantinos Psounis[‡] and Ahmed Helmy*

*Dept. of Computer and Information Science and Engineering, University of Florida, Gainesville, Florida 32611-6120

[†]Computer Engineering and Networks Lab, ETH Zurich

[‡]Dept. of Electrical Engineering, University of Southern California, Los Angeles, California 90089-2560
Email: wjhsu@ufl.edu, spyropoulos@tik.ee.ethz.ch, kpsounis@usc.edu, helmy@ufl.edu

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 time correlations. We achieve the goal with, on the one hand, the concept of communities that leads to skewed location visiting preferences, and, on the other hand, time periods that allow us to model time dependent behaviors and periodic re-appearances 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 nodal spatial distribution, the average node degree, and the hitting and meeting times for the TVC model, and provide examples of how to utilize this theory to guide design decisions in routing protocols.

I. INTRODUCTION

In recent years, there has been an exponential growth in the popularity of portable computation and communication devices. Advances in wireless communication technologies and standards have made ubiquitous communication an emerging reality. With the ever expanding adoption of these wireless-capable devices, there is an increasing interest in new communication paradigms and applications that do not necessarily rely on the infrastructure.

Mobile ad hoc networks (MANETs) [9] are self-organized, infrastructure-less networks that could potentially support many applications, such as vehicular networking (VANET) [30], [31], disaster relief [29], wild-life tracking [32], [33], and Internet provision to rural areas [3], to name a few. 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 [8], routing performance [16], and cluster maintenance [28]. 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) it should first capture *realistic* mobility patterns of scenarios in which one wants to eventually operate the network; (ii) at the same time it is desirable that the model is *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 [5], [6], [8], [26]; (iii) finally, it should be *flexible* enough to provide qualitatively and quantitatively different mobility characteristics by changing some parameters of the model, yet in a repeatable and scalable manner; designing a new mobility model for each existing or new scenario is undesirable.

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. The most widely used mobility models are *random mobility models* such as random walk, brownian motion, random direction, and random waypoint [11]. Their strength is the theoretical tractability but their weakness is the lack of realism. More complicated *synthetic mobility models* (e.g., [21], [48], [49]) improve the realism, but most of the time at the expense of theoretical tractability. More recently, a large number of *empirical mobility traces* from real mobile users have been collected [12], [13], [14], [15], [26]. Although one can use such traces directly in an evaluation with excellent realism, these traces are usually rather inflexible and provide only a single snapshot of the underlying mobility process. To address these two issues, *trace-based mobility models* [18], [19], [20], [24] have been proposed (i.e. larger, more flexible synthetic traces created from the smaller empirically collected ones). Yet, most of these models do not possess the necessary flexibility to match mobility characteristics of traces other than the ones on which they are based.

The main contribution of this paper is the proposal of a *time-variant community mobility model*, referred to as the TVC model, which is *realistic, flexible, and mathematically tractable*. The model captures several important mobility characteristics we, and other researchers, have observed empirically from various WLAN traces [17]. One of the salient characteristics is *location preference*. In the TVC model, we extend the concept of communities from [22] to serve as pop-

ular locations for the nodes. Another important characteristic is the *time-dependent, periodical behavior* of many nodes. To capture this, we implement time periods in which the nodes move differently [1]. To our best knowledge, this is the first *synthetic mobility model* that captures non-homogeneous behavior in both *space* and *time*.

To establish the flexibility of our TVC model we show that we can match its two prominent properties, *location visiting preferences* and *periodical re-appearance*, with *multiple* WLAN traces, collected from environments such as university campuses [15], [17] and corporate buildings [12]. 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[26]. In the former case, we observe that *location visiting preferences* and *periodical re-appearance* are also prominent characteristics in vehicular movements. In the later case, we are able to match our TVC model with some other mobility characteristics, namely the inter meeting time and encounter duration between different users/devices. These latter quantities are particularly important for encounter-based (or “delay-tolerant” [2]) protocols. Despite these characteristics are not explicitly incorporated in our model by its construction, they can be still realistically reproduced.

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 *nodal spatial distribution*, the *average node degree*, the *hitting time* (time required for a mobile node to hit a randomly selected coordinate) and the *meeting time* (time required for two mobile nodes to come within communication range of each other). These quantities are 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 measurement sets (traces) collected from multiple scenarios, and has also been theoretically treated to the extent presented in this paper. Due to its strengths in both flexibility and theoretical tractability, the TVC model has two major applications: to generate realistic mobility patterns under a wide range of different scenarios, as we show in section IV, and to facilitate performance analysis and prediction, as we show in section VII. We also make the code of the TVC model available at [37].

The remaining of the paper is organized as follows: In Section II we discuss some related work. Our *TVC model* is then introduced in Section III. In Section IV, we show how to generate realistic mobility scenarios with matching mobility characteristics in various traces. Then, in Section V, we embark to present our theoretical framework and derive generic expressions of various quantities under the TVC model. The accuracy of these expressions is validated against simulations in Section VI. Additionally, in Section VII, we

motivate our theoretical framework further, by applying our analysis to provide guidelines and 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. For example, although the per node capacity of static networks goes to zero as network density increases [7], it is shown that mobility can improve the scaling law for the capacity to $O(1)$ [8]. Different underlying mobility models also change the performance ordering of various MANET routing protocols [16]. More recently, mobility has been utilized as the enabling factor for message delivery in delay tolerant networks (DTNs[2]), where a complete path from the source node to the destination node does not exist at any time instant, broadening the scenarios in which communication networks can be established.

A wide variety of mobility models are available in the research community (see [11] 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. They are not only easy to generate, tune and scale, but also amenable to mathematical analysis. A number of important properties for these models have been studied, such as the stationary nodal distribution [44], the hitting and meeting times [22], and the meeting duration [45]. These quantities in turn enable routing protocol analysis to produce performance bounds [4], [5], [6]. 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 performances 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 [21], stay on designated paths for movements [49], or move as a group [48]. More variants of mobility rules can be found in various models [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., [12], [14], [15]) or cellular phone networks (e.g., [13]), to track user mobility by monitoring their locations (through their associations with base stations). Such traces can be replayed as input mobility patterns for simulations of network protocols [46], [47]. More recently, DTN-specific testbeds [26], [27], [30], [32], [33] aim at collecting encounter

events between mobile nodes instead of the actual mobility patterns. Some initial efforts to mathematically analyze some properties of these traces can be found in [26], [51]. Yet, due to the experimental nature of these studies, 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 [18], [19], [20], [24]. Based on the collected traces, these models discover the underlying mobility rules that lead to the observed properties (such as the nodal distribution, 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 a previous work of ours [17] in which we identify several prominent properties that are common in multiple WLAN traces. Large-scale deployments of WLANs in university [14], [15] and corporate [12] campuses provide excellent platforms in which huge amount of user data can be collected and analyzed. Resources for measurement-based study are available at two network trace archives [34], [35]. We leverage these traces to understand empirical user mobility, and propose the TVC model based on the prominent characteristics observed. This model extends the concept of communities proposed by us in [22] and also introduces time-dependent behavior. A preliminary version of the model has been presented in [1].

We differentiate our work from other trace-based models [18], [19], [20], [24] 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 [24], 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 that this additional requirement will largely facilitate the application of our model for performance prediction of various communication protocols. Finally, although we start from mobility characteristics observed from WLAN traces, we propose a mobility model that has wider applicability - in addition to WLANs, it can be tuned to match with other types of traces, such as a vehicle mobility trace [36], and even with other characteristics in other traces of human mobility (e.g., the encounter duration and the inter-meeting time [26]).

As a final note, the concept of community is also mentioned in [25] in a different context. 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 follow location-based preference to make movement decisions, and each node moves independently of the others (nonetheless, node sharing the same community will exhibit mobility correlation). Mobility models with inter-node depen-

ency require a solid understanding of the social network structure, which is an important area under development. In future work, we intend to include such mobility behaviors motivated by factors other than location preference (e.g. social bonds) into our model.

III. TIME-VARIANT MOBILITY MODEL

A. Mobility Characteristics Observed in WLAN Traces

The main objective of the paper is to propose a mobility model that captures the important mobility characteristics observed in daily life. To better understand this mobility, we have chosen to study a number of wireless LAN traces collected by several research groups (e.g., traces available at [34] or [35]). The reason for this choice is that WLAN traces log information regarding large numbers of nodes, and thus are more reliable for statistical analysis. After analyzing a large number of traces, we have observed two important properties that seem to be recurrent in all of them: *skewed location visiting preferences* and *time-dependent mobility behavior* [17].

First, by *location visiting preference* we mean the amount of time that a node spends associated with a given access point (AP). In Fig. 1(a) we calculate for various traces the fraction of the total online time an average node spends with its most favorite AP, its second favorite AP, etc., up to its least favorite AP. (This is essentially the probability density function of the association time of a node with an AP, with the APs sorted in descending order of total association time.) It is clear from the plot that a node on average spends more than 65% of its online time associated with its favorite AP, and more than 95% of its online time at only five APs. We refer to this behavior by saying that the location visiting preference (or in short “location preference”) of nodes is *skewed*.

Second, by *time-dependent mobility behavior* we refer to the fact that nodes tend to behave differently during different times of the day (or even during different days), and most specifically to exhibit some amount of periodicity in terms of the locations they choose to visit. In Fig. 1(b) we plot the probability of a node appearing in the same location at some time in the future, as a function of the difference in time. It is evident from the plot that nodes appear at the same AP with a higher probability after a time-gap of integer multiples of days. This creates the saw-tooth pattern in the curves. A slightly stronger weekly correlation could also be observed in some plots (see for example the slightly large peak in the MIT curve for a time gap of seven days). It is thus clear that nodes behave differently in different periods in time, and that similar behaviors tend to be repeated on a daily basis.

Unfortunately, most existing mobility models fail to capture these two properties. For simple random models, like random direction, random waypoint, random walk, etc., there is obviously no preference in both space and time. This is demonstrated in Fig. 1 by a straight line (uniform distribution) for the Random Direction model for the respective probabilities. Even for more sophisticated models that try to capture other aspects of mobility, such as group mobility in the RPGM model [48] or a model considering obstacles and pathways [49], these two properties would also be straight

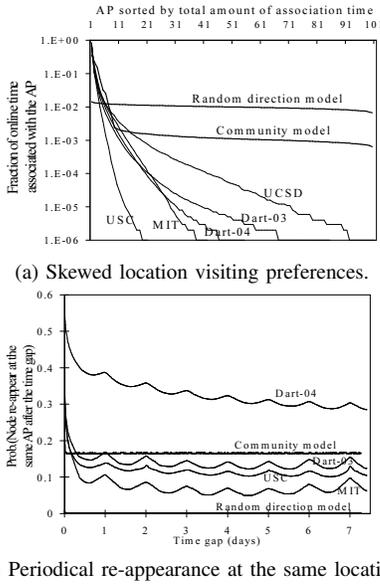


Fig. 1. Two important mobility features observed from WLAN traces. Labels of traces used: MIT: trace from [12], Dart: trace from [15], UCSD: trace from [14], USC: trace from [17].

lines in the plots as spatial and temporal preference is not a part of these models¹. There do exist some more recent models (e.g., [22], [21], [18], [19], [20]) that aim at capturing spatial preference explicitly. An example of such a model is the simpler community model of [22]. As is shown in Fig. 1(a), 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 re-appearance* property cannot be reproduced, as shown by the flat curve labeled *community model* in Fig. 1(b).

It is our goal to design a mobility model that successfully captures both of these two properties, observed in the majority of traces. One could argue that a potential shortcoming of this approach is that WLAN traces do not register continuous movement of the devices, but rather associations of users/nodes with specific APs. What is more, some devices are not always on, and typically there are some gaps in the coverage of access points in these networks. However, we believe that the two main properties we observed, namely *skewed location preference* and *time-dependency*, are prevalent in real-life mobility. This belief is further supported by observing typical daily activities of humans: most of us tend to spend most of the time at a handful of frequently visited locations, and a recurrent daily or weekly schedule is an inseparable part of our lives. As a result, a model supporting location-preference and time-dependent mobility should be able to capture human mobility in many contexts, if carefully designed. Comparison with non-WLAN traces in Section IV confirms our argument.

¹In the case of the obstacle-based model, some locations are not allowed to be visited at all; yet among all the permissible ones, no particular preference is assigned to any node.

TABLE I
PARAMETERS OF THE TIME-VARIANT COMMUNITY MOBILITY MODEL

For all parameters, we follow the convention that the subscript of a quantity represents its community index, and the superscript represents the time period index.

| | |
|--------------------------------|--|
| 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 |
| $Comm_j^t$ | The j -th community during time period t |
| π_j^t | Probability that the next epoch is performed in community j during time period t |
| $v_{min}, v_{max}, \bar{v}$ | Minimum, maximum, and average speed ² |
| $D_{max,j}, D_j$ | Maximum and average pause time after each epoch ² |
| 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 |
| P_j^t | Fraction of time the node is in state j ($P_j^t = P_{move,j}^t + P_{pause,j}^t$) |
| 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 |
| Ω^t | The set of all possible relationships between a target coordinate and the communities in time period t |
| $P(w^t)$ | Probability of a given relationship w^t in time period t |
| $A(w^t)$ | Area corresponding to the relationship w^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 | Unit-time meeting probability |
| P_M | Meeting probability for a time period t |
| $HT(case)$ | Expected hitting time under the given "case" |
| $MT(case)$ | Expected meeting time under the given "case" |

B. Construction of the Time-variant Community Model

Skewed location preferences arises naturally due to extended stay at locations that bear importance to us, such as homes and offices, cafeterias and libraries. To capture this phenomenon, we construct popular location(s) for the nodes in the simulation field. These locations, or rather geographical areas, we call *communities*, and we make a node visit its own community more often than other areas. Different nodes can pick different communities, creating nodes with very diverse behaviors. Furthermore, multiple communities could be defined if a node tends to visit multiple locations with high probability, some of which could also be shared by more than one nodes (e.g. people working in the same building, libraries, etc.).

Periodical re-appearance at a given location is related to omnipresent schedules in our lives. Almost everyone follows recurrent daily or weekly schedules, and different behaviors based on time-of-day have been observed in many contexts. To capture time-dependent behaviors, we introduce structure in the time domain by the use of *time periods*. For a given node, we assign several *time periods* during which it behaves differently. For example, a node may have different communities during different periods or the same communities but different mobility parameters to move between them. To further ensure periodicity, the time period assignment follows a *recurrent structure*, with the same "time-period" and its respective statistical characteristics occurring, say, for all weekday mornings.

We illustrate the model with an example in Fig. 2. We also

use this example to introduce the notations we use (see Table I) in the rest of the paper. As shown in the example, there are $V = 3$ (where V denotes the total number of time periods) time periods TP1, TP2, and TP3 (of duration T^1 , T^2 , and T^3 , respectively). During each time period, there are a number of communities, that is, geographical areas that are heavily visited. These communities can be chosen differently in each time period, as shown by the three sub-plots. Within a given time period t , the j -th community is denoted as $Comm_j^t$.² This is a square geographical area of edge length C_j^t . Note that by construction the communities can overlap (as in TP1 in Fig. 2), or one community can even contain the other (as in TP2 in Fig. 2). Finally, the number of communities in each time period may vary. For example, there are 3 communities in total in the first period, 2 in the second one, and 4 in the third³. The total number of communities in period t is denoted as S^t . This construction allows for maximum flexibility when designing the simulation setup for nodes with different behaviors. As for the structure in time domain, we need to arrange time periods in a re-occurring sequence (see Fig. 2 or Fig. 3) that corresponds to a daily or weekly schedule.

We now describe how a node moves inside the above construction. Node movement consists of a sequence of *epochs*. Each of these epochs is a Random Direction movement⁴. In a typical Random Direction epoch, a node chooses at the beginning its speed uniformly in $[v_{min}, v_{max}]$, and a direction (angle) uniformly in $[0, 2\pi]$; it also chooses the length (distance) of movement (usually distributed exponentially with average in the order of the network dimension), and moves towards this direction with the chosen speed and for the chosen distance; at the end of the epoch, the node picks a pause time randomly and then proceeds to the next epoch.

The difference between our community model and the Random Direction model is that in addition to all other parameters, in the community model the node also chooses randomly the community in which the next epoch will be performed. That is, with probability π_j the next epoch takes place inside the node's j -th community ($\sum_{j=1}^{S^t} \pi_j = 1$), rather than moving around the whole simulation area randomly, as in the standard Random Direction model. (Note that we usually add superscript t in the notation, i.e. π_j^t , to denote that these probabilities might change between time periods.) We say the node is in *state* j when it has an epoch in the j -th community. Further, to ensure that a local move is compatible with the local community size, we also scale the local epoch length by drawing it from an exponential distribution with average

²For 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. Note that all parameters used in the paper can be set differently for *each node*. 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 .

³To allow a node to move randomly among the whole simulation field sometimes, we often allocate one community to be the whole simulation field (e.g. $Comm_3^1$ in period TP1 in Fig. 2).

⁴Note that we could also choose random waypoint or random walk models for the type of movement during each epoch.

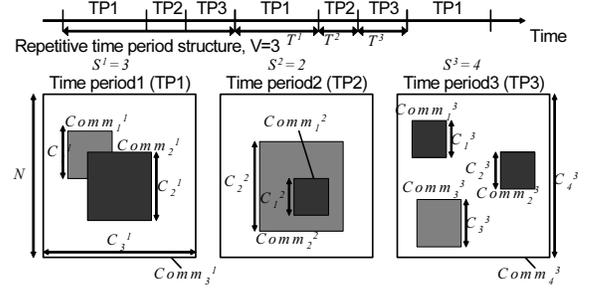


Fig. 2. Illustration of a generic scenario of time-variant mobility model, with three time periods and different numbers of communities in each time period.

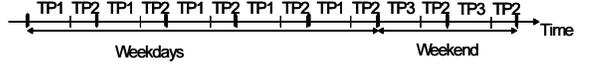


Fig. 3. An illustration of a simple weekly schedule, where we use time period 1 (TP1) to capture weekday working hour, TP2 to capture night time, and TP3 to capture weekend day time.

length \overline{L}_j , that is, in the order of the given community size⁵. It is important to note that a node can still perform some of its epochs in the whole simulation area, by assigning an additional community that corresponds to the whole simulation field (e.g. $Comm_3^1$ in period TP1 in Fig. 2). We refer to such epochs as *roaming epochs*. Finally, after an epoch, a node pauses for a time uniformly chosen in $[0, D_{max,j}]$, where the maximum pause time is again dependent on the community.

As a final note, one may argue that capturing location preference and time-dependencies could plausibly be achieved with a mobility model constructed with different ways than the one we propose. However, our choices are largely guided by the fact that most of the building blocks we utilize to create our mobility model (e.g. random direction epochs, communities, etc.) are easy to understand, and have been shown to be amenable to theoretical analysis [22]. This 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 FOR SIMULATION

The TVC model described in the previous section provides a general framework to model a wide range of mobility scenarios, and provides a powerful tool for simulation-based protocol or service evaluations in MANETs. We have made our mobility trace generator available at [37]. The tool provides mobility traces in both ns-2[38] compatible format and time-location (i.e., (t, x, y)) format.

In this section, our aim is twofold: (i) first, we would like to demonstrate the model's flexibility and how it can be configured to generate mobility instances that are representative of various target wireless networks such as WLANs, VANETs, etc.; (ii) at the same time, we would like to validate

⁵To avoid boundary effects, we assume that if the node hits the community boundary it is re-inserted from the other end of the area (i.e., the boundaries are "torus" boundaries).

the model’s “realism” or “accuracy” by explicitly comparing mobility instances produced by our model with real mobility instances captured in well-known, publicly-available 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.

We first outline here some general guidelines about how to use the model in order to construct specific mobility scenarios. Then, we show how to explicitly configure the TVC model in order to match the mobility characteristics observed in three case studies: a wireless LAN trace, a vehicular trace, and a human-encounter trace. All the parameter values we use in the examples in this section are also available at [37].

STEP 1: Determine the structure in Space and Time The first step to construct the TVC model for a given scenario is to setup the communities and the time period structure. If the map of the target environment is available, one should observe the map and identify the points of attractions in the given environment and how they vary with respect to time (e.g., restaurants on campus during lunch time, hotels in an amusement park during nights), and assign the communities/time periods in the TVC model accordingly. If the map is not available, alternatively, one could use the general mobility characteristics observed in typical traces for the particular target network (as shown, for example, in Fig. 1) as guidelines to assign the structure in space and time. For example, from location preference curves like the one in Fig. 1(a), one can determine the number of communities one needs to explicitly create; as a very simple example, if in most WLAN traces it is observed that the typical node spends say 95% of its time at around 2 to 5 preferred locations (depending on the node), then one could assign each node to have from 2 to 5 local communities in the network (with the actual number and locations of communities randomly chosen for each node), with a larger (roaming) community representing the rest of the 5% of mobility time⁶. Similarly, from curves like the ones in Fig. 1(b), one may observe the *re-appearance* periodicity and decide on the time period structure accordingly. If a finer time granularity is necessary (e.g. time-of-day) one could additionally observe the mobility characteristics (e.g. location preferences) on an hour-by-hour basis and identify clear changes in a node’s daily behavior.

STEP 2: Assign Community-related Parameters Ideally, for a given environment, once the communities are identified, the related parameters (e.g., $\pi_j^t, \overline{D}_j^t, \overline{L}_j^t$, which represent the probability, average pause time, and average epoch length, at community j during time period t) could be assigned according to the mobile nodes’ behavior in each community (e.g., how long does a typical person spend at the cafeteria for lunch?). Nevertheless, in most cases such information is not available, or extremely difficult to obtain. Hence, one could

⁶In reality, one may be able to capture more complex structures with more communities or structure between them, by combining knowledge from the actual network area (map), generic mobility characteristics, and other information about the network.

again resort to measured statistics from typical traces to guide the assignment of the parameters. It is not difficult to see that, typically, the attraction of the communities (π_j^t) and the time spent in each community (related to $\overline{D}_j^t, \overline{L}_j^t$) determine the shape of the location visiting preference curve. Thus, one can use basic probability theory to calculate the expected fraction of time a node spends in a given community as a function of these parameters, and derive from it the values needed to obtain a given location preference profile observed in a curve in Fig. 1(a). We calculate these community stay probabilities later in Lemma 5.1.

STEP 3: Assign User On-off Behavior Note that the mobility trace generated by the TVC model is an “always-on” mobility trajectory of the mobile node (i.e., the node is always present somewhere in the simulation field). Depending on the target environment, this always-on behavior may not be realistic. In many empirically collected traces, not all nodes are present all the time (i.e., some of the nodes are “off” or not in the observed area sometimes). This is the case in two of the scenarios we discuss below - in the WLAN traces, nodes are “on” only when they are not moving; in the vehicle mobility trace, nodes are “on” when they are moving. Thus, before producing the final synthetic trace, assumptions about when the user is considered “on” should sometimes be made and superimposed to the TVC mobility traces generated. We have applied this step to the traces of the two scenarios mentioned above.

Next, we look into three specific case studies, namely a set of WLAN traces, a vehicular trace, and a trace of inter-node encounters. We show how to apply the fore-mentioned procedure in each case, and show that synthetic mobility traces produced by the TVC model successfully match the characteristics observed in the real traces.

1) *Matching Mobility Characteristics with WLAN Traces:* In the first example, we show that the TVC model can re-create the *location preferences* and *re-appearance probability* curves observed in WLANs. We construct our synthetic trace from the TVC model with the following steps: (STEP1) We divide the simulation area into a 10-by-10 grid and use these 100 grid cells as the locations for the purpose of measuring mobility statistics for the simulated nodes. For each node, we assign some of the grids to serve as the node’s communities during each time period, according to the method described earlier. (STEP2) We use the mobility characteristics obtained from the WLAN traces (i.e., curves in Fig. 1(a)) to calculate the attraction from the communities (π_j^t) and the pause times of the node (\overline{D}_j^t) to shape a decaying tail of *location visiting preference*. (STEP3) Since devices are usually turned off when users move them in the real WLANs, we make a similar assumption that the mobile nodes are considered “on” only when they are not moving. When the simulated node moves, we assume that it is not associated with the grid. Note that the curves in Fig. 1(b) represent the probability of an “on” node associated with the same community after the given time gap, and the peaks appear when the considered points in time are in the same type of time period. Therefore, the peak value is $\sum_{j=1}^{S^t} (\pi_j^t)^2 (P_{on,j}^t)^2$, where $P_{on,j}^t$ denotes the probability

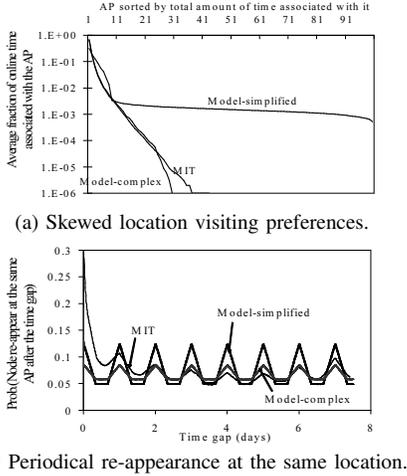


Fig. 4. Matching mobility characteristics of the synthetic traces to the MIT WLAN trace.

a node is considered “on”. Hence, the fraction of time nodes spend on moving ($\overline{L_j^t}/\bar{v}$) and pause ($\overline{D_j^t}$) can be adjusted (note that in this case, $P_{on,j}^t = \overline{D_j^t}/(\overline{D_j^t} + \overline{L_j^t}/\bar{v})$) to change the peak values in the curve of *periodical re-appearance* property to match with the curves in Fig. 1(b).

We use the MIT WLAN trace (first presented in [12]) as an example to display the match between the synthetic trace derived from the TVC model and the real trace. We also achieved good matching with the USC[17] or the Dartmouth[15] traces, but do not show it here due to space limitations (see [37]). We show the *skewed location visiting preferences* and the *periodical re-appearance* properties in Fig. 4 (a) and (b), respectively. We first try a simple synthetic model (labeled as *model-simplified*, using the parameters of *Model-1* in Table II) with one community in two time periods. While this simple model captures the major trends in the mobility characteristics, there are several noticeable differences. First, since there is only one community, the tail in the *model-simplified* curve in Fig. 4(a) is “flat” as opposed to the exponentially diminishing tail of the *MIT* curve (notice the Y-axis in Fig. 4(a) is in log scale). Second, the peaks in the *model-simplified* curve in Fig. 4(b) are of equal heights, due to the simple two-alternating-time-period structure, as opposed to varying peak values of the *MIT* curve. We can improve the matching between the synthetic trace and the real trace by adding complexity in both space (using more communities) and time (using more complex schedule, such as the weekly schedule shown in Fig. 3). In a refined model labeled as *Model-complex* in Fig. 4, we show that the resulting mobility characteristics match very closely with the MIT trace. This also demonstrates the flexibility of our model - the user can adjust its complexity by choosing the number of communities and time periods needed to achieve a desired level of matching with the mobility characteristics.

2) *Matching Mobility Characteristics with Vehicle Mobility Traces*: In this example we display that *skewed location visiting preferences* and *periodical re-appearance* are also prominent mobility properties in vehicle mobility traces. We obtain a vehicle movement trace from [36], a website that

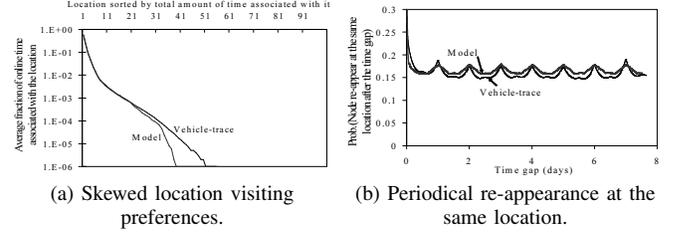


Fig. 5. Matching mobility characteristics of the synthetic trace to the vehicle mobility trace.

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. We obtain the mobility characteristics of the taxis by the following steps. For each taxi, we first identify its movement range within the 40-day period, then draw a rectangular area that bounds the movement of the taxi, and divide this area into equal-sized 10-by-10 grids. We tally the mobility statistics of the taxis using these 100 grids as locations, and show the results in Fig. 5 (a) and (b), respectively, with the label *Vehicle-trace*. It is interesting to observe that the trend of vehicular movements is very similar to that of WLAN users in terms of these two properties.

We further show that, using the outlined procedures, we can generate a synthetic trace with similar mobility characteristics as the vehicle mobility trace. After observing the trace closely, we discover that the taxis are offline (i.e., not reporting their locations) when not in operation. Hence in the synthetic trace we make the corresponding assumption (in STEP3) that the nodes are associated with the current grid they reside in only when they are moving; we then consider the pause times as breaks in the taxi operation (hence $P_{on,j}^t = (\overline{L_j^t}/\bar{v})/(\overline{D_j^t} + \overline{L_j^t}/\bar{v})$ in this case), from which we can calculate or adjust the respective model parameters. The curves in Fig. 5 with label *Model* correspond to the mobility characteristics of the synthetic trace. 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 or human mobility) without losing its basic properties. We defer this for future work.

3) *Matching Contact Characteristics with Encounter-based Traces*: In this example, we show that the TVC model is generic enough to also reproduce the distributions of the inter-meeting time and the encounter duration observed from a human encounter trace [26], by setting up its parameters properly. Specifically, we tune our mobility model to mimic the behaviors observed in an experiment performed at INFOCOM 2005 [26]. In this experiment, wireless devices were distributed to 41 participants of the conference, with appropriate software installed that could 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

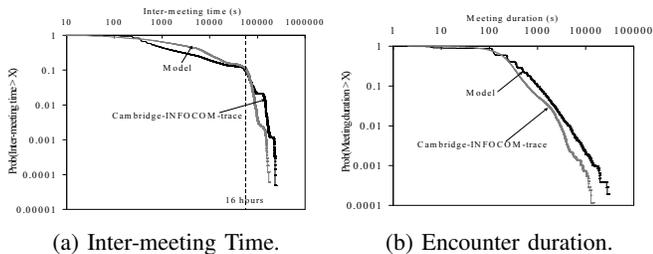


Fig. 6. Matching *inter-meeting time* and *encounter duration* distributions with the encounter trace.

recurrent schedule with five different types of time periods (STEP 1): technical sessions, coffee breaks, breakfast/lunch time, evening, and late night (see [37] 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 mimic the behavior of patrons skipping part of the conference. It is interesting to note that the inter-meeting time distribution has a sharp drop (the “knee” in the curve) at 16 hours, which is approximately the time gap between the end of the day and the beginning of the subsequent day at the conference. This suggests the nodes (naturally) meet with lower probability during the nights, and thus the time-dependent mobility provided by our TVC model is appropriate. We can naturally achieve this by assigning nodes to disjoint communities (i.e., the “hotel rooms”) during the nights. 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 π_r^t) would increase the meeting probability (see, e.g., Eq. (15)), hence reducing inter-meeting times. Since the devices used to collect the encounter traces are always-on, we do not apply any changes to the synthetic trace in STEP 3. We randomly generate 820 pairs of users and obtain their corresponding distributions of the inter-meeting time and the encounter duration. These distributions are shown 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, once we observe the target environment closely and come up with the right underlying parameters, the TVC model is able to capture the consequent mobility characteristics 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 strength and flexibility of the TVC model in terms of its ability to reproduce the properties

observed in real mobility traces. Although other proposed models have also managed to match some sets of collected measurements [18], [19], [20], [24], none of the existing works has been shown to capture the variety of qualitatively different traces (e.g. WLAN, vehicles, inter-contacts) that the TVC model does. 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 will first derive the various properties and then validate their accuracy against simulation results. Then, we will provide some examples of how the various theoretical quantities thus obtained can be used to calculate or optimize the performance of popular protocols, in realistic mobility settings.

We start here by deriving the theoretic expressions of various properties of the proposed mobility model. We first calculate the *nodal spatial distribution*. This can be represented as a two-dimensional probability density function of nodes at any given point in the space. The expected number of nodes in an area is then calculated by integrating the density function. This property provides a basic demographic profile of the simulation area, and can also help to evaluate whether the model reflects the target environment. Then, we derive the *average node degree*, which is the average number of nodes residing within the communication range of a given node. This is a quantity of interest due to its implication on the success rate of various tasks in mobile ad hoc networks [41], [42]. Finally, we derive the expected *hitting* and *meeting times* for our model. 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 simulation 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. Note that this definition is different from *inter-meeting time* used in [23], [26], which accounts for the time duration between a meeting event and the subsequent one⁷. These two quantities are of interest due to their close relationship to the performance of DTN routing protocols, or in general the performance of processes that rely on node encounters. Knowing the meeting time for a mobility model is, for example, crucial when using a “mobility-assisted” or “store-carry-and-forward” protocol to deliver a message [4], [5], [6], while hitting times might be needed if some nodes in the network are static (e.g. sensors, base stations, etc.).

We note that a preliminary version of some of the theoretical derivations presented here appear for a special case of our TVC model only in [1] (that model included one community and two time periods only). Here, we generalize all derivations for any community and time-period structure. Moreover, we present some additional results regarding the spatial distribution and the average node degree that are relevant to various wireless communication protocols, as we show in Section VII. We start with a useful lemma that calculates the probability

⁷Although not shown in this paper, we could use similar techniques as in [45] to derive inter-meeting times

of a node to reside in a particular state.

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

$$P_{move,j}^t = \frac{\pi_j^t (\overline{L_j^t}/\overline{v_j^t})}{\sum_{k=1}^{S^t} \pi_k^t (\overline{L_k^t}/\overline{v_k^t} + \overline{D_k^t})}, \quad (1)$$

$$P_{pause,j}^t = \frac{\pi_j^t \overline{D_j^t}}{\sum_{k=1}^{S^t} \pi_k^t (\overline{L_k^t}/\overline{v_k^t} + \overline{D_k^t})}. \quad (2)$$

Proof: The 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 each state, weighted by the probabilities of choosing the state. ■

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 Nodal Spatial Distribution

We start with the derivation of the nodal spatial distribution. This becomes relatively straight-forward after we observe that a node follows a basic *random mobility model* (i.e., random direction) in each community. Hence, when a given node is in state j , it appears equally likely in any point within $Comm_j$.

Theorem 5.2: For a given area A , the probability for a node to appear in A at any given time instant during time period t is

$$\iint_A p(x, y) \, dx dy, \quad (3)$$

where the function $p(x, y)$ is the spatial density function,

$$p(x, y) = \sum_{\{j | (x, y) \in Comm_j^t\}} P_j^t / C_j^{t^2}. \quad (4)$$

Proof: A node could appear at a given point in space when it is in state j if and only if the j -th community includes the point. Within the community, the appearance probability of the node is uniformly distributed. Considering a given point (x, y) , the probability for a node to appear at the point is the sum of the contributions from all of its communities that contain the point. ■

Note that the nodal spatial distribution for each time period is independent, hence can be calculated separately with the above Theorem.

B. 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 5.3: 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 $A(a_j^t, b_k^t)$, as illustrated in Fig. 7. Then, the contribution of node b to the average node degree of node a , when a resides

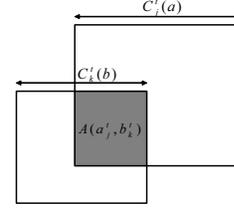


Fig. 7. Illustration of overlapping communities in general case.

in its j -th community and b resides in its k -th community, is given by

$$\frac{\pi K^2}{C_j^{t^2}(a)} \frac{A(a_j^t, b_k^t)}{C_k^{t^2}(b)}, \quad (5)$$

where K is the communication range of the nodes.

Proof: Since nodes are uniformly distributed within each community, 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 5.3, we include all community pairs and arrive at the following Theorem.

Theorem 5.4: The average node degree of a given node a is

$$\sum_{\forall Comm_j^t(a)} P_j^t(a) \sum_{\forall b} \sum_{\forall Comm_k^t(b)} P_j^t(b) \frac{\pi K^2}{C_j^{t^2}(a)} \frac{A(a_j^t, b_k^t)}{C_k^{t^2}(b)}. \quad (6)$$

Proof: Eq. (6) 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. ■

Corollary 5.5: In the special case when all nodes choose their communities uniformly at random among the simulation field, Eq. (6) degenerates to

$$\begin{aligned} & \sum_{\forall b} \sum_{\forall Comm_k^t(b)} P_k^t(b) \frac{\pi K^2}{C_k^{t^2}(b)} \frac{C_k^{t^2}(b)}{N^2} \\ &= \sum_{\forall b} \frac{\pi K^2}{N^2} \sum_{\forall Comm_k^t(b)} P_k^t(b) = \sum_{\forall b} \frac{\pi K^2}{N^2}. \end{aligned} \quad (7)$$

Proof: This result follows from that a randomly chosen community is anywhere in the simulation field equal likely. If nodes pick their communities randomly and independently, the actual location of node a would not make any difference in its average node degree. Regardless of the location of node a , it falls within the k -th community of node b with probability $C_k^{t^2}(b) / N^2$. Within the community, node b appears uniformly, contributing probability $\pi K^2 / C_k^{t^2}(b)$ of appearance within node a 's communication range. Note that the equation reduces to each node b contributing $\pi K^2 / N^2$ to the average node degree, the same as if node b roams around the whole simulation area without any preference in space (i.e., communities). ■

Similar to the nodal spatial distribution, the average node degree can be calculated for each time period separately.

C. Derivation of the Hitting Time

The sketch of the derivation of the hitting time is as follows: (1) We first condition on the relative location of the target coordinate with respect to a node's communities (e.g. target inside community i , target outside community, etc.). We thus have to derive the hitting time for each sub-case separately. (2) We then derive the *unit-step* probabilities of hitting a target, P_h , for a given sub-case. 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). In other words, we approximate the continuous mobility with a discrete version of it where nodes move in discrete steps. It has been shown in [22] that the latter provides a good approximation for the continuous version, and is easier to analyze for our purposes. (3) The expected hitting probability for a whole time period, P_H , is then calculated for each sub-case from the unit-step probability, by assuming "hitting" occurs independently in each time step⁸. (4) Finally, taking the weighted average of each sub-case (i.e. weighted by probability of a given target being located inside a given community) we get the overall hitting time.

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 Ω^t . Note that the cardinality of set Ω^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). Also, not all of the 2^{S^t} combinations are always valid. For example, in the set up of time period 2 in Fig. 2, the communities are overlapped, hence if the target is within $Comm_1^2$ it must be within $Comm_2^2$.

Lemma 5.6: By the law of total probability, the average hitting time can be written as

$$HT = \sum_{w^1 \in \Omega^1, \dots, w^V \in \Omega^V} P(w^1, \dots, w^V) HT(w^1, \dots, w^V), \quad (8)$$

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 Eq. (8), 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 5.7: If the target coordinate is chosen independent of the communities and the communities in each time period are chosen independently from other periods, then

$$P(w^1, \dots, w^V) = \prod_{t=1}^V P(w^t), \quad (9)$$

⁸This assumption of independence is shown in [22] to be a good approximation, when the expected length of an epoch is in the order of the square root of the area of the community the epoch takes place in.

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: $\{\text{target} \in \text{comm}_1^t\}, \{\text{target} \notin \text{comm}_2^t\}, \dots, \{\text{target} \in \text{comm}_S^t\}$.

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 5.8: For a given time period t and a specific scenario w^t ,

$$P_h^t(w^t) = \sum_{j=1}^{S^t} I(\text{target} \in \text{Comm}_j^t | w^t) P_{\text{move},j}^t 2K\bar{v}_j^t / C_j^{t2}, \quad (10)$$

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

Proof: The overall unit-time hitting probability is the sum of the hitting probabilities contributed by epochs in each state. Note that the hitting event can only occur when the node is physically moving, and the node can hit the target when it is moving in its j -th community only if the target coordinate is within the community⁹. When a node moves with average speed \bar{v}_j in community 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 $2K\bar{v}_j / C_j^{t2}$ [22]. Hence the contribution to the unit-time hitting probability by movements made in state j is $P_{\text{move},j}^t 2K\bar{v}_j^t / C_j^{t2}$, i.e., when the node moves in community j and the target is in the newly covered area in the time unit. ■

Note that the movement made in each time unit does not increase or decrease the probability of hitting the target in the subsequent time units, therefore each time unit can be considered as an independent Bernoulli trial with success probability given in Eq. (10). The corollary below immediately follows.

Corollary 5.9: The probability for at least one hitting event to occur during time period t under scenario w^t is

$$P_H^t(w^t) = 1 - (1 - P_h^t(w^t))^{T^t}. \quad (11)$$

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 5.10:

$$HT(w^1, \dots, w^V) = \sum_{t=1}^V HT(w^1, \dots, w^V | \text{first hit in period } t) \cdot P(w^1, \dots, w^V, \text{first hit in period } t), \quad (12)$$

where the probability for the first hitting event to happen in

⁹We 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 .

time period t is

$$P(w^1, \dots, w^V, \text{first hit in period } t) = \frac{\prod_{i=1}^{t-1} (1 - P_H^i(w^i)) \cdot P_H^t(w^t)}{P}, \quad (13)$$

and the hitting time under this specific condition is

$$HT(w^1, \dots, w^V | \text{first hit in period } t) = \sum_{i=1}^V T^i \cdot \left(\frac{1}{P} - 1\right) + \sum_{i=1}^{t-1} T^i + \frac{1}{P_h^t(w^t)}, \quad (14)$$

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

Proof: Eq. (13) 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 Eq. (14) 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 Eq. (14) 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. ■

D. 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 5.10.

Similar to Lemma 5.8, we add up the contributions to the meeting probability from all community pairs from node a and b in the following Lemma.

Lemma 5.11: 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 , as illustrated in Fig. 7. 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

$$\begin{aligned} P_m^t(a_j^t, b_k^t) = & P_{move,j}^t(a) P_{move,k}^t(b) \hat{v} \frac{2K\bar{v}}{A(a_j^t, b_k^t)} \frac{A(a_j^t, b_k^t)}{C_j^{t^2}(a)} \frac{A(a_j^t, b_k^t)}{C_k^{t^2}(b)} \\ & + P_{move,j}^t(a) P_{stop,k}^t(b) \frac{2K\bar{v}}{A(a_j^t, b_k^t)} \frac{A(a_j^t, b_k^t)}{C_j^{t^2}(a)} \frac{A(a_j^t, b_k^t)}{C_k^{t^2}(b)} \\ & + P_{stop,j}^t(a) P_{move,k}^t(b) \frac{2K\bar{v}}{A(a_j^t, b_k^t)} \frac{A(a_j^t, b_k^t)}{C_j^{t^2}(a)} \frac{A(a_j^t, b_k^t)}{C_k^{t^2}(b)}. \end{aligned} \quad (15)$$

Proof: Equation (15) consists of two parts:

(I) Both of the nodes are moving within the overlapped area. This adds the first term in Eq. (15) to the meeting probability.

The two ratios, $\frac{A(a_j^t, b_k^t)}{C_j^{t^2}(a)}$ and $\frac{A(a_j^t, b_k^t)}{C_k^{t^2}(b)}$, 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 $\frac{2K\bar{v}}{A(a_j^t, b_k^t)}$, 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 the (higher) relative speed between the two. This difference is captured with the multiplicative factor \hat{v} [22].

(II) One node is moving in the overlapped area, and the other one pauses within the area. This adds the remaining two terms in Eq. (15) to the unit-time meeting probability. These terms follow similar rationale as the previous one, with the difference that 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 . ■

Note that (15) is the general form of Equation (13) and (14) in [1]. 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 5.12: The probability for at least one meeting event to occur during time period t is

$$P_M^t = 1 - \sum_{\forall(j,k)} \{P_{ov}(a_j^t, b_k^t) \cdot (1 - P_m^t(a_j^t, b_k^t))^{T^t}\}, \quad (16)$$

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 [1]. Due to space constraint, the Lemma is not reproduced here.

Finally, similarly to Theorem 5.10, the expected meeting time can be calculated using the results in the Lemmas in this section.

Theorem 5.13: The expected meeting time is

$$MT = \sum_{t=1}^V MT(\text{meet in period } t) P(\text{meet in period } t). \quad (17)$$

Where the quantities in the above equation are calculated by

$$P(\text{meet in period } t) = \frac{\prod_{i=1}^{t-1} (1 - P_M^i) \cdot P_M^t}{Q}, \quad (18)$$

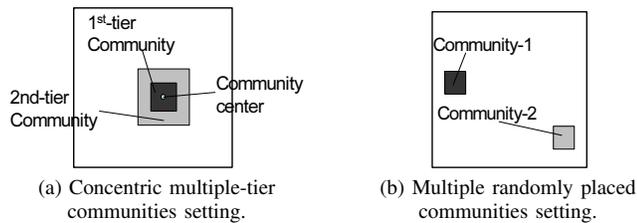


Fig. 8. Illustration of the community setup for the generic cases of time-variant community model.

$$MT(\text{meet in period } t) = \sum_{i=1}^V T^i \cdot \left(\frac{1}{Q} - 1\right) + \sum_{i=1}^{t-1} T^i + \frac{1}{P_m^t}, \quad (19)$$

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 5.10.

Eq. (18) holds as each cycle of time periods follows the same repetitive structure, and for the first meeting event to occur in time period t it must not occur in time period $1, \dots, (t-1)$. The first term in Eq. (19) corresponds to the expected duration of full time period cycles until the meeting event occurs. Since for each cycle the success probability of the two nodes meeting each other is Q , in expectation it takes $1/Q$ cycles for them to meet for the first time, and there are $1/Q - 1$ full cycles. The second term in Eq. (19) is the sum of duration of time periods before the time period t in which the meeting event occurs in the last cycle. Finally, the third term is the fraction of the last time period before the meeting event occurs. Note that the last part is an approximation which holds if the time periods we consider are much longer than unit-time. ■

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 parameter settings. Through extensive simulations with multiple scenarios and parameter settings, we establish the accuracy of the theoretical framework.

We summarize the parameters for the tested scenarios in Table II. Table II (a) lists the parameters we use for a simplified model (two time periods with two communities in each time period, where one of the communities is the whole simulation field). For more complex models, we try out the setup of tiered communities (as shown in Fig. 8 (a)) and multiple randomly placed communities (as shown in Fig. 8 (b)). In the tiered communities layout, a randomly chosen point in the simulation field serves as the *center* of the communities, and multiple tiers of communities with different sizes share the same center. This construction is suggested by a common observation from our daily lives: People visit the vicinity area of locations that bear importance to them more often than roam far away. When we assign the tiered community structure, it naturally makes sense to have the node visit the outer tiers less frequently than the inner tiers, although this is not required for the theoretical derivation. In the simulations, we use two alternative time periods with a two-tier local community in each time period, and the parameters are listed in Table II (b). In the multiple

randomly placed communities layout, multiple communities are instantiated randomly to show that our theory is not limited to a single community. We use two time periods with two randomly placed communities each for this scenario. Other than the difference in community setup and sizes, we again use the parameters in Table II (b) for this case. Our discrete-time simulator is written in C++, and nodes move as described in Section III. More details about the simulator, as well as the simulator code, can be found at [37].

A. The Nodal Spatial Distribution

To observe the nodal spatial distribution, we divide the simulation area into a 20-by-20 grid and count the average number of nodes in each grid block during the simulation. The results presented in this subsection is the average of 5000 runs of independent simulations.

If the communities are randomly chosen, the node should appear at each of the 400 evenly divided grid equal-likely, with probability $1/400$. We observe that the spatial distribution of node varies a bit about this value in the simulation, as shown in Fig. 9 (a). The minor discrepancy is due to the finite number of samples. To make the scenario more interesting, we also generate the spatial distribution for nodes when the communities are fixed. We use the parameter sets of *Model-1* (one community in each time period) and *Model-5* (two-tier community in each time period) from table II, and assign the center of the community at either $(300, 300)$ or $(700, 700)$ with one half probability. The resulting nodal spatial distributions are shown in Fig. 9 (b) and (c), respectively. The node appears with higher probability where the communities are assigned. From Eq. (3), for the scenario in Fig. 9 (b), the node appears in the community with probability 0.0864 and in other area with probability 0.0008, respectively. For the scenario in Fig. 9 (c), the node appears in the first-tier community, the second-tier community, and the other area with probability 0.0759, 0.0039, and 0.0004, respectively. In both cases the simulation results follow the theoretical results reasonably well, within about 10% error for the area in the communities.

B. 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 runs last for 60000 seconds in this subsection.

As we show in Corollary 5.5, 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 mobility parameters. We illustrate this scenario in Fig. 10 (a), where we compare the evolution of the theoretical average node degree versus the communication range (K) to the simulation results for some of the models listed in Table II. The simulation curves follow the prediction

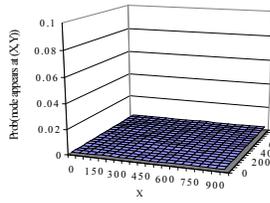
TABLE II
PARAMETERS FOR THE SCENARIOS IN THE SIMULATION

Common parameters: For simplicity, 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. In the simple model we use a single local community (with subscript l) in each time period. For the generic model, we test with two different configurations: (1) A two-tier community in each time period, as illustrated in Fig. 8 (a). In this scenario the inner tier community and the outer tier community has edge length C_{l1} and C_{l2} , respectively. (2) Two randomly placed communities in each time period, as illustrated in Fig. 8 (b). In this scenario the communities both have edge length C_{l1} , but the parameters correspond to the two communities are different (i.e., correspond to subscript $l1$ and $l2$ in the table). In all cases, there is also a roaming state (with subscript r) in which the node moves about the whole simulation area (i.e. the whole simulation area is a community).

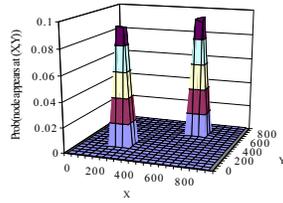
| (a) The simple model. | | | | | | | | | | | | | | |
|-----------------------|-------------------------------|------|------------|------------|---------------|---------------|---------|---------|-----------|-----------|-----------|-----------|-------|-------|
| Model name | Description | N | C_{l1}^1 | C_{l1}^2 | $D_{max,l}^1$ | $D_{max,r}^1$ | L_l^1 | L_r^1 | π_l^1 | π_r^1 | π_l^2 | π_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 |

| (b) The generic model. | | | | | | | | | | | | | |
|------------------------|------|------------|------------|------------|------------|----------------|----------------|---------------|----------------|----------------|---------------|-------|-------|
| Model name | N | C_{l1}^1 | C_{l2}^1 | C_{l1}^2 | C_{l2}^2 | $D_{max,l1}^1$ | $D_{max,l2}^1$ | $D_{max,r}^1$ | $D_{max,l1}^2$ | $D_{max,l2}^2$ | $D_{max,r}^2$ | T^1 | T^2 |
| Model 5 | 1000 | 100 | 300 | 100 | 300 | 25 | 15 | 1 | 30 | 25 | 3 | | |
| Model 6 | 1000 | 150 | 450 | 150 | 450 | 50 | 20 | 15 | 30 | 20 | 30 | | |
| Model 7 | 1200 | 160 | 480 | 160 | 480 | 50 | 20 | 15 | 30 | 20 | 30 | | |

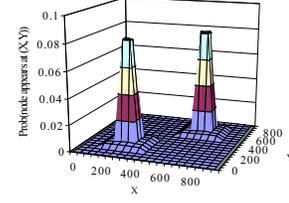
| Model name | L_{l1}^1 | L_{l2}^1 | L_r^1 | L_{l1}^2 | L_{l2}^2 | L_r^2 | π_{l1}^1 | π_{l2}^1 | π_r^1 | π_{l1}^2 | π_{l2}^2 | π_r^2 | T^1 | T^2 |
|------------|------------|------------|---------|------------|------------|---------|--------------|--------------|-----------|--------------|--------------|-----------|-------|-------|
| Model 5 | 300 | 500 | 1000 | 200 | 300 | 1000 | 0.6 | 0.3 | 0.1 | 0.85 | 0.1 | 0.05 | 5760 | 2880 |
| Model 6 | 100 | 300 | 1000 | 200 | 500 | 1000 | 0.5 | 0.35 | 0.15 | 0.7 | 0.2 | 0.1 | 5760 | 2880 |
| Model 7 | 140 | 600 | 1500 | 200 | 500 | 1600 | 0.8 | 0.15 | 0.05 | 0.7 | 0.2 | 0.1 | 5760 | 2880 |



(a) Randomly placed community.

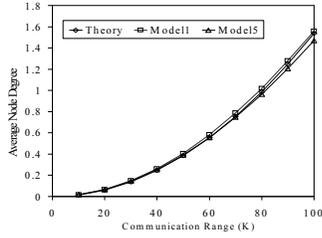


(b) Single-tier community centered at (300, 300) or (700, 700) with one half probability.

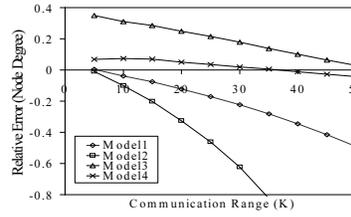


(c) Two-tier community centered at (300, 300) or (700, 700) with one half probability.

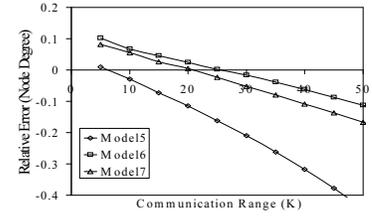
Fig. 9. Spatial distribution of the node (shown as the probability for a node to appear in each 50x50 grid block).



(a) Randomly placed community.

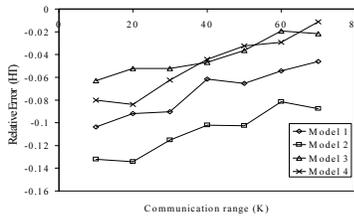


(b) Relative error for scenarios with fixed single-tier community.

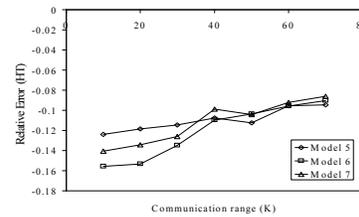


(c) Relative error for scenarios with fixed two-tier communities.

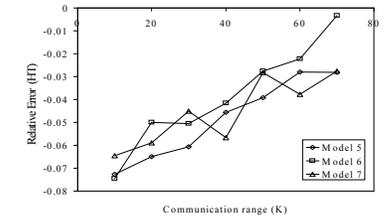
Fig. 10. Comparison of theoretical and simulation results (the average node degree).



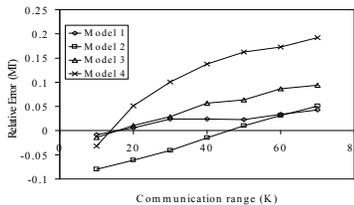
(a) Hitting time, simple model.



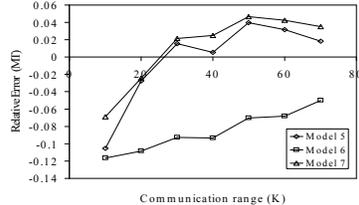
(b) Hitting time, multi-tier communities.



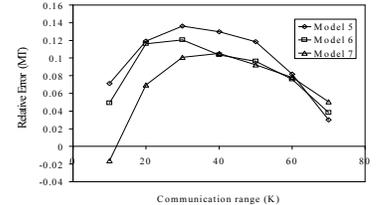
(c) Hitting time, multiple random communities.



(d) Meeting time, simple model.



(e) Meeting time, multi-tier communities.



(f) Meeting time, multiple random communities.

Fig. 11. Relative error between theoretical and simulation results (the hitting time and the meeting time).

of the theory well. Other configurations we tried (not listed here) also show similar trends.

Again, 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 listed in Table II. Models 1 through 4 correspond to scenarios with single-tier communities in each time period, and models 5 through 7 correspond to scenarios with multi-tier communities. We show the relative errors, calculated as $Error = (Theory - Simulation) / Simulation$, in Fig. 10 (b) and (c). A positive error indicates the theoretical value is larger than the simulation result, while a negative error indicates the converse. In the simulations, when the communication ranges are small as compared to the edge of the communities, the relative errors are low, typically below 10% except for *Model-3*, 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 Eq. (5) is no longer accurate since the communication disk extends out of the overlapped area in most cases. That is the reason we start to observe the discrepancies between the theory and simulation. Besides *Model-3*, we observe at most 20% of relative 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.

C. 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 listed in Table II, and compare the average results with the theoretical values derived from the corresponding equations (i.e. (8) and (17)). 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 the relative errors between the theoretical values and the simulation results for various scenarios in Fig. 11. We see that for all the scenarios, the relative errors are within acceptable range. These results display the accuracy of our theory under a wide range of parameter settings. The absolute values for the error are within 16% for the hitting time and within 20% for the meeting time. For more than 70% of the tested scenarios, the error is below 10%. The errors between the theoretical and simulation results are mainly due to some of the approximations we made in the various derivations. For example, there exist some border effects with respect to the hitting and meeting probabilities within a community. 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. Furthermore, the approximation of the hitting and meeting processes with discrete Bernoulli trials is valid only for the epochs that are large enough (in the order

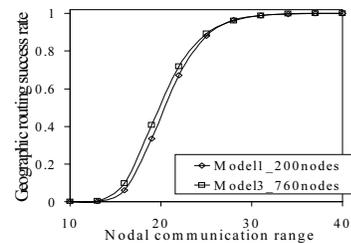


Fig. 12. Geographic routing success rate under different mobility parameter sets and node numbers.

of community size). 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 [42]. Thus, using the analytical results of Section V-B 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-B, 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 [43]). 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. 12 (each point is the percentage of success out of 2000 trials). If we assume now that the mobility model was different, say *Model-3*, we would like to know how many nodes we would need to achieve similar performance. Using Eq. (6) 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*. Comparing the resulting message delivery ratio for this scenario to the original scenario (200 nodes with *Model-1*) in Fig. 12, we see that similar success rates are achieved in both scenarios 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 connectivity is only intermittent (often referred to as Delay Tolerant Networks) [39]. It has been shown that message propagation under epidemic routing can be modeled with sufficient accuracy (assuming the number of nodes is large enough) using a simple fluid-based model [40]. (Note that its performance has also been analyzed using Markov Chain [33], [23] and Random Walk [5] models.) This fluid model has been borrowed from the Mathematical Biology community, and is usually referred to as the SI (Susceptible-Infected) epidemic model. 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, β (the implicit assumption there of course being that nodes meet independently). This contact rate in the SI model is equivalent to the unit-step meeting probabilities calculated in Section V-D. Thus, one could in essence plug-in these meeting probabilities into the SI model equations and calculate the delay for epidemic routing. 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 example, nodes with the same or overlapping communities tend to meet much more often than nodes in far away communities. For this reason, we extend the basic SI model to a more general scenario that is applicable to the TVC model.

We consider the following setup in the 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 the time needed until it reaches all other nodes in the network using epidemic routing is calculated. The propagation of the message can be described by the following equations:

$$\begin{cases} \frac{dI_1(t)}{dt} = \beta_{ov}I_1(t)S_1(t) + \beta_{no_ov}I_2(t)S_1(t) \\ \frac{dI_2(t)}{dt} = \beta_{ov}I_2(t)S_2(t) + \beta_{no_ov}I_1(t)S_2(t) \\ S_1(t) + I_1(t) = M/2 \\ S_2(t) + I_2(t) = M/2. \end{cases} \quad (20)$$

where $S_x(t)$ and $I_x(t)$ denote the number of susceptible and infected nodes at time t in group x , respectively. Parameters β_{ov} and β_{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 Eq. (15) to obtain these quantities. This model is an extension from the standard SI model [40] and similar extensions can be made for more than two groups [50]. 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

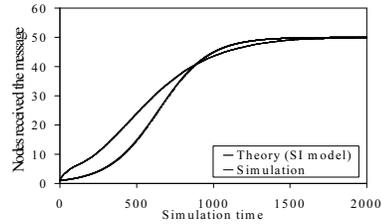


Fig. 13. Packet propagation with epidemic routing. The total population is divided into 2 groups with different community.

($I_2(t)$). We can solve the system of equations in (20) to get the evolution of the total infected nodes in the network. As can be seen in Fig. 13, the *theory* curve closely follows the trend in *simulation* curve (the non-perfect matching between the two curves, is due to the fact that fluid models become more accurate approximations of the actual stochastic spreading for large numbers of nodes). This indicates first that scenarios generated by our mobility model are still amenable to fluid model based mathematical analysis (SI), despite the increased complexity introduced by the concept of communities. It also shows that results produced thus 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. [22], [5], [40]), for a large range of mobility scenarios that can be captured by the TVC model.

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 re-appearance at the same location*. 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 [37]. In addition to providing realistic mobility patterns, the TVC model can be mathematically analyzed to derive several quantities of interest: the *nodal spatial distribution*, the *average node degree*, the *hitting time* and the *meeting time*. Through extensive simulation studies, we have verified the accuracy of our theory.

The TVC model can be easily generalized to provide scenarios in which nodes display more heterogeneous behavior. Nodes may have different set of parameters and even the time period structure can be different for different nodes. With these extensions, we have a mobility model to describe an environment including users with diverse mobility characteristics. We believe such a model is a very important building block for understanding protocol performances in real-life settings. To demonstrate this last point, we also provide some examples

of how our theory can be used in practice to predict protocol performance and guide design decisions.

In the future we would like to further analyze the performance of various routing protocols (e.g., [5], [6]) 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 communities and time periods of nodes 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 [46]).

ACKNOWLEDGEMENT

The authors thank Michal Piorkowski for pointing us to the cab spotting website and sharing archived data. This work is supported by NSF awards CNS-0520017 and career award 0134650.

REFERENCES

- [1] W. Hsu, T. Spyropoulos, K. Psounis, A. Helmy, "Modeling Time-variant User Mobility in Wireless Mobile Networks," In Proceedings of IEEE INFOCOM, May 2007.
- [2] Delay tolerant networking research group. <http://www.dtnrg.org>.
- [3] S. Jain, K. Fall, and R. Patra, "Routing in a delay tolerant network," In Proceedings of ACM SIGCOMM, Aug. 2004.
- [4] R. C. Shah, S. Roy, S. Jain, and W. Brunette, "Data mules: Modeling and analysis of a three-tier architecture for sparse sensor networks," Elsevier Ad Hoc Networks Journal, 2003.
- [5] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, "Efficient Routing in Intermittently Connected Mobile Networks: The Single-copy Case," to appear in ACM/IEEE Transactions on Networking, Feb. 2008.
- [6] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, "Efficient Routing in Intermittently Connected Mobile Networks: The Multi-copy Case," to appear in ACM/IEEE Transactions on Networking, Feb. 2008.
- [7] P. Gupta and P. Kumar, "Capacity of wireless networks," Transactions on Information Theory, 46(2), 2000.
- [8] M. Grossglauser and D. Tse, "Mobility increases the capacity of ad hoc wireless networks," IEEE/ACM Transactions on Networking, 10(4), 2002.
- [9] C. Perkins, "Ad Hoc Networking," Addison-Wesley, published Dec. 2000.
- [10] S. Ross, "Introduction to Probability Models," 8th edition, Academic Press, published Dec. 2002.
- [11] 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.
- [12] M. Balazinska and P. Castro, "Characterizing Mobility and Network Usage in a Corporate Wireless Local-Area Network," In Proceedings of MOBISYS 2003, May 2003.
- [13] N. Eagle and A. Pentland, "Reality mining: sensing complex social systems," in Journal of Personal and Ubiquitous Computing, vol.10, no. 4, May 2006.
- [14] M. McNett and G. Voelker, "Access and mobility of wireless PDA users," ACM SIGMOBILE Mobile Computing and Communications Review, v.7 n.4, Oct. 2003.
- [15] T. Henderson, D. Kotz and I. Akyoz, "The Changing Usage of a Mature Campus-wide Wireless Network," In Proceedings of ACM MOBICOM, Sep. 2004.
- [16] F. Bai, N. Sadagopan, A. Helmy, "The IMPORTANT Framework for Analyzing the Impact of Mobility on Performance of Routing for Ad Hoc Networks", AdHoc Networks Journal - Elsevier, Vol. 1, Issue 4, pp. 383-403, November 2003.
- [17] W. Hsu and A. Helmy, "On Important Aspects of Modeling User Associations in Wireless LAN Traces," In Proceedings of the Second International Workshop On Wireless Network Measurement, Apr. 2006.
- [18] C. Tudeuce and T. Gross, "A Mobility Model Based on WLAN Traces and its Validation," In Proceedings of IEEE INFOCOM, Mar. 2005.
- [19] R. Jain, D. Lelescu, M. Balakrishnan, "Model T: An Empirical Model for User Registration Patterns in a Campus Wireless LAN," In Proceedings of ACM MOBICOM, Aug. 2005.
- [20] D. Lelescu, U. C. Kozat, R. Jain, and M. Balakrishnan, "Model T++: An Empirical Joint Space-Time Registration Model," In Proceedings of ACM MOBIHOC, May 2006.
- [21] 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 Computing and Communication Review, Jan 2005.
- [22] T. Spyropoulos, K. Psounis, and C.S. Raghavendra, "Performance Analysis of Mobility-assisted Routing," In Proceedings of ACM MOBIHOC, May 2006.
- [23] R. Groenevelt, P. Nain, and G. Koole, "The Message Delay in Mobile Ad Hoc Networks," In Proceedings of PERFORMANCE, Oct. 2005.
- [24] M. Kim, D. Kotz, and S. Kim, "Extracting a mobility model from real user traces," In Proceedings of IEEE INFOCOM, Apr. 2006.
- [25] M. Musolesi and C. Mascolo, "A Community Based Mobility Model for Ad Hoc Network Research," In Proceedings of the Second International Workshop on Multi-hop Ad Hoc Networks (REALMAN), May 2006.
- [26] 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 Proceedings of IEEE INFOCOM, Apr. 2006.
- [27] J. Su, A. Chin, A. Popivanova, A. Goel, and E. de Lara, "User Mobility for Opportunistic Ad-Hoc Networking," In Proceedings of the 6th IEEE Workshop on Mobile Computing Systems and Applications, Dec. 2004.
- [28] A.B. McDonald and T.F. Znati, "A Mobility-based Framework for Adaptive Clustering in Wireless Ad Hoc Networks," In IEEE Journal on Selected Areas in Communications, vol. 17, issue 8, pp. 1466-1487, Aug. 1999.
- [29] R. Aldunate, S. Ochoa, F. Pena-Mora, and M. Nuaabaum, "Robust Mobile Ad Hoc Space for Collaboration to Support Disaster Relief Efforts Involving Critical Physical Infrastructure," In Journal of Comp. in Civil Engineering, vol. 20, issue 1, pp. 13-27, 2006.
- [30] J. Burgess, B. Gallagher, D. Jensen, and B. Levine, "MaxProp: Routing for Vehicle-Based Disruption-Tolerant Networks," In Proceedings of IEEE INFOCOM, Apr. 2006.
- [31] F. Bai, T. Elbatt, G. Hollan, H. Krishnan, V. Sadekar, "Towards Characterizing and Classifying Communication-based Automotive Applications from a Wireless Networking Perspective," in Proceedings of the 1st IEEE Workshop on Automotive Networking and Applications (AutoNet 2006), Nov. 2006.
- [32] 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 ASPLOS-X conference, Oct. 2002.
- [33] T. Small and Z. Haas, "The Shared Wireless Infostation Model: a New Ad Hoc Networking Paradigm (or where there is a whale, there is a way)," In Proceedings of the 4th ACM MOBIHOC, June 2003.
- [34] MobiLib: Community-wide Library of Mobility and Wireless Networks Measurements. <http://nile.usc.edu/MobiLib>.
- [35] CRAWDAD: A Community Resource for Archiving Wireless Data At Dartmouth. <http://crawdad.cs.dartmouth.edu/index.php>.
- [36] Cab Spotting, a project that tracks taxi mobility in the San Francisco Bay Area. Trace available at <http://cabspotting.org/api>.
- [37] Simulation codes used in this work and its detailed description are available at http://nile.cise.ufl.edu/~weijjehs/TVC_model
- [38] The Network Simulator - NS-2. <http://www.isi.edu/nsnam/ns/>
- [39] A. Vahdat and D. Becker, "Epidemic Routing for Partially Connected Ad Hoc Networks," Technical Report CS-200006, Duke University, April 2000.
- [40] X. Zhang, G. Neglia, J. Kurose, and D. Towsley, "Performance Modeling of Epidemic Routing," in Proceedings of IFIP Networking 2006.
- [41] P. Krishna, N.H. Vaidya, M. Chatterjee, and D.K. Pradhan, "A Cluster-based Approach for Routing in Dynamic Networks," In ACM SIGCOMM Computer Communication Review, vol. 27, issue 2, pp. 49-64, Apr. 1997.
- [42] K. Seada, M. Zuniga, A. Helmy, B. Krishnamachari, "Energy-Efficient Forwarding Strategies for Geographic Routing in Lossy Wireless Sensor Networks," In Proceedings of ACM Sensys, Nov. 2004.
- [43] B. Karp and H. Kung, "GPSR: greedy perimeter stateless routing for wireless networks," In Proceedings of ACM MobiCom, Aug. 2000.
- [44] C. Bettstetter, G. Resta, P. Santi, "The Node Distribution of the random waypoint mobility model for wireless ad hoc networks," in IEEE Trans. on Mobile Computing, vol. 2, issue 3, pp. 257-269, Jul. 2003.
- [45] A. Jindal and K. Psounis, "Fundamental Mobility Properties for Realistic Performance Analysis of Intermittently Connected Mobile Networks," In Proceedings of IEEE PerCom Workshop on Intermittently Connected Mobile Ad Hoc Networks (ICMAN), 2007.
- [46] W. Hsu and A. Helmy, "On Nodal Encounter Patterns in Wireless LAN Traces," the Second International Workshop On Wireless Network Measurement (WinMee 2006), April 2006.

- [47] J. Leguay, T. Friedman, and V. Conan, "Evaluating Mobility Pattern Space Routing for DTNs," in Proceedings of IEEE INFOCOM, April, 2006.
- [48] X. Hong, M. Gerla, G. Pei, C. and Chiang, "A group mobility model for ad hoc wireless networks," in Proceedings of the 2nd ACM international workshop on modeling, analysis and simulation of wireless and mobile systems, August, 1999.
- [49] A. Jardosh, E. Belding-Royer, K. Almeroth, S. Suri. "Towards Realistic Mobility Models for Mobile Ad hoc Networks." in Proceedings of ACM MobiCom, Sep. 2003.
- [50] S.Tanachaiwiwat and A. Helmy, "Encounter-based Worms: Analysis and Defense", IEEE Conference on Sensor and Ad Hoc Communications and Networks (SECON) 2006 Poster/Demo Session, September 2006.
- [51] T. Karagiannis, J-Y Le Boudec, M. Vojnovic, "Power Law and Exponential Decay of Inter Contact Times between Mobile Devices," In Proceedings of ACM MobiCom, Sep. 2007.