A Mobility Model Based on WLAN Traces and its Validation

Cristian Tuduce and Thomas Gross
ETH Zurich, Switzerland

Abstract— **The simulation of mobile networks calls for a mobility model to generate the trajectories of the mobile users (or nodes). It has been shown that the mobility model has a major influence on the behavior of the system. Therefore, using a realistic mobility model is important if we want to increase the confidence that simulations of mobile systems are meaningful in realistic settings.**

In this paper we present an executable mobility model that uses real-life mobility characteristics to generate mobility scenarios that can be used for network simulations. We present a structured framework for extracting the mobility characteristics from a WLAN trace, for processing the mobility characteristics to determine a parameter set for the mobility model, and for using a parameter set to generate mobility scenarios for simulations. To derive the parameters of the mobility model, we measure the mobility characteristics of users of a campus wireless network. Therefore, we call this model the *WLAN mobility model*. Mobility analysis confirms properties observed by other research groups. The validation shows that the WLAN model maps the real-world mobility characteristics to the abstract world of network simulators with a very small error.

For users that do not have the possibility to capture a WLAN trace, we explore the value space of the WLAN model parameters and show how different parameters sets influence the mobility of the simulated nodes.

I. Introduction

As the availability and popularity of wireless networks increases, the research community strives to offer new systems (network architectures, networking protocols, services, etc.) that take into account the user's mobility. In the early stages of research, such mobile systems are simulated. A simulator offers a cheap and easy platform on which mobile systems can be studied.

To generate the trajectories of the mobile entities (users, nodes, etc.), the simulators rely on a mobility model. The mobility model has a major influence on the performance of a mobile system [1], [2], [3]. Therefore, results obtained with an unrealistic model may not reflect the true performance of a system (be it protocol or application) in real environments.

Several mobility models have been proposed recently [1], [4], [3]. These models are based on assumptions about the node's mobility (e.g., nodes move in random directions at random speeds). The problem with these mobility models is that they lack validation against real environments. Using a validated mobility model, however, is important to increase the confidence that simulations of future systems are meaningful.

This work was funded, in part, by the NCCR "Mobile Information and Communication Systems", a research program of the Swiss National Science Foundation, and by a gift from Intel's Microprocessor Research Laboratory.

One method of measuring real-life mobility characteristics is monitoring the user mobility in a wireless network (WLAN) setting. However, WLAN traces are not directly suited for trace driven simulations, because they lack detailed location information. Yet, we show that WLAN traces contain enough location information to train a mobility model for network simulations. We develop a mobility model framework and show how this framework can be used to obtain a mobility model that yields the same mobility characteristics as the WLAN users.

Several research studies focus on the analysis of wireless networks. We unify the results obtained in wireless network analysis and propose a mobility model based on measured mobility characteristics. The model captures the mobility characteristics of users during working hours. We validate the WLAN mobility model using cross-validation and show that the model captures the real-life mobility characteristics with a very small error.

II. The mobility model

A mobility model is a set of rules used to generate trajectories for mobile entities. Mobility models are used in network simulations to generate network topology changes due to node movement. A network simulator must know the position of a mobile node at any one time. Using the exact node position the simulator can compute signal fading from one node to another and take actions based on the current network topology (e.g., determine the set of nodes that will receive a certain packet).

A mobility model uses an environment description to define the bounds of the simulated world. In addition to the bounds, the environment description can include obstacles or restrictions within the simulated environment (e.g., walls, streets). These restrictions directly influence the way nodes move: simulated humans must not walk trough walls, simulated cars must stay on the streets, etc.

At a high level of abstraction, mobility has two components: a spatial component and a temporal component. The spatial component describes *where* the mobile entity is moving, and the temporal component describes *when* an entity is moving and at which speed. Thus, when developing a mobility model, the two components of mobility must be clearly defined.

The general set of parameters required by a mobility model to build the simulated world contains: the simulated population size N, the simulation time t_{sim}, the environment description, the spatial mobility characteristics, and the temporal mobility

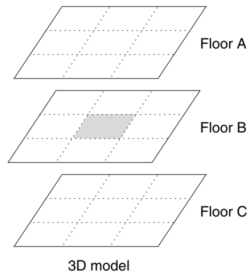

Fig. 1. The 3D model of the WLAN environment.

characteristics. The mobility model produces node trajectories that can be fed into a network simulator.

This paper presents a mobility model that imposes on the modeled nodes the same mobility characteristics as measured in a real wireless LAN. We use traces recorded in a WLAN setting to extract the spatial and temporal mobility characteristics required by a mobility model. The next sections describe our mapping of the real-world characteristics to the mobility model.

A. WLAN environment

We model the environment in which the nodes move as a set of *cells*. We call the area covered by an access point a cell. In the most general case, cells are modeled by cubic volumes. A widely used transmission range for wireless network simulations is 250m. We measure the transmission range of several IEEE 802.11 network adapters. The measurements show that the 250m setting is very optimistic. In an unobstructed space, the maximum distance for which we obtained reliable communication was 175m. The cube included in a sphere with a radius of 175m has a width of 202.07m. Therefore, we set the default cell width to $C_{width} = 200m$.

Wireless LANs can spread over buildings with multiple floors, each floor being covered by multiple cells. A simple 3D model of a 3 floor building is depicted in Figure 1. The dotted lines delimit the cells covering a floor. In this environment a cell may have up to 26 neighboring cells (like the shaded cell). Because of the relatively high number of users that roam to a neighboring cell on another floor (64% in our WLAN trace), we model the environment as a three dimensional volume. However, current network simulators simply ignore node elevation. To use the WLAN mobility model with such a network simulator, the 3D model of a building could be simply flattened out. The disadvantage of this method is that transitions between floors can be modeled only at the edge of a floor. Another way to cope with the flat environment of the current simulators is to use flattened building models and a transition function that maps the different stair-cases as jumps from cells on one floor to the corresponding cells on another floor. Although this method looks like random teleportation, it will map the 3D model of a building to the 2D simulated world.

We leave, however, the use of building models with two or three dimensions, and/or the choice of a 3D network simulator up to the user of the model.

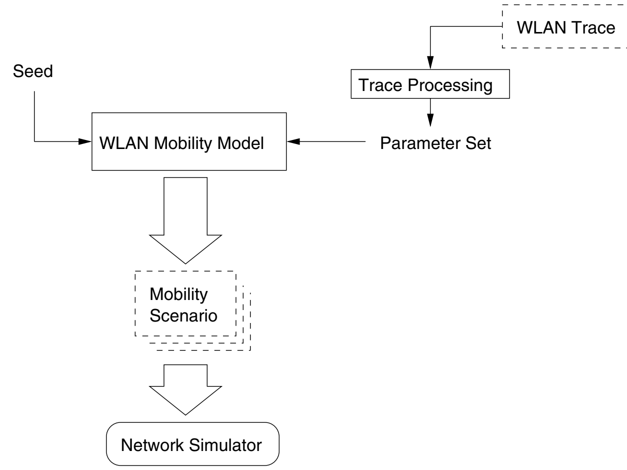

Fig. 2. The framework of the WLAN mobility model.

B. WLAN user behavior

The behavior of the nodes is modeled with active-inactive cycles. The terms *active* and *inactive* refer to the networking activity. In active state, a node is connected to the wireless network and is associated with an access point (or cell). A node in inactive state is not connected to the network. During this state, the node moves from one cell to another. A seamless movement between cells is modeled by an active-inactive-active sequence with an inactive period of zero seconds. These active-inactive cycles model the behavior of a real WLAN user: open device, work, close device, move.

III. WLAN MOBILITY MODEL FRAMEWORK

The values of the mobility model parameters are obtained by processing a WLAN trace. The exact processing of the WLAN trace is described in Section IV. The parameter values are used by the WLAN mobility model to generate different mobility scenarios. For a set of parameter values and different seeds, the mobility model generates different mobility scenarios. These mobility scenarios can be used by a network simulator to study the behavior of a mobile system. Our implementation of the WLAN mobility model generates mobility scenarios for the *ns-2* simulator. The framework of the WLAN mobility model is depicted in Figure 2.

A mobility model for network simulations can be seen as the collaboration between two processes: a spatial process $\mathcal{P}_S$ and temporal process $\mathcal{P}_T$. The result of this collaboration is the generation of node trajectories in a described environment. The environment is defined by the boundaries of the physical space in which mobile nodes can move, together with restrictions on the way nodes are allowed to move. The environment and its restrictions can be specified in multiple ways. For example, the Random Waypoint [5] model defines the rectangular area in which nodes are allowed to move and an empty set of restrictions (i.e., nodes are allowed to move to any point within the defined area). Another example of defining the

```
load environment description
for every simulated node do
  time := 0
  while time < t_sim do
    call P_S {select next destination}
    call P_T {generate timing}
    move to next destination
    time := time + current_session
  end while
end for
```

Fig. 3. The algorithm used by the WLAN mobility model to generate node trajectories.

environment and its restrictions is given by the Manhattan [6] mobility model. In this model, nodes are allowed to move only on grid-like streets. At an intersection, the nodes are only allowed to turn right or left or to proceed straight-ahead.

The first process used by the mobility model, the spatial process ($\mathcal{P}_S$), defines the spatial behavior of the mobile nodes. If we consider the simulated space as a volume of points, the spatial process defines the subset of points to be visited by a specific node during the life-time of a simulation. For example, the next destination of a node can be selected using a uniform random distribution, as implemented by the Random Waypoint model.

The temporal process ($\mathcal{P}_T$) defines the time component of the mobility model. Part of the temporal component of a model is, say, the time in which the node must reach the destination point (i.e., the speed of the node).

The movement of the nodes is modeled as a series of movement sessions between different points in the environment. The next destination of a node is selected by an invocation of the spatial process $\mathcal{P}_S$, and the duration of the current session is generated by the temporal process $\mathcal{P}_T$. The environment restrictions are enforced by the spatial process. After both spatial and temporal components of the next movement are defined, the node moves to the next destination. This iterative session generation is repeated for the entire duration of the simulation. The algorithm used to generate node trajectories is summarized in Figure 3.

IV. The model parameter sets

In this section we describe what information users of the WLAN mobility framework must record and the way she they map this information to a network simulator. We illustrate how the framework can be used to generate the realistic mobility characteristics of a WLAN.

To capture the spatial and temporal characteristics required by the mobility model, we analyze user mobility using two mobility metrics: prevalence and persistence, as introduced by Balazinska and Castro [7]. Prevalence and persistence characterize mobility independent of the duration of the mobility trace and the time users spend on the network. Prevalence captures the spatial component of user movement; it is the fraction of time a user spends at a given location. The temporal behavior of the users is captured by persistence, which measures the amount of time a user spends at a given location. The isolation of the spatial and temporal characteristics of the WLAN mobility helps us define the spatial and temporal processes needed by the mobility model.

As described in Section II-A, the environment in which the modeled nodes move is described as a set of cells within a building. The width of a cell is $C_{width} = 200m$. The dimension of the building is specified by the number of cells that cover the three physical dimensions C_x, C_y, C_z.

A. WLAN traces

We gather mobility data from a wireless network spread across 32 buildings. These buildings host seminar and lecture halls as well as offices. The university campus does not include student dormitories.

The WLAN infrastructure is built using Cisco Aironet 340 and 350 access points, configured to run in infrastructure mode. There are 166 access points installed throughout the campus (i.e, there are 166 cells). With the exception of seven lecture halls that are heavily used (occupied), the access points are placed to minimize the overlapping of the cells; these seven lecture halls are covered by two access points each.

1) Methodology: Mobility data was gathered in two traces. The first trace captures user mobility information over nine weeks, from May 19th 2003 to July 9th 2003. The second mobility trace captures mobility information over nine weeks, from April 1st 2004 to May 31st 2004. Depending on the WLAN infrastructure, a user can gather the mobility information in various ways. For instance, to gather the mobility information for the first trace, we query the wireless access points (APs) using the SNMP protocol. The APs were polled every minute for user association information. For the second trace, the access points were configured to send syslog messages to a server whenever a host was either connecting to the wireless network, disconnecting from the network, or roaming between access points. Using this information gathering technique, we can record the mobility events with high timing accuracy. The access points send three types of events: association, disassociation, and roaming messages. Association and disassociation messages indicate the moments when WLAN users connect or disconnect to/from the network. Roaming messages indicate a seamless movement of a WLAN user from one AP to another AP. Roaming messages are followed by an association event from the new AP and a disassociation message from the old access point.

We consider each unique MAC address as a separate user. In reality, it is possible that a user has multiple network cards, or that users exchange cards among them. From our observations, students swap network cards very rarely, therefore, the assumption that each MAC address is associated with one user seems to be correct. To ensure user privacy, the MAC addresses and the access point names were anonymized.

2) Analysis: We analyze the WLAN traces to validate them against traces gathered by other researchers. The comparison relies on the temporal behavior of the WLAN users.

The first trace captures 3073 users that connected to the wireless network 97575 times. 414 users connected to the network just once; this accounts for 13.47% of the user population. For the duration of the second trace 4762 users connected to the wireless network 343626 times. 32 users connected to the network just once during the nine weeks of the trace; this accounts for 0.65% of the user population. The most active user was seen 3859 times, visiting 25 access points. For the time period between the two traces (almost one year), the user population increased with 54.96%, while the networking presence (by means of session numbers) increased with 252.16%. These trends indicate an increase in the popularity of the wireless network in this campus.

We analyze the patterns in the number of users accessing the network simultaneously. The number of connected users follows the weekly trends expected in a facility where users access the network mostly during the working hours. The number of connected users is significantly lower on Saturdays and Sundays. Within a day, the number of users increases starting with 6 am and diminishes slowly after 6 pm, with a slight decrease over lunch time. The cyclic pattern of the number of connected users confirms the observations made in other networks used during working hours [7], [8], [9].

3) Traces used: To study the degree of similarity between multiple mobility scenarios derived from different traces recorded in the same environment, we split both WLAN traces into multiple traces. Since both WLAN traces account for about two month, we split both traces in two halves. In Section V we show that this arbitrary split does not influence the choice of the model's parameter space. The four traces account for the WLAN mobility during the following periods: mid May to mid June 2003, mid June to July 2003, April 2004, and May 2004.

B. Spatial node distribution

As described earlier in this paper, the spatial mobility characteristics are defined by the spatial process ($\mathcal{P}_S$) of the mobility model. Before the spatial process is first invoked for a specific node, the initial place of each node must be determined. The nodes are placed into cells using a uniform random distribution. This is the initial place from which each node will start its trajectory.

When the spatial process is first invoked for a node, the spatial process must already have an idea of how much of the simulated space the node will cover during the life-time of the simulation. Therefore, before the simulation begins, each node is assigned a set of cells the node will visit for the duration of the simulation. We call this set the *checkpoint* set (V).

The size of the checkpoint set is the number of cells a node must visit $|V| = n_c$. This number is determined by analyzing the WLAN traces. We use a distribution global to all nodes to determine the number of cells each node will visit (n_c). The set of visited cells $A = \{\alpha_i | 1 \leq i \leq A_{max}\}$ stores the distribution of the node population over the number of visited cells. α_i is the fraction of the population visiting i cells (for α_i of the simulated population $n_c = i$). Using different α_i values results in generating mobility scenarios with different mobility characteristics. The maximum number of cells a node can visit is A_{max}. The maximum value of A_{max} is the total number of cells in the environment.

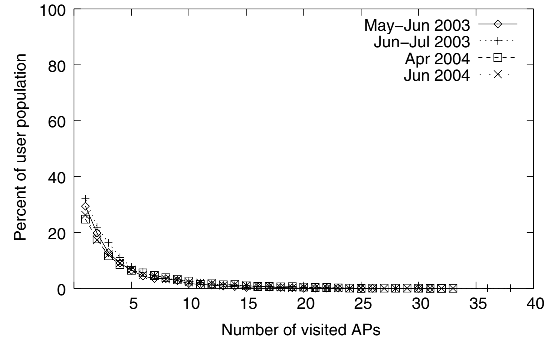

Fig. 4. The distribution of the number of visited access points.

The distributions ($A = \{\alpha_i\}$) observed in the WLAN traces are depicted in Figure 4. The traces are similar and we observe that 30%-32% of the user population visited just one access point. The distribution of the users over the number of visited access points follows a power law distribution with only a very low percentage of users visiting more than 20 access points. Therefore, we set $A_{max} = 20$ and we use the distributions depicted in Figure 4.

The checkpoint set of a node is built iteratively, starting with the cell where the node was initially placed. The size of the checkpoint set is equal to the number of cells a node must visit $|C| = n_c$. This set is built iteratively, starting with the cell where the node was initially placed. At each iteration, the model makes a probabilistic decision to choose the next cell of the node. The next checkpoint can be either the same cell, one of the neighboring cells, or a non-neighboring cell. The probability that a node moves from a cell (x, y, z) to a neighboring cell $(x \pm 1, y \pm 1, z \pm 1)$ is p_{neigh}. The node moves to a non-neighboring cell with the probability p_{non_neigh} or remains in the same cell with the probability p_{same}. The transition probabilities must sum up to 1:

$$p_{same} + p_{neigh} + p_{non\_neigh} = 1.$$

We use the WLAN trace to determine the transition probabilities. To measure the number of movements to neighboring or non-neighboring cells we rely on information about the topology of the access points. This information can be extracted from the WLAN trace without the need of processing a separate map of the WLAN. The roaming messages delivered by the access points offer a powerful clue about access point placement. A roaming message contains the access point *from* which the node is roaming and the AP *where* the node is moving. For a roaming event to occur, the two access points must be neighbors.

In the May-June 2003 trace, for example, 30% of the movements had a neighboring cell as destination. 34% of the movements were to non-neighboring cells. The rest of

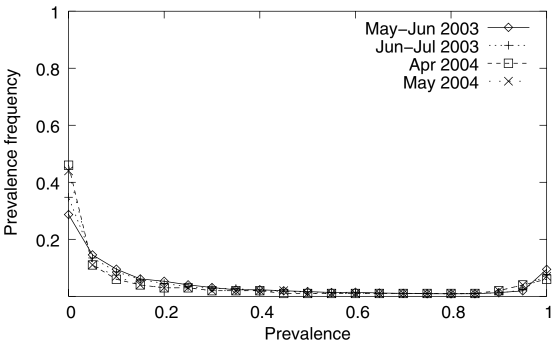

Fig. 5. Prevalence distribution for different WLAN traces.

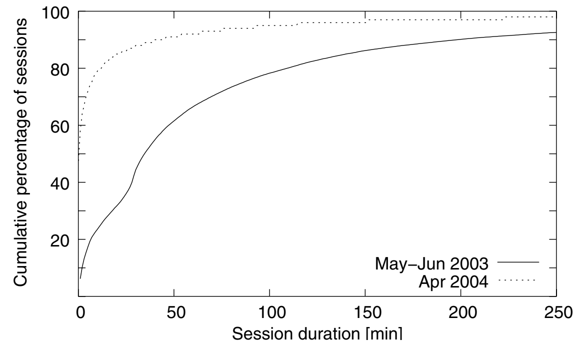

Fig. 6. The session length distribution.

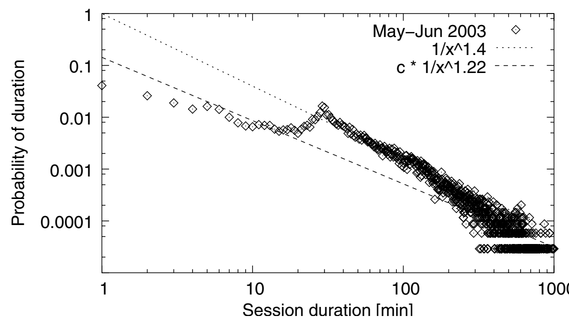

Fig. 7. Session length probability for the May to mid June 2003 trace .

36% were consecutive sessions at the same location. Based on this information, we set the three movement probabilities to $p_{same} = 0.36$, $p_{neigh} = 0.30$ and $p_{non\_neigh} = 0.34$.

The spatial mobility characteristics are captured by prevalence. Prevalence is the fraction of time a user spends at a given location. Each user has a non-zero prevalence value (fraction of time spent at an access point) for each access point the user ever visited. The prevalence distribution is the histogram of the prevalence values, normalized by the number of users. To ease the analysis, prevalence values are captured in bins of 0.05.

The prevalence distribution of the four WLAN traces is depicted in Figure 5. For prevalence values between 0.01 and 0.85, the prevalence frequency follows a power-law distribution. Prevalence values between 0.90 and 1 correspond to the mostly stationary nodes (nodes that spend 90%–100% of their time in the same cell). The distribution of prevalence values shows that the fraction of time users spend in different cells is unbalanced. Users spend most of their time in one cell and visit other cells very shortly. This aspect is common to all four wireless traces.

To sum up the preparations for the spatial process invocations, each node was assigned the number of cells it must visit n_c and the checkpoint set V (the set of cells the node will visit for the duration of the simulation). At each invocation, the spatial process $\mathcal{P}_S$ selects a cell from the checkpoint set as the new destination of the node. The new destination is selected from V using a uniform random number generator. In Section V we show that although the selection of the next cell is done by a random process, the prevalence distribution of the modeled nodes is induced by the checkpoint set selection (and follows the distribution depicted in Figure 5).

C. Session length distribution and movement decisions

For the duration of the simulation, the nodes follow inactive-active cycles. During the active state, a node is associated with a cell and it is actively participating in the network. For the duration of the inactive state, a node is not part of the wireless network. The timings the model generates are alternatively active and inactive periods. The temporal process $\mathcal{P}_T$ can therefore be seen as a process alternatively invoking an active timer process and an inactive timer process.

The duration of the inactive state is modeled by a uniform random distribution. The nodes move during the inactive period to the next cell. The duration of the inactive period is adjusted to the travel distance, so that a node does not move faster than a maximum speed v_{max}. For our model we choose $v_{max} = 3m/s$; this speed is an approximation of the speed of a running human.

To capture the temporal aspect of WLAN user mobility, we analyze the persistence of network sessions; the persistence measures the amount of time a user spends at a given location. Figure 6 depicts the cumulative percentage of user sessions against session duration. The four WLAN traces are pairwise similar. The graphs capture 92% of the sessions for the 2003 trace and 97% of the sessions for the 2004 trace. The rest of the sessions are very long sessions. However, most of the captured sessions are very short. For the 2003 trace, 50% of the sessions are up to 35 minutes; 66% of the sessions are shorter than 1 hour. In the 2004 trace, 75% of the sessions are up to 7 minutes; 92% of the sessions are shorter than one hour.

Figures 7 and 8 depict the probability with which a session has a certain length. The probability that a session has the duration S is computed as the sum of all sessions of length S divided by the total number of sessions. The probability that a session lasts longer drops following the shape of a power-law function with a low exponent. The traces are again pairwise similar. For the 2003 traces, the probability that a session lasts longer than 30 minutes follows the shape of the function $1/x^{1.4}$. Although this function is a good

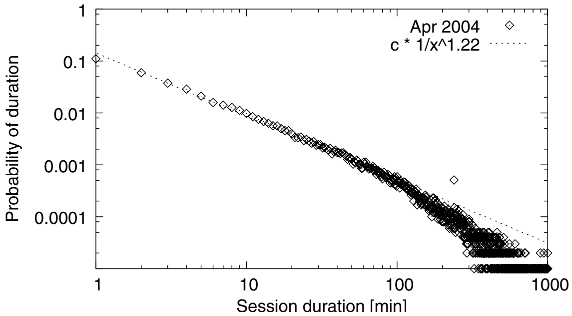

Fig. 8. Session length probability for the April 2004 WLAN trace.

approximation for long sessions, the probability that a session is shorter than 30 minutes is better approximated by the function that approximates the 2004 traces: $c \cdot 1/x^{1.22}$, where $c = 1/7$.

Therefore, the duration of the modeled active state (the session time) is generated using a general Pareto distribution with a shape parameter of 0.22 and a scale parameter of 0.15. This distribution is fitted from the WLAN trace and approximates the probability function $c \cdot 1/x^{1.22}$ (depicted in Figure 8). Using this distribution, short sessions appear with higher probability than long sessions.

To sum up, the user of a model must define the spatial and temporal processes used by the WLAN mobility model. The parameters of the model are summarized in Appendix A. Three of the parameters (population size, simulation time and maximal speed) are common parameters to almost every mobility model. Specific to the WLAN mobility model are the distribution of visited access points and the movement probabilities. An in-depth comparison of the parameters used by different mobility models is presented in Section VII.

The mobility characteristics measured in our network are close to the characteristics reported by Balazinska and Castro [7] and by Balachandran et al. [10]. Prevalence and persistence follow power-law distributions with low exponents. Users spend most of their time at one location, and short periods of time at other locations. Most of the sessions are short, and the probability that a session lasts longer decreases rapidly.

V. WLAN MODEL VALIDATION

The validation of the WLAN mobility model can be done at the mobility scenario level or at the network simulator level. To validate the mobility model at the simulator level, we should both implement (for the real-world) and simulate a mobile system, and then compare the behavior of the simulated system with the behavior of the real system. The drawback of this method is that network simulators still lack accuracy in simulating some aspects of the real world (e.g., propagation properties of the obstacles). Therefore, we validate the WLAN model at the mobility scenario level.

Using the parameters extracted from the WLAN traces, the four parameter sets can be used to generate node trajectories for modeled users. To investigate at which extent the four parameter sets capture the same mobility characteristics as recorded in the WLAN traces, we compare the spatial and temporal characteristics of the generated trajectories with the recorded mobility characteristics of the WLAN users.

A. Validation methodology

We use cross-validation to validate the prevalence and persistence distributions. The simplest type of cross-validation, called *holdout method*, splits the data set in two sets: a training set and a test set. The model is fitted using the training set only. The model is then used to predict the values for the data in the testing set. A penalty function is used to quantitize the error, which is the difference between the predicted values and the testing set. The errors are accumulated to give the mean test error. This error is used to assess the accuracy of the model. One of the commonly used penalty functions for cross-validation is the mean squared error. The predicted value is compared with the test data and their difference is squared. The error is cumulated and averaged over the number of data points:

$$E = \frac{1}{n} \sum_{i=1}^{n} (P_i - T_i)^2$$

where P_i is the value predicted by the model and T_i is the value from the training set. A perfect model matches the training data in every measurement point ($E = 0$). Lower values of the error are better than higher values, since they indicate a better matching model.

B. WLAN mobility model parameter sets

To perform the cross-validation, we split each WLAN trace in two parts. A part of the WLAN trace is used to extract the parameters of the mobility model. These parameters are used to generate simulated node trajectories. The second part of the trace is then used to validate the generated spatial and temporal distributions.

The size of the testing set varies from one measurement (called *leave-one-out* method) to an arbitrary fraction of the data set. Since the leave-one-out method is meaningless in the context of the WLAN mobility model, we search for the smallest reasonable testing set. We search for the smallest user population that still maintains the mobility characteristics of the entire population. We use binary search and therefore iteratively halve the user population. At each step, the nodes are selected at random from the entire trace. For each population size we analyze the mobility characteristics of the reduced population. The mean squared error for each iteration is depicted in Figure 9. We observe that the error increases dramatically after the 4th iteration. At this iteration the population size is 191 users. We round this figure up to 200 users and use it as the size of the WLAN testing sets.

For each of the four WLAN traces, we quarantine 200 randomly selected users and set their traces aside for testing. We use the remaining of the traces to extract the parameters needed by the mobility model. The parameters used by the WLAN mobility model and the way these parameters are extracted from the traces is described in Section IV.

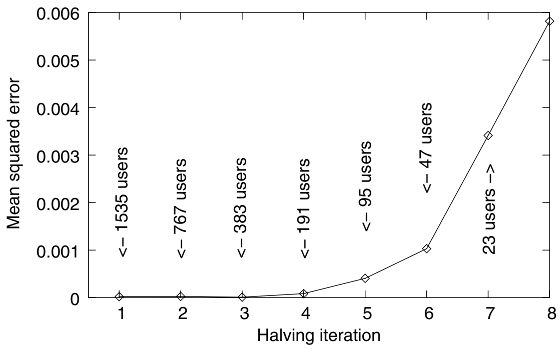

Fig. 9. The mean squared error for each population halving iteration.

TABLE I

THE TRANSITION PROBABILITIES OF THE FOUR PARAMETER SETS.

Trace	p_{same}	p_{neigh}	p_{non_neigh}
May – June 2003	0.3609	0.3079	0.3312
June – July 2003	0.4128	0.3076	0.2795
April 2004	0.4539	0.4752	0.0710
May 2004	0.4692	0.4653	0.0654

After isolating 200 users for the test sets, the distribution of the user population (in the training set) over the number of visited access points is the same with the distributions depicted in Figure 4. These distributions are part of the parameter set used by the mobility model (distribution A described in Section IV-B). To determine the checkpoint set of each node we have to extract from the traces the probabilities with which nodes will choose the next destination. This operation relies on the graph of the access points placement; we extract this graph from the WLAN trace as well. We use the AP graph to compute the transition probabilities of the mobility model and we summarize the values in Table I. The difference between the traces recorded in 2003 and 2004 are given by the increase of the user population (from 3073 in 2003 to 4762 in 2004) and the increase in the size of the WLAN (from 143 APs in 2003 to 166 in 2004).

To fit the session length distribution of the mobility models we use a general Pareto distribution with the shape parameter $a = 0.22$ and a scale parameter $b = 0.15$. These values are given by the test distributions depicted in Figures 7 and 8.

We simulate 200 nodes for 66000 seconds, which yields roughly the same number of sessions as in the test traces. For evaluation purposes the simulation time is determined experimentally, such that we obtain about the same number of sessions as recorded in the test set traces. When evaluating a mobile system, t_{sim} is determined by the simulated mobile system properties and the evaluation strategy.

C. Spatial distribution evaluation

Once the values of the parameters are set, we generate node trajectories using the WLAN mobility model. We simulate 200 mobile nodes, moving in a space containing 150 cells ($C_x = 10$, $C_y = 5$, $C_z = 3$).

We use the traces generated by the mobility models and compare them with the WLAN test sets. We generate the prevalence distributions for the generated nodes and for the test set users and apply the mean squared average as a penalty function. The difference between the modeled nodes and the test set users is summarized in Table II. The errors are small and have the same magnitude with the errors used to stop the halving iteration process when we searched for the test set size (see Figure 9).

TABLE II

THE MEAN SQUARED ERROR BETWEEN THE PREVALENCE DISTRIBUTIONS OF THE GENERATED NODES AND WLAN USERS.

Trace	Mean Squared Error
May – June 2003	0.00143
June – July 2003	0.00575
April 2004	0.00033
May 2004	0.00034

D. Session length distribution validation

To compare the session time distributions, we sort the generated and test data and plot them on the same graph. The sessions generated by the mobility model trained with the April 2004 data are depicted in Figure 10. Ideally the generated sessions should have the same duration as the sessions in the test set and, therefore, be located on the dotted median. We observe that the mobility model estimates pretty well the sessions shorter than 120 minutes and slightly underestimates longer sessions.

To quantitize the fitness of the session length distribution we apply a penalty function on the generated and test set distributions. We use the error average instead of the mean squared errors, to express the error with the same unit as used in the distribution (minutes). The deviation of the generated session length from the test data is depicted in Figure 11. The deviation of the two distributions (generated and test set) is generally small. For the 2004 traces the mean error is less than 3 minutes for sessions of up to 400 minutes. Since 97% of the sessions recorded in the 2004 traces are shorter than 250 minutes (see Figure 6), we consider this error acceptable. For the 2003 traces, the mean error increases quicker than the error of the 2004 traces. At session length of 400 minutes, the models generate an error of 14 and 18 minutes (for the two 2003 test sets).

To sum up, in this section we showed that the WLAN mobility model maps the real-world mobility characteristics to the abstract world of network simulators with a very small error. The four parameter sets of the mobility model are able to generate sessions for the modeled nodes with very small error rates, both for the spatial component of the mobility, as well as for the temporal component.

VI. THE VALUE SPACE OF THE MODEL PARAMETERS

In this section we analyze the parameter space of the WLAN mobility model and study how different values of the parameters influence the mobility of the modeled nodes. To facilitate the comparison of the WLAN mobility model with

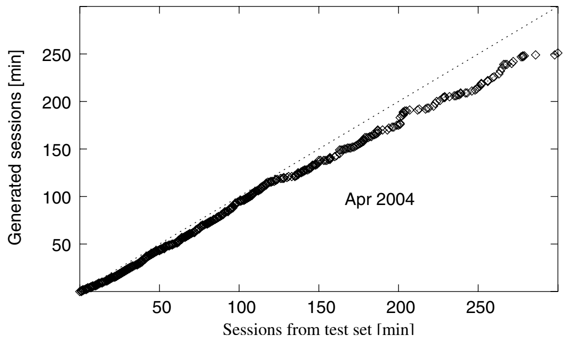

Fig. 10. Comparison between session length generated by the WLAN mobility model and session length from the test set.

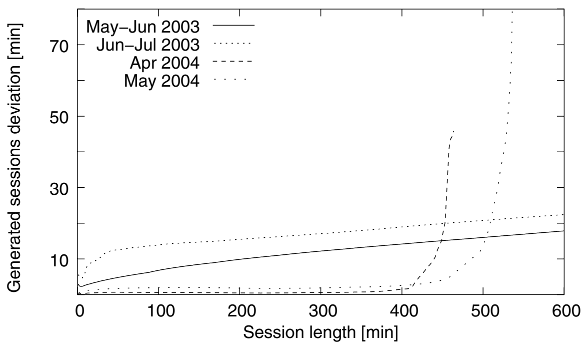

Fig. 11. The deviation of the generated sessions against session duration.

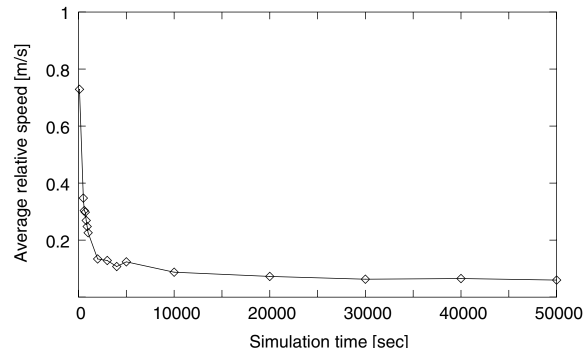

Fig. 12. The average relative speed as a function of simulation time.

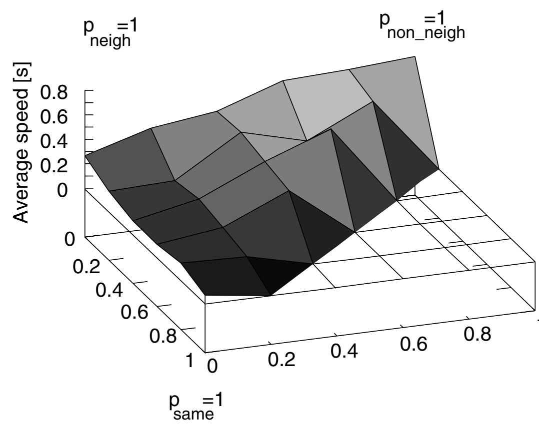

Fig. 13. The average speed depending on the movement probabilities.

other mobility models presented in the literature, we use the average speed and the average relative speed. The average speed is computed as the average speed of all modeled nodes at every time unit during the simulation. The average relative speed uses the standard definition from physics.

To determine the parameter space of the WLAN mobility model, we analyze the effect of each parameter on the mobility of the generated scenario. The first parameter we consider is the simulation time t_{sim}. The mobility characteristics of the WLAN model should be independent of the simulation time. We observe however, that the common simulation time of 900 seconds, used in most of the mobile system studies, is too short for the model to display the mobility characteristics of the WLAN. Depicted in Figure 12 is the average relative speed of the modeled nodes plotted against simulation time. The average relative speed stabilizes at about 3000-4000 seconds of simulation time.

Since the probabilities used to select the next destination of a node ($p_{same}, p_{neigh}, p_{non\_neigh}$) have to sum up to 1, we consider them together in our analysis. To assess the influence of the three probabilities on the mobility of the simulated nodes, we vary each of the probabilities from 0 to 1 (in 0.2 increments). We generate mobility scenarios for each of the probability values and compute the average speed of the modeled nodes. The average speed plotted against probability variance is depicted in Figure 13. The measurements show that for $p_{same} = 1$ the average speed is low. For this parameter setting, the mobility is low because the nodes choose a location in the same cell as their new destination. The average speed increases when $p_{neigh} = 1$, because the next destination of a node will always be a neighboring cell. The average speed reaches its peek when $p_{non\_neigh} = 1$. In this case, the next destination of a node is a non-neighboring cell.

If we consider the three probabilities as the three dimensions of a space, the value space of all three parameters are points on the plane that cuts the three axes at the value of 1. The value space of the three probabilities and the abstract way they influence node mobility is highlighted in Figure 14.

The population size (N) together with the environment size (C_{width}, C_x, C_y, C_z) determine the density of the simulated nodes. Since the node trajectories are generated independently for each node, the mobility should not depend on the node density. To study the influence of node density on the mobility, we generate mobility scenarios for 200 nodes moving in increasingly smaller environments (from 140 cells to 40 cells). This setting generates nodes densities commonly used in network simulations: 20-80 nodes per square kilometer. As expected, the average speed of the nodes for different node densities did not show a clear trend. Depicted in Figure 15 is the average speed plotted against node density, for two runs of the WLAN mobility model with the same parameter set.

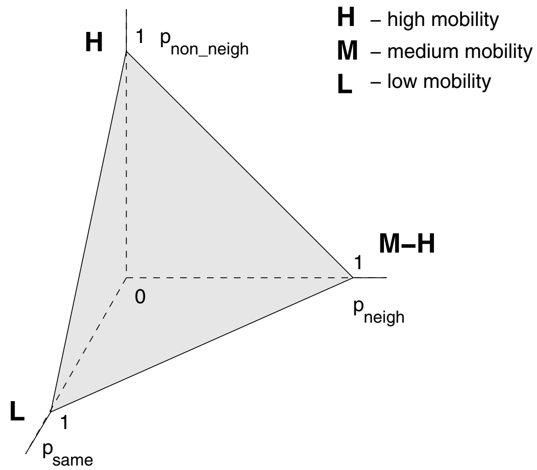

Fig. 14. The mobility degree depending on the next destination probabilities.

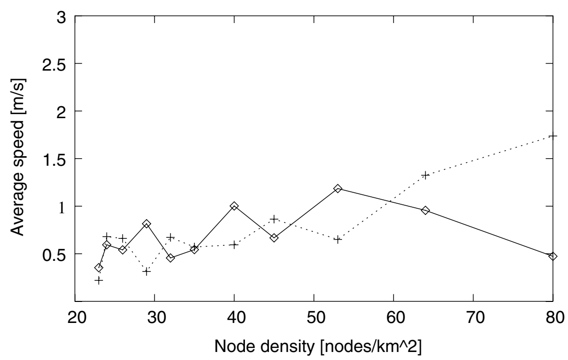

Fig. 15. Average speed vs. node density (two runs of the WLAN mobility model).

VII. Related work

The work presented in this paper builds on previous work done in two research areas: mobility modeling for network simulations and wireless network analysis. We first present other mobility models for network simulation; we briefly describe the functionality and the parameter sets of each model. For an in-depth study of various mobility models, we refer the interested reader to the survey done by Camp et al. [1].

One of the most popular mobility models for mobile networking research is the Random Waypoint (RW) model [5]. This model is based on the assumption that the user's movements follow a walk-and-pause cycle. The environment in which the nodes move is described as a flat surface without restrictions. The spatial and temporal processes are uniform random number generators. A user chooses a random destination and moves to that destination at a randomly chosen speed. Once at the destination, the user stops for a pause time. The walk-pause cycle is repeated for the entire time of the simulation. The parameters of the model are the population size, the simulation time, the environment size, the maximum speed, and the maximum pause time. The number of parameters is relatively low, since the only restrictions on the node movement are the maximum speed and the bounds of the environment. Variants of the RW model use different components of the physical mobility (e.g., direction and travel time) to determine the next destination. However, the main idea remains the same: random destinations are approached at random speeds. Based on observation from the WLAN trace, we show that the mobility in a WLAN setting is not modeled properly by a completely random model. The parameters that are not modeled properly by random distributions are session time and next destination.

Hong et al. [4] propose the Reference Point Group Mobility Model (RPGM). This model is based on the assumption that real users tend to move in groups. The path of a group is determined by the movement of a logical center. The members of a group move randomly in the neighborhood of the logical center. The path of a group is defined explicitly by a set of checkpoints along with the corresponding time intervals. Besides simulation time and population size, the parameters of the RPGM model are a group motion vector and a random motion vector. The random motion vector determines how the nodes move around the center of the group. The random motion vector is basically a uniform random number generator that selects the next destination of a node as a random point within a specified radius from the group center. The group motion vector determines the path of the logical center of the group. This vector is not defined by the mobility model itself, leaving open the path the group will follow.

Jardosh et al. [3] propose a mobility model in which the modeled users walk around predefined (rectangle) obstacles. A Voronoi diagram is used to determine the path of the mobile nodes. The Voronoi diagram is formed by the corners of the defined obstacles; it's a planar graph whose edges are line segments that are equidistant from two obstacle corners. However, users still move between randomly selected destinations at randomly selected speeds. This model can be seen as a variation of the Random Waypoint model, where the environment limits the trajectories of mobile nodes to the Voronoi graph.

The second research area our work builds on is wireless network analysis. Along with the increased availability and popularity of wireless network installations, WLAN analysis was used to determine the usage and usability of WLANs. Tang and Baker [9] traced the mobility of 74 users in a campus network for 12 weeks and a metropolitan-area network with 24773 radios [11]. The authors examined the network traces for overall user behavior, network traffic and load characteristics. The user mobility is analyzed, but no steps were taken in the direction of modeling user mobility.

Kotz and Essien [8] examined a campus wireless network for eleven weeks. They captured a large user population – 1706 distinct users (network cards) – connecting to 476 access points, spread over 161 buildings. Their work builds on the analysis done by Tang and Baker [9] and confirms previous observations for a larger WLAN setup.

Balachandran et al. [10] analyzed user mobility and network usage during the three days of a major conference. They examined the behavior of 195 users connecting to four access points. The authors study the behavior of the wireless user by means of traffic and mobility. Balachandran et al. do a first step in the direction of WLAN user mobility modeling by proposing a simple mobility model. The user arrival times are modeled by a Markov-Modulated Poisson Process and the session length by a Pareto distribution. Given the relatively small WLAN setting (4 access points), no modeling of the spatial mobility component is done. We model the temporal as well as the spatial characteristics of the WLAN mobility.

Balazinska and Castro [7] analyzed user mobility and network usage for a corporate network. The network spans over three buildings with 177 access points. The authors captured the behavior of 1366 distinct users, connected to the network over a period of four weeks. In their WLAN analysis, the authors use persistence and prevalence as metrics to model user mobility. No attempt is made in modeling the mobility of the WLAN users. The overall analysis of the WLAN trace confirms the observations from previous studies: WLAN users spend most of their time at one location, visiting other locations briefly.

Schwab and Bunt [12] investigate the usage of a campus wireless network of 18 access points. The authors captured the activity of 134 unique users during a one week period (in January 2003). They present a new trace collection methodology based on the LEAP authentication system. The authors analyze where, when and how much the wireless network is being used. Although their analysis represents an important step towards understanding WLAN mobility, the authors do not attempt to model the mobility.

Based on the insights provided by previous WLAN analysis, we provide a method to extract the mobility characteristics recorded in a WLAN trace and use them (in the form of model parameters) to generate mobility scenarios for network simulations.

VIII. Discussion

In this section, we compare the WLAN mobility model with other mobility models. To compare the mobility models, we use the IMPORTANT framework [6] described by Bai et al.. IMPORTANT is a framework to systematically analyze the impact of mobility on the performance of routing protocols. Bai et al. define the mobility parameters of the connectivity graph, and study their impact on the routing protocols. The parameters that have an influence on the mobile system's behavior are the average relative speed (aRS), the average degree of spatial dependence (aSD), and the average link duration (aLD). For the relative speed, the authors use the standard definition from physics. The degree of spatial dependence is the extent of similarity of the velocities of two nodes not too far apart. The average link duration is the duration of a link averaged over node pairs.

We compare the WLAN mobility model with the Random Waypoint (RW) model and the Reference Point Group Model (RPGM). Bai et al. compare the Random Waypoint model, the RPGM model, the Manhattan model and the Freeway model.

TABLE III
Comparison of the WLAN mobility model with the Random Waypoint model (RW) and the RPGM model.

Model	Rel. speed [m/s]	Link duration [s]	Sp. dep.
WLAN model	0.13	42.67	0.03
RW model	4	100	0.025
RPGM model	1	900	0.5

We do not consider the Manhattan and the Freeway models because they are not suited for WLAN-like environments. For RW and RPGM models, we use the results published by Bai et al. [6] in their analysis. The authors model the mobility of 40 nodes over a time period of 900 seconds in an 1000x1000m area. We compare the characteristics of the WLAN model with the characteristics of the other models when the maximum speed for the other models is set to 5m/s. For the WLAN mobility model, we model the movement of 200 nodes over 10000 seconds in an environment of 1000x2500m. We model a higher number of nodes because our evaluation showed that 200 nodes is the smallest population that still reflects the mobility characteristics of the WLAN users. To preserve the node density, we increase the size of the environment. We model node movement over a longer time period because a simulation time of 900 seconds is too short to display the mobility characteristics of the WLAN (as shown in Figure 12).

The values of the mobility parameters for the compared models are summarized in Table III. The data shows that the average relative speed (aRS) of the WLAN mobility model is one order of magnitude lower than the aRS of the other models. The average link duration (aLD) is half of the aLD of the Random Waypoint, and about 20 times lower than the RPGM model. Because the nodes modeled by the RPGM model move together and the topology within a group changes rarely, the aLD value of the RPGM model is high. The average spatial dependency (aSD) of the WLAN model is comparable with the aSD of the Random Waypoint model, and is one oder of magnitude lower that the aSD of the RPGM model. Again this high spatial dependency for the RPGM model is explained by the coordinated group movements.

Overall, the WLAN mobility model yields lower mobility characteristics than the mobility models used in today's network simulations. This is an indication that the systems designed for WLAN-like environments and simulated with today's mobility models could grow in complexity to cope with mobility rates that will not be reflected in reality.

IX. Conclusions

Simulations of wireless systems make many simplifying assumptions to obtain a setup that is computationally tractable. The movement of the nodes is one aspect (other important aspects like signal propagation are beyond the scope of this paper). This paper describes an executable mobility model that uses parameters extracted from real-life mobility and generates mobility scenarios based on these parameters. We define a parameter set that captures the mobility characteristics of a WLAN and extracts these parameters from WLAN traces. To allow others to compare measurements on their WLAN with our observations, we present a framework that guides the user from recording the wireless trace to generating mobility scenarios using the WLAN mobility model.

We validated the WLAN mobility model by cross-validation, using traces recorded with two different methods, during the time frames May-July 2003 and April-May 2004 in an university campus. The validation shows that the proposed WLAN mobility model is successful in capturing the mobility characteristics of the real-world movement with a very small error.

For users that want to use the WLAN mobility model without recording a WLAN trace in advance, we explore the parameter value space and show how different parameter sets influence the mobility of the modeled nodes.

The mobility model has a major influence on the behavior of the simulated systems. Basing the parameters of the model on real-life measurements is a challenge, but is a necessary step towards better simulators. If we compare our results with other mobility models we see that the commonly used simulation time of 900 seconds is too short to display the mobility characteristics of the kinds of WLAN settings that we investigated. Furthermore, some models like the Random Waypoint and RPGM models operate with nodes that are more mobile than those that we observed in a wireless network. While there may be environments that exhibit the high mobility presumed by these models, for other setups these models may guide the protocol designer to make decisions that are not warranted by real-life situations. Understanding the consequences and tradeoffs of mobility is a difficult topic, but a model as presented here presents a way to characterize environments and to compare protocols based on some aspects of real networks. We look forward to see characterizations of other networks with different properties (parameters) and hope to direct protocol designers to spend their resources on those parts of the design space that have a direct influence on end-user performance.

ACKNOWLEDGMENTS

We appreciate feedback, comments, and advice by Gregg Pottie, Ed Knightly, and the INFOCOM 2005 referees. We thank Armin Brunner and Derk-Jan Valenkamp of the ETH facilities group for providing us access to the campus network information.

REFERENCES

[1] Tracy Camp, Jeff Boleng, and Vanessa Davies, "A Survey of Mobility Models for Ad Hoc Network Research," *Wireless Comm. Mobile Comp. (WCMC): Special issue on Mobile Ad Hoc Networking: Research, Trends and Applications*, vol. 2, no. 5, pp. 483–502, 2002.

[2] Dmitri D. Perkins, Herman D. Hughes, and Charles B. Owen, "Factors Affecting the Performance of Ad Hoc Networks," in *Proc. Int. Conf. Comm. (ICC '02)*, Apr.–May 2002, vol. 4, pp. 2048–52.

[3] Amit Jardosh, Elizabeth M. Belding-Royer, Kevin C. Almeroth, and Subhash Suri, "Towards realistic mobility models for mobile ad hoc networks," in *Proc. Int. Conf. Mobile Comp. Networking (MobiCom '03)*. Sep. 2003, pp. 217–29, ACM Press.

[4] Xiaoyan Hong, Mario Gerla, Guangyu Pei, and Ching-Chuan Chiang, "A Group Mobility Model for Ad Hoc Wireless Networks," in *Proc. Int. Workshop Model. Anal. Sim. Wireless Mobile Sys. (MSWiM '99)*. Aug. 1999, pp. 53–60, ACM Press.

[5] David Johnson and David Maltz, "Dynamic source routing in ad hoc wireless networks," in *Mobile Computing*, T. Imelinsky and H. Korth, Ed. 1999, pp. 153–181, Kluver Academic Publishers.

[6] Fan Bai, Narayanan Sadagopan, and Ahmed Helmy, "IMPORTANT: A framework to systematically analyze the Impact of Mobility on Performance of RouTing protocols for Adhoc NeTworks," in *Proc. IEEE Infocom 2003*, 2003, pp. 825–835.

[7] Magdalena Balazinska and Paul Castro, "Characterizing Mobility and Network Usage in a Corporate Wireless Local-Area Network," in *Proc. Int. Conf. Mobile Sys., App., Serv. (MobiSys '03)*, May 2003.

[8] David Kotz and Kobby Essien, "Analysis of a Campus-wide Wireless Network," in *Proc. Int. Conf. Mobile Comp. Networking (MobiCom '02)*. Sep. 2002, pp. 107–18, ACM Press.

[9] Diane Tang and Mary Baker, "Analysis of a Local-Area Wireless Network," in *Proc. Int. Conf. Mobile Comp. Networking (MobiCom 2000)*. Aug. 2000, pp. 1–10, ACM Press.

[10] Anand Balachandran, Geoffrey M. Voelker, Paramvir Bahl, and P. Venkat Rangan, "Characterizing User Behavior and Network Performance in a Public Wireless LAN," in *Proc. Int. Conf. Meas. Modeling Comp. Sys. (SIGMETRICS '02)*. Jun. 2002, pp. 195–205, ACM Press.

[11] Diane Tang and Mary Baker, "Analysis of a Metropolitan-Area Wireless Network," *Wireless Networks*, vol. 8, no. 2/3, pp. 107–20, Mar.–May 2002.

[12] David Schwab and Rick Bunt, "Characterising the Use of a Campus Wireless Network," in *Proc. IEEE Infocom 2004*, 2004.

APPENDIX

A. List of model parameters

Parameter	Description
t_{sim}	simulation time
N	size of the simulated population
C_{width}	width of a simulated cell
(C_x, C_y, C_z)	number of cells for the three directions of the environment
A_{max}	maximum number of cells a node can visit
$A = \alpha_i$	distribution of the number of visited cells
V	(computed) set of cells a node will visit
n_c	(computed) the number of cells a nodes will visit ($\|V\| = n_c$)
p_{same}	probability to remain in the same cell
p_{neigh}	probability to move to a neighboring cell
p_{non_neigh}	probability to move to a non neighboring cell
v_{max}	maximum speed of a node