Empirical Modeling of Campus-wide Pedestrian Mobility: Observations on the USC Campus

D. Bhattacharjee, A. Rao, C. Shah, M. Shah, A. Helmy
Electrical Engineering Department
University of Southern California
Los Angeles , CA 90089. USA.
{dbhattac, ashwinvr, chintans, manansha, helmy @ usc.edu}

Abstract - **Mobility is one of the main factors affecting the performance of ad hoc networks. Obtaining realistic mobility models is essential for the proper evaluation of ad hoc protocol performance. In our paper, we develop a hybrid mobility model, by observing the actual movement patterns of people on-campus and then post processing the data to form their spatio-temporal distributions. Our model differs from the synthetic and wireless network trace based models in that the model is empirical in nature and devoid of the restrictions as observed in some of the existing mobility models. Our model has been implemented by a structure similar to that of a Finite State Machine (FSM), using a tool that operates on our traces to generate another trace file which is readable by the Network Simulator (NS-2).**

Keywords: ***Ad-hoc Networks, MANET, Mobility, Modeling, Pedestrian Mobility***

I. INTRODUCTION AND RELATED WORK

All the mobility models in the research done on ad hoc networks are classified into synthetic models and models based on traces obtained from Wireless network usage. We now briefly explain both.

A. Synthetic models

Initially when mobility was to be introduced into the development of ad hoc networks, researchers looked into building some fundamental models like the Random Waypoint model [1], wherein a mobile node is able to move spatially in a random manner, losing out on some specifics of movement in the real world. For example the RWP model selects velocities uniformly from a predetermined set of velocities [V_{min} to V_{max}]. The movements of the node over various epochs (time intervals) are at a constant speed. Direction of movement during successive epochs is also random and not correlated to movement direction in the previous epoch. However these initial deficits were fixed in models like the Gauss-Markov model, which correlates the speed of a mobile node in one epoch to that in the previous epoch. Also to have more realistic speeds, the Smooth RWP model talks about selecting a set of "preferred" velocities from [V_{min} to V_{max}]. This makes the movement during epochs tend towards realistic representation of the node mobility. Also synthetic models deal with pathways and obstructions which restrict the movement of nodes in the simulation area (just like in reality, where people cannot walk through buildings and have to go around them).

A. Helmy was supported by grants from NSF Career, Intel and Pratt & Whitney Institute for Collaborative Engineering (PWICE).

The above stated models incorporate the improvements over the RWP model individually in each paper. In addition, recently several studies [1],[2],[5] have introduced several synthetic mobility models that are useful in simulations yet not based on actual mobility measurements. There has not been much work done in incorporating all the parameters of the above mentioned models to build a "hybrid" mobility model. Also, all the work done in mobility is not based on observations of actual movement patterns of people collected from a region of interest where a future ad hoc network could possibly be deployed. Instead, mathematical emulations to more realistic movements have been made. Our model incorporates the above 2 aspects.

B. Trace based models

The models in this category are built based on wireless network usage traces. These models are built on data collected by using SNMP polling, tcpdump [2], [3], [4] etc. at the access points (APs). Thus these movement patterns inherently show clustering around APs, which is definitely not representative of actual movement patterns. These studies are limited to people using pervasive devices like laptops and PDAs especially in the studies conducted at UCSD [3]. Capturing snapshots of accesses made to APs may not always give path details of a particular node. However it can help in forming certain group statistics and temporal/spatial correlations of people accessing these APs. The sample population in our study is not restricted to use of any such mobile devices, thus proving to be more general when trying to put forth such a mobility model.

We suggest a very different approach to modeling mobility in ad hoc networks. We concentrate on ad hoc networks which may be deployed on campus like environments in the future. Every student on-campus is a potential ad hoc node with the ever increasing popularity of pervasive mobile devices. Our mobility model works on "real traces" collected from the campus, observing the actual movement of students. The way we do this is by recording data like number of people observed on campus pathways at different locations and at different times of the day/week and form spatial/temporal movement distributions. We believe this model captures the real world aspects like pathways, obstructions, directional preferences of students, temporal and spatial variations of their movement, velocities and group /subgroup statistics, all into one single hybrid model. We now explain the paper sections.

Section II explains the methodology we developed to collect, organize and analyze the movement traces, section III presents some of the results of the study and its discussion; section IV mentions how the current work could be extended and section V finally concludes the paper.

II. METHODOLOGY

In order to address the above problem we first propose and describe a hybrid mobility model the parameters of which are estimated from the traces. The trace collection methodology is then described that enables a systematic, reproducible, and sound tracing of people movement on campus. We show the step-by-step procedure for our methodology.

A. The Proposed Hybrid Model equation

Broadly, we intend to use the statistical parameters we obtain from the traces along with parameters borrowed from synthetic models.
A direction equation is intended follows:

$$Dir_{Next} = Dir_{Current} + Dir_{update\_factor} \qquad (3.1)$$

The "Dir_{update_factor}" will be discussed in the section on Methodology.

B. Methodology

First, obtain a campus map and categorize buildings into five types. Namely, Residential, Cafeterias, Classes, Libraries and Computing facilities. We choose this system of classification; since a student would usually visit these places during his/her stay on campus. Some administrative buildings may be considered too. Mark the categorized buildings *uniformly* over the campus map. Join these buildings via shortest pathways (to avoid triangularization) and the intersections observed in the process form our Observation Locations (OL). The reason to select the shortest pathway is because usual human behavior is to take the shortest path between two buildings assuming he/she does not make a detour visit. The OLs for our analysis are *KOH, JEP, LVL, OHE, Tommy Trojan Statue, BRI, Carl's Jr. and PED* (Refer map). The data which is collected at these OLs have been tabulated as shown below.

TABLE I. DATA COLLECTION TABLE

Date	Time	Group Size	Dir. of group	Subgroup Size	Dir	Mode

A Group is defined as a cluster of people not necessarily socially related. The method of selecting this cluster is to lock on to a particular group and then considering every other person within a range of 1.5 meters from this group as member of the same group. A subgroup is defined as a full set or a subset of a group that changes direction. Note the direction of each subgroup (N, S, NE etc.) as per the table above. Note the mode of movement of each subgroup as such walk, cycling etc.

The OL observation time slot is divided by the number of pathways at that OL which is observation time of each pathway.

When observing each pathway, observe a group until one of the following events occur:

1) Group moves beyond the line of sight of the observer.
2) Original group splits into subgroups.
3) When the original group stops. Note the pause time in the comment column. When the group resumes movement follow the above steps.
4) Do not observe the splitting of subgroups although observe the direction in which they are moving.

While adopting this methodology, the alternatives were to follow each person and obtain the path traversed by him/her. This would have given a very accurate movement pattern. But, it was not clear as to how long should a person be followed, which persons to follow etc. Also, while pursuing one person, we would have lost statistics of many people in the vicinity.

III. RESULTS AND DISCUSSIONS

A. The Real Traces

The trace collection methodology has been discussed in the previous section. We now present the trace collection *scenario* as under:

- Trace period - February 25 to April 25, 2004
- # of people - 6389 over trace period.
- # of groups - 1758 over trace period
- # of subgroups - 2382 over trace period

Details of groups and subgroups at each OL are attached in appendix I.

Distances were measured by a rectangular coordinate system developed with KOH being set as the origin of the system. The distances of the various OLs with respect to the origin are as under:

TABLE II. ACTUAL DISTANCES OF OLs FROM ORIGIN

OL	X co-ordinate (in meters)	Y co-ordinate (in meters)
LVL	0	245.41
JEP	0	170.72
KOH	0	0
OHE	170.2	0
PED	117.37	85.36
TOMMY	117.37	170.72
CARL'S JR.	170.2	138.71

The numbers in meters is the *actual* distance between the OLs in the campus. Thus, no scaling is required while conducting analysis which utilizes these distances (for example the traces generation for NS).

The temporal distributions of people at two of the OLs; namely; JEP and TOMMY are shown in Fig. 1.. We divided the 9 am to 5 pm period into 4 time slots, each of 2 hours duration. The distributions are with respect to each time slot, where 1 is the first time slot namely, 9 to 11 am and 4 is the last one i.e. 3 to 5 pm.

As observed from the graphs, the distributions are dissimilar. However, some interesting observations are made. LVL shows a drop in the number of people from the 11 to 1 pm to 1 to 3 pm transition.. This is so because in the 11 to 1 pm slot, people visit the library more to complete assignments and access computing facilities for various purposes like checking email. However, the 1 to 3 pm slot corresponding to lunch time, people move out of the LVL OL. This can also be observed from the distribution at the JEP OL where it observes an increase in the number of people observed in the corresponding time slots. JEP being at one of the major exits to the University, observes more people between 1 to 3 pm corresponding to lunch time. Even TOMMY observes a drop in the same time slots the reason being the same. We observed that the probability of movement from TOMMY to JEP in the 1 to 3 pm time slot is the second highest (0.27).

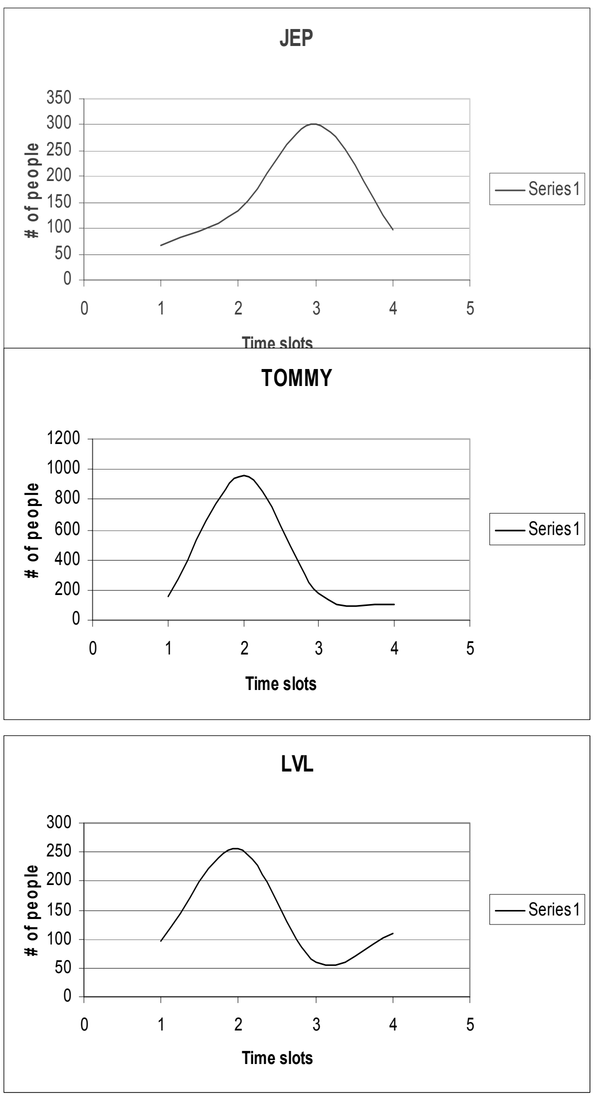


Figure 1. Temporal distributions of people at JEP ,Tommy, Leavy

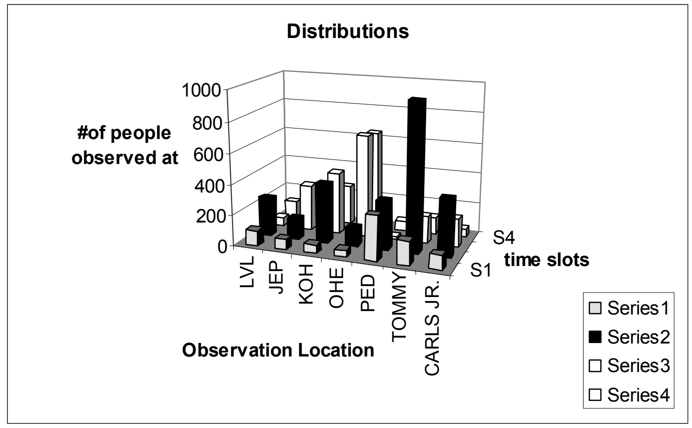


Figure 2. Temporal-spatial distributions at the various OLs.

Observe from the distribution in Fig. 2 that there is an increase in the number of people observed from PED to TOMMY in the second time slot i.e. S2. We also observed that the probability of movement from PED to TOMMY is 0.36, the highest at that OL and this corresponds with the distribution above. Thus, we establish that such a spatial distribution can be used to obtain the direction update factor in the direction equation mentioned above. Specifically, the update factor would be a function of the maximum probability at the OL and this probability itself is a function of the time slot in which it is calculated. Thus, this in effect, takes in to account the temporal distribution of people at a particular OL. In other words, the update factor is a function of the maximum probability of movement in a particular *direction* and this probability measure is itself a function of location (OL) ***and*** the time (time slot) at which it was calculated. The function (update factor) returns the (x,y) coordinates of the next location *based* on its *current* location and these coordinates are a function of the spatio-temporal characteristics of the *current* location and this is a desirable property of a mobility model. Analytically, the generic direction equation may be implemented as below:

$$(x,y)_{next} = (x,y)_{current} + x_i \tag{4.1}$$

; where $x_i = f\{\max_i[p_i\,(OL_{current}\,, t)]\}$

;i = 1,2,3,4 corresponding to each direction N,S,E,W; pi =prob. of corresponding i.

This analytical equation is implemented in C code; where the probabilities are obtained from our traces (method to calculate probability attached in appendix I).The code essentially “reads in” our trace file and generates another trace file which can be read by the Network Simulator (NS-2).Alternatively, the direction equation may be implemented by a FSM as mentioned before.

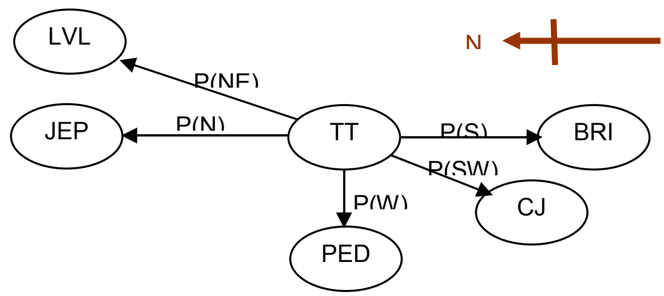


FIGURE 3. The FSM Model

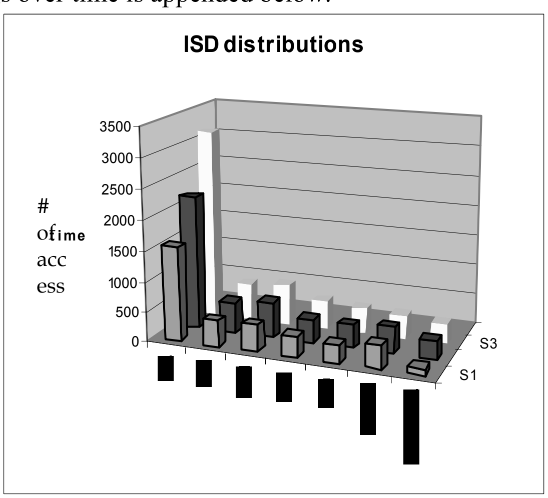
An example FSM model for the Tommy Trojan (TT) OL can be seen in Fig. 3.

It must be noted that this is a "first order" FSM i.e. a FSM for a *particular* OL (TT in this case) at a *particular* time slot (depending on from which time slot the probabilities are taken). In a *particular time slot,* we can obtain FSM for each OL and "join" them in succession to form a higher order FSM model for the entire campus. However, this will be valid only for one time slot and 4 such higher FSMs will have to be obtained to describe mobility in the 4 time slots considered.
Thus, the FSM provides a very good "pictorial" view of the probable movement of people on campus at a particular time slot. For the purpose of simulations we use the C code implementation.

As mentioned, the FSM gives a very good pictorial view of the mobility model of the campus in a particular time slot. Essentially, it diagrammatically represents equation (4.1) and thus can be used for comprehension and presentation of the proposed model. The C code implementation generates a NS-readable trace file and thus can be used for stand-alone simulations. It can also be used for protocol testing over the proposed mobility model under a simulation environment. Lastly, it can also be used to gain insight into the mobility characteristics if the campus and such insight or knowledge may be applied to better routing protocol design.

B. The ISD (USC Wireless Network Service Provider) Traces

The distribution of the number of user accesses at the specified buildings over time is appended below:



As observed, the distribution is very different form that obtained from the real traces. This is because the above distribution (obtained by method 2) observes movements only around buildings which have Access Points (APs). However, our trace is independent of any such restrictions. Hence, the difference in the observed distributions.

However, from the real traces we could estimate the number of people at a particular location in campus and from the ISD traces estimate the usage of wireless network if any. If the mobility exhibited by users (as derived from our trace) is a lot and the usage of wireless network is more, then we could conclude that we night need seed relay stations of link stability of an ad-hoc network were to be implemented in campus in the future. In such a way, we could utilize both the real and ISD traces. Now we briefly state the observations made in implementing a mobility model with the ISD traces versus that with the real traces.

For the ISD traces we observe the following:
<u>Advantages:</u>

- Can help capture group statistics.

<u>Disadvantages:</u>

- Clustering around buildings does not capture true movement patterns from mobility viewpoint.
- Does not give path details.
- Limited to people with laptops and handhelds

A CDF derived from the above table is appended below:

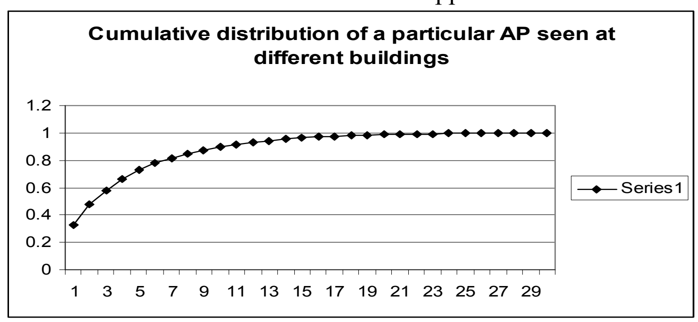


Figure 5. Cumulative Distribution of a particular AP seen at different buildings

As seen, 50% of the users access 3 or less buildings during the trace period. This fact raises questions on the correctness of the mobility model if derived from these statistics.

The observations of the real trace method are appended below:
<u>Advantages:</u>

- Can help capture group statistics in more details.
- Actual user direction pattern over time and place.
- Gives actual path information.
- Not limited to laptop or PDA users.

<u>Disadvantages</u>**:**

- Difficult to scale; esp. in a large campus.
- Requires sufficient data for confidence tied to scalability in terms of time.

IV. FUTURE WORK

- Nodes in the C implementation are restricted to campus. They are not allowed to move in or out. Generate a code to support this enhancement.
- Allow for a pause varying pause time when a node moves "out". For example, a person might go for a 60 minute lunch break outside campus. In this case, the pause time is 60 min.

V. CONCLUSIONS

We have devised a method to collect real traces of movement of people on campus that can be reproduced. With such traces a mobility model in analytical and graphical terms was developed. The analytical model was implemented in C code.

Results and statistics discussion showed that the model reflects mobility as observed on campus.
However, from the real traces we could estimate the number of people at a particular location in campus and from the ISD traces estimate the usage of wireless network if any. If the mobility exhibited by users (as derived from our trace) is a lot and the usage of wireless network is more, then we conclude we might need relay stations for the stability of an ad hoc network.

REFERENCES

[1] F. Bai, N. Sadagopan and A. Helmy. "Important: A framework to systematically analyze the impact of mobility on performance of routing protocols for ad hoc networks". *Proc. (INFOCOM 2003). April 2003.*

[2] M. Balazinska. "Characterizing Mobility and Network Usage in a Corporate Wireless Local-Area Network". *MIT Lab. for Computer Science* and Paul Castro *IBM T.J. Watson Research Center*.

[3] M. McNett and G. M. Voelker. "Access and Mobility of Wireless PDA Users". *Dept. of Computer Science and Engineering, University of California, San Diego.*

[4] D. Kotz & K. Essien. "Characterizing usage of a campus-wide wireless network". *Technical Report TR2002-423, CS Dept., Dartmouth College*, *March 2002.*

[5] A. Jardosh, E. Belding-Royer et al. "Simulation and implementation issues: Towards realistic mobility models for mobile ad hoc networks". *Proc. Of Mobile computing and networking*, *September 2003.*

APPENDIX I. CALCULATION OF PROBABILITIES

Consider the trace below:

TABLE I. EXAMPLE TRACE

Group Size	Direction of Group	Subgroup size	Direction of subgroup
2	N	1	S
		1	E
2	S	2	N
2	W	2	W
3	W	2	E
		1	W
4	E	3	E
		1	N
2	E	2	N
2	W	1	N
		1	E
3	S	3	E
2	N	2	N
3	N	2	N
		1	E
4	W	2	E
		1	W
		1	N

The above table is a sample trace which we have shown here to explain the probability calculation method. Now the aim is ultimately to get the probabilities with which a student would be traversing a path at a certain time-slot in the day. We want to get the directional probabilities. So let's calculate the number of people traveling East in the observation time-window. Definitely from the groups column we get 6 people traveling East. However from the subgroup column when we calculate the number traveling East we should not consider those people who were already traveling East (group column) and still continue East (Subgroup column).

Hence # of people going East = 6 + 10 = 16
0f people going West = 11 + 0 = 11
0f people going North = 7 + 7 = 14
0f people going South = 5 + 1 = 6
number of people at OL = 47
Hence P(E) = 16/47; P(W) = 11/47; P(N) = 14/47; P(S) = 6/47

TABLE II. GROUP/SUBGROUP STATISTICS:

	9 - 11 AM		11 - 1 PM	
	Group	Subgroup	Group	Subgroup
LVL	32	15	101	63
JEP	**X**		54	15
KOH	20	28	108	88
OHE	17	25	23	42
PED	99	37	112	86
TOMMY	32	81	261	330
CARL'S JR.	24	47	103	155
TOTAL GR	**224**		**762**	
TOTAL SUB		**233**		**779**

APPENDIX II. USC CAMPUS MAP WITH OL POSITIONS

1)KOH, 2) JEP, 3) LEAVY, 4) TOMMY TROJAN, 5) PED, 6) OHE, 7) CARL'S JR.

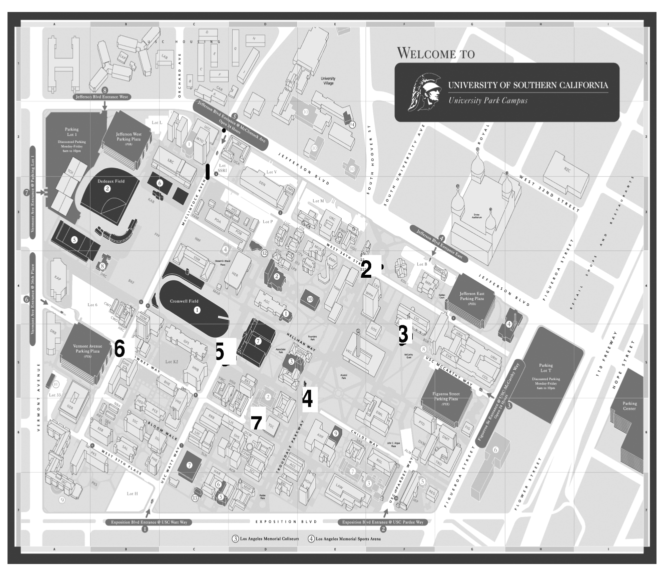


Figure 6. USC Campus Map with the OL Positions