Towards a Model of User Mobility and Registration Patterns

Ravi Jain **Anupama Shivaprasad** **Dan Lelescu** **Xiaoning He**

{jain, anupama, lelescu, xiaoning}@docomolabs-usa.com

DoCoMo USA Labs, 181 Metro Drive, San Jose, CA 95110, USA

The evaluation of a great deal of research on ad hoc networks, as well as cellular networks, depends on models of user mobility. Many models have been developed and utilized, such as the random walk and random waypoint models. These are simple to implement and analyze but unlikely to be realistic. We develop a model based on extensive experimental data from a campus Wi-Fi LAN installation, representing traces from about 6000 users over a period of about 2 years. This data does not enable us to develop a user mobility model directly. However, as a first step, we develop a model of the time and sequence of locations at which user devices register. Note that this can be very useful, for instance to evaluate protocols that attempt to manage routing or resource allocations at different nodes. This paper reports work in progress on developing a user registration model. It shows the key time domain as well as space domain features we have extracted from the data. In particular, we show that the time features indicate heavy-tailed, although not power-law, distributions. The spatial features strongly indicate registration localization and hierarchy. The model itself can be represented as a set of probability distributions for various parameters. The modeler, for example a protocol designer, can then generate traces that conform to these distributions while varying the scale of the model in terms of the number of users. We close with a brief discussion of further work to refine and extend the model.

I. Introduction

The design of protocols and algorithms for ad hoc networks is often evaluated by simulation, with a user mobility model as input. Several mobility models have been developed, such as random walk, random waypoint [1] and the obstacle model [4]. However, while they are simple and easy to implement, they are unlikely to model real user mobility very accurately.

Instead of trying to make existing conceptual models more sophisticated (and more complex) so as to approximate real user mobility better, we start with experimental data to develop an abstracted model. The idea is that the model could be used in the following way. The modeler (e.g. a protocol designer) would specify a relatively small number of parameters, and the model would probabilistically generate movement traces consistent with those parameters. In particular the modeler would specify the number of users to be modeled. We also assume that location is represented discretely, e.g. as geographical zones, rather than in terms of continuous motion in some coordinate system; the modeler would specify the number of locations to be modeled. Additional parameters will be described later.

To develop our model we use a large set of user mobility traces collected from the Dartmouth College campus Wi-Fi network, which recorded the time and identity of over 500 wireless cells visited by more than 6000 users over a period of over 2 years [5].

This data has some limitations for our purposes, but is the best available to us at present. The main limitation is that it does not record geographical locations of users, but their registrations at access points as they move around campus. Thus, as a first step, we develop a user registration model rather than a mobility model; we believe that the registration model can be refined heuristically to develop an estimated mobility model as the next step. Besides, the registration model is useful in its own right for many system applications, such as those attempting to improve mobility or QoS management.

The Dartmouth installation itself is not an ad hoc network, but to our knowledge there is no large data set available about users in ad hoc network settings. We do believe that users move largely for external reasons (e.g. going for lunch to the cafeteria) rather than based on the type of wireless network they use, so the type of network will not affect the model significantly. In any case, as our initial results show, the model is likely to be far more realistic than simple models such as random walk and random waypoint.

Our work thus has the goal of making two contributions: (1) a mobility model based on an extensive empirical data set, and (2) a methodology for taking new data sets, possibly for different environments and with true location information, and developing models.

While we have not fully attained our goal, this summary describes work in progress and some interesting results already obtained. In section 2 we describe the timing features we have extracted from the data. In section 3 we briefly summarize the spatial features, which are a major focus of existing models. In particular we show that the patterns we have found show a localized and hierarchical structure, and do not have a random walk pattern. Finally we end with some brief remarks about further work.

Terminology. When a user disconnects from the network (e.g. powers off a laptop), this is called the OFF state, and is regarded as a virtual AP. A *transition* is a registration at one AP followed by a registration at a different AP.

II. Modeling Time Features

The gross features of the Dartmouth data are described in [5] and are not repeated here. We model three key features, over all users, for representing the time characteristics of registration patterns. These are the mean time that a user registers at any given AP (called the residence time), the mean time that a user is OFF (the OFF time), and the number of OFF states per day per trace. Observe that these quantities are sufficient to model the time features and generate the time features of the traces probabilistically.

We find that the mean residence time and mean OFF time distributions are highly skewed. We fit the data to well-known probability distributions as follows. The parameters of the candidate CDFs describing the data are estimated using Maximum Likelihood Estimation; the candidate with the lowest fitting error with parameters within 95% confidence is selected. In particular, the pdf for the mean residence time *r* fits a Weibull distribution; the corresponding cdf is shown Figure 1:

$$p_{a,\,b}(r) = ab\ r^{b-1}\,exp(-ar^{b})\ I_{(0,\,\infty)}(r)$$

The experimental parameters are $a = 141.448$, $b = 0.505$, and I is the indicator function; the confidence interval values are (134.231, 149.053) for a, (0.497, 0.513) for b. The root mean square error (RMSE) is 0.0039 and *R-squared* is 0.9520, indicating a very good fit.

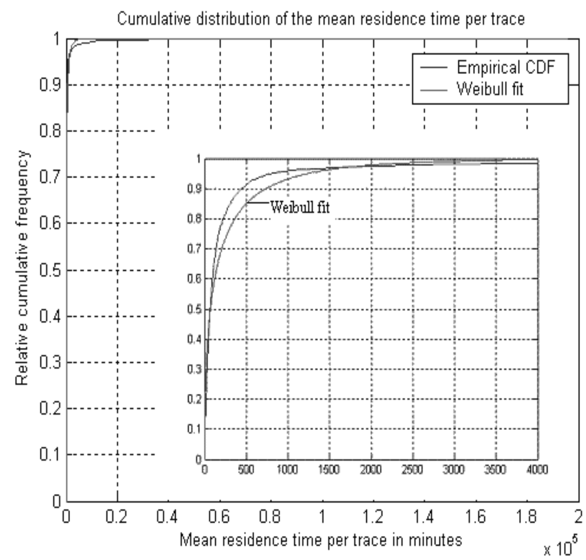

Figure 1: Mean residence time (inset plot is zoomed-in version of the same plot)

We point out that this distribution is *heavy tailed*. However, unlike other heavy-tailed distributions that are much discussed recently e.g. [3], the data can be shown to not obey a power law.

The pdf for the mean OFF time *s* fits a log-normal distribution with $\mu = 7.437$, $\sigma = 1.602$ with confidence intervals (7.395, 7.478) and (1.573, 1.632) respectively. The RMSE is 0.0012 and *R-squared* is 0.9855.

$$p_{\mu,\sigma}(s) = \frac{exp(-\,(\ln s - \mu)^2 / 2\sigma^2)}{s\sigma\sqrt{2\pi}}$$

The cdf of the data and the curve fit are shown in Figure 2.

The pdf of the number of OFF states per day per trace fits a Weibull distribution, and the parameters are *a=0.3107* with 95% confidence interval (0.3084,0.313), and *b=1.709* with 95% confidence interval (1.698, 1.719). The RMSE is 0.02979 and the *R-squared* is 0.9901. The data and curve fits are shown in Fig. 3.

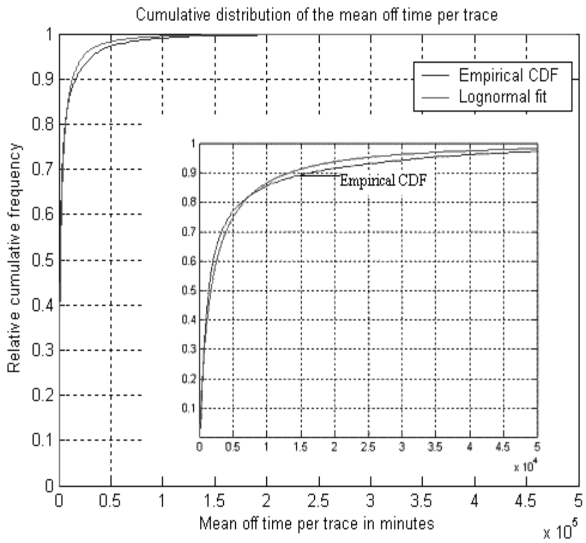

Figure 2: Mean OFF time

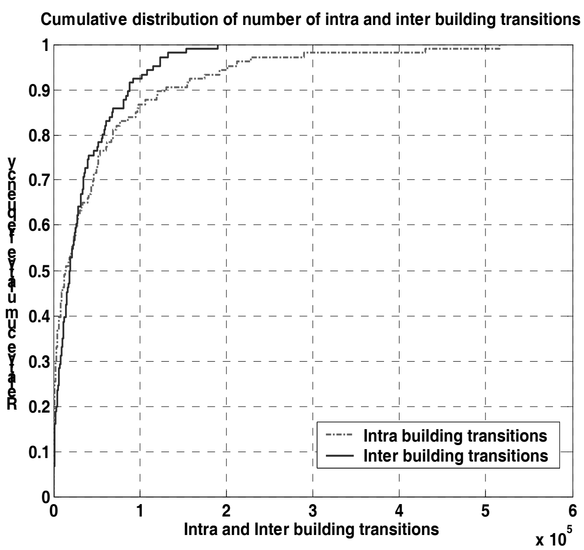

Figure 4: Inter-building and Intra-building transitions

Previous models such as random walk often use residence times that are drawn from a uniform distribution, and generally do not model the OFF state at all.

III. Modeling Spatial Features

The Dartmouth campus covers over 200 acres and almost 200 buildings. Our first observation about spatial features is that transitions show strong locality based on buildings. Considering all the transitions over all the traces, the number of transitions that occur between APs in the same building (an intra-building transition) far exceeds the number of inter-building transitions. The CDFs of intra-building and inter-building transitions are shown in Figure 4. (This is consistent with, and inspired by, our previous work [6]).

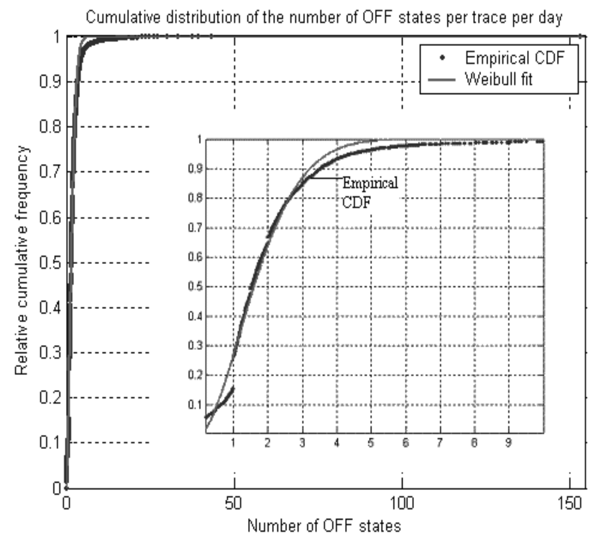

Figure 3: Number of OFF states per day per trace

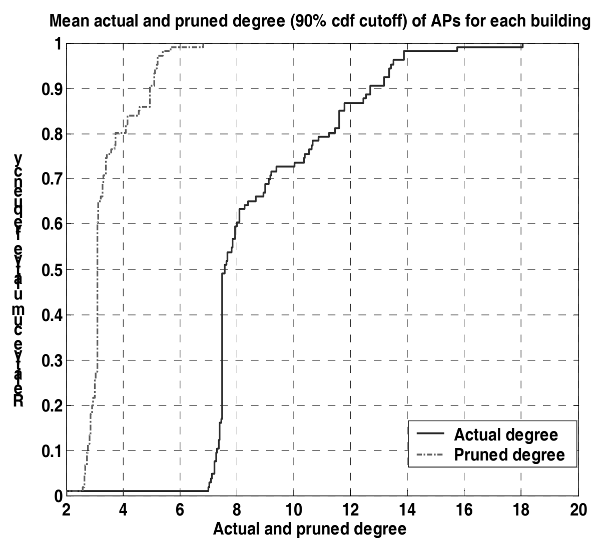

Figure 5: Intra-building pruned degree

This indicates that it is important to get the intra-building model right first. The number of APs per building varies from as low as 1 to as many as 22. If we build a *transition graph* for each building, where each vertex represents an AP and each edge a transition, we find it is a complete mesh, i.e., the degree of each node in a graph for a building with *m*

APs is *m - 1*. To investigate further, we label each edge *(u, v)* with the probability that a user transitions from *u* to *v*; we find that the transition probabilities are highly skewed. We prune the degree for each vertex by deleting all outgoing edges that, combined, contribute less than 10% probability. We find the pruned degree is much less than the degree as shown in Figure 5.

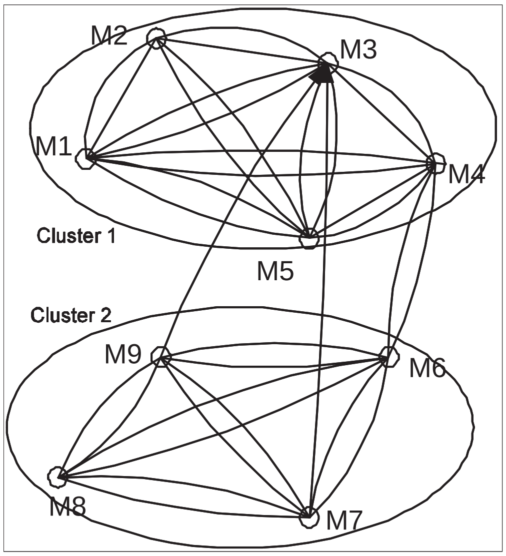

Figure 6: Hierarchy within intra-building transitions

The pruned degree results indicate that a random walk model or even a random waypoint model is not likely to be a good choice when considering correlation with real data.

We further analyze the transition graph alone, after pruning, for buildings with many APs. We apply the graph clustering tool MCL [2] and find strong evidence of hierarchical structure within buildings. For example, a particular building with 9 APs is found to have 2 clusters of size 5 APs and 4 APs respectively. This is shown in Figure 6.

IV. Further work

We are developing distributions for transition probabilities to complete the modeling of the spatial features. With that the initial registration model is complete. In further work we will generate traces from the model and compare against the data. We also have several heuristics that we are experimenting with to estimate a mobility model from the registration model.

Acknowledgements

We are indebted to David Kotz, Tristan Henderson, and their colleagues at Dartmouth College for making their experimental data available, as well as providing additional information (e.g. AP locations) and assistance. We are also grateful to Eric van den Berg of Telcordia for his advice.

References

[1] T. Camp, J. Boleng, and V. Davies, Mobility Models for Ad Hoc Network Simulations, *Wireless Communication & Mobile Computing*, 2, 5, 483-502, 2002.

[2] S. van Dongen, Graph clustering by flow simulation, PhD Thesis, Univ. of Utrecht, May 2000.

[3] C. Faloutsous, P. Faloutsous, and M. Faloutsous, On power-law relationships of the Internet topology, *Proc. SIGCOMM*, Sep. 1999.

[4] A. Jardosh, E. M. Belding-Royer, K. C. Almeroth, and S. Suri, Towards realistic mobility models for ad hoc networks, *Proc. MobiCom*, Sep. 2003.

[5] D. Kotz and K. Essien, Analysis of a campus-wide wireless network, *Proc. Mobicom*, Sep. 2002.

[6] L. Song, D. Kotz, R. Jain, and X. He, Evaluating next-cell predictors with extensive Wi-Fi mobility data, *Proc. Infocom*, Mar. 2004.