Understanding Human Movement Patterns via Trace-based Analysis

Rashmi Manchanda  
M.tech Scholar/RPIIT  
TECHNICAL CAMPUS  
KARNAL (INDIA)/KUK

Sandeep Garg  
Asstt. Professor, RPIIT  
TECHNICAL CAMPUS  
KARNAL (INDIA)/KUK

Abstract—To simulate a mobile ad hoc network, several factors play a crucial role, including, mobility of mobile nodes, protocol used for routing information among mobile nodes. Since the movements of mobile nodes significantly affect the network connectivity, modeling mobility is highly critical. In other words, for credible evaluation of routing protocols for a MANET, the movement of mobile nodes must be simulated realistically.

Index Terms—About four key words or phrases in alphabetical order, separated by commas.

I. INTRODUCTION

A mobile ad hoc network (MANET) is an infrastructure-less network with wireless nodes moving from one location to another. Instead of using a central base station to which all the computer must be connected and communicate, it identifies a particular network where every mobile node acts as base station (or router) and can move around. Due to the mobile nature of nodes, frequent updates are required to maintain connectivity information among mobile nodes in the network. In other words, it is a self-configuring network. Typical applications for these networks include: information exchange in disaster situations, establishing connectivity in situations with networks on the go. To simulate a mobile ad hoc network, several properties need to be modeled, e.g., mobility of mobile nodes. Since the movement of mobile nodes significantly affects the connectivity of network, modeling mobility is highly critical. In other words, for credible evaluation of routing protocols for a MANET, the movement of mobile nodes must be simulated realistically. Simulation-based studies offer several advantages over real (tested) experiments; specifically, simulations are scalable, reproducible, and highly cost efficient. However, simulation environments generate with several challenges. In particular, assumptions made during simulations may lead to misleading and, thus, incredible results. Therefore, recent efforts in modeling mobility of nodes in an ad hoc network are based on reverse engineering. Specifically, human movement traces are collected from real scenarios, analyzed for statistical features that may be present in human movement, and then used in the development of a new mobility model based on the features extracted from the real mobility traces. In this paper, we follow the same approach and analyze two sets of mobility traces collected from real conference scenarios for their statistical features.

Research shows that human movements commonly have a set of seven statistical features [5]. With this work, we analyze the two real scenarios and validate if these seven statistical features are indeed found in those two traces. The remainder of the paper is organized as follows. Section II presents the related work done in this area. In Section III, we highlight the recent trends in modeling mobility and discuss our choice of mobility traces as well as a mobility model for mimicking these two scenarios. Section IV includes results of our analysis. Finally, in Section V, we conclude our findings and discuss future directions for research in this area.

II. RELATED WORK

A. Background

Mobility models can be broadly categorized into: synthetic and trace-based mobility models. A Synthetic mobility model attempts to model mobility in a simplistic manner. Specifically, traces generated by a synthetic mobility model are based on random movement patterns. Random Waypoint (RWP) was the first synthetic model proposed [22], in which mobile nodes move independent of each other. In this model, mobility is modeled as a completely random phenomenon.

A trace-based mobility model, on the other hand, is based on movement patterns found in real datasets. By analyzing real data, researchers can better understand the statistical features present in human walk. Traces analysis [13,14] for different scenarios generates different results. As a result, features obtained from the analysis of one dataset may not be applicable to another dataset. Research community has been contributing mobility traces collected from real scenarios to several repositories like CRAWDAD [19], UNC/forth [20], and MobiLib [21].

B. Recent Trends in Mobility Models

Simulation results based on synthetic mobility models may not be accurate due of random moments and, thus, do not match with realistic scenarios. On the other hand, trace based mobility models provide high accuracy, as these models are
developed based on movements in real world. Several mobility models have been proposed in the recent past and based on the analysis done in [5], SMOOTH provides a better match for statistical features extracted from real traces than other mobility models, including SWIM [6], SLAW [8], TLW [11], and RWP [22]. As a result, SMOOTH clearly differentiates among protocols that make routing decisions from contact history of mobile nodes.

C. Trace-based Analysis

Statistical features extracted from a real movement trace depend on the trace collection and analysis method used. In particular, a trace collection may be either coarse-grained or fine-grained. Similarly, analysis method may require input parameter values that affect the result of the analysis [12,14]. Therefore, results obtained from one dataset cannot be applied to other datasets. It may be, however, interesting to explore how different values for input parameters used for analysis can affect the statistical features extracted from real movement traces.

III. OUR CONTRIBUTION

In this work, we are interested in analyzing a real movement trace and validate the list of statistical features listed in [5]. Munjal et al. analyzed a set of seven mobility traces collected from a wide range of scenarios and analyzed the traces for seven statistical features commonly found in human movement. These traces can be categorized into location-based and contact-based traces [13]. Authors extracted the contact information from each trace and plotted against the distributions obtained from the real traces. The comparison, however, is only visual. In addition, the distributions obtained from real traces are not analyzed to determine if the distribution in fact best fits power-law. Thus, it is hard to determine the statistical significance of this analysis. Our goal is to perform statistical analysis of these traces and validate if the distributions extracted from real traces in fact satisfy the statistical features listed in [5]. For this purpose, we have selected two real movement traces, Infocom’05 and Infocom’06, used in [5]. Both of these traces are collected from conference scenarios. We have restricted our choice to this set of traces due to the following reasons:

1. Due to a widely available Bluetooth connectivity in mobile devices, it is reasonable to assume that contact-based traces are easy to capture.
2. These movement traces (collected from two conference scenarios) have been widely used in trace-based analysis conducted in recent papers.
3. Contact information obtained from these traces can be highly useful in predicting route information among mobile devices engaged in an ad hoc network and, thus, will help improve the throughput of the underlying network as well as decrease the network latency involved in delivery of packets.

To simulate the real scenario, we have chosen the SMOOTH mobility model. We believe that our choice for SMOOTH is ideal, as highlighted below:

1. The SMOOTH mobility model has previously analyzed the contact-based traces used in this work. Thus, the model provides a series of results that can be used as a benchmark for the analysis done in this study.
2. As mentioned in Section II.B, compare to other recently proposed mobility models, SMOOTH is the only model that has been validated with both contact-based and location-based traces. In other words, results obtained from SMOOTH are more credible compared to other mobility models.

Compare to other traces used in [5], we note that these two traces have been widely used in analysis done by other researchers, are of longer duration, have large number of nodes and, thus, are more reliable. Location-based traces used in [5] are of short duration and may be biased, leading to incorrect results.

A. EasyFit Software

To fit theoretical distributions to our dataset, we use the EasyFit software available at [4]. Easy Fit allows us to fit a large number of distributions and, thus, finds the best fit to the data. Easy Fit allows fitting a large number of distributions to data allows to easily and quickly selecting the probability distribution which best fits to data, and perform specific calculations using the best fitting model. EasyFit uses three goodness of fit tests, namely Kolmogorov-Smirnov [1], Anderson-Darling [2], and Chi-Squared [3], to compare the empirical data with several other possible distributions and ranks the fitted distributions based on the statistics obtained via these three tests. Specifically, EasyFit provides the parameter values for each of the fitted distributions as well as indicates if the fitted distribution provides a good fit.

IV. ANALYSIS OF CONTACT-BASED TRACES

The output from EasyFit software provides three outputs, as briefly described below.

1. Goodness of fit, which signifies if the fitted distribution provides a good fit for the data,
2. Summary, which lists the parameter values for each fitted distribution, and
3. Graphs, which plots the graphs for each fitted distribution vs. the input data.

EasyFit provides over 55 theoretical distributions that are available for data fitting. Similar to the distributions used in [14], we choose four distributions, namely, Exponential [15], Gamma [16], Levy [17], and Pareto2 (a type of power-law) [18]. Results obtained from this analysis are presented in Figures 1-18. Next, we describe the statistical features extracted from these two traces and discuss the fitting process.
As discussed in [5], contact information among mobile nodes can help us understand their movement patterns and, thus, predict the neighbor that will help a fast delivery of packet to its intended destination. To predict the contacts made by a mobile node, we analyze these two traces and extract the following three distributions:

1. ICT is defined as the time between two subsequent contacts for a pair of mobile nodes.
2. CD is defined as the duration for which two mobile nodes remain connected when in contact.
3. CN is defined as the number of times two mobile nodes contacted each other during the trace duration.

A. Infocom’05

Table 1 lists the parameters and their value used to describe the Infocom’05 scenario. The original trace and the related documentation are available at the CRAWDAD repository [23]. The format of the trace is as follows:

\(<\text{node1}, \text{node2}, \text{startTimeOfMeeting}, \text{endTimeOfMeeting}, \text{numberOfContacts}, \text{interContactTime}>\)

where,

- \(\text{node1}\), lists the identity of the first node,
- \(\text{node2}\), lists the identity of the second node,
- \(\text{startTimeOfMeeting}\), lists the time of contact,
- \(\text{endTimeOfMeeting}\), lists the time when contact is lost,
- \(\text{numberOfContacts}\), lists the number of times contact is made, and
- \(\text{interContactTime}\), lists the time between current and last contact.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration</td>
<td>3 days</td>
</tr>
<tr>
<td>Number of Nodes</td>
<td>41</td>
</tr>
<tr>
<td>Conference Area</td>
<td>2500 m²</td>
</tr>
<tr>
<td>Bluetooth Range</td>
<td>25 m</td>
</tr>
</tbody>
</table>

We analyze this trace to extract the CNs, CDs, and ICTs distributions. We then import each distribution as a .csv file to the EasyFit software. We make a selection of the theoretical distributions of interest. The software fits these theoretical distributions to our imported dataset and ranks each of the fitted distributions. Figures 1-9 present the results of our analysis for CNs, CDs, and ICTs for the Infocom’05 scenario. As shown in figures, CNs, CDs, and ICTs best fit power law and, thus, are in accordance with the statistical features listed in [5].

**Contact Numbers (CNs):**

![Figure 1: Goodness of fit for CNs for Infocom’05 scenario.](image1)

![Figure 2: Parameter values of the fitted distributions for CNs for Infocom’05 scenario.](image2)
Figure 3: Comparison of the CNs distribution obtained from the Infocom’05 scenario and the fitted Pareto2 distribution.

Contact Durations (CDs):

Figure 4: Goodness of fit for CDs for Infocom’05 scenario.

Figure 5: Parameter values of the fitted distributions for CDs for Infocom’05 scenario.

Figure 6: Comparison of the CDs distribution obtained from the Infocom’05 scenario and the fitted Pareto2 distribution.
Inter-Contact Times (ICTs):

**Figure 7:** Goodness of fit for ICTs for Infocom’05 scenario.

**Figure 8:** Parameter values of the fitted distributions for ICTs for Infocom’05 scenario.

**Figure 9:** Comparison of the ICTs distribution obtained from the Infocom’05 scenario and the fitted Pareto2 distribution.

**B. Infocom’06**

Table 1 lists the parameters and their value used to describe the Infocom’05 scenario. The original Infocom’06 trace and the related documentation are available at the CRAWDAD repository [24]. The format of this trace is similar to the format of Infocom’05 trace, as mentioned in Section IV.A.

### Table II: List of input parameters and their values used for simulating Infocom’06 scenario

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration</td>
<td>3 days</td>
</tr>
<tr>
<td>Number of Nodes</td>
<td>78</td>
</tr>
<tr>
<td>Conference Area</td>
<td>2500 m²</td>
</tr>
<tr>
<td>Bluetooth Range</td>
<td>25 m</td>
</tr>
</tbody>
</table>
Figures 10-18 present results obtained from the analysis of the Infocom’06 trace with the EasyFit software. We obtain the distribution for CNs, CDs, and ICTs for the Infocom’06 scenario in a similar manner as discussed in Section IV.A. We fit the data with four distributions (same set of the distributions used in Section IV.A). As shown in Figures 10-18, we note that both CDs and ICTs obtained from the Infocom’06 trace best fit Pareto2 (a form of power law) distribution. CNs, however, best fit exponential distribution.

**Contact Numbers (CNs):**

![Figure 10: Goodness of fit for CNs for Infocom’06 scenario.](image1)

![Figure 11: Parameter values of the fitted distributions for CNs for Infocom’06 scenario.](image2)

![Figure 12: Comparison of the CNs distribution obtained from the Infocom’06 scenario and the fitted Exponential distribution.](image3)

**Contact Durations (CDs):**

![Figure 13: Goodness of fit for CDs for Infocom’06 scenario.](image4)
Figure 14: Parameter values of the fitted distributions for CDs for Infocom’06 scenario.

Figure 15: Comparison of the CDs distribution obtained from the Infocom’06 scenario and the fitted Pareto2 distribution.

Inter-Contact Times (ICTs):

Figure 16: Goodness of fit for ICTs for Infocom’06 scenario.

Figure 17: Parameter values of the fitted distributions for ICTs for Infocom’06 scenario.

Figure 18: Comparison of the ICTs distribution obtained from the Infocom’06 scenario and the fitted Pareto2 distribution.
V. CONCLUSIONS

We analyzed two contact-based traces collected from a conference scenario. Our analysis shows that both traces satisfy the statistical features listed in [5]. While these traces validate the existence of statistical features commonly found in human movement, we note that the results obtained from our analysis cannot be generalized. Therefore, it is highly critical to understand the statistical features of mobility before simulation studies are performed. In other words, the MANET research community requires a more in-depth analysis of a variety of movement traces. Such analyses will help us standardize the techniques that should be used for trace collection, analysis, as well as movement via mobility models.

VI. FUTURE WORK

In this work, we have validated statistical features found in human movement via analysis of movement traces collected from real scenarios. Our future work in this direction will include a more in-depth analysis of the real movement traces and developing new techniques that will help researchers in this field create standardized techniques for trace collection, analysis, and development of new mobility models that mimic the real scenarios well.

REFERENCES