On Nodal Encounter Patterns in Wireless LAN Traces

Wei-jen Hsu and Ahmed Helmy

Department of Electrical Engineering, University of Southern California
Email: {weijenhs, helmy}@usc.edu

***Abstract*— In this work we study WLAN traces from five different sources and focus on investigation of encounter patterns between users. We find that typical wireless LAN users encounter with a small portion of the whole population (no more than 60% in all traces, and on average between 1.88% to 6.70%). Total encounters of MNs follow BiPareto distribution. These few encounters are sufficient to build a connected relationship network, which is a Small World graph. We further investigate the potential of node-to-node information diffusion, and find that the richness of encounter pattern provides a reliable platform on which information diffusion without infrastructure is feasible and robust.**

I. Introduction

Most of recent research work on analyzing wireless LAN (WLAN) traces focused on individual user behavior [1], [2], [3]. Such previous work provides good understanding of WLAN users, and have made useful WLAN traces available to the research community (e.g. from [1], [2], [3], [13]). Thus far, most research works utilizing these traces focused on individual behavior of mobile nodes (MNs[1]). Understanding of individual behavior is important in itself, but it does not reveal how MNs interact with one another in the real traces. In this paper we go beyond the level of individual users, and start to look into a simple yet important interaction event among MNs: *Encounters*. Encounters are important events in wireless networks as they provide chances for MNs to directly communicate, even without an infrastructure. By studying the encounters of MNs in realistic settings, we develop an understanding of the properties of potential infrastructure-less networks in such environments. We seek to understand encounter patterns of MNs by applying WLAN traces from university and corporation campuses in this paper, in addition to real encounter traces collected at a recent INFOCOM conference [11]. We compare and contrast our observations for the various traces to distill and explain the commonalities and differences observed.

Specifically, we try to quantify the distribution of encounter events a MN has, and look into the encounter patterns of all MNs to understand the relationship between MNs formed by encounters. This is a research topic that received less attention in the past, but can be useful and sometimes essential for classes of future mobile networking protocols. For example, encounter histories are used to discover routes in ad hoc network routing protocols (e.g. MAID[15], EASE[14]), and encounters are used directly in delay tolerant networks (DTNs) to propagate packets. We first define an *encounter* between two users as the event of their association with the same AP for overlapping time intervals. From all the WLAN-based traces we studied, we find that the distribution of encounters is highly asymmetric, indicating a heterogeneous user population. Surprisingly, we find that a user, on average, only encounters between 1.8% and 6.7% of the network user population within a month. We also establish that total number of encounters for each MN follows BiPareto distribution, the parameters of which are environment specific. We further utilize the Small World model [5] to understand the characteristics of the *encounter-relationship graphs* formed by WLAN users, in which two nodes are connected by a link if they ever encounter. We find that although direct encounters of individual nodes just cover small portions of the whole population, WLAN users form connected Small World graphs via encounters, and the metrics of the formed Small Worlds (i.e. disconnected ratio, clustering coefficient, and path length) converge quickly to its long-term values in most cases. Finally, we propose information diffusion experiments to understand how information could be spread among users *without* the help of an infrastructure. We show that even with a relatively high percentage of users being selfish (i.e., not participating in information propagation), the information still spreads and reaches most of the population, indicating the richness of encounter pattern in current WLAN users. Also, if encounters with short time duration are not exploited, the performances of information diffusion also do not degrade significantly.

The contribution of this work is two-folds: First, by investigating the inter-node encounters and utilizing the concept of Small World, we provide new methodologies to understand underlying user behaviors in wireless networks. The understanding gained by studying distributions of encounter events and the *encounter-relationship graphs* reveals how a network can be formed between MNs given the usage pattern in the studied environments. It could be utilized to design better protocols or applications in the future. Second, by experimenting information diffusion with current WLAN traces, we display the potential for the success of information diffusion by the participation of only wireless users (i.e. without infrastructure). We consider these as important findings and they warrant

[1]In this paper we use the terms *user*, *node*, and *mobile node (MN)* interchangeably. We assume that one MAC address in the trace corresponds to a unique device (MN), and a MN is always tied to the same user.

further study.

In section II we discuss the related works. We briefly introduce the traces used in the work and related issues in section III. We study the encounters between MNs in section IV and introduce the Small World approach to explain the encounter-relationship graph in section V. The information diffusion experiment is explained in section VI. We provide some discussions and conclude the paper in section VII.

II. Related work

In this paper we utilize the WLAN traces available to the research community (e.g. [1], [2], [3]) and look into the encounter patterns of the users in those traces. These traces were collected from university campuses and corporations with different characteristics. Former studies on the traces focused mostly on either averaged individual user behavior or global statistics about the network usage. In this work we take one step further to study the *relationships* between users in the traces. We bring up new perspectives to study the WLAN traces by looking into encounter distributions and utilizing Small World theory to describe the encounter relationship graph. Small World graph model is proposed in [5] and widely utilized to describe various networks in many areas, such as social networks, Internet topology, and electrical power networks [6]. In [7] the author applied the concept of Small World to devise a contact-based resource discovery scheme in wireless networks. At the same time, we also perform the same tasks on the encounter traces collected at a conference [11], and compare the findings.

In [15] the authors found that, under mobility models with homogeneous behaviors (i.e. Each node follows exactly the same model with some randomness), eventually each node encounters with all other nodes in the network (i.e. achieving 100% encounter ratio). However, the empirical observations from large WLAN traces show very different behaviors, with most nodes encountering only a very small portion of the whole population, during a time frame as long as a month. This observation indicates that the user populations in larger environments, such as university campuses, are actually *not* homogeneous.

In recent years, packet forwarding in sparse, frequently disconnected ad hoc networks received increasing attention from the research community. In such network scenarios, a complete end-to-end path from source node to destination node is usually unavailable. Therefore, mobile nodes have to store packets and forward copies of the packet to others during encounters, and the packet propagates in the network through nodal mobility and encounter. Most of the previous work in this area focus on designing packet forwarding heuristics [8] , [9], [10]. In general, different degrees of knowledge of mobility pattern is assumed [8], or a homogeneous mobility model is used [10]. In this work, we complement the above studies from an empirical point of view, and investigate the issue of whether the store-and-forward model is potentially feasible under current usage pattern of wireless devices by deriving encounter patterns of nodes from the WLAN traces and applying simple forwarding strategy on the derived encounters. Our findings, as we show in section VI, are encouraging. In [20], the authors also use WLAN traces and develop a routing strategy based on the frequent association pattern of MNs. They make similar assumption about encounters as we do in deriving communication opportunities between MNs from the WLAN traces, but the work in [20] is more focused on a routing scheme utilizing estimation of MN pairs with higher encounter probability by comparing involved MN's individual association histories. We, on the contrary, focus more on understanding of encounter pattern itself.

There are also research works explicitly focused on collecting encounter traces using small mobile devices[11], [19] with similar motivation as ours. Utilizing always-on, easy-to-carry devices such as iMOTES or PDAs, in those works researchers are able to capture human encounters with high accuracy. However, encounters in human mobility do not always translate into encounters in computing devices and hence communication opportunities, as the usage patterns (such as devices turning on/off, preferred locations to operate the devices, etc.) are not captured. The WLAN traces, on the other hand, capture the usage pattern (i.e. Time and locations the devices are used) truthfully, but may miss encounters out of the coverage region of access points. The two approaches have strengths in different aspects and can be complementary. In addition, deriving encounter information from WLAN traces has the advantage of obtaining a much larger data size (e.g. thousands of nodes) as compared to the experiment approach (e.g. tens of nodes) taken in [11], [19]. Further more, the WLAN traces have been collected for years in university campuses or corporations (in which encounter patterns may be different from a conference setting studied in [11]) and they provide opportunities to understand encounter patterns across longer time period. We also hope that, by deriving encounter patterns using WLAN traces, we would be able to find the most important, longer duration encounter events, which are more useful for information exchange. In that respect the derived encounter traces are suitable for investigation of information diffusion performances.

In addition to taking the available traces from the community, we also make our own WLAN traces collected at USC campus available at [21], together with many pointers to existing WLAN trace archives.

III. Target environment and trace collection methods

In this study we mainly focus on wireless traces collected from university campuses and corporations. We obtain wireless LAN traces from four different sources, including totally over 12,000 distinct users and over 1,300 APs. To our best knowledge this is the most extensive data set analyzed so far. Among the traces, the USC trace is collected specifically for the purpose of our studies, while Dartmouth [3], UCSD [2], and MIT [1] traces were collected by other research groups. We also analyze the encounter trace collected at INFOCOM

TABLE I

STATISTICS OF STUDIED TRACES

Trace source	Unique users	Unique APs	Unique buildings	Trace duration	User type	Environment	Analyzed part in this paper	Users in analyzed part	Labels used in graphs
MIT[1]	1,366	173	3	Jul. 20 ’02 to Aug. 17 ’02	Generic	3 Engineer buildings	Whole trace	1,366	MIT
Dartmouth[3]	10,296	623	188	Apr. ’01 to Jun. ’04	Generic	Whole campus	Jul. 2003	2,518	Dart-03
							Apr. 2004	5,582	Dart-04
UCSD[2]	275	518	N/A	Sep. 22 ’02 to Dec. 8 ’02	PDA only	Whole campus	Sep. 22 ’02 to Oct. 21 ’02	275	UCSD
USC	4,548	79 ports	73	Dec 03-Now (trap) Apr 20 05-Now (detail)	Generic	Whole campus	Apr. 20, ’05 to May. 19 ’05	4,528	USC
Cambridge	41 internal nodes	N/A	N/A	Mar. 7 ’05 to Mar. 10 ’05	iMOTE	conference	Whole trace	41	Cambridge

by Cambridge group [11]. For the Cambridge trace, we focus on the internal nodes[2] only.

These five traces are chosen to represent different environments, user populations, location granularity, and trace-collection methods. The traces were collected with different methodologies, but we can derive the association history information for each user from all four WLAN-based traces. By association history, we mean the timestamps of events related to changes of user association, including a user starts/ends association with an AP, or re-associates to another different AP. We could further derive the time duration a user associated with each AP from its association history. Time granularity of events is one second for Dartmouth traces as they are collected using syslog server. For USC trace we collect user online/offline event at the switch ports, which also gives time granularity of one second. For UCSD and MIT, polling methods were used to collect the traces, hence the time granularity is limited by the polling intervals, which are 20 seconds and 5 minutes, respectively. Encounter trace can be derived from the association history as detailed in the next section. For the Cambridge trace, each node explicitly collects its encounters by sending neighbor inquiries every 120 seconds. We use the encounter information directly.

In order to make the results we get below comparable between traces, we only analyze selected one-month chunks from the longer Dartmouth and UCSD traces. All these traces, except UCSD trace, collect association events of generic wireless network users with various WLAN capable devices. UCSD trace is from a specific study about PDA users. All the traces, except MIT trace, are collected from the entire campus wireless network. MIT trace is collected from three engineering buildings, hence its user population is not as diverse as the other traces, and the geographic scope of trace collection is smaller. USC trace is the only one that has coarser, per switch port location granularity. Each switch port aggrgates the traffic from multiple APs, with the coverage approximately corresponds to a building on campus. All the others WLAN traces have per AP location granularity. Cambridge trace is the only short trace in our analysis. It is a four-day trace at a conference setting (INFOCOM 2005). The encounter information is directly collected, not derived, for the Cambridge trace. We summarize the important characteristics of these traces, including the time periods we analyze in the paper, in Table I.

[2]Internal nodes refer to the iMOTES distributed by the researchers in their experiments [11]

IV. ENCOUNTERS BETWEEN NODES

Nodal encounters in mobile network are important events as they provide opportunities for involved nodes to build up some relationship or to communicate directly. Here, for WLAN traces, we define an *encounter event* as the duration of two MNs associate with the same AP during overlapping time intervals. The wireless LAN traces provide sequences of AP (switch port for USC trace) association history for MNs in the network. We can derive when MNs encounter with each other by replaying the traces and finding MNs associated with the same AP simultaneously. The approximation may be not completely accurate, as there can be nodes covered by the same AP but unable to communicate directly, nodes being able to communicate while associated with different APs, or nodes encountering outside the coverage of any AP. However, we believe that the encounter events derived from WLAN traces capture the major portion of MNs within direct communication range under current usage pattern.

The distribution of these encounter events is the first step to understand the structure of inter-MN relationship in the traces. The direct questions to ask about the encounter events are: How many other MNs does a user meet? Do nodes meet with each other repeatedly or not?

Fig. 1 shows the CCDF of fraction of MNs a given MN has encountered through the whole trace period. From the figure we observe that all the nodes in WLAN traces encounter only at most about 50% of the user population within a month, with UCSD trace being the only exception. This may be partly due to the fact that the 275 PDA users in UCSD trace were all selected from freshman class, and they tend to stay in several common dorms as stated in [2]. In all the other traces, **on average a MN encounters with only 1.88% (Dart-03) to 6.70% (Dart-04) of the whole user population** within the 30-day trace period. The small average encounter ratio is a combined result of several reasons: (1) Most MNs are not always on, and (2) Most MNs do not visit many APs[16], hence they can only meet with those who also visit this small set of APs.

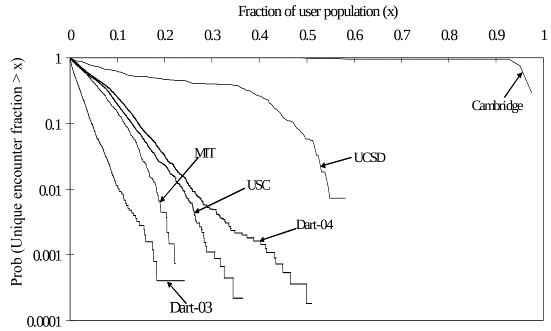


Fig. 1. CCDF of unique encounter fraction

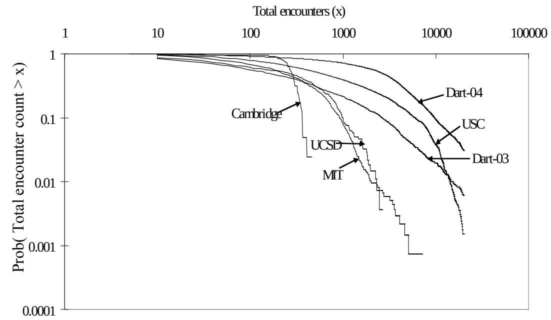


Fig. 2. CCDF of total encounter count

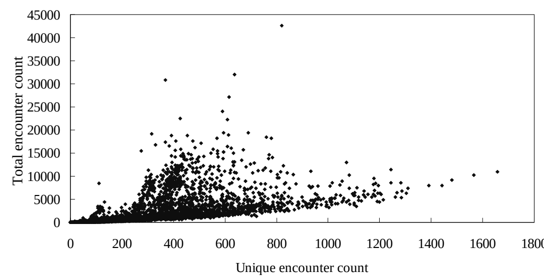


Fig. 3. Unique encounter count versus total encounter count, USC.

On the other hand, from the Cambridge trace, most of the 41 users meet with majority of others during the short trace duration (4 days). Specifically, there are 12 MNs who meet with all other 40 MNs, and 39 out of 41 MNs meet at least 38 other nodes. The curve in Fig. 1 is mostly a horizontal line at high probability until unique encounter fraction reaches 0.95. This high unique encounter fraction may be due to the environment setting (a conference, at which people are supposed to meet) or the fact that the selection of participants are related (i.e. People who are interested in study of mobility patterns and wireless networks in general) rather than randomly picked from the conference attendees.

We also show the CCDF of the total encounter events a MN has throughout the trace period in Fig. 2. We observe the **total encounter counts for MNs in each trace span across several orders**. There are both MNs with extremely few or many encounters. This is an evidence of **heterogeneous behavior** among MNs. The actual number of total encounters depends on the size of population in the traces. Large traces (i.e., USC and Dartmouth traces) tend to have more encounters than small traces (i.e., UCSD, Cambridge traces). However, regardless of the size of population, **the curves for total encounter count derived from WLAN traces seem to follow BiPareto distribution**. We try to fit BiPareto distribution curves to the empirical distribution curves, and use Kolmogorov-Smirnov test [12] to examine the quality of fit. The resulting D-statistics for all traces are between 0.068 and 0.025, which indicate we have a reasonably good fit between the BiPareto distribution curves and the empirical distribution curves. The details about Kolmogorov-Smirnov test and the parameters of the fitted BiPareto distribution curves are listed in Appendix A. For the Cambridge trace, the total encounter counts for MNs are not as diverse as those in WLAN traces. This may be due to the fact that most nodes participate the conference actively throughout the whole trace period (4 days), but this is unlikely for the longer, one-month WLAN traces. BiPareto distribution does not show a good fit for the Cambridge trace, as its total encounter distribution drops sharply at a "knee" around 250.

A closer investigation of the relationship between unique encounter count and total encounter count of the same MN reveals that **high unique encounter count does not always imply high total encounter count**. The correlation coefficients between unique encounter count and total encounter count for various traces are from 0.732 to 0.195. Except for the UCSD trace, all other traces have correlation coefficient below 0.6. As an illustration, we show the scatter plot of unique encounter count versus total encounter count for USC trace in Fig. 3. We observe that some nodes have not many unique encounter counts, but high total encounter counts. This indicates that some node pairs may have a lot of repetitive encounters, suggesting some node pairs have closer relationship than other pairs. This point warrants further study.

V. Encounter-relationship graph

In section IV, we see that **MNs have low percentage of unique encounters among the whole population**. Given this fact, We raise a question regarding the possibility of establishing campus-wide relationships among majority of MNs via encounters alone. That is, do encounters link MNs on campus into one single community, or just many small cliques?

To investigate this question, we define a *static encounter-relationship graph* (ER graph) as follows: Each MN is represented by a node in the *ER graph*, and an edge is added between two nodes if the two corresponding MNs have encountered at least once during the studied trace period. By construction of the *ER graph*, we collect all encounters of MNs within a time period and collapse them on a static graph. The exact timing of encounters are ignored, but we focus on the structure of interconnections built between nodes by available encounter events during that period of time. In other words, the concept of *ER graph* is introduced to capture the potential of establishing a connected network among MNs based on direct encounters alone, and understand the structure of such network.

We use three important metrics to describe the characteristics of encounter-relationship graphs, defined as follows:

- **Clustering coefficient** (CC) is used to describe the tendency of nodes to from cliques in the graph. It is formally defined as:

$$CC = \frac{\sum_{n=1}^{M} CC(n)}{M}$$

where

$$CC(n) = \frac{\sum_{A,B \in N(n)} I(A \in N(B))}{|N(n)| \cdot (|N(n)| - 1)}$$

n, A, B are nodes. $N(n)$ is the set of neighbors of node n and $|N(n)|$ is its size. $I(\cdot)$ is the indicator function. M is the total number of nodes in the graph.
Intuitively, clustering coefficient is the average ratio of neighbors of a node that are also neighbors of one another. Higher *CC* indicates higher tendency that neighbors of a given node are also neighbors to each other.

- **Disconnected ratio** (DR) is used to describe the connectivity of ER graph. It is defined as:

$$DR = \frac{\sum_{A=1}^{M} (M - |C(A)|)}{M(M-1)}$$

where $C(A)$ is the set of nodes that are in the same connected sub-graph with node A. *DR* indicates that, on average, what is the percentage of unreachable node starting from a given node in the graph.

- **Average path length** (PL) is used to describe the degree of separation of nodes in the *ER graph*. It is defined as:

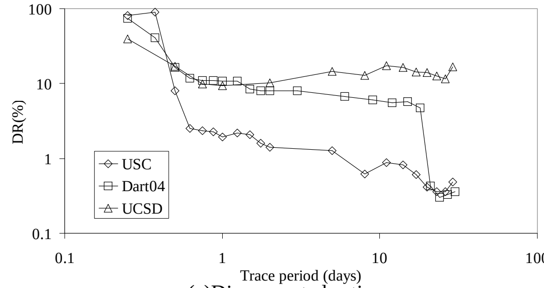
$$PL = (1 - DR) \cdot PL_{con} + DR \cdot PL_{disc}$$

Where PL_{con} is the average path length among the connected part of the *ER graph*, defined as:

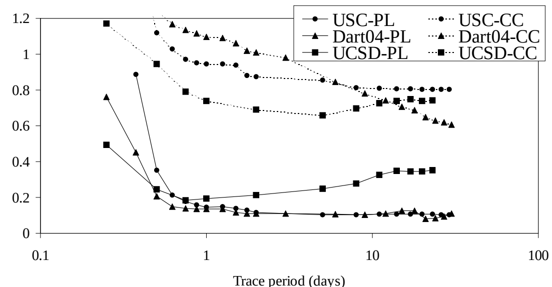
$$PL_{con} = \frac{\sum_{A=1}^{M} \sum_{B \in C(A)} PL(A,B)}{\sum_{A=1}^{M} |C(A)|}$$

$PL(A,B)$ is the hop count of the shortest path between node pair (A,B) in the *ER graph*. Note this path is not the same as shortest spatial path between node pair (A,B), which may never exist. PL_{disc} is the penalty on average path length for *disconnected* node pairs in *ER graph*. In the following we use the average path length of regular graphs (defined later) with the same node number and average node degree for PL_{disc}.

We study how the above metrics evolve for the *ER graphs* derived from various studied period of WLAN traces. Taking USC trace, Dartmouth trace (Dart-04), and UCSD trace as examples, we show the evolution of the three metrics with respect to various studied trace periods in Fig. 4 (a)-(b). The graphs for other traces show very similar trends, and we leave them in Appendix B to maintain conciseness here.



(a)Disconnected ratio



(b)Normalized clustering coefficient and average path length. The figure is cut from above to show the details between 0 and 1 on Y-axis.

Fig. 4. Change in the *ER graph* metrics with respect to trace period

From Fig. 4 (a) we note that given sufficient long trace durations, the *ER graphs* **have low** *DR* **(not larger than** 10% **for traces longer than one day in most cases), which implies that nodal encounters are sufficient to provide opportunities to connect almost all nodes in a single community**, even though each node encounters with only a small subset of MNs directly. This is an encouraging result that points out the feasibility of building a large, wide-reach network relying only on direct encounters. We further point out that although *DR* starts out very high with very short trace period (i.e. trace duration under one day), since MNs have not moved around to create encounters yet, it decreases rather quickly as trace period increases. Within one day, *DR*'s reduces to around 10%. Although the numbers of MNs in the *ER graph* keep increasing as we look at longer trace periods, in most cases *DR* does not change significantly after one day.

Another interesting finding is revealed by taking a further look at the other two metrics, clustering coefficient (*CC*) and average path length (*PL*). To highlight a unique property of these *ER graphs*, we also calculate *CC* and *PL* for regular graphs and random graphs with the same corresponding total node number M and average node degree d. In regular graphs, nodes are first arranged on a circle and each node is connected to d closest neighbors on the circle. In random graphs, d randomly chosen nodes are assigned as neighbors for each node. Typically, regular graphs have high *CC* and *PL* while random graphs have low *CC* and *PL*. They are the two extreme cases on the spectrum. In Fig. 4 (b), we show the normalized *CC*'s and *PL*'s of the *ER graphs* for various trace periods. These normalized metrics represent, on the scale from 0 (corresponding to random graph) to 1 (corresponding to regular graph), where do the metrics obtained from the *ER* *graph*s fall. They are defined as:

$$CC_{norm} = \frac{CC - CC_{rand}}{CC_{reg} - CC_{rand}}$$
$$PL_{norm} = \frac{PL - PL_{rand}}{PL_{reg} - PL_{rand}}$$

where CC_{norm} and PL_{norm} represent normalized CC and PL, respectively. The subscripts reg and $rand$ imply that the corresponding metric is obtained from the regular graph and random graph, respectively, with the same total node number and average node degree.

We observe that *ER graphs* display **high normalized CC's** which are close to those of corresponding regular graphs (i.e., Normalized *CC*'s being close to 1, and in some cases even higher than 1), and **low normalized PL's** which are close to those of corresponding random graphs. This highlights that a special pattern of encounters exists in all WLAN traces: Nodes visiting similar set of APs are highly likely to encounter with all others and introduce highly connected clusters among these nodes, leading to high *CC*. This phenomenon is especially obvious for very short traces, since most the MNs do not change its association to create many encounters, the *ER graphs* for short trace periods feature many small disconnected cliques, each of them being a full-mesh formed by MNs associate with the same AP for that trace period. As we look at longer traces, some of the nodes in one cluster also have random encounters with nodes in other clusters, and those links serve as "shortcuts" in the *ER graphs* that reduce *PL*. In previous literature, graphs with high *CC* close to regular graphs and low *PL* close to random graphs are referred as Small World graphs [5], [6]. By looking at various traces, we indicate that the *ER graphs* **formed by encounters among nodes using wireless network appear to be Small World graphs**. We also observe that **both *PL* and *CC* converges to its final values rather quickly in about one day for most traces**, although the size of *ER graphs* keep increasing as more nodes appear in longer traces.

For the Cambridge trace, we look into similar metrics. We find that for even a small period of time (e.g., 1 day) the 41 MNs encounter most of other nodes. Hence, the *CC* is very high (above 0.91 even if we take only the first day into consideration), and the *PL* is low (less than 1.1). Actually, the 41 MNs presented in the trace almost form a fully-connected mesh, and the *DR* is 0. This may be partly due the nature of the conference setting from which the trace was collected. People move around to meet more often than in their regular daily life at universities or corporations, hence the encounter pattern at a conference seems to be richer than in regular environments. The well-connected *ER graph* may also come from the fact that conference is held in a place much smaller than a university campus or a corporate building. The single clique in the Cambridge trace may in fact correspond to one of the cliques observed in the WLAN traces (i.e., The MNs visiting similar sets of APs). However, the above arguments need further validation, by more thorough study of encounters in different settings.

VI. Information diffusion using encounters

In addition to establishing relationship between nodes, encounters can also be utilized to diffuse information throughout the network. In this model, information is spread with nodal mobility and encounters, where nodes exchange information when they encounter each other directly. The speed and reachability of information diffusion among the nodes are determined by the actual pattern and sequences of encounters. In this section we seek to answer the question of whether the *current* encounter pattern between MNs in wireless networks are rich enough to be utilized for information diffusion. If the answer is yes, what is the delay incurred in such information diffusion scheme, and how robust is it?

In this section, we first understand the optimistic expectation of the potential performance of information diffusion under ideal assumptions in subsection VI-A. We remove some of the assumptions and evaluate the performance in more realistic settings in subsequent subsections.

A. Ideal performance of information diffusion

As a first step to understand the potential of information diffusion under realistic encounter pattern, we make the following ideal assumptions: (1) There are sufficient bandwidth and reliable communication between MNs, and sufficient storage space on all MNs. (2) MNs discover the communication opportunities immediately when they encounter other MNs, and (3) every MN in the network is willing to participate in forwarding information for others. In this study, we focus more on analyzing how the encounter pattern itself influences the performance of information diffusion. The experiments in subsection VI-B and VI-C deal with changing encounter patterns when some of the above assumptions are removed. However, we do not address the technology limitations on the devices itself (i.e., storage capacity, power constraint, etc.).

The diffusion mechanism we use is the following: When a source node has information to send, it simply transmits it to all nodes it encounters if they have not received the information yet. All intermediate nodes cooperate in information diffusion, keeping a copy of received information and forwarding it the same way as the source node does. This simple approach is known as epidemic routing in the literature [9]. Under perfect environment with sufficient resources, it achieves lowest delay and highest delivery rate.

In all the simulations, we use a traffic pattern in which the source node has some information it wishes to send to all other nodes. The source starts to "diffuse" the information when it is first online. As time evolves, nodes encounter with each other and increasing portion of the whole population receive the information. We study the percentage of nodes that received the information with various trace periods (i.e., the MNs that received the message over total MNs that appeared during the trace period under discussion) and show the results in Fig. 5, using USC, Dart-04, Dart-03, and MIT traces as examples. Each point in the figures of this section is an average value of

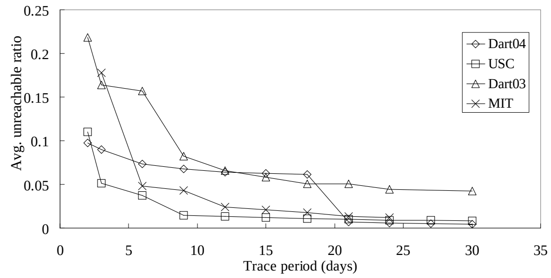


Fig. 5. Unreachable ratio of broadcast messages using epidemic routing

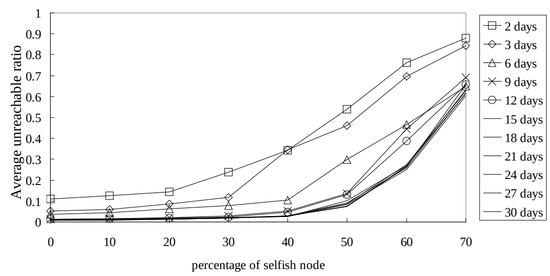


Fig. 6. Unreachable ratio with various selfish node percentage and trace period

multiple experiments. In each experiment we start information diffusion from a different source node. We choose to use 30% of the nodes that appear earliest in the corresponding trace period as sources.

From Fig. 5 we observe that even within a short trace period (e.g., two days) the information can reach moderate portion of the population as the unreachable ratio is less than 25% in all traces. As the trace period increases, reachability also improves. In all except the Dart-03 trace, **the unreachable ratios are less than 2% if we allow one month for the information diffusion**. Given that most nodes encounter a small portion of the whole population (Fig. 1), this result is perhaps beyond our original expectation. **It gives a positive confirmation that it is potentially possible to deliver information relying only on encounters**, in a campus environment with high success rate, under *current* user behavioral pattern.

B. Performance of information diffusion with selfish users

After studying the ideal case, we consider more realistic setup. We first relax the ideal assumption (3) above. In some cases, a portion of nodes may not be cooperative to propagate the information. To understand how uncooperative users potentially influence the feasibility of information diffusion, we carry out the following experiment: We make a portion of users *selfish* such that they never forward information for other sources, and we study the performance degradation under this setup. For each of the trace periods used, we increasingly make a certain percentage of nodes selfish, starting from those with *highest unique encounter counts*. By making nodes with high unique encounter counts selfish first, we eliminate more transmission opportunities than if we pick selfish nodes randomly, hence we expect to observe greater impact on performance.

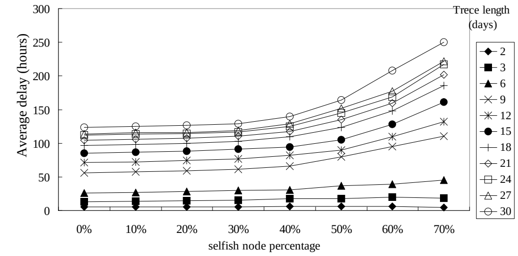


Fig. 7. average message delay with various selfish node percentage and trace period

The relationship between percentage of selfish node and the unreachable ratio for USC trace is shown in Fig. 6. For the sake of conciseness, we only show figures for USC trace here. The figures for other traces display similar trends and they are shown in Appendix C. The result is very surprising: For all trace period tested, the unreachable ratio does not increase significantly before at least 20% of nodes are selfish. The performance is even more robust if we take longer period of trace. This implies that **even a significant portion of users are not willing to propagate information for others, the underlying nodal encounter pattern is rich enough for the information to find an alternative way through**. Hence delivery rate is quite robust for up to an intermediate percentage of selfish nodes. Note that we make the MNs with most unique encounters selfish first, hence the performance of information diffusion is robust even if the nodes with *most* chances to propagate the information are not cooperative. We further show how average delay of information diffusion changes with increasing selfish node percentage in Fig. 7 for USC trace. In the figure, average delay increases for longer trace duration because information that is not deliverable in shorter trace periods becomes deliverable. More interestingly, for all tested trace durations, average delay does not increase significantly before more than 40% of nodes are selfish. This implies **average delay is also robust against selfish user behavior up to an intermediate percentage**.

C. Performance of information diffusion with long encounters only

Another ideal assumption we made is that MNs can communicate with each other successfully regardless of the duration of encounters. This may not be true in realistic scenario due to the following reasons: First, short encounters that last only for seconds may not be sufficient to complete full message exchange due to bandwidth limitation and protocol overheads. To address this issue, one can simply remove the encounters for which the time-bandwidth product is too small to permit useful information exchange. Second, MNs do not discover encounter events instantly. Periodic beacon signals are typically used to discover neighbors, and the broadcast period cannot be set too short in order to save power. Hence, short encounters may not even be noticed by the participating parties. In the following experiment, we address these limitations by removing encounters with short durations

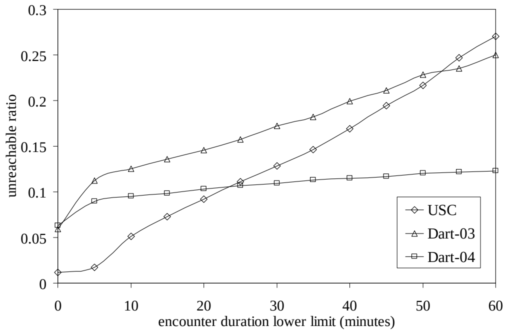


Fig. 8. Unreachable ratio with short encounters under the duration lower limit removed

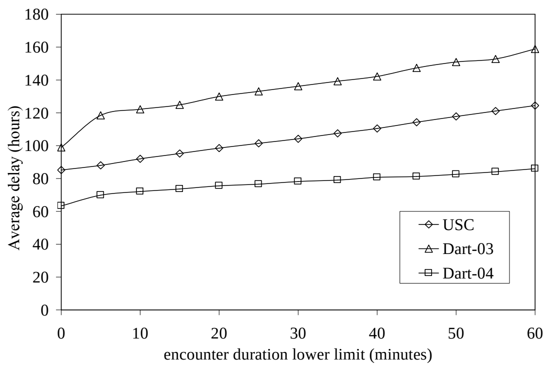


Fig. 9. Delay with short encounters under the duration lower limit removed

from all the MNs in the trace, and re-evaluate the performance of information diffusion.

In Fig. 8, we show the relationship between unreachable ratio versus the lower limit of encounter duration (i.e. We remove all encounter events that have shorter duration than the value), using 15-day traces from USC and Dartmouth as examples. From the graph we observe that, unreachable ratio increases almost linearly as we increase the lower limit of usable encounter duration. There is no obvious point at which the performance suddenly degrades severely. We carry out the experiments up to the shortest usable encounter set at one hour, a rather demanding scenario. Even in such cases, the unreachable ratio is below 30%. This implies **removing encounters with short duration does not cause abrupt degradation in the performance of information diffusion**, in terms of both reachability and delay (the graph is shown in Fig. 9). In other words, short encounters are not the key reason for the success of information diffusion. The encounter events with long duration are also rich enough to be utilized for message propagation.

VII. Discussion and Conclusion

In the paper we investigate the encounters between MNs in WLAN traces from four sources. We find that MNs encounter with only a small subset of other nodes (on average between 1.88% to 6.70%), and the total encounter counts follow BiPareto distribution. In spite of low percentage of unique encounters, the relationship graph constructed using encounters alone connects most of the MNs. Further more, such encounter-relationship graphs display Small World graph characteristics, and its metrics converge to its long-term value within only short time periods. Using simulation study with a simple protocol, we also display the potential for information diffusion without relying on the infrastructure, utilizing encounters and mobility of MNs alone.

Low encounter percentage as shown in the traces is not observed in any of the simulation scenarios used for performance evaluation in the literature. In typical synthetic mobility scenarios, as those summarized in [4], all nodes follow the same model to make movement decision, albeit with randomness, and eventually encounter with all other nodes [15]. The encounter pattern from real wireless network traces reflects that university campus is a *heterogeneous* environment rather than a homogeneous one constructed by synthetic mobility models. To better understand how protocols perform in such heterogeneous environment, using homogeneous synthetic models [4] is not sufficient.

Comparison of the above findings to the encounter trace collected at a conference shows that these findings are more applicable to large, heterogeneous environments. The conference trace shows higher, almost 100% unique encounters. The total encounter counts of MNs do not differ as drastically as in the WLAN traces. The *ER graph* of the conference trace is connected, almost complete graph. In other words, for smaller environment with users showing similiar mobility characteristics, such as a conference, synthetic models may be adequate. It warrants further study to understand the differences between the environments.

The Small World approach to understand *ER graphs* and the result of information diffusion experiments both highlight positive potential of building a campus wide network without infrastructures. The robustness of information diffusion brings up two interesting points: (1) For message delivery, the delivery ratio and delay are not affected significantly, even if we can not choose the shortest paths due to non-cooperative users or unutilized short encounters. (2) On the other hand, it would be difficult to prevent diffusion of harmful or malicious messages, such as computer worms or viruses, from propagating through encounters. Both observations are due to the richness in underlying encounter pattern providing abundent chances for message delivery. The performance of information diffusion under various information delivery schemes and potential methods to prevent malicious information from spreading are both directions for future work. In the future we plan to apply the understandings gained from studying WLAN traces to provide more realistic user models (some initial results are available in [16]), and design/evaluate information diffusion protocols that utilize the understanding of user behaviors from the traces.

Acknowledgments

We would like to thank Brian Yamaguchi and Carl Hayter at USC ISD division for helping us collected the WLAN traces on the USC campus for the past 2 years. We would also like to extend our thanks to Dr. David Kotz, Dr. Tristan Henderson, Dr. Marvin McNett, Dr. Magdalena Balazinska, and Dr. James Scott for helping us obtain the Dartmouth, UCSD, MIT, and Cambridge traces, respectively.

We are also thankful to USC HPCC and Jabed Faruque for providing high quality computing facility. In addition, we would like to thank Fan Bai, Sapon Tanachaiwiwat, and Shaocheng Wang for providing useful feedback on this work.

This work was funded in part through the NSF CAREER Award 0134650.

REFERENCES

[1] M. Balazinska and P. Castro, "Characterizing Mobility and Network Usage in a Corporate Wireless Local-Area Network," In Proceedings of MobiSys 2003, pp. 303-316, May 2003.

[2] M. McNett and G. Voelker, "Access and mobility of wireless PDA users," ACM SIGMOBILE Mobile Computing and Communications Review, v.7 n.4, October 2003.

[3] T. Henderson, D. Kotz and I. Abyzov, "The Changing Usage of a Mature Campus-wide Wireless Network," in Proceedings of ACM MobiCom 2004, September 2004.

[4] T. Camp, J. Boleng, and V. Davies, "A Survey of Mobility Models for Ad Hoc Network Research," Wireless Communication & Mobile Computing (WCMC): Special issue on Mobile Ad Hoc Networking: Research, Trends and Applications, vol. 2, no. 5, pp. 483-502, 2002.

[5] D. J. Watts and S. H. Strogatz. "Collective Dynamics of 'Small-World' Networks," Nature, vol. 393, pp. 440-442, 1998.

[6] R. Albert and A. Barabasi, "Statistical mechanics of complex networks," Review of modern physics, vol. 74, no. 1, pp. 47-97 , Jan. 2002.

[7] A. Helmy, "Small Worlds in Wireless Networks," IEEE Communications Letters, pp. 490-492, Vol. 7, No. 10, October 2003.

[8] K. Fall, "A Delay-Tolerant Network Architecture for Challenged Internets," In Proceedings of ACM SIGCOMM, August 2003.

[9] A. Vahdat and D. Becker, "Epidemic Routing for Partially Connected Ad Hoc Networks," Technical Report CS-200006, Duke University, April 2000.

[10] A. Spuropoulos, K. Psounis, and C. Raghavendra, "Single-copy routing in intermittently connected mobile networks." IEEE SECON, October 2004.

[11] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, and C. Diot, " Pocket Switched Networks and the Consequences of Human Mobility in Conference Environments," Workshop on Delay Tolerant Networking, Aug. 2005.

[12] R. Hogg and E. Tanis, "Probability and Statistical Inference," sixth edition, Prentice Hall, 2001.

[13] W. Hsu and A. Helmy, "IMPACT: Investigation of Mobile-user Patterns Across University Campuses using WLAN Trace Analysis," unpublished USC technical report. Available at http://nile.usc.edu/MobiLib/Trace_analysis_TR.pdf

[14] M. Grossglauser and M. Vetterli, "Locating Nodes with EASE: Mobility Diffusion of Last Encounters in Ad Hoc Networks," In Proceeding of IEEE INFOCOM, April 2003.

[15] F. Bai and A. Helmy, "Impact of Mobility on Mobility-Assisted Information Diffusion(MAID) Protocols in Ad hoc Networks," USC Technical Report. Available at: http://nile.usc.edu/important/MAIDreport.pdf

[16] W. Hsu and A. Helmy, "On Important Aspects of Modeling User Associations in Wireless LAN Traces," the Second International Workshop On Wireless Network Measurement (WiNMee 2006), April 2006.

[17] C. Nuzman, I. Saniee, W. Sweldens, and A. Weiss, "A Compound Model for TCP Connection Arrivals for LAN and WAN Applications," Computer Networks, 40:319V337, October 2002.

[18] M. Papadopouli, H. Shen, and M. Spanakis, "Characterizing the Duration and Association Patterns of Wireless Access in a Campus," 11th European Wireless Conference 2005, Nicosia, Cyprus, April 10-13, 2005.

[19] J. Su, A. Chin, A. Popivanova, A. Goel, and E. de Lara, User mobility for opportunistic ad-hoc networking, in Proceedings of the 6th IEEE Workshop on Mobile Computing Systems and Applications (WMCSA04), 2004.

[20] J. Leguay, T. Friedman, and V. Conan, "Evaluating Mobility Pattern Space Routing for DTNs," to be appear in INFOCOM 2006.

[21] MobiLib: Community-wide Library of Mobility and Wireless Networks Measurements (Investigating User Behavior in Wireless Environments). USC WLAN trace and pointers to many WLAN trace archives available at http://nile.usc.edu/MobiLib.

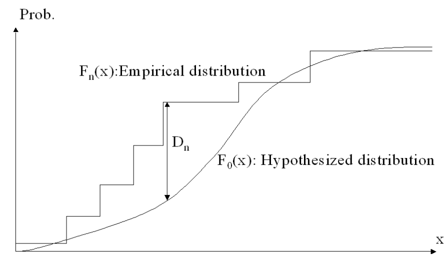


Fig. 10. Illustration of D-statistics and K-S test

APPENDIX A. BIPARETO DISTRIBUTION AND KOLMOGOROV-SMIRNOV TEST

In this section we first briefly introduce Kolmogorov-Smirnov test and BiPareto distribution, and then list the detail numerical results of using BiPareto distribution curves to fit total encounter distributions obtained in section IV.

BiPareto distribution is used in [17] to fit the number of connections per user TCP session and mean connection inter-arrival time in a TCP session. Later, BiPareto distribution is again used in [18] to fit the distribution of association session length in wireless LAN. In this work, we use it to fit the distribution of encounters a MN has in WLANs. The CCDF of BiPareto distribution is as follows:

$$Prob(X > x) = (\frac{x}{k})^{-\alpha}(\frac{x+c}{k+c})^{\alpha-\beta},\ x > k$$
$$Prob(X > x) = 1,\ x \leq k$$

The left part of CCDF curve of BiPareto distribution on log-log scale is a straight line with slope $-\alpha$. As the x variable comes close to the turning point, c, the slope of the CCDF curve gradually changes from $-\alpha$ to $-\beta$. In our study of total encounter distributions, we choose $k = 1$ for all curves.

Kolmogorov-Smirnov test is used to determine whether the hypothesized distribution (in our case, the BiPareto distribution) adequately fit the empirical distribution. K-S test is not sensitive to the binning of data set, unlike Chi-square test. Therefore we choose K-S test in our study.

Referring to Fig. 10, in K-S test the distances between the hypothesized distribution and the empirical distribution are measured throughout the range of random variable x, and the maximum of the measured distances is called D-statistics. More formally, D-statistics is defined as:

$D_n = sup_x[\mid F_n(x) - F_0(x) \mid]$

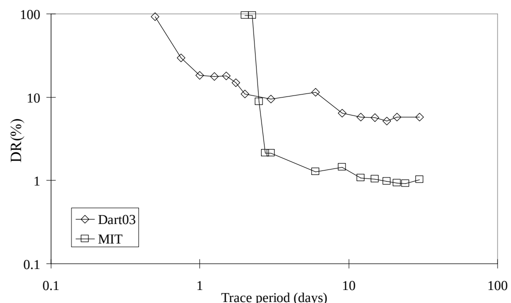
where $F_n(x)$ and $F_0(x)$ are the empirical and hypothesized distributions, respectively. Intuitively, D-statistics measure the maximum difference between the two distribution curve. A smaller D-statistic indicates a better fit of the hypothesized distribution to the empirical distribution.

We use minimum squared error method to find the best fit of BiPareto distribution curves to the empirical total encounter

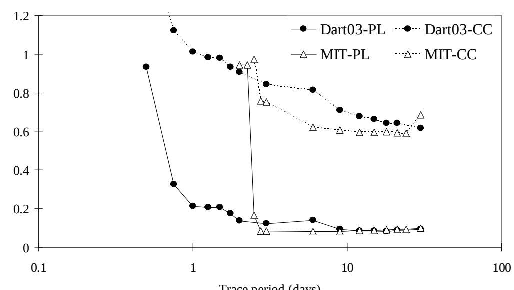
TABLE II

BIPARETO DISTRIBUTION FITTING TO TOTAL ENCOUNTER CURVES AND D-STATISTICS FOR K-S TEST

Trace name	BiPareto parameters			D-statistics
	α	β	c	
MIT	0.027	9.8	4000	0.036
UCSD	0.062	16.3	9900	0.068
USC	0.019	0.83	550	0.049
Dart-03	0.0723	0.81	290	0.049
Dart-04	0.0285	4.43	11850	0.025



(a)Disconnected ratio



(b)Normalized clustering coefficient and average path length

Fig. 11. Change in the *ER graph* metrics with respect to trace period

distributions for various traces. The parameters are listed in Table II. From the table we observe that the D-statistics are no larger than 0.05 except for UCSD trace (0.07), indicating a reasonable fit of the BiPareto distribution.

APPENDIX B. ADDITIONAL GRAPHS FOR ENCOUNTER-RELATIONSHIP GRAPHS METRICS

In addition to the figures shown in section V, we also obtain the same metrics for MIT and Dart-03 traces. The figures have similar trends as discussed in section V (Fig. 11). One interesting observation here is that for MIT trace, disconnected ratio is very high until day 3 in the trace. A further investigation reveals that MIT trace collection was started on a Saturday, and for a pure working environment Saturdays and Sundays are the least active days. The disconnected ratio is almost 100% until day 3 because the MNs that are on during the weekend are mostly stationary ones. We observe a jump of number of node in the trace, a sudden decrease in *DR*, and an abrupt change in both *CC* and *PL* on day 3.

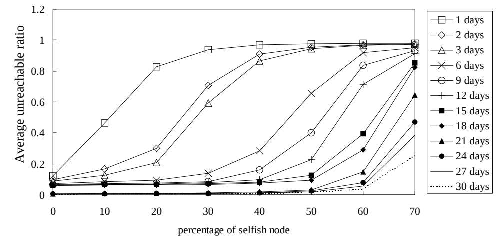


Fig. 12. Information delivery ratio with various selfish node percentage and trace period, using Dartmouth trace April 2004

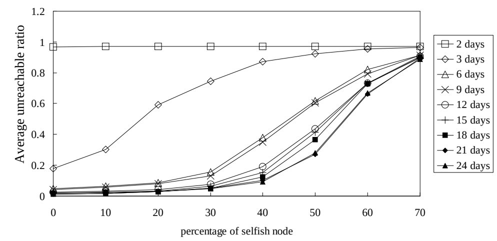


Fig. 13. Information delivery ratio with various selfish node percentage and trace period, using MIT trace

APPENDIX C. ADDITIONAL GRAPHS FOR INFORMATION DIFFUSION EXPERIMENTS

In addition to the USC trace, we further perform similar information diffusion experiments on adding selfish user behavior to the Dartmouth and MIT traces. The experiment setup is the same as described in subsection VI-B.

The results for the average unreachable ratio are shown in Fig. 12, 13, and 14 for Dart-04, MIT and Dart-03 traces, respectively. Due to space constraints, we cannot show figures for delay, but they are not very different from Fig. 7 in subsection VI-B. Interested readers can refer to [13] for the figures. The trends for Dart-04 and MIT-rel traces are similar to those shown in subsection VI-B. For longer trace periods (above 9 days), the unreachable ratio does not change significantly for up to 20% of selfish nodes, and the robustness of performances increases if longer trace periods are used. This confirms that the robustness of information diffusion under *current* encounter pattern is not an artifact of coarse location granularity in the USC trace. In Dart-03 trace, the performance of information diffusion is less robust than other traces, since it has the smallest encounter ratio (c.f. Fig. 1) among all the traces. The unreachable ratio for Dart-03 trace increases faster as compared to other traces when we make users selfish.

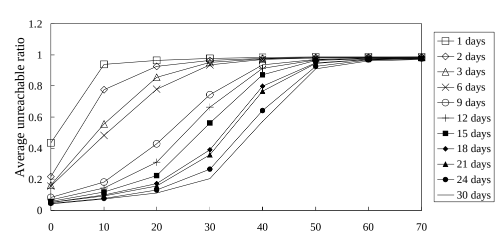


Fig. 14. Information delivery ratio with various selfish node percentage and trace period, using Dartmouth trace July 2003