Capturing User Friendship in WLAN Traces

Wei-jen Hsu and Ahmed Helmy

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

I. Introduction

Most of recent research works on analyzing wireless LAN (WLAN) traces focused on individual user behaviors [1], [2], [3]. These previous works provide good understandings on WLAN users, and have made vast amount of WLAN traces available to the research community (e.g. from [1], [2], [3], [6]). However, we know from daily lives that we do not make random movement decisions. Usually, WLAN users show preferences in their visits to a small set of the campus. As a result, mobile nodes (MNs[1]) in WLAN traces are in fact not uniformly distributed across campus, and users with similar preferences show up at the same access point (AP) more frequently. We look into this issue and try to identify the closeness (i.e. *friendship*) between node pairs, and understand its influences on network connectivity if we make connections between nodes based on their friendship.

Specifically, we give several intuitive definitions about friendship between MNs, utilizing traces about their association to APs in a WLAN. These friendship indexes capture the observed closeness between the involved MNs from the trace. Although such closeness may or may not reflect friendship in social context, it reveals the closeness between wireless devices as displayed in their association patterns. Empirical distribution of these friendship indexes mostly follow exponential distribution, with few node pairs showing high friendship index.

We further utilize the Small World model [4] to understand the characteristics of the *encounter-relationship graphs (ER graphs)* formed by WLAN users, in which two nodes are connected by a link if they ever associate with the same AP during overlapped time intervals. We find that WLAN users form connected Small World graphs via encounters. Furthermore, we investigate the issue of how friendship influence the characteristics of ER graphs. We find that if nodes with high friendship indexes are used in ER graph, the resultant graph displays higher clustering coefficient and average path length. In other words, it is more inclined toward a *regular graph*. On the other hand, if we use nodes with low friendship index in ER graph, it displays lower clustering coefficient and average path length. This finding points out, similar to social networks, close friends in WLANs often form cliques and random friends are keys to wide-reached connectivity in a network.

[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.

II. WLAN Traces

In this paper we utilize the WLAN traces available to the research community (e.g. [1], [2], [3]), which were collected from university campuses 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, by defining friendship indexes and analyzing its distributions from the traces.

In this study we mainly focus on wireless traces collected from university campuses. 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. The traces were collected with different methodologies, but we can derive the association history information for each user from all four traces, and further derive other metrics based on that. These four traces are chosen to represent different campus environments, user populations, location granularity, and trace-collection methods. 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. We cannot go into the details about trace processing due to space constraints. Please refer to [6] for more details.

We bring up new perspectives to study the WLAN traces by utilizing Small World theory to describe the *encounter relationship graph (ER graph)*. Small World graph model is proposed in [4] and widely utilized to describe various networks in many areas, such as social networks, Internet topology, and electrical power networks.

III. Friendship between Nodes

In our daily lives, we are bound to meet with colleagues and friends much more often than others. In this section we try to investigate using the wireless LAN traces whether such uneven distribution of closeness among MN pairs exists in WLAN traces. We define *encounters* between MNs as the time periods they associate with the same AP in the WLAN trace. The likelihood or duration of encounters between two MNs captures the *friendship* between them. This ”friendship” in WLAN trace may or may not reflect social friendship, which is impossible to validate from anonymized traces. We propose to identify friendship between MN pairs based on three different dimensions: Encounter duration, encounter count, and encounter AP count, with the following definitions:

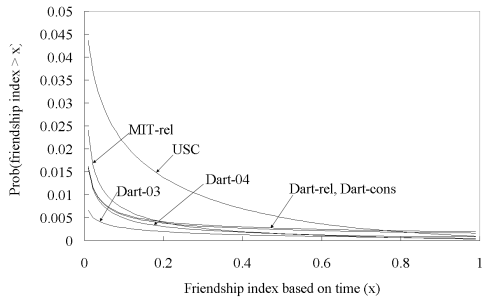


Fig. 1. CCDF of friendship index based on time

- **Friendship based on encounter time**: We define friendship index based on duration of encounter as $Frd_t(A,B) = E_t(A,B)/OT(A)$, which is the ratio of sum of encounter durations between node A and B, $E_t(A,B)$, to total online time of node A, $OT(A)$. This is an index for how good a friend node B is to node A based on duration of encounters. Note that in general $Frd_t(A,B) \neq Frd_t(B,A)$ and $0.0 \leq Frd_t(A,B) \leq 1.0$ for any node pair A and B.
- **Friendship based on encounter count**: The friendship index based on encounter count is defined as $Frd_c(A,B) = E_c(A,B)/S(A)$, which is the ratio between association sessions of node A that contains encounter events with node B, $E_c(A,B)$, to total association session count of node A, $S(A)$.
- **Friendship based on encounter AP count**: The friendship index based on encounter AP count is defined as $Frd_{AP}(A,B) = E_{AP}(A,B)/AP(A)$, which is the ratio between number of APs at which node A has encounters with B, $E_{AP}(A,B)$, to total APs node A visits, $AP(A)$.

We first observe how friendship indexes distribute among all node pairs in the traces. As shown in Fig. 1, the CCDF curves of friendship indexes based on encounter time follow exponential distributions for all campuses. We use Kolmogorov-Smirnov test [7] to examine the quality of fit. The resulting D-statistics for all traces are between 0.0356 and 0.0052, which indicates we have a reasonably good fit between the exponential distribution curves and the empirical distribution curves. Please see [6] for more detailed results.

Exponential distribution of friendship index is an indication that majority of nodes do not have tight relationship with one another. In all the traces, only less than 5% of ordered node pairs (A,B) have friendship index $Frd_t(A,B)$ larger than 0.01. This reveals the fact that for node pairs that do encounter with each other, most of them do not show tight relationship. Among all node pairs with non-zero friendship index, only 4.47% of them has friendship index larger than 0.7, and another 11.85% of them with friendship index between 0.4 to 0.7. Friendship indexes based on encounter frequency or encounter AP count also show similar exponential distributions.

We next look into the issue of whether friendship index for an ordered node pair $Frd_t(A,B)$ and its reversed tuple $Frd_t(B,A)$ are symmetric. We calculated the correlation coefficients for all the traces for three definitions of friendship indexes. The resulting correlation coefficients between ordered node pair (A,B) and (B,A) are low in most cases (ranging from 0.415 to -0.024, the only exception being 0.629 for friendship index based on encounter time for Dartmouth 2004 trace), implying high asymmetry in friendship indexes.

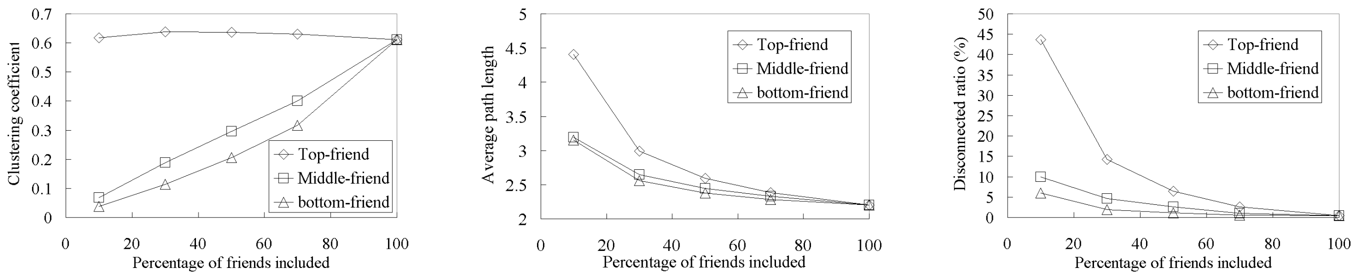
IV. Encounter-relationship graph with friends

Encounters between nodes can be viewed as opportunities for them to exchange messages. Based on this assumption, 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 small pieces of cliques? Furthermore, how does friendship influence the encounter patterns of nodes?

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. The concept of *ER graph* is introduced to capture potential for establishing relationships based on direct encounters. Typically, a MN may maintain relationship selectively only with those MNs that are considered "trust-worthy". For example, a MN may choose to trust those MNs with which it has high friendship indexes. The criteria of choosing the nodes to keep a relationship may influence the structure of *ER graphs*. This issue is the main focus of investigation in this section. Specifically, we try to include friends with various degree of closeness in the *ER graph*, and see how it influences the structure of the graph. We use friendship index based on time as an example to show how different friendship levels of included links can change the structure of *ER graph* significantly.

We sort the list of nodes that a node A has encountered according to friendship index, $Frd_t(A,B), \forall B \ni Frd_t(A,B) \neq 0$. After sorting, each node picks a certain percentage of nodes from the list with which to establish a link on *ER graph*. We choose nodes from top, middle, or bottom of the list and with various percentages, and obtain the corresponding metrics for the new *ER graphs* that include only the links to the chosen nodes. Note that the links in *ER graphs* are directed links when we consider friendship, as friendship is asymmetric between a given node pair. We observe the following metrics from the ER graphs formed with selected links based on the friendship indexes:

- **Clustering coefficient** (CC) is used to describe the tendency of nodes to from cliques in the graph. It is formally defined as:
  $CC = \frac{\sum_{node i=1}^{M} CC(i)}{M}$
  where $CC(i) = \frac{\sum_{A \in F(i)} \sum_{B \in F(i)} I(A \in F(B))}{Frd(i) \cdot (Frd(i)-1)}$



(a) Clustering coefficient (b)Average path length (c)Disconnected ratio

Fig. 2. Metrics of encounter-relationship graph by taking various percentage of friends

$I(\cdot)$ is the indicator function, $Frd(i)$ is the number of friends node i chooses to include in the graph, $F(i)$ is the set of chosen friends of node i, and M is the total number of nodes in the graph. Note that friendship is a asymmetric relationship, so $B \in F(A)$ does not imply $A \in F(B)$, and vice versa. Intuitively, clustering coefficient is the average ratio of neighbors of a node that are also neighbors of one another.

- **Disconnected ratio** (DR) is used to describe the connectivity of ER graph. It is defined as the average percentage of nodes that are within the same connected sub-graph of a given node.
- **Average path length** (PL) is used to describe the degree of separation of nodes in the *ER graph*. It is defined as the average of path lengths (in unit of hops) for all node pairs on the ER graph. If a node pair is disconnected, a higher penalty (the average path length of *regular graphs* with the same node number and average node degree) is accounted as the path length for that pair of nodes.
  When calculating average path length and disconnection ratio, the paths must follow the direction of edges on the ER graph.

Following the above definitions, we obtain the metrics when including given percentages of all encountered nodes from the top, middle, or bottom of the sorted encounter node list according to friendship index based on time. The figures are shown in Fig. 2. We use USC trace as an example, and similar results are also observed in other traces.

The figures show a clear trend that if neighbors ranked high in friendship index are included, the resultant ER graph shows stronger clustering, and the average path length is much higher. The result stems from the fact that top friends of a given node are also likely to be top friend between one another, forming small cliques in the graph. Clustering coefficient remains high due to these cliques. Disconnection ratio and average path lengths are high due to the lack of links between different cliques. On the other hand, when low-ranked friends are included in the graph, the links included are distributed in a more random fashion, reflected by the low clustering coefficient and low average path length. Similar results are also observed in social science study of friendship between pupils [5]. As larger portion of friends are included in the graph, all three metrics converge to the values when all encounters are included.

V. DISCUSSIONS AND CONCLUSION

In this paper we proposed friendship indexes to capture closeness between MNs from the WLAN traces. We find that friendship indexes are asymmetrically distributed among all MNs. Although it is not possible to establish the exact reason behind the closeness of MN pairs, this information may be utilized in several applications, such as better algorithms for cluster-forming in ad hoc networks, or finding a node to temporarily store a packet with higher probability to deliver it later to the final recipient. Protocols that are aware of social relationship among MNs may be an interesting direction in the future.

Using the concept of encounter-relationship graph (*ER graph*), we establish that it is possible to create a campus-wide community based solely on nodal encounters. Generally, in social-relationship aware mechanisms, one tends to trust top-ranked friends more than the others. However, as we see in section IV, using top-ranked friends only results in an *ER graph* with high clustering coefficient and average path length, and may lead to a disconnected network. In order to remain connected to a larger community, one should also use some randomly-chosen users (or middle friends) to reduce the degree of separation in underlying *ER graph*.

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] D. J. Watts and S. H. Strogatz. "Collective Dynamics of 'Small-World' Networks," Nature, vol. 393, pp. 440-442, 1998.
[5] A. Rapoport and W. Horvath, "A Study of a Large Sociogram," Behavioral Science 6, 279-291, 1961.
[6] 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
[7] R. Hogg and E. Tanis, "Probability and Statistical Inference, 6th Ed.," Prentice Hall, 2001.