MobiCom Poster Abstract: On the Structure of User Association Patterns in Wireless LANs

Wei-jen Hsu[a] *wjhsu@ufl.edu*
Debojyoti Dutta[b] *ddutta@usc.edu*
Ahmed Helmy[a] *helmy@cise.ufl.edu*
[a]Department of Computer and Information Science and Engineering, University of Florida, Gainesville, FL
[b]Department of Computational Biology, University of Southern California, Los Angeles, CA

I. Introduction

There has been a rapid increase in wireless LAN (WLAN) deployments, users and traffic in recent years. Thus there is an urgent need to mine the fine-grained analysis of the structures and trends in wireless user association patterns. Such study could have significant impact on the design of behavior-aware services and protocols to create social networks based on similar behavioral patterns. In this paper, we propose new methods to understand such trends of user associations in a campus WLAN, and ways of quantifying those trends. Our goal is to quantify repetitive and consistent patterns of user association behavior.

In this paper, we choose to study the traces of user's *daily* association patterns with WLANs at the USC and Dartmouth campuses. Specifically, we focus on the following questions in this paper: (1) What kind of modality do users display with respect to association patterns? (2) Is it possible to summarize the user association patterns for multiple days or weeks in a concise fashion? (3) How do we quantify users as having similar or dissimilar association patterns? Can we utilize such metrics to partition the whole user population into clusters of similar users? Due to space constraints, we only present brief discussions of each question here. Please refer to [4] for more details.

Our primary contribution is the methods we use to systematically analyze the association patterns. To this effect, we define novel features that can be extracted from traces, similarity metrics using singular value decomposition and we answer the questions we ask using unsupervised learning methods such as clustering [1]. To illustrate the usage of our tools and methods, we analyze the association patterns of $5,000$ users at USC's campus WLAN[2], and $6,582$ users at Dartmouth[3]. Note the proposed methods are not limited by the choice of data set and can be applied to study other WLAN traces.

II. Multi-modal user AP association

To show the multi-modal behavior of users we represent the user's *association vector* for each day as an n-entry vector, $(a_1, a_2, ..., a_n)$, where a_i represents the *fraction* of online time during the day the user spends at location i. Given the daily association vectors of a user for the trace duration, we utilize clustering techniques [1] to identify whether the user association pattern is single-modal or multi-modal. By *single-modal users*, we refer to those who display similar association vectors (i.e., the distance between association vectors are small), and *multi-modal users* refer to the converse. We use Manhattan distance as the distance measure between the association vectors.

We apply two ways to calculate inter-cluster distance (complete-link and average distance). Regardless of the clustering threshold chosen, the association patterns for many users fall into multiple clusters. Approximately 80% of users are classified as multi-modal (i.e., with more than two clusters) under intermediate clustering threshold, as shown in Fig. 1, and few users are single-modal in nature.

III. Summarizing association trends

We design a succinct way to express the major trend of user association pattern during the trace period for each user. The proposed method generates a summary using a small number of descriptive vectors with a quantitative measure of the importance for each vector. We concatenate user association vectors in columns of an *association matrix*, and apply singular value decomposition. We find that although user behavior is mostly multi-modal, the intrinsic dimensionality for these association matrices is actually low, indicating that there exists *dominating* association behavior for most users. Only a few eigenvectors and corresponding eigenvalues are needed to capture a high target percentage of power in the association matrices, as shown in Fig. 2: For more than 99% of users, we can use at most 7 eigenvectors and eigenvalues to

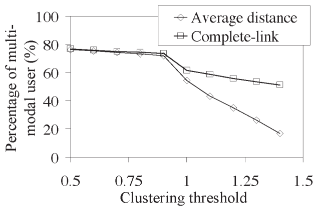

Figure 1: Percentage of multi-modal user under various clustering thresholds and distance calculation methods.

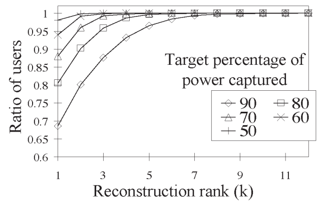

Figure 2: Low association matrices dimensionality: High percentage of power is captured with low reconstruction rank.

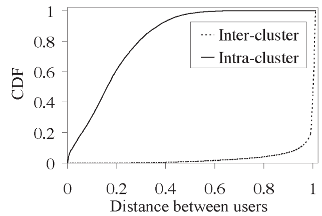

Figure 3: CDF of distances for inter-cluster and intra-cluster user pairs.

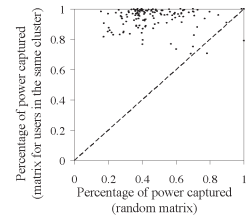

Figure 4: Cumulative power captured in top four eigenvectors of random (X) versus feature-based distance (Y) clustering.

capture more than 90% of power in their association matrices. That implies, with only a few eigenvectors and eigenvalues, we can summarize the association matrices with low reconstruction errors. At most, the top eight eigenvalues and eigenvectors can be used to reconstruct those association matrices with an average error of 5%, in terms of the L_1 and L_2 matrix norms.

IV. Quantifying user similarity

In order to identify similar users for grouping them, we propose two novel metrics to quantify the similarity between the user association patterns. In a *straight-forward* approach, we derive the metric based on detailed, comprehensive comparison of the association vectors of the users. In our *feature-based* approach, we use the eigenvectors and eigenvalues obtained from the association matrix as a feature set to determine the similarity. These two distance metrics are intimately related, with a correlation coefficient of 0.9119, but our *feature-based approach* is computationally simpler than the *straight-forward approach*. Finally, we utilize these metrics to obtain a partition of the user population by clustering. We demonstrate that the distribution of inter-cluster distances of the resulting partition is well-separated from that of the intra-cluster distances, as shown in Fig. 3 for the *feature-based* distance. The *straight-forward* distance shows similar results.

We further verify that indeed we have grouped similar users together by composing and analyzing the *joint association matrix*. We concatenate the association vectors of a cluster of similar users in a larger matrix. The percentage of power captured by the top eigenvectors of this *joint association matrix* should be high, as the association vectors in the matrix follow similar trends. By contrast, if association vectors of users with different trends are put in one *joint association matrix*, the percentage of power captured by its top eigenvectors should be much lower. For the clusters identified with the similarity metrics, we show that the cumulative power captured in the top four eigenvectors are much higher than if the clusters are formed randomly in Fig. 4, indicating that the clusters contain users with similar association trends.

V. Discussion and Potential Usage

The singular value decomposition based representation of user association patterns can help the network administrator to efficiently identify the major trends in user association patterns, and provide personalized location aware-services (e.g., storing user's email and files on machines close to the frequently visited locations; behavior aware advertisements). Further analysis of the cluster sizes with distinct association trends could reveal the dominant behavioral trends in the network and provide guidelines for network and service planning. Association behavior analysis also relates to establishing norms of the network operation and abnormal behavior detection. In addition, a user can profile oneself based on the association pattern and utilize such information in a larger social context for expanding social networks based on behavioral pattern. Association profiles can be used in designing context-aware routing protocols in encounter-based networks.

References

[1] A. Jain, M. Murty, and P. Flynn, "Data Clustering: A Review," ACM Computing Surveys, vol. 31, no. 3, September, 1999.

[2] MobiLib: http://nile.usc.edu/MobiLib.

[3] D. Kotz, T. Henderson and I. Abyzov, CRAWDAD data set dartmouth/campus/movement/01_04 (v. 2005-03-08), March 2005.

[4] Longer version of technical report available at http://arxiv.org/abs/cs.NI/0606002