Poster Abstract: Gender-based Feature Analysis in Campus-wide WLANs

Udayan Kumar *ukumar@cise.ufl.edu*
Nikhil Yadav *nyadav@cise.ufl.edu*
Ahmed Helmy *helmy@cise.ufl.edu*
Department of Computer and Information Science and Engineering, University of Florida, Gainesville, FL

Rapid WLAN deployment has led to various research challenges and scenarios. In this paper we come up with a technique to classify users into social groups and then use this information to investigate the usage behavior of these groups. Grouping of users can be done based on diverse parameters like gender, major or other interest groups. In this paper we show the general methodology used to accomplish WLAN user groupings with an example of grouping by gender in a major university campus. The usage patterns of Males and Females are contrasted based on this grouping. Possible applications of this work encompass user profiling; studying gender gaps in social sciences; announcement/advertisement customization to target groups.

I. Introduction

WLAN deployment across university campuses has risen rapidly in recent years. There has been a marked increase in mobile users and traffic as a result. Analyzing usage of WLAN is currently a major research issue. Previous works have analyzed association patterns in traces [4][3], but have not studied specific group or gender behavior. In this paper we present a novel technique to analyze WLAN usage based on gender. This study examines the difference between WLAN usage patterns of males vs. females. Further, we can investigate gender gaps in WLAN usage. Gender bias in technology adoption in the Internet [1] has been studied before.

One of the major challenges in conducting this study is how to categorize users into groups based on gender, major, interest, etc. as WLAN traces seldom provide this information [5]. They generally provide information about associations of MAC addresses with access points (AP). To categorize users based on gender, we propose the following novel technique. Most universities have Sororities and Fraternities as social organizations. Sororities are female organizations while Fraternities have males. Given the physical location of APs on campus, APs located in sororities and fraternities are identified, and the users associated with them as are classified as female or male.

We use a month long WLAN trace collected from a major University campus[1]; 12 fraternities and 7 sororities on campus were considered for this study. As fraternities and sororities have visitors, our classification needed further refinement. Since visitors are infrequent users, their associated number sessions is, in general, less than that of regular users. These visitors are excluded for improved accuracy, using the following steps: ***i.*** Extract the number of sessions per MAC for each fraternity or sorority AP. ***ii.*** Vary the min. session duration (as a threshold for regular users) and observe its effect on the number of sessions and distinct users. ***iii.*** Obtain a suitable threshold for the session duration and session count to classify users above these limits as being either males or females. On performing the above procedure for the studied trace, we show the session count curve for sorority and fraternity users in Figure 1 in descending order with respect to the number of sessions. An interesting feature of the Figure 1 is the presence of sharp bend (knee) as the number of sessions per MAC decreases. Intuitively, this means that MACs below the knee have an order of magnitude less number of sessions (accounting for

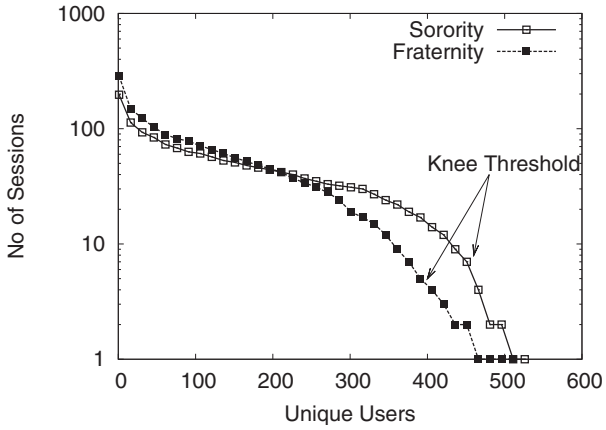

Figure 1: Session count for fraternities and sororities

[1]MobiLib USC WLAN traces[2]

the difference between a regular user and a visitor). All users below the knee were classified as visitors and removed from the study. Changing the session duration had no effect on the shapes of the curves [6]. The session duration threshold here was the average session duration of sororities and fraternities for all the users.

The final sets of MAC addresses obtained after the exclusion have a higher probability of belonging either to males or females. On doing this extraction, out of 777 MACs associated to fraternities and 687 sorority associated MACs, we classify as regular users 463 MACs associated with sororities and 452 with fraternities. These users are used for further analysis.

II. Analysis of Male and Female WLAN Usage

Once we have the MAC addresses of users classified as males and females, we can look for those MAC addresses in the whole campus trace and analyze their behavior. For this we use a data mining approach to extract features of interest in our research using SQL queries. In this paper we concentrate on the following questions:

1. WLAN Usage and Gender Distribution: What are the trends in WLAN usage across different (buildings) areas on campus?

2. Average online time: are there trends in the average online times of users and can differences be spotted based on gender and (building) areas within the campus?

3. Manufacturer preferences: Which device vendors do different genders prefer?

We also analyze two more one month long traces from consecutive semesters. We see that trends are consistent across the semesters. More details are available here [7], due to space constraints we are not including them in this paper.

II.A. WLAN usage by area

Figure 2 shows the usage distribution per building type. Communications buildings show a higher population of male users, social science buildings have a higher count of female users. It is interesting to see that there are more female WLAN users than males in Engineering buildings.

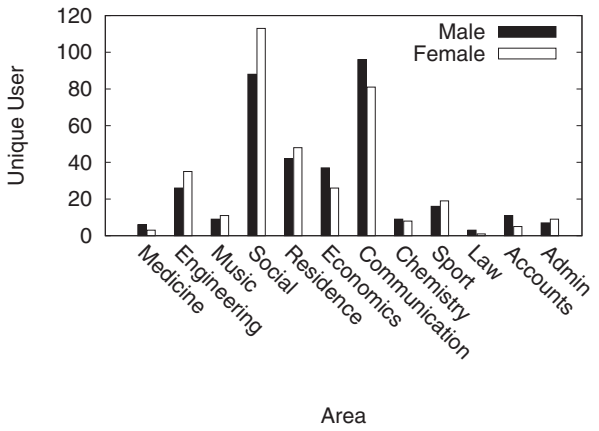

Figure 2: Average duration in different Areas

II.B. Average session duration

We study the different average session duration for male and female users across the campus. Figure 3 shows this. From Figure 3 we can deduce that males spend more time online than females in most of the areas. Females show dominant usage in the Social Science and Law areas across campus. We can deduce from this that on average, male users tend to stay - as WLAN users - at certain places for longer times than females.

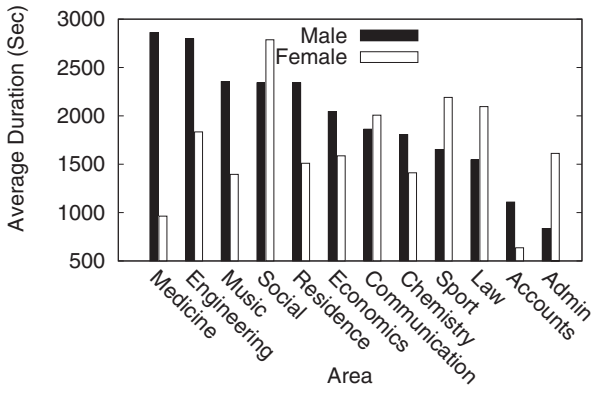

Figure 3: Average duration in different Areas

II.C. Manufacturer Preferences

In Figure 4, it is interesting to note that Apple computers are more popular amongst females than males. Intel devices are more popular amongst males. A statistical significance test, *Chi-Square*, shows with 90% confidence that there is a bias among genders for vendors.

This consistency in preferences is seen in the other two samples of one month as well.

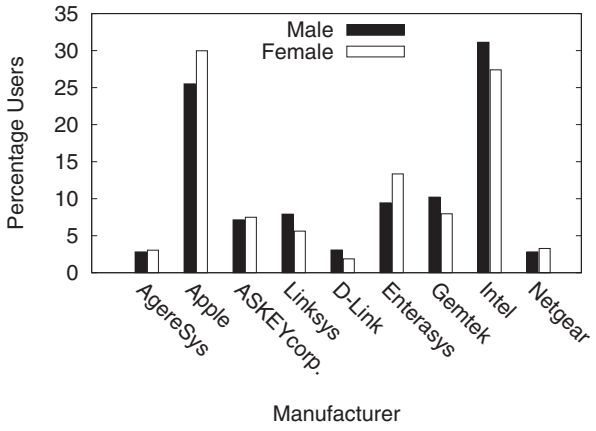

Figure 4: Device Distribution by Manufacturer

III. Applications

Analysis of gender based usage preferences can be used to profile users. The extent of WLAN adoption amongst genders is obtained and like [1] can be of interest to social scientists studying socio-economic and socio-cultural differences between genders. Trends in mobile social networking for these groups can be used to provide services. Announcements and advertisements on campus can be directed based on the general psyche of males and females. The areas these users frequent more could serve as good places to advertise related interests, services or products.

IV. Conclusions and Future Work

A study of similar characteristics across different campuses is planned. This could encompass a detailed time based analysis of gender mobility based on different time periods of the day (morning, afternoon, evening hours). Our methodology of gender classification and use of SQL queries on the WLAN traces are generic. The results from this research are based on a sample of the user population, since gender may be identified based on sorority and fraternity AP associations. The concept of an area is used, which includes majors as well as different buildings on campus to improve the richness of data set. To conclude, there is a distinct difference in WLAN usage patterns for different genders even with similar population sizes. Females seem to dominate in WLAN usage in areas of Social Science and Law and prefer Apple over Intel. Males have longer session durations than females in most cases. We hope for this study to open the door for other mobile social networking studies and profile-based service designs.

References

[1] Ruby Roy Dholakia et al. *The Internet Encyclopedia*, chapter Gender and Internet Usage. Wiley, 2003.

[2] Wei-Jen Hsu and Ahmed Helmy. MobiLib: Community-wide Library of Mobility and Wireless Networks Measurements (InvestigatingUser Behavior in Wireless Environments). Downloaded from http://nile.cise.ufl.edu/MobiLib.

[3] Wei-Jen Hsu and Ahmed Helmy. On Modeling User Associations in Wireless LAN Traces on University Campuses. In *IEEE WiNMee, Boston MA*, April 2006.

[4] Wei-Jen Hsu and Ahmed Helmy. On Nodal Encounter Patterns in Wireless LAN Traces. In *IEEE WiNMee, Boston MA*, April 2006.

[5] David Kotz, Tristan Henderson, and Ilya Abyzov. CRAWDAD data set dartmouth/campus. Downloaded from http://crawdad.cs.dartmouth.edu/dartmouth/campus.

[6] Udayan Kumar and Nikhil Yadav. http://www.nile.cise.ufl.edu/socnet.

[7] Udayan Kumar, Nikhil Yadav, and Ahmed Helmy. Gender Based Grouping of Mobile Student Societies. Technical Report, Download from http://www.cise.ufl.edu/∼ukumar/techrepgender.pdf.