Analyzing Gender-gaps in Mobile Student Societies

Udayan Kumar, Nikhil Yadav, Ahmed Helmy
Computer and Information Science and Engineering, University Of Florida
Email: {nyadav, ukumar, helmy}@cise.ufl.edu

1. 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 [1], 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 investigate gender gaps in WLAN usage. Gender bias in technology adoption in the Internet [2] has been studied before but not in wireless networks.

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. They generally provide information about associations of MAC addresses with access points (APs) [4]. 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 represent male organizations. Given the physical location of APs on campus, APs located in sororities and fraternities are identified, and the users associated with them are classified as female or male.

Choice of trace: In order to carry out a meaningful analysis of the various groups of WLAN users, specifically targeting and differentiating between genders, the traces need to provide the following information: ***i.*** comprehensive syslog (or SNMP) logs for various buildings, dorms and sororities/fraternities for the general population of students (without population bias) and for extended periods of time (30 days or more), ***ii.*** mapping between the APs (or point of collection) to specific building's designation. ***iii.*** differentiation between individual users and ability to track users (without necessarily knowing their identity). We have investigated WLAN traces collected at USC, UNC and Dartmouth[4]. Dartmouth traces do not provide AP-to-building mapping, which makes it difficult to do this kind of study. UNC traces on the other hand have limited number of APs in sororities and fraternities. We chose the USC traces for our study. It provides ~10 fraternities and 10 sororities are included in WLAN traces and the AP-to-building mapping is also available.

As fraternities and sororities have visitors, our classification needed further refinement. Since visitors are infrequent users, their associated number of 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.

After performing the above procedure for the studied trace, we plot the session count curves for sorority and fraternity users in Fig. 1 in descending order with respect to the number of sessions. An interesting feature of the figure is the presence of a sharp bend (*knee*) as the number of sessions per MAC decreases. MACs below the knee have an order of magnitude less number of sessions (intuitively accounting for 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 [3]. The session duration threshold here was the average session duration of sororities and fraternities for all the users. (no significant change in the analysis was found when no threshold was used).

The final sets of MAC addresses obtained after the exclusion can be classified as either to males (or females) with more confidence. On doing this extraction, out of 777 MACs associated with fraternities and 687 sorority associated MACs, we classify 463 MACs associated with sororities and 452 fraternity associated MACs as regular users. These users are used for further analysis

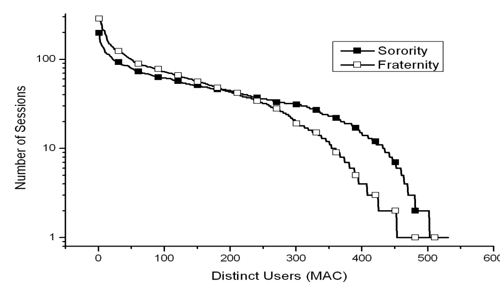

Fig 1: Session count for Sorority and Fraternity users

2. Analysis of Male and Female WLAN Usage

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:

a) WLAN Usage and Gender Distribution: What are the trends in WLAN usage across different (buildings) areas on campus?
b) 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?
c) Manufacturer preferences: Which device vendors do different genders prefer?

2.1 WLAN usage by area

Fig 2 shows the usage distribution per building type. *Communications/Journalism* (Comm) 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, which is counter-intuitive.

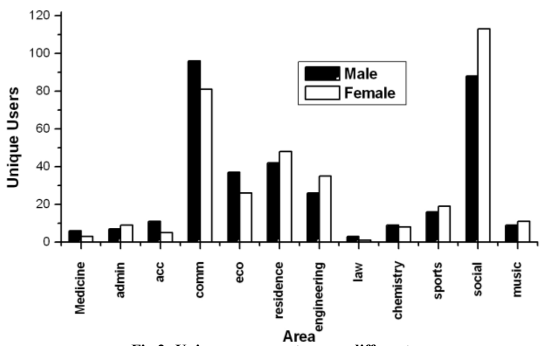

Fig 2: Unique user count across different areas

2.2 Average session duration

We study the different average session duration for male and female users across the campus.

From Fig. 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.

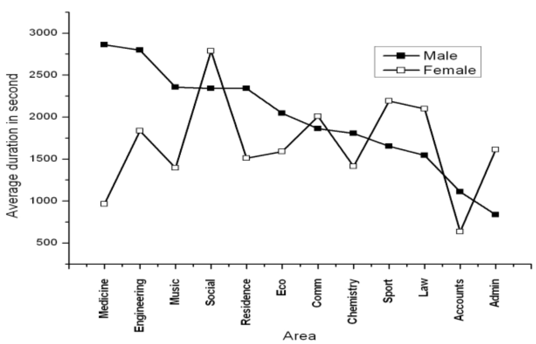

Fig 3: Average Session Duration by Area

2.3 Manufacturer Preferences

The preference of manufacturer (based on the type of wireless card traced) is shown In Fig. 4. It is interesting to note that *Apple* computers are more popular amongst *females* than males.

$$\frac{\#\text{Female apple user} - \#\text{Female intel user}}{\#\text{Female apple user} + \#\text{Female intel user}} = 0.045$$

$$\frac{\#\text{Male intel user} - \#\text{Male apple user}}{\#\text{Male intel user} + \#\text{Male apple user}} = 0.099$$

Intel devices are more popular amongst males. For this study, only major vendors were considered.

A similar analysis was done for 3 days, 1 week and 2 weeks, showing consistent trends (with the above 30-day trace analysis). A more extensive *time evolution* study is planned.

3 Applications

Analysis of gender based usage preferences can be used to *profile* users. The extent of WLAN adoption amongst genders is obtained and like [2] 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. This is part of a paradigm our group is designing called *Profile-Cast.*

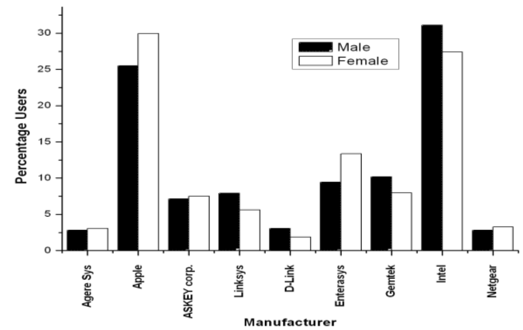

Fig 4: Device distribution by manufacturer

4 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. In this research we have been able to categorize users based on anonymous traces. Potential applications include advertising, etc. but it also raises questions about privacy and security. How much can one reverse-engineer from these traces? This is still an open question under investigation. 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] W. Hsu, A. Helmy, "On Nodal Encounter Patterns in Wireless LAN Traces", *IEEE WiNMee, April 2006*

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

[3] http://www.nile.cise.ufl.edu/socnet

[4] http://crawdad.cs.dartmouth.edu/data.php