Association Analaysis in Wireless Network using Trace data.

L.Velmurugan
Sr.Lecturer, MCA
EBET group of Institutions
Kangayam – 638 108, INDIA

Dr.P.Thangaraj
Head, Dept. of CSE,
Bannariamman Institute of technology
Sathyamangalam -638 401, INDIA

***Abstract*— With the explosive growth in wireless devices like smart phones, PDA's, netbooks, WLAN networks are usually overloaded at workplaces, social and home environments. Though a large number of mobility models have been proposed, a better way is to study wireless mobility traces which gives more accurate movement information for a population. An accurate location predictor can significantly improve the performance or reliability of wireless network protocols and the wireless network infrastructure itself. These improvements lead to a better user experience to a more cost-effective infrastructure or both. In this work, it is proposed to study mobility traces and generate rules for mobility of users which was previously unknown in campus wide wireless network. We also study analyze and show that at certain locations access points are very rarely used.**

***Keywords*-WLAN, Mobility trace, Association rule, Wireless network management, Access point.**

I. Introduction

Extensive research based on the data collection from the wireless local area network has helped us to study the behavior of the network like traffic transmitted, client session duration. Even though the network behavior can be understood from these studies, it does not explain why the network behaves in this particular way [1]. The network adapter in most of the modern laptops can access one or more types of IEEE 802.11 network, but the PDAs, printers, and audio players are some of the new wireless devices which lead to changes in the way that WLANs are used. For example, a wireless PDA can have different usage patterns than a wireless printer; a PDA could be more mobile as its user traverses a WLAN-enabled campus, whereas the printer may be immobile serving wireless clients. The immense growth of WLANs facilitates the development of new applications, which may also exhibit new usage characteristics. For instance, Quality-of-service (QoS) requirements in Real-time multimedia applications may become difficult to be fulfilled in a shared-medium WLAN. Simultaneously, some of these new applications and devices may emerge as seen in many wireless PDAs which are sold equipped with streaming audio or video software.

Software developers who create new wireless and location-aware applications, designers who develop new high-throughput and multimedia-friendly wireless networking standards, and providers who deploy and manage WLANs should thoroughly understand the usage, and the trends in the usage of these new devices and applications.

University campus, airports, hospitals and office environment have successfully deployed Wireless networks[2]. Understanding the behavior of such environments is very crucial. The important method to investigate in modeling association patterns of individual users to the access points is established due to the large amount of wireless local area network (WLAN) traces available to the research community (APs)[3][4].

In this paper we focus on the user mobility inside a college campus by deriving association rules among the access points. For our work we select a one month wireless mobility trace collected by Dartmouth college and available publicly in the crawdad web site.

This paper is divided into the following sections; section II describes the mobility trace data used in our work, section III deals with association rules, section IV describes the experimental setup and results obtained and section V draws conclusion of our work.

II. Dataset used in our work

We use the Dartmouth college mobility traces which were collected over a period of three years inside the college campus. The Dartmouth College campus spread over 200 acres has about 190 buildings. Dartmouth College has over 6500 students and faculty together, and during the data capture there were over 3200 undergraduates on campus. 476 Cisco 802.11b access points (APs) were installed in 2001 covering most of the campus. Over time, APs have been added to increase coverage and AP's have also been installed in new buildings coming up. Wireless clients roam seamlessly between APs as all APs share the same SSID. On the other hand, a building's APs are connected to the building's existing subnet. The 188 buildings with wireless coverage span 115 subnets, so clients roaming between buildings may be forced to obtain new IP addresses.
Time stamp was added by thee syslog-recording server for all the message that arrives. Each message contains the access point name, the card's MAC address and the type of message[5]: Key messages are given below.

Authenticated. This message is ignored as all cards that use the network must be authenticated.

Associated. After authentication, a card chooses one of the in range access points and associates with that AP; all traffic to and from the card goes through that AP.

Reassociated. The card monitors periodic beacons from the APs and may choose to reassociate with another AP. This feature supports roaming. Some cards do not reassociate but always associates.

Roamed. When a card reassociates with a new AP, the new AP broadcasts that fact on the Ethernet; upon receipt, the old AP emits a syslog “Roamed” message. This message is also ignored as it depends on an inter-AP protocol below the IP layer, it only occurs when a card roams to another AP within the same subnet.

Disassociated. When the card no longer needs the network, it disassociates with its current AP.

The data obtained was sanitized so that there is a consistent MAC address mapping and a consistent IP address mapping across all of the traces. For example the sanitized IP address 111.777.333.999 will correspond to the same raw IP address in all of the traces. Similarly, the sanitized MAC address aa:bb:xx:dd:yy:ff will always correspond to the same raw MAC address in all of the traces.

The sample traces of a single user is as shown below

1043953158	AcadBldg29AP1
1043953493	OFF
1044315333	AcadBldg19AP3
1044326960	OFF
1044367564	AcadBldg19AP3
1044370690	OFF
1044379076	AcadBldg19AP3

The first column represent the unix time stamp. The unix time stamp is a way to track time as a running total of seconds. This count starts at the Unix Epoch on January 1st, 1970. The second column represents the access point the user was present at a particular time. Each access point name has been 'blinded'. The actual AP names are meaningless to people outside Dartmouth, so it has been replaced by names with a descriptive name of the form Building#AP#;

For example ResBldg100AP2 is AP#2 in the Residential Building #100.

The types of building are:
AdmBldg = administrative buildings
AthlBldg = athletic buildings (stadiums, ski lodge, etc)
LibBldg = libraries
OthBldg = other (test APs)
ResBldg = residences (dorms, fraternities, graduate housing, etc)
SocBldg = social buildings (communal centers, eating areas etc)

There is a special AP name in the mobility data (OFF) which means the card is off the wireless network. This was determined by the syslog message "Disauthentication" from the last associated AP with reason of "Inactivity."

For our analysis we considered one month trace from Ist March 2002 to 31^{st} March 2002. Statistics of the data obtained is given in table 1.

Total Number of Users	374
Total Number of APs accessed	413
Total number of APs accessed less than 10 times	116
Total number of syslog recordings	31191
Syslog recordings with “OFF”	8911

Table 1. Descriptive statistics of selected data set

The frequency distribution of top 25 Access points used is shown in figure 1.

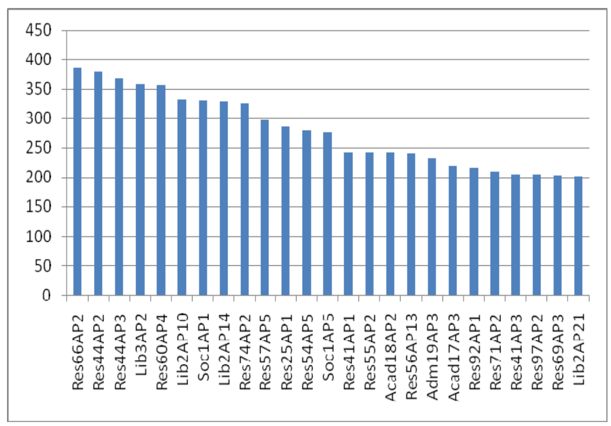

Figure 1.Access points used the most during the month of March 2002.

III. MINING AND RULE DISCOVERY

In this paper we propose an association rule based approach to find user movement within the campus and hence enhance user experience using data mining techniques extensively. Association rule mining over the *basket data model* was proposed by Agrawal *et al* [6]. It allows analysts to infer useful information on movement of an user, and data usage pattern. The web trace essentially consists of a large number of individual records called *transactions* and each transaction is a list of access points accessed by a particular user in a given period of time. The goal of association rule mining is to discover rules of the type, “whenever a transaction includes a particular access point W , it is likely to contain a specific access point I $\notin$ W".

In the case of access points in the campus, such rules can be used to plan the location of new access points and enhance existing points where the load is high. Informally, the input to association rule mining consists of the collection of transactions and two parameters $\theta \leq 1$, the required *support* and $\gamma \leq 1$, the desired *confidence*. It consists of two steps, namely *frequent itemset mining* in which itemsets with frequency of at least θ are identified, and *association rule mining* in which the association rules of the type $W \Rightarrow I$, with I $\notin$ W, are identified. The itemset *W U I* should have a support of at least θ , and of all the transactions containing *W*, the fraction of the transactions that contain *I* should be at least γ . Agrawal and Srikant [7] present the *Apriori* algorithm for frequent itemset mining and *FastGenRules* heuristic to generate the association rules.

IV.EXPERIMENTAL RESULT AND ANALYSIS

The data was prepared using the following rules.

RULE I : Access points which were sparsely used by users is eliminated by the following rule:

For $I \in W$ and < 4% of total user access.

RULE II : Devices which are rarely mobile is eliminated. This is identified using the rule:

For $U \in K$ and < 1.5% of total handovers.

Using the above principle we obtain 109 access points for which rules has to be discovered. Similarly eliminating very sparse users we obtain 134 users based on whom the rule is to be discovered.

Since the network usage is growing and prediction is to be done for future planning we used a minimum support of 25% and a confidence of 75%. Rules discovered for the above support and confidence are

Soc1AP5==> Soc1AP1	conf:(0.81)
Soc1AP1 ==> Soc1AP5	conf:(0.8)
Soc1AP2 ==> Soc1AP1	conf:(0.76)

Rules that were discovered for APs not used by the user was also measured. Here we used a higher minimum support of 84% and confidence of 85%. Some of the rules discovered are

Acad6AP2==> Acad6AP3	conf:(0.99)
Res83AP1==> Res83AP7	conf:(0.98)
Acad18AP7==> Acad18AP12	conf:(0.97)
Acad6AP4==> Acad6AP3	conf:(0.97)
Acad6AP3==> Acad6AP2	conf:(0.97)
Acad18AP12==> Acad18AP7	conf:(0.93)
Acad6AP6==> Acad18AP12	conf:(0.93)
Acad30AP3==> Acad21AP2	conf:(0.93)

V. CONCLUSION

In this paper we study the wireless network traces from Dartmouth college with focus on mining rules of users movement. We propose rules to describe data selection and capture various aspect of modeling user association pattern in WLANs. The findings that can be listed include large percentage of users are offline time most of the time. Secondly users tend to follow a specific pattern and imited number of APs are visited in the network and large proportion of online time spent at very few of its most visited APs.

REFERENCES

[1] Tristan Henderson, Denise Anthony, David Kotz "Measuring wireless network usage with the Experience Sampling Method" In Proceedings of the First Workshop on Wireless Network Measurements (WiNMee 2005) , April 2005

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

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

[4] W. Hsu and A. Helmy, "IMPACT: Investigation of Mobileuser Patterns Across University Campuses using WLAN Trace Analysis," unpublished USC technical report. Available at

[5] http://crawdad.cs.dartmouth.edu/meta.php?name=dartmouth/campus2004, September 2004.

[6] R. Agrawal and R. Srikant. Fast algorithms for mining association rules in large databases. In *VLDB*, pages 487-499, 1994.

[7] R. Agrawal and R. Srikant. Fast algorithms for mining association rules in large databases. In *VLDB*, pages 487-499, 1994