Identifying wireless user groups and association patterns in a campus

Francisco Chinchilla, Maria Papadopouli, Billy Saelim
Dept. of Computer Science, University of North Carolina at Chapel Hill

Diane Pozefsky
IBM Corporation

Harshavardhan Sabbineni
Dept. of Electrical & Computer Engineering, Duke University

***Abstract*— Our goal is to understand the movement patterns of wireless users and producing models that describe client associations. This can be beneficial in many domains, including coverage planning, resource reservation, supporting location-dependent applications and applications with real-time constraints, and produce models for simulations. We conducted a 77-day measurement study of wireless users and their association patterns on a major university campus using the IEEE 802.11 wireless infrastructure.**

We cluster wireless clients based on their inter-building mobility and the characteristics of their movement while they are *continuously* connected to the infrastructure. We further distinguish wireless clients based on their inter-building mobility, their visits to access points (APs), and the features of their continuous walks in the wireless infrastructure. Such measurements can benefit protocols and algorithms that aim to improve the performance of the wireless infrastructures by load balancing, admission control, and resource reservation across APs.

I. Introduction

Recently IEEE 802.11 networks became widely available in universities and corporations to provide wireless Internet access. There are very few studies investigating the user access and mobility patterns. Research on this issue can impact several areas, such as simulation studies on wireless networks, deployment and administration of wireless infrastructures, protocol design for intelligent and robust wireless infrastructures, and user access and traffic characterization. Currently, most of the simulation studies on wireless networks and protocols consider simplistic communication and association patterns for the wireless users. There is a need for more realistic models of the user communication and association patterns to drive simulations of wireless networks. User association patterns can be advantageous in resource reservation, admission control, and capacity planning at the APs. This can improve both the performance of user access as well as wireless bandwidth utilization.

The UNC wireless infrastructure provides coverage for nearly every building in the 729-acre campus and includes a diverse academic environment of university departments, programs, administrative, activities, and residential buildings. These buildings support 26,000 students, 3,000 faculty members, and 9,000 staff/administrative personnel. Of the 26,000 students, 61% are undergraduates, and more than 75% of these own a wireless laptop. The majority (232) of the access points (APs) on campus were configured to send syslog events to a server in our department between 12:00:00 am on February 10, 2003 and 11:59:59 pm April 27, 2003. During this trace period we recorded 8,158,341 syslog events for 7,694 clients, and 222 APs distributed among 79 buildings. You can find more details about the measurement testbed, definitions and rules in Section 3 of our earlier paper [3]. In this paper we will use the term *session* to identify *continuous* wireless associations of a client. Each client may have several sessions. More specifically, we use the following terms:

Visit: A client begins a visit to an AP when a (re)association message is received from that AP for that client and ends when *any* syslog message from *any* different AP is received for that same client or a disconnection message. The difference in the timestamp of these two messages defines the duration of the visit to the AP.

Session: A session is a continuous sequence of visits to APs. A session begins when a currently disconnected client receives a (re)association message and ends when the next disconnection message is received. The difference in the timestamps between the disconnection message and the first (re)association message defines the duration of the session.

AP Path: An AP path is a sequence of *continuous* inter-AP transitions. For example, if an wireless client that was originally disconnected connects to APs 1, 2, 1, 1, 10, 3, 4, 5, 1, and 6 before disconnecting, its path is $1 \rightarrow 2 \rightarrow 1 \rightarrow 10 \rightarrow 3 \rightarrow 4 \rightarrow 5 \rightarrow 1 \rightarrow 6$. The length of this AP path is eight. The *building path* is similarly defined. For example, if APs 1 and 2 belong to building A, AP 10 to building B, APs 3 and 4 to building C, AP 5 to building D, and AP 6 to building E, the building path of the above example is $A \to B \to C \to D \to A \to E$ and its length is five. We can also compute the *longest* (with respect to the number of inter-building transitions) *cycle-free subpath (LCFS)* of a session. A cycle in a building path is a return to a building already visited in the session. In this example, the LCFS of the session is four and corresponds to the subpath $B \to C \to D \to A \to E$. We also have position information for all building in campus. Given this positioning information, we define the *building path distance* to be the sum of the euclidean distances of each two consecutive buildings in the path. The *diameter* of a path is the maximum pair-wise euclidean distance of two buildings in the path. We can compute the duration of each inter-AP transition and therefore the duration of the entire session of a client. Similarly, we can define the *pace* of a path as the ratio of the building path distance to the path duration.

The key issue that drives this study is the identification and clustering of wireless clients based on the characteristics of their sessions. More specifically, we consider several metrics such as the session duration, building path distance, building path length, longest cycle-free subpath, diameter, and pace. Each client may have several paths that exhibit different characteristics with respect to these metrics. Our goals are to determine the important features that identify the sessions and cluster the clients based on their characteristics. We apply statistical techniques, including multidimensional clustering and visualization to detect patterns in usage and mobility.

This research extends our earlier study [3], the studies by Kotz and Essien [4], Balachandran *et al.* [1], Tang and Baker [5], and Balazinska and Castro [2] by focusing more closely on the association and mobility patterns of individual clients rather than on the entire population of mobile clients and in a finer time granularity. We monitor the behavior of each wireless user with respect to its association patterns and carry out user-behavior analysis more accurately. Similarly, Balazinska and Castro study the wireless association patterns of users in a corporation considering aggregate information about their associations at every AP. They do provide a clustering of users based on their duration at one AP. However, they only poll the APs at every 5 min to get the current association information and do not capture path related information for these associations. To the best of our knowledge, this is the only study using wireless traces that tracks the associations of individual wireless users and focuses on such clustering issues.

In this paper, we have identified a broad range of clients with different characteristics. We have classified the wireless clients into main categories based on their inter-building mobility, duration at each building, and frequency of their sessions in the trace. Depending on the inter-building mobility, we can distinguish the *short-range* clients (i.e., the clients with zero-length building path but may visit multiple buildings on different sessions), *stationary* clients (a subset of the short-range clients that consists of clients that visit APs in only one building), and the *mobile* clients (all remaining clients). The LCFS of a session is an indication of the mobility of the client. We found 2,635 mobile clients and the 2.58% of them have an average LCFS path length equal or greater than two. The mean of the maximum LCFS path length of the mobile clients is 2.08. We found 18 mobile clients with a maximum LCFS path length equal or greater than 10. Table I presents the mean and median values on the average and maximum diameter and building path length of mobile users (considering all their sessions).

Based on the duration of their visits at different buildings of their session, we identify *transient* and *non-transient* sessions. For example, a possible definition for the transient sessions are the ones that do not have any visits to a building that last more than 30 min. Of our 228,563 total sessions, 131,633 sessions are transient. Clients that had only a few sessions in the trace are classified as *visitors* whereas the remaining as *regulars*. For example, 25% of the clients have less than four sessions. We found that the average user changes buildings every 17 min and 21 sec while in a session. If we only consider the mobile clients, their mean and median inter-building transition correspond to 7 min and 25 sec (while in a session) and 10 min and 39 sec (while in a session), respectively. We also investigate if sessions that started during different time periods (such as weekday, weeknight or weekend) or in different type of buildings exhibit unique characteristics. We would like to refine this classification and better understand the behavior of different wireless users. Although the percentage of mobile users is not very high, we believe that in the next few years it will grow fast. We also plan to conduct another study in the next few months to see how the percentage of mobile users has changed.

REFERENCES

[1] Anand Balachandran, Geoffrey Voelker, Paramvir Bahl, and Venkat Rangan. Characterizing user behavior and network performance in a public wireless lan. In *Proceedings of the ACM Sigmetrics Conference on Measurement and Modeling of Computer Systems*, 2002.

Metric	Mean (Mean, Median)	Maximum (Mean, Median)
Diameter	55, 18	429, 225
Building path distance	279, 45	2568, 485

TABLE I

PATH STATISTICS (IN METERS) ON ALL MOBILE CLIENTS.

[2] Magdalena Balazinska and Paul Castro. Characterizing mobility and network usage in a corporate wireless local-area network. In *First International Conference on Mobile Systems, Applications, and Services (iMobiSys)*, May 2003.

[3] Francisco Chinchilla, Mark Lindsey, and Maria Papadopouli. Analysis of wireless information locality and association patterns in a campus. In *Proceedings of the Conference on Computer Communications (IEEE Infocom)*, Hong Kong, March 2004.

[4] David Kotz and Kobby Essien. Analysis of a campus-wide wireless network. Technical Report TR2002-432, Dept. of Computer Science, Dartmouth College, September 2002.

[5] Diane Tang and Mary Baker. Analysis of a local-area wireless network. In *ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom)*, pages 1–10, Boston, Massachusetts, USA, August 2000.