Persistence and Prevalence in the Mobility of Dartmouth Wireless Network Users

Clara Lee
Senior Honors Thesis
Dartmouth College Computer Science
Technical Report TR2003-455
Advisor: David Kotz
May 2003

Abstract

Wireless local-area networks (WLANs) are increasing in popularity. As more people use WLANs it is important to understand how these users behave. We analyzed data collected over three months of 2002 to measure the persistence and prevalence of users of the Dartmouth wireless network.

We found that most of the users of Dartmouth's network have short association times and a high rate of mobility. This observation fits with the predominantly student population of Dartmouth College, because students do not have a fixed workplace and are moving to and from classes all day.

1 Introduction

Since 2001, we have collected extensive data to study the use of a wireless network, specifically the wireless network at Dartmouth College. Analyzing the usage patterns will help developers to build and maintain wireless networks that are better suited to the user's needs.

This paper explores the mobility and association time of users of the Dartmouth wireless network, based on data for over two thousand users. With an understanding the behavior of a wireless network user, developers can design more effective applications that simultaneously take advantage of the wireless network and the behavior of the wireless users, network operators can more effectively deploy and manage networks, and engineers can design more effective network protocols and products.

2 Background

Dartmouth College deployed its wireless network in Spring 2001. By Spring 2002, there were about 500 Cisco Systems Aironet model 350[1] access points providing 11 Mbps coverage [4]. Because Dartmouth College has a relatively small campus, the wireless network covers all the buildings on campus and most of the outdoor spaces as well. The wireless network covers dormitories, classrooms, offices, and public spaces, including the Green, a large grassy area in the center of campus [4]. There are usually several access points per building, so there are often regions where multiple access points can be heard.

Students, staff, and professors use this network. All undergraduates at Dartmouth are required to own a computer. At the time we collected the data for this paper, about

[1] Specifications at www.cisco.com.

40% of students owned a laptop. Today, about 60-80% of students own a laptop, and 80-90% of new students choose a laptop as their primary computer.

2.1 Definitions

Throughout this paper we employ the term *user* to refer to a single wireless card, identified by its unique MAC address. Although there are certainly some devices used by more than one person, and some people who use multiple devices, our approach should be a close approximation.

Within this paper we use the term *association* to mean the duration that a user is communicating with an access point. Our data only records the fact that a user is associated with an access point. A wireless card associates with one access point at a time, and occasionally reassociates with another access point when the user moves or when the radio signal otherwise appears stronger at the new access point.

Balazinska and Castro [2] provided the terms and ideas behind prevalence and persistence. Consider two users, Alice and Bob. Alice spends one day associated with access point A and a second day associated with access point B. Bob spends both days switching between access point A and access point B every ten minutes.

Prevalence is the percentage of time a user spends associated with an access point. This time does not have to be consecutive. A user's prevalence is the fraction of their time that they are in a particular location. Because Alice and Bob connect to the same access points for the same amount of time, they have the same prevalence values: 50% for access point A, 50% for access point B.

Persistence is the length of time that a user stays at an access point before either leaving the network or transferring to another access point. Each association counts separately. Therefore, Alice has two persistence values, one for the first day when she associated with access point A for the whole day, and one for the second day when she associated with access point B for the whole day. Alice's average persistence is twenty-four hours. Bob, on the other hand, has many (6 x 24 x 2) persistence values, with an average persistence value of ten minutes.

2.2 Collection Methods

David Kotz collected the data for this paper in Spring Term 2002, specifically March 25th, 2002 to June 9th, 2002, by polling all of the access points around campus. He polled each access point approximately every five minutes using SNMP, the Simple Network Management Protocol.

2.3 Processing Methods

The data contains the results of each poll of an access point. For each poll, we have one entry for each MAC address that was associated at that time. We used this information to determine how long a user was at an access point. We calculate the duration of a user's association by subtracting the time the user arrives at an access point from the time the user departs that access point. A user's arrival time is the time of the first poll they were seen at the access point, and a user's departure time is the time of the poll before the first poll where they are missing from the access point. Of course, given the five-minute polling interval, the user may have arrived nearly five minutes prior to

the arrival tiem we measure, and may have left just prior to the first poll where they are missing, so out computation may be up to ten minutes short. This approach mirrors Balazinska and Castro [2].

This approach can be problematic when an access point fails to respond to a poll. If we fail to connect to an access point or fail to get an answer from the access point twice in a row we count all associations as having ended. Any users that are associated with the access point when we can poll it again count as starting completely new associations.

A user's persistence is derived from the set of those measurements about the duration of each association. We present the probability distribution of persistence values, that is, the probability that the user's association would last for a certain amount of time. The probability of a given persistence is the number of values we see for that particular persistence divided by the total number of persistence values.

To calculate the prevalence we totaled the time that a MAC address is at an access point throughout the trace and divided that by the total time that the MAC address is on the network. Because there were 512 access points and 2300 MAC addresses active on the wireless network when we collected this data, we do not display the full prevalence matrix here.

3 Results

In this section, we report and discuss the persistence and prevalence data collected from the wireless network.

3.1 Persistence

The distribution of user persistence is shown in Figure 1. It appears that most of the action is taking place in less than five minutes: we see that the largest bar by far is at zero to five minutes. These were the users that did not show up in any consecutive polls, so their actual association time could have been just over zero (if they associated just before this poll and left just after this poll) to just under ten minutes (if they had associated just after the previous poll and disassociated just before the next poll).

The majority of users have short association durations. This data reflects the fact that most users are students who move around a lot. Given this data, and other data [4], it is clear that many other visits happen entirely between two polls and therefore are not included in this graph. In addition, this data reflects the fact that some wireless cards are extremely aggressive about changing access points when the signal strength changes [4].

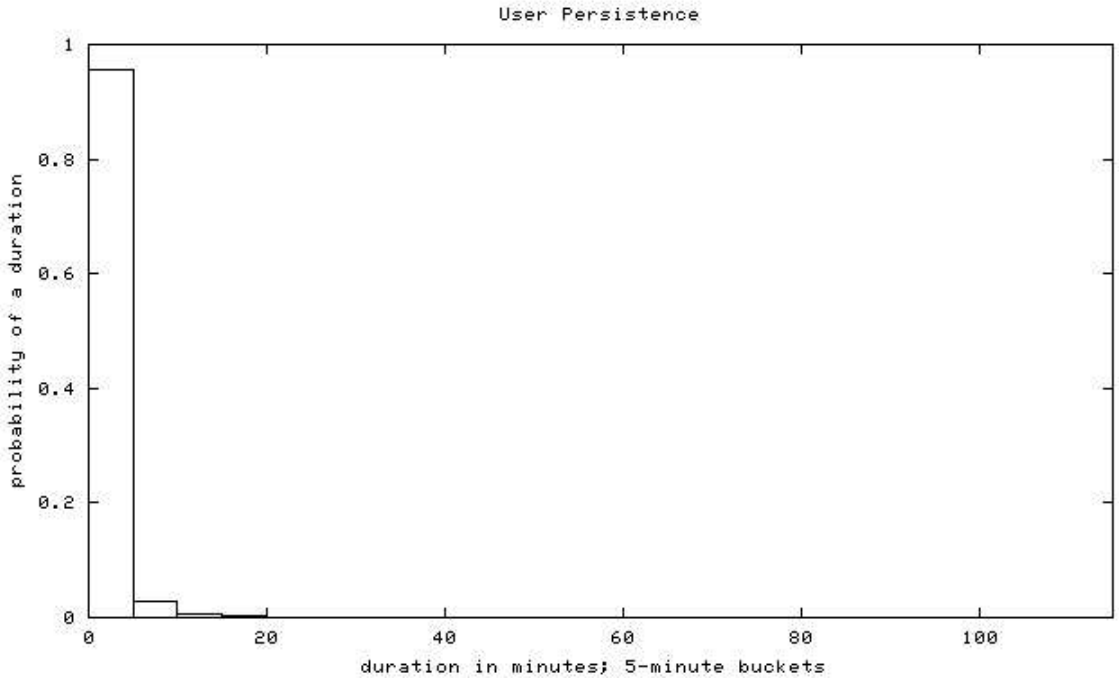

Figure 1: Probability of persistence

Figure 2 is a graph of the same data as Figure 1, however, Figure 2 has a much smaller y-axis and therefore shows the probability of persistence for over 5 minutes much more accurately.

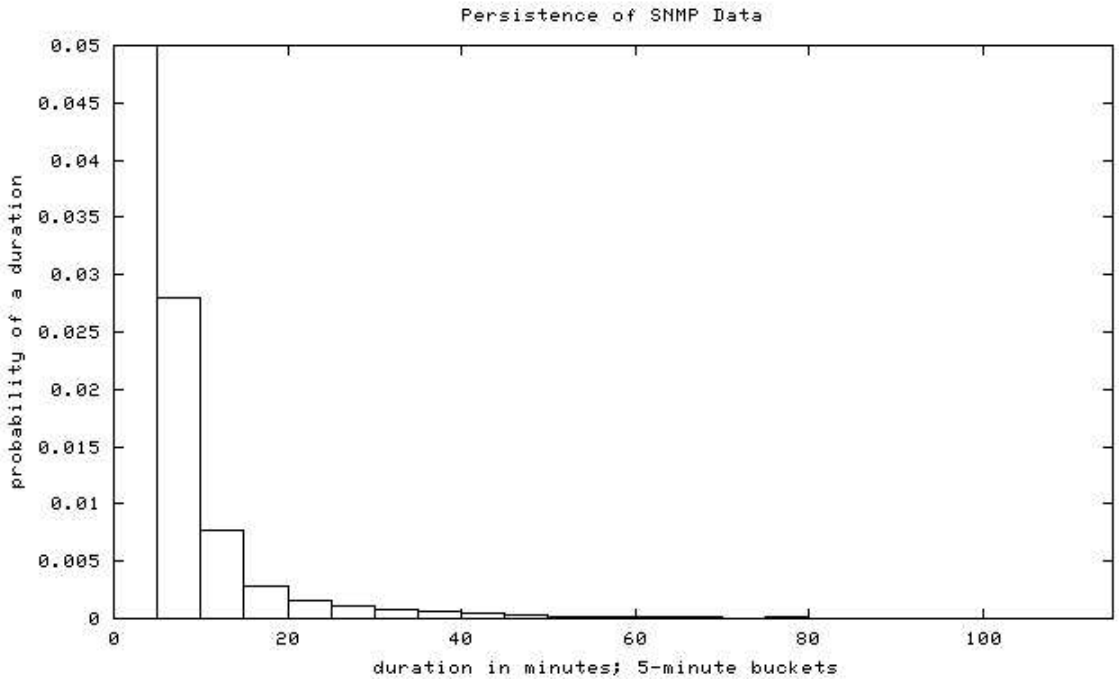

Figure 2: Expansion of probability of persistence

3.2 Prevalence

We represented prevalence as a matrix, with each row representing an access point and each column representing a MAC address (user). The value of each cell is the fraction of the user's time that the user spent at that access point [2]. The full prevalence matrix is 512 access points down by 2300 MAC addresses across. For each user, we computed that user's maximum and median prevalence across all access points. Then we gathered the users into a table with high, medium, and low maximum and median prevalence. We displayed the percentage of users that fall into each group in that cell of the table. This table follows that used by Balazinska [2].

In the table below, we see that 60% of the users at Dartmouth have a medium or high maximum prevalence and a low median prevalence. They spend over one third of their time at one access point, but otherwise, they do not spend a large amount of time in any single place. This data seems to reflect a standard student behavior of using a laptop in a dorm most of the time and randomly using it around campus occasionally. This data could also reflect the aggressive reassociation of some wireless cards, which artificially deflates the median prevalence.

Maximum Prevalence	**Median Prevalence**		
	Low [0,0.25)	Medium [0.25,0.5)	High [0.5,1]
Low [0,0.33)	20% highly mobile	0.3% mostly mobile	0%
Medium [0.33,0.66)	37% somewhat mobile	3.1% regular	2.5% semi-stationary
High [0.66,1]	23% occasionally mobile	0.3% somewhat stationary	14% stationary

Table 1: User categorization based on prevalence

The 14% of users that have both a high maximum prevalence and a high median prevalence are stationary. Those users are probably either students who for some reason do not take their laptops to class, or an administrator or professor whose job does not allow or give cause for much movement.

About 20% of users are highly mobile; they have low maximum prevalence and low or medium median prevalence. These users do not have any place where they usually stay for a long period of time. They are not spending a lot of time in their dorm or place of work. In fact, they do not have any place where they stay for longer than one third of their time.

Mostly mobile users have a low maximum prevalence and a medium median prevalence and make up 0.3% of the user population. These users alternate between several access points.

Users with a medium maximum prevalence and a high median prevalence make up 2.5% of the population. These users divide the majority of their time between two access points.

Users that are somewhat stationary make up 0.3% of the population. These users have a maximum prevalence and a medium median prevalence. These users have a primary place where they do their work, but also visit one or two other access points.

It is also interesting to see that only 3.1% of users have medium maximum prevalence and medium median prevalence. These users have two or more places where they spend the bulk of their time.

Semi-stationary, regular, mostly mobile, and somewhat stationary behaviors appear to be atypical behavior for users at Dartmouth. Users seem to either move around a lot (low maximum prevalence and low median prevalence) or have one place where they spend a large amount of their time and are mobile the rest of the time, if at all.

3.3 Limitations of our Analysis

One problem with this work is that it is based on polls at five-minute intervals. As we can see from looking at Figure 1 there is a lot of activity going on in short amounts of time. An increase in polling frequency would be an important improvement for any new data collection. Alternatively, including other types of data that could keep track of user mobility would be helpful.

Another problem is that we only looked at individual access points. This means that any movement between access points within a building counts as a move; it may not be that the user is actually moving, it could be that the wireless card re-associated for a better signal. The next analysis of this data should consider all the access points in a building together. This issue could have skewed the median prevalence towards a low median prevalence.

4 Related Work

Kotz and Essien [4] also analyzed data collected at Dartmouth College. The usage patterns of users were the main interest in this study, as they were in this paper, but they did not study persistence or prevalence.

Balazinska and Castro [2] introduced the concepts of persistence and prevalence. They collected data from a corporate setting. They also found that the majority of users are somewhat or occasionally mobile. Their data showed a higher incidence of stationary users, however, and a much lower incidence of highly mobile users.

Hutchins and Zegura [3] collected data from the Georgia Institute of Technology. They found many short associations, as we do. They found that almost half of users never move, however, which is much less mobile than the Dartmouth wireless network.

Balachandrian, Voelker, Bahl, and Rangan [1] collected data from users at the ACM SIGCOMM'01 conference. They found short association durations: 60% of all associations were less than ten minutes long. They also found significant mobility in their user population, even with only four access points. About 75% of their users visited more than one access point per day. This study only covered one room with four access points, however, so the mobility aspects of the data are not particularly meaningful.

5 Conclusion

We examine the persistence and prevalence of all wireless network users in eleven weeks at Dartmouth College during Spring 2002 when 2300 MAC addresses

appear in the SNMP data. Dartmouth's ubiquitous network coverage and large wireless population make this data particularly interesting.

Users of the Dartmouth wireless network show a large amount of mobility. Almost 80% of users have a median prevalence under 25%. Therefore, most of the users are highly mobile, even if they spend a significant amount of their time in one place. This mobility fits in with the model of students taking their laptops around campus to class and to study with friends in out of the way locations.

We found that most of the activity was taking place in less than five minutes. Future SNMP polling of wireless networks should poll more frequently than every five minutes to obtain more detailed mobility data.

6 Acknowledgements

I would like to thank several people for helping me through this thesis: Professor David Kotz for his patience and sympathy throughout the course of this paper; the staff of Academic Computing, especially Rick Barton and Susan Schwarz, for their help with SPSS and Matlab; the developers of CoralReef, tcpdump, gnuplot, and gawk; and all my friends for their support.

7 References

[1] Anand Balachandran, Geoffrey M. Voelker, Paramvir Bahl, and P. Venkat Rangan. Characterizing user behavior and network performance in a public wireless LAN. In *Proceedings of the 2002 ACM SIGMETRICS Conference on Measurement and Modeling of Computer Systems*, pages 195-205, Marina Del Rey, June 2002. ACM Press.

[2] Magdalena Balazinska and Paul Castro. Characterizing Mobility and Network Usage in a Corporate Wireless Local-Area Network. In *Proceedings of the* 2003 *International Conference on Mobile Systems, Applications, and Services (MobiSys)* pages 303-316. USENIX Association, May 2003.

[3] Ron Hutchins and Ellen W. Zegura. Measurements from a campus wireless network. In *Proceedings of the IEEE International Conference on Communications (ICC)*, volume 5, pages 3161-3167, New York, April 2002. IEEE Computer Society Press.

[4] David Kotz and Kobby Essien. Analysis of a campus-wide wireless network. In *Proceedings of the Eighth Annual International Conference on Mobile Computing and Networking (MobiCom)* pages 107-118. ACM Press, September 2002.