

Mobility Patterns in Microcellular Wireless Networks

Suttipong Thajchayapong and Jon M. Peha
Department of Electrical and Computer Engineering
Carnegie Mellon University
5000 Forbes Ave, Pittsburgh, PA 15213
Emails: st@andrew.cmu.edu, peha@cmu.edu

Abstract— This study investigates mobility patterns in micro-cellular wireless networks, based on measurements from the 802.11-based system that blankets the Carnegie Mellon University campus. We characterize the distribution of *dwelt time*, which is the length of time that a mobile device remains in a cell until the next handoff, and *sign-on interarrival time*, which is the length of time between successive sign-ons from the same mobile device. Many researchers have assumed that these distributions are exponential, but our results based on empirical analysis show that dwell time and sign-on interarrival time can be accurately described using heavy-tailed arithmetic distributions that have infinite mean and variance. We also show that the number of handoffs per sign-on can be modeled accurately with a heavy-tailed distribution.

I. INTRODUCTION

To analyze or simulate micro-cellular wireless networks, as is necessary when determining how much capacity is needed or whether a given protocol is effective, a researcher must make assumptions about mobility patterns. As this paper will show, these assumptions have generally been unrealistic. Realistic models can only be found by analyzing measured data. In this study, we perform an empirical analysis on the data collected from the Wireless Andrew network, the enterprise-wide broadband micro-cellular wireless network that blankets the Carnegie Mellon University (CMU) campus [1,2,3]. CMU was a pioneer of 802.11-based networks, which have recently become extremely popular, so it is likely that many future enterprise networks will look like Wireless Andrew.

Mobile devices in the system we observed generate a sign-on message to establish a connection with the micro-cellular wireless network, and a handoff message when they move from cell to cell. *Sign-on interarrival time* is the length of time between successive sign-ons from the same mobile device. *Dwell time* is the period that a mobile device maintains a connection with a cell until the next handoff. We will characterize both distributions

Many researchers have assumed that dwell time and sign-on interarrival time are exponentially distributed, because this is intuitively reasonable and easy to analyze. We will show that this is also incorrect, at least for these microcellular networks. Some researchers have instead sought to derive these distributions using intuitively reasonable assumptions about the speed and direction of device movement, and the geometry of cells. For example, a researcher might assume that movement of mobile devices is Brownian [4], or that a device periodically changes its direction by an angle that is uniformly

distributed from $-a$ to a [5], or that direction periodically changes with probability 20% [6,7]. Deriving distributions from known assumptions has an advantage; in addition to telling you what a distribution is, it tells you why. The difficulty with this approach is that each of these intuitively reasonable assumptions about movement may lead to different results, and it is hard to tell whether a result is accurate without verifying the assumptions through measurement.

The most straight-forward approach to determining dwell time and sign-on interarrival time is to measure them directly. Assembling accurate data is difficult, so this is the least common approach. There has been some empirical analysis of mobility in commercial cellular networks. Measurements in the British Columbia Telephone Mobility Cellular network revealed a lognormal distribution for dwell time [8]. Since few institutions have long experience with vast micro-cellular systems like Wireless Andrew, there has been little measurement-based analysis of mobility of these networks until now.

In Section II, we briefly describe the Wireless Andrew network and our empirical data. Sections III and IV contain our analysis of dwell time and sign-on interarrival time, respectively. Section V, describes our analysis on the number of handoffs per sign-on, which is defined as the number of times that a mobile performs handoffs before signing on to the network again. Finally, Section VI gives a conclusion.

II. DESCRIPTION OF WIRELESS ANDREW AND THE DATA

Wireless Andrew is an enterprise-wide broadband micro-cellular wireless network built to provide a high-speed wireless service to Carnegie Mellon University campus. The project began in 1993 as the world's first large micro-cellular wireless network. Wireless Andrew consists of many Access Points (AP). Each access point supports a wireless link to mobile devices within range. It also serves as a bridge to the university's wired network [2].

The data examined in this study was measured from the Wireless Andrew network in 1997, when there were approximately 90 access points covering 6 buildings and approximately 100 users. The network operated at 2 Mb/s in the 915MHz band using a precursor to the IEEE 802.11 standard. Most of the mobile devices were laptop computers that ran all of the software on a typical desktop computer. (Wireless Andrew has since grown to 650 access points, 3 million square feet, 3000-4000 users, and wireless links of up to 11 Mbps [2].)

A mobile device initially signs on with the access point that offers the strongest signal. A connection is maintained with the current access point until the received signal level becomes sufficiently weak and there is another access point that offers a stronger signal [9,10]. From June 1997 to December 1997, the packets associated with every sign-on or handoff were captured along with a timestamp. 56% of devices were mobile in this period. The other devices were excluded from this study, along with one outlier device [11], which we believe was used for testing.

In this paper, we consider a handoff or sign-on to occur at the instant when an access point informs the mobile device that the handoff or sign-on was successful (even if the acknowledgement does not successfully reach the mobile for some reason). Sign-on interarrival time is the time between a sign-on and the previous sign-on while dwell time is the time between a handoff and the previous handoff or sign-on.

III. ANALYSIS OF DWELL TIMES

A. Distribution of Dwell Times

Table 3.1: Descriptive Statistics of Dwell data

Mean	3004.79 seconds (\cong 50 minutes)
Median	3 seconds
Standard Deviation	70,753.28 seconds (\cong 20 hours)
Max	5,115,456 seconds (\cong 58.83 days)
Min	0 seconds
Total number of samples	9105 samples

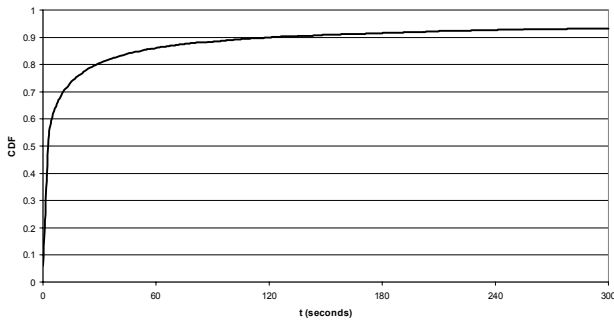


Figure 3.1: The Cumulative Distribution Function of Dwell Times

Figure 3.1 shows the distribution of dwell time. Most dwell times are small: approximately 54.62% are three seconds or less. We believe that many of these small dwell times occur when the mobile devices are receiving signals of comparable strength from multiple access points. This causes the mobile devices to frequently switch connection from one access point to another.

Despite the fact that 90% of all dwell times are less than two minutes, the mean was over fifty minutes. This shows that the tail of the dwell time distribution dominates the mean.

B. Estimations of the Distribution of Dwell Times

In this section, we determine closed-form expressions of the distribution for dwell time, which is denoted by $f(t)$. In our analysis, $f(t)$ is derived using the measured PMF $p(i)$, which describes the probability that dwell time rounded to the second

equals i . We considered a number of possible shapes for the PDF. We show results with the exponential distribution ae^{-bt} , which many researchers have assumed to be representative of dwell time, and the Pareto distribution at^{-b} . In each case, the constants a and b are set to minimize least mean squared (LMS) error. In order to determine how well our equations fit the distributions, in our plots, we also show high and low error bars. Each point on an error bar curve is twice the sample standard deviation above or below probabilities, which in this case is $p(i)$. The sample standard deviation has a value of $\sqrt{p(i)/N}$ where $N = 9105$ is the total number of samples.

1) Estimation of Dwell Time Distribution in Seconds

We estimate the PDF $f(t)$ of dwell time for small values of t . For the first three seconds, $f(t)$ increases with respect to t . The exponential and Pareto distribution with constants selected to minimize LMS errors from $p(i)$ between 4 seconds and 300 seconds are shown in Table 3.2.1 and in Figure 3.2. The Pareto distribution yields a much lower error than the other distributions we tried, including the exponential.

Table 3.2.1: Equations that fit Dwell Time Distribution

Types of Estimated Equations	Optimal Equations	LMS Errors
Pareto	$0.374t^{-1.44}$	3.247×10^{-9}
Exponential	$0.009e^{-0.03t}$	5.131×10^{-8}

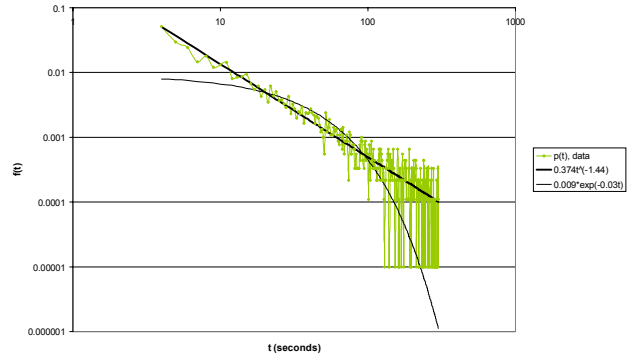


Figure 3.2a: The Estimation of Dwells Distribution using Pareto and Exponential

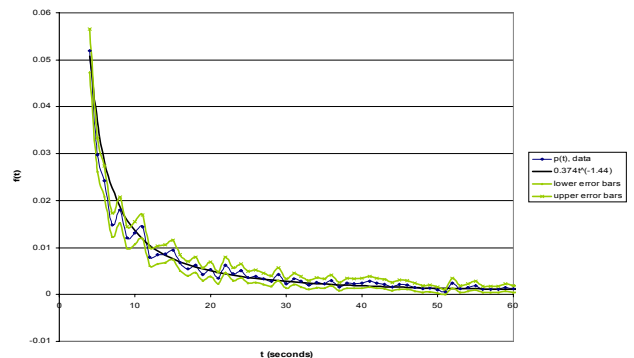


Figure 3.2b: The Estimation of Dwells Distribution using Pareto

2) Estimation of Dwell Time Distribution in Minutes

While the distribution in Section 3.2.1 accurately describes dwell time within the first five minutes, which account for 93.34% of the dwells we observed, the mean is dominated by

the larger dwell times. In order to represent the tail of the distribution more clearly, we define a new probability mass function $q(j)$, where $q(j)$ is the probability that the dwell time rounded up to the nearest minute equals j , i.e.

$$q(j) = \sum_{i=60j-59}^{60j} p(i)$$

We determine the distributions $f(t)$ that

minimize LMS error from $q(j)$ for periods of two minutes to 1000 minutes, where $q(j)$ corresponds to $f(j-0.5)$ minutes). The results are shown in Table 3.2.2 and in Figure 3.3. Figures 3.3a and 3.3b show that the Pareto is close to the measured data and well within error bars for dwell times up to 300 minutes, and the same is true for dwell times of many minutes [11]. Once again, the Pareto distribution closely fits the empirical data while the exponential and other distributions we have tried do not.

Table 3.2.2: Equations that fit Dwell Time Distribution

Types of Estimated Equations	Optimal Equations	LMS Errors
Pareto	$0.065t^{-1.445}$	1.749×10^{-9}
Exponential	$0.004e^{-0.062t}$	1.468×10^{-7}

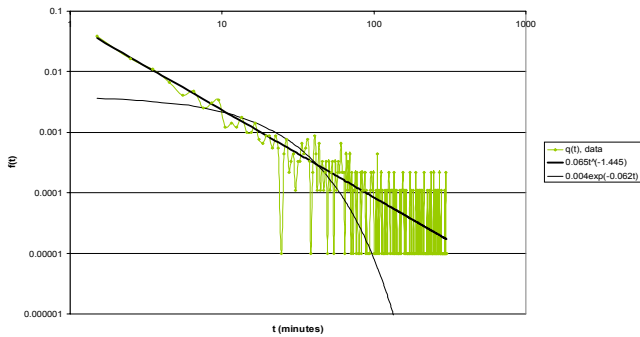


Figure 3.3a: The Estimation of Dwells Distribution using Pareto and Exponential

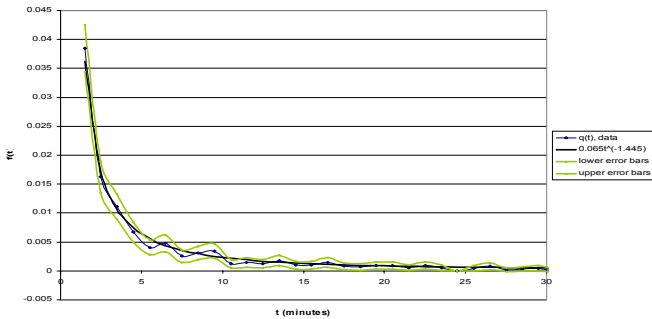


Figure 3.3b: The Estimation of Dwells Distribution using Pareto

C. Predictions using the Distribution of Dwell Time

Given that it has been μ since last handoff, we can find the probability $P(k, \mu)$ that a handoff will occur within the next period of duration k from our empirical data. We also calculate the same probability $F(k, \mu)$ from the estimation of the distribution. The graphs of $P(k, \mu)$ and $F(k, \mu)$ for $k = 1$ minute, 5 minutes and 1 hour are shown in Figure 3.4. The estimated closed-form expressions seem to work well for this prediction.

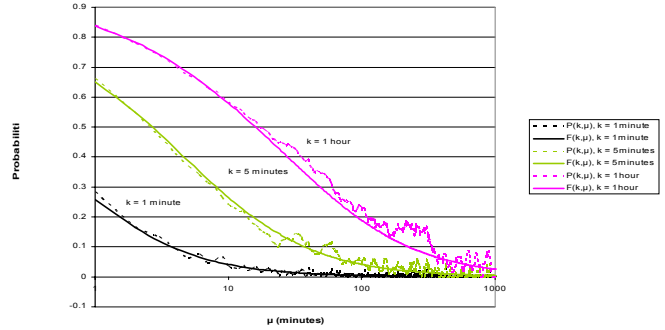


Figure 3.4: Predictions of the Probabilities that Dwell ends within 1 minute, 5 minutes and 1 hour using Empirical Data and Estimations of Dwell Time Distribution

Let $E(K)$ and M denote the expected value and median of time until the next handoff, respectively, given that it has been μ since the last handoff. Figure 3.5 shows $E(K)$ and M as a function of μ , based on our empirical data. It shows that the longer it has been since the last handoff, the longer it is likely to be until the next.

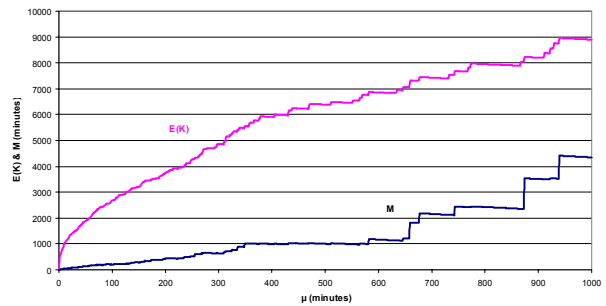


Figure 3.5: Predictions of the Expected Values ($E(K)$) and Medians (M) of period that dwell will end

IV. ANALYSIS OF SIGN-ON INTERARRIVAL TIMES

A. Distribution of Sign-on Interarrival Times

Table 4.1: Descriptive Statistics of Sign-on Interarrival data

Mean	31,162.57 seconds (\cong 8.65 hours)
Median	23 seconds
Standard Deviation	257,698.57 seconds (\cong 71.58 hours)
Max	8,013,965 seconds (\cong 92.75 days)
Min	0 seconds
Total number of samples	13785 samples

Figure 4.1 shows the distribution of sign-on interarrival time. The table and the graphs show that most sign-on interarrival times are small: approximately 50.4% are 23 seconds or less. For example, this can occur when mobile devices are near the edge of the network, so they frequently make and lose connection to the closet access points.

The sample mean of sign-on interarrival time was over 520 minutes, while 80% of all sign-on interarrival times are less than 8 minutes. This shows that the tail of the distribution dominates the mean as it did with dwell time.

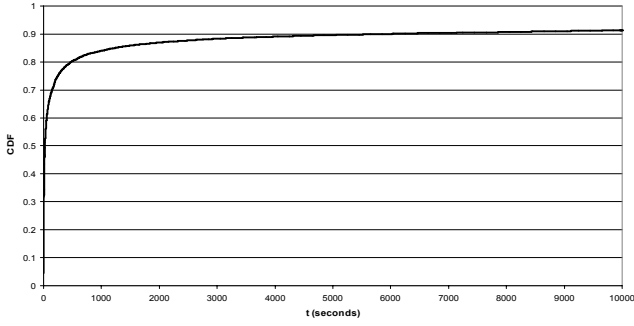


Figure 4.1: CDF of Sign-on Interarrival Times

B. Estimations of Sign-on Interarrival Time Distribution

In this section, we find closed-form expressions for the distribution for sign-on interarrival time, denoted by $g(t)$. In our analysis, $g(t)$ is derived using the measured PMFs $w(i)$ which describes the probability that sign-on interarrival time rounded to the second equals i . More specifically, we assume that $w(i)$ corresponds to $g(i \text{ seconds})$. We considered a number of possible shapes for the PDF. In this paper, we show results with the exponential distribution ae^{-bt} , which many researchers have assumed to be representative of sign-on interarrival time, and the Pareto distribution at^{-b} . In each case, the constants a and b are set to minimize least mean squared (LMS) error. As we did with dwell time, we also show high and low error bars which are two sample standard deviations ($\sqrt{w(i)/N}$) from $w(i)$, where $N = 13785$ is the number of samples.

1) Estimation of Sign-on Interarrival Distribution in Seconds

We first estimate the PDF $g(t)$ of sign-on interarrival time for small values of t . For the first two seconds, $g(t)$ increases with respect to t . The exponential and Pareto distribution with minimum LMS errors from $w(i)$ between 3 seconds and 300 seconds are shown below in Table 4.2.1 and in Figure 4.2. The Pareto distribution yields a much lower error than the exponential and other distributions we tried. In addition, it is clear from the Figures that the distribution fits the data well.

Table 4.2.1: Equations that fit Sign-on Interarrival Times Distribution

Types of Estimated Equations	Optimal Equations	LMS Errors
Pareto	$0.205t^{-1.165}$	2.356×10^{-9}
Exponential	$0.004e^{-0.011t}$	1.069×10^{-7}

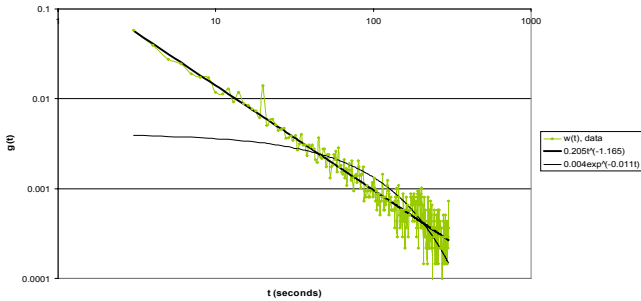


Figure 4.2a: The Estimation of Sign-on Interarrivals Distribution using Pareto and Exponential

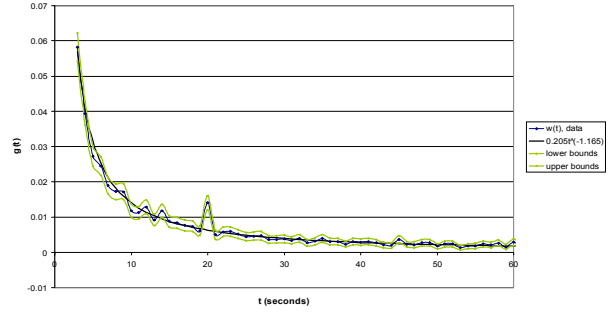


Figure 4.2b: The Estimation of Sign-on Interarrivals Distribution using Pareto

2) Estimation of Sign-on Interarrival Distribution in Minutes

In order to represent the tail of the distribution more clearly, we define a new probability mass function $v(j)$, where $v(j)$ is the probability that the sign-on interarrival time rounded up to the nearest minute equals j , i.e. $v(j) = \sum_{i=j-59}^{60j} w(i)$. We assume that $v(j)$ corresponds to $g(j - 0.5 \text{ minutes})$, and determine the distribution that minimizes LMS error from $v(j)$ for periods ranging from 2 minutes to 10000 minutes. The results are shown below in Table 4.2.2 and in Figure 4.3 for periods from 2 minutes to 720 minutes (=12 hours). It can be seen that the distribution is still heavy-tailed and has infinite mean and variance.

Table 4.2.2: Equations that fit Sign-on Interarrival Time Distribution

Types of Estimated Equations	Optimal Equations	LMS Errors
Pareto	$0.131t^{-1.345}$	1.052×10^{-8}
Exponential	$0.006e^{-0.04t}$	5.945×10^{-7}

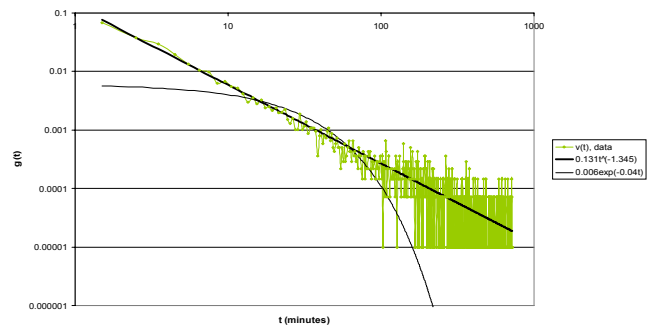


Figure 4.3a: The Estimation of Sign-on Interarrivals Distribution using Pareto and Exponential

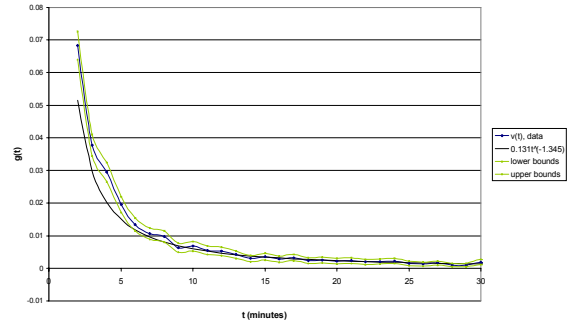


Figure 4.3b: The Estimation of Sign-on Interarrivals Distribution using Pareto

3) Estimation of Daily Cycle

In order to reduce noise and present the distribution of large sign-on interarrival times even more clearly, we define another probability mass function $y(d)$ for integer d corresponding to successive six-hour periods, i.e.. $y(d) = \sum_{j=360d-359}^{360d} v(j)$. We

find that the distribution has a wave-like curve. The peaks of this curve correspond to multiples of 24 hours. This shows a daily cycle pattern of users' behaviors, which happens when users sign-on at approximately the same times of the day. For example, a user arrives at her work place and signs-on once and then comes back and signs-on again at approximately the same time the next morning. The three-day peak is slightly higher than the two-day peak, presumably because people leave the office on Friday and return Monday.

We employ a distribution of the form $h(t) = 0.131t^{-1.345} \bullet s(t)$ where $0.131t^{-1.345}$ is the distribution that worked well for a 12-hour period and $s(t)$ is a sine wave with a 24-hour (1440 minutes) period. We further assume that the sine wave has a peak at time 0 so that $s(t) = m \sin(2\pi t/1440 + \pi/2) + n$. The constant m and n are selected to minimize LMS error with $y(d)$, the fraction of sign-on interarrivals occurring in the d^{th} 6-hour period. We assume that $y(d)$ corresponds with $h(d \bullet 6 \text{ hours} - 3 \text{ hours})$. The plot of $y(d)$ and the estimation $h(t)$ is shown in Figure 4.4.

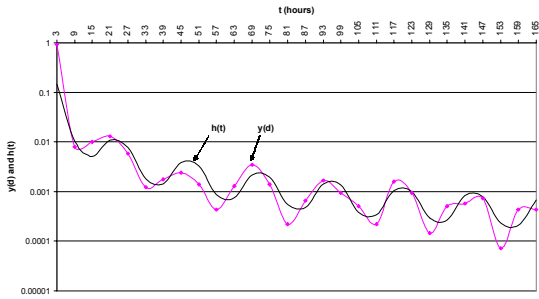


Figure 4.4: The Estimation of Daily Cycle Pattern of Sign-on Interarrival Times Distribution using a sine function and a Pareto as its envelope

C. Predictions using Sign-on Interarrival Time Distribution

Let $E(S)$ and H denote the expected value and median of time until the next sign-on, respectively, given that it has been β since the last sign-on. Figure 4.5 shows that, as with dwell time, the longer it has been since the last sign-on, the longer it is likely to be until the next. However, this effect seems to greatly diminish after a day without a sign-on.

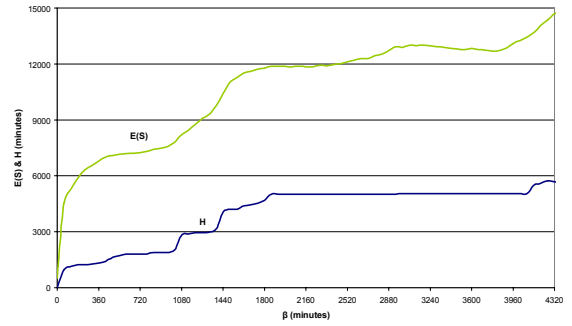


Figure 4.5: Predictions of the Expected Values ($E(S)$) and Medians (H) of period until the next sign-on

V. ANALYSIS OF THE NUMBER OF HANDOFFS PER SIGN-ON

We define the probability mass function of the number of handoffs per sign-on, $\psi(z)$, where z is the number of handoffs per sign-on. We included all sign-ons that were followed by one or more handoffs. Moreover, as described in Section 3.1, in the system we employed, the majority of dwell times were short, presumably because a device that is receiving signals of comparable strength from multiple access points sometimes oscillates between them. Since these rapid handoffs are probably not indicative of actual user mobility, we exclude those handoffs that follow dwell times less than four seconds.

Table 5.1: Descriptive Statistics of Number of Handoff per sign-on data

Mean	2.44 handoffs per sign-on
Median	1 handoff per sign-on
Standard Deviation	3.26 handoffs per sign-on
Max	44 handoffs per sign-on
Min	1 handoff per sign-on
Total number of samples	1583 handoffs per sign-on

We estimate the distribution using the geometric ab^z , and Zeta az^{-b} distributions, where a and b are found from minimizing least mean squared (LMS) error from $z = 1$ to $z = 10$. Figure 5.1 and Table 5.2 show the resulting equations, and their corresponding LMS errors. It can be seen that the Zeta distribution is a close fit, while the geometric is not. Indeed, even though LMS was only minimized for z in the range of 1 to 10, Figure 5.1 shows that the Zeta is still a good fit for larger values of z .

The Zeta distribution with exponent -2.165 is heavy-tailed with finite mean and infinite variance. We note that $\sum_{z=1}^{\infty} 0.805 z^{-2.165} = 1.218$, so this expression cannot exactly represent the distribution for all z . This may be caused by the uncertainty in our data for large values of z due to the fact that we have only 1583 samples, or it may be because the tail of the PMF decays faster when $z > 10$ than it does when $z < 10$.

Table 5.2: Equations that fit Number of Handoffs per Sign-on Distribution

Types of Estimated Equations	Optimal Equations	LMS Errors
Zeta	$0.805z^{-2.165}$	0.007656
Geometric	$1.131(0.478)^z$	0.026636

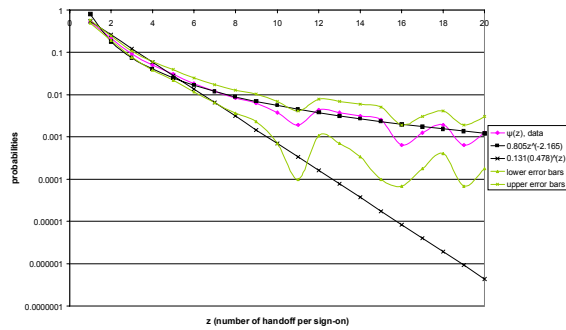


Figure 5.1: Estimation of the Distribution of Number of Handoffs per Sign-on using Geometric and Zeta

VI. CONCLUSIONS

Using empirical data from Carnegie Mellon University's Wireless Andrew network [1,2,3], which is representative of the enterprise-wide 802.11-based broadband micro-cellular wireless networks that are rapidly growing in popularity, we have derived distributions for dwell time, sign-on interarrival time, and the number of handoffs per sign-on.

For both dwell time and sign-on interarrival time, the sample mean exceeded the sample median by roughly three orders of magnitude, which shows that the tails of these distributions dominate the means. Indeed, both distributions can be well characterized with an arithmetic tail, and parameters that imply infinite variance and even infinite mean. This surprising result contrasts greatly with the common assumptions that dwells and sign-on interarrivals can be modeled with a light-tailed exponential distribution. We have demonstrated that the exponential works poorly in both cases. We have also shown that dwell times in the microcellular network are nothing like the lognormal distribution observed in the BC cellular network [8].

The PDF curve for dwell time increases from 0 to 3 seconds, which is the median. We believe that the median is low because there are circumstances where a device is receiving signals of comparable strength from multiple base stations, and it oscillates among them whenever signal strengths or interference levels change. From four seconds to roughly 17 hours [11], the Pareto distribution with an exponent of -1.445 fits well with the observed distribution of dwell time.

For sign-on interarrival time, the PDF curve similarly increases as time goes from 0 to 2 seconds, and then it follows a Pareto distribution for durations of a few seconds, minutes, and even hours. In addition, we find a daily cycle of users' behavior. The long-term distribution can be described well by multiplying the same Pareto function with a sine wave that has a 24-hour cycle.

Because these distributions are arithmetic and heavy-tailed, the longer it has been since the last handoff or sign-on, the longer it is likely to be until the next. This fact could be exploited by algorithms that predict mobility, perhaps to estimate the likelihood that the number of devices in a cell will exceed capacity. This could be useful in admission control [12], frequency assignment and other algorithms.

The fact that the length of time each mobile device is in a particular 802.11 LAN micro-cell is heavy-tailed may imply that the period when a device is sending traffic to that LAN is also heavy-tailed, which could add to long-range dependence of traffic in micro-cellular networks. This long-range dependence has already been observed in wide-area cellular networks [13].

Finally, a Zeta distribution is a reasonably accurate estimate of the number of handoffs per sign-on. With the parameters we found, this distribution is also heavy-tailed with infinite variance, but finite mean.

REFERENCES

- [1] Hills, Alex, "Wireless Andrew [mobile computing for university campus]", *IEEE Spectrum*, June 1999 Volume: 36 Issue: 6, Pages 49 - 53
- [2] Hills, Alex, "Larger Scale Wireless LAN Design", *IEEE Communications Magazine*, Volume: 39 Issue: 11, Nov. 2001, Pages 98 -107
- [3] Bennington B. J.; Bartel C R., "Wireless Andrew: Building a High Speed, Campus-Wide Wireless Data Network", *ACM Mobile Networks and Applications*, 9-22, January - February, 2001
- [4] Jugl, E. ; Boche, H, "Dwell Time Models for Wireless Communication Systems" *Vehicular Technology Conference, 1999, Fall. IEEE VTS 50th, Volume: 5, 1999, Pages 2984 -2988*
- [5] Zonoozi, M.M.; Dassanayake, P., "User mobility modeling and characterization of mobility pattern" *Selected Areas in Communications, IEEE Journal on, Volume: 15 Issue: 7, Sept. 1997, Pages 1239 -1252*
- [6] Kourtis, S.; Tafazolli, R., "Evaluation of handover related statistics and the applicability of mobility modelling in their prediction", *Personal, Indoor and Mobile Radio Communications, 2000. The 11th IEEE International Symposium on, Volume: 1, Pages 665 -670*
- [7] Kourtis, S.; Tafazolli, R., "Modelling cell residence time of mobile terminals in cellular radio systems", *Electronics Letters, Volume: 38 Issue: 1, 3 Jan. 2002, Pages 52 -54*
- [8] Jedrzycki, C.; Leung, V.C.M., "Probability distribution of channel holding time in cellular telephony systems", *Vehicular Technology Conference, 1996, Mobile Technology for the Human Race, IEEE 46th, Volume: 1, 1996, Pages: 247 -251*
- [9] "Data Manual WaveLAN Air Interface", AT&T, July 11, 1995
- [10] "Roaming for WaveLAN -WaveAROUND-", AT&T
- [11] Thajchayapong S., "Mobility Patterns in Micro-Cellular Wireless Networks", *Master's Thesis, Carnegie Mellon University, May 2002*
- [12] Peha J. M.; Sutivong A., "Admission Control Algorithms For Cellular Systems," *ACM Wireless Networks, March-April, 2001, Volume 7, No. 2, pp 117-125. Also at www.ece.cmu.edu/~peha/papers.html*
- [13] Jiang M.; Nikolic, M.; Hardy, S.; Trajkovic, L., "Impact Self-similarity on Wireless Network Performance," *IEEE International Conference on Communications, June 2001, pp. 477-81.*