Characterization of CDMA2000 Cellular Data Network Traffic

Carey Williamson   Emir Halepovic   Hongxia Sun   Yujing Wu
Department of Computer Science
University of Calgary
{carey,emirh,sunh,ywu}@cpsc.ucalgary.ca

Abstract

This paper describes the analysis of low-level measurements from a `CDMA2000 1x` *cellular data network. The network traces record detailed information about wireless Internet packet data call activity on the network, including mobile station identity, call initiation, burst behaviour, supplementary channel usage, soft handoffs, and call termination. The analysis in this paper focuses on one continuous week-long trace data set, representative of cellular data network activity. The results from the analysis illustrate the burstiness of the packet call arrival process and the diurnal patterns of cellular data users. The results also characterize the activity per cell site, activity per user, data burst activity, user mobility, and the density of cellular network coverage. Several observations reinforce known results about heavy-tailed properties in wired Internet traffic, while others show interesting differences in wireless versus wireline traffic.*

1. Introduction

Many cellular network providers offer wireless Internet data services to their subscribers. These services include electronic mail, Web browsing, text messaging, camera phones, gaming, and more. Emerging services include low-bandwidth wireless video streaming, home security monitoring, and peer-to-peer file sharing.

New data services can have a dramatic impact on the usage of the cellular network, particularly as emerging applications grow in popularity and the user base expands. Understanding the current data traffic characteristics is important to the capacity planning of future wireless/cellular data networks [7, 8, 10].

There are few published measurement studies about cellular data network traffic in the literature [1, 2, 9, 11, 14]. Adya *et al.* [1, 2] studied Web browsing behaviour, and identified power-law relationships in object popularities. Kunz *et al.* [11] discussed the differences between WAP traffic and Web browsing traffic. Shankaranarayanan *et al.* [14] presented measurements of cellular voice and data users, and used this information to parameterize simulation models for a capacity planning study. Sharp *et al.* [15] study traffic patterns in public safety networks, while Vujicic *et al.* [18] present a detailed statistical characterization of voice traffic in a "push to talk" mobile radio system.

In this paper, we present a detailed analysis of the data traffic characteristics observed on a CDMA2000 1x cellular data network. The measurements were collected with the cooperation of a cellular network provider, with help from the vendors of the equipment used in the cellular network.

The network traces record low-level information about the packet data call events occurring between the mobile stations (i.e., cell phones), the base station (i.e., cell site), and the base station controller. The events indicate mobile station identifiers, as well as the start time, end time, cell site, sector id, and carrier frequency used for each packet data call. Within each packet data call, the trace also records information about fundamental channel and supplementary channel usage, including the data rate and duration for each supplementary channel data burst in the forward and reverse directions.

To the best of our knowledge, our study is the first to use such fine-grain measurement data to characterize wireless Internet packet data call traffic. The analysis in this paper uses one continuous week-long trace of cellular data network traffic, representing typical network usage. We use a locally-written analysis program to process the traces and summarize the statistical properties of the traffic observed.

There are three main observations from our traffic analysis. First, the packet call arrival process is very different from the Poisson arrival process traditionally associated with cellular voice traffic [7]. Second, the daily usage patterns of cellular data network users have a diurnal structure, but the peak usage tends to be in the evening hours. Finally, our results identify heavy-tailed workload characteristics in several aspects of cellular data network traffic, including the activity per user and the burst activity on the supplementary channel. These characteristics are consistent with prior results characterizing wired Internet traffic [1, 2, 4, 13].

Our results suggest that skewed distributions and power-law characteristics are ubiquitous properties of Internet traffic in both wired and wireless networks. These workload characteristics arise from the behaviours of the human users, rather than from the specific network technologies being used. Modeling these characteristics is important for capacity planning of cellular networks, wireless LANs, wireless mesh networks, and any emerging hybrid wireless network technologies.

The remainder of this paper is organized as follows. Section 2 summarizes our data collection methodology and the traces used in this paepr. Section 3 presents the results from our network traffic analysis. Section 4 concludes the paper, and discusses our ongoing work.

2. Methodology

The data sets analyzed in this paper were collected from an operational CDMA2000 1x cellular data network in March 2004. The measurements were collected by instrumenting vendor equipment to report all cellular network events involving mobile stations, the base station, and the base station controller. Event timestamps are recorded with 20 ms granularity.

A total of 43 traces were collected from the cellular data network. The traces were collected from several different measurement points, over the time span of several weeks. The individual traces range in duration from about 1 to 24 hours, depending on the location and the time of day.

The aggregate set of traces represents over 480,000 packet data calls from over 10,000 cellular network users. The aggregate data set provides a large sample for the statistical analysis of network traffic characteristics.

In this paper, we restrict our attention to one continuous week-long portion of the trace data from one measurement location, as a representative example of the cellular data network activity. We select this dataset because it has the highest level of traffic activity observed among the traces collected, and because it provides the longest duration of continuous trace collection.

Table 1 summarizes the characteristics of the specific traces used in this paper. These traces record aggregate traffic from over 100 cell sites in the network.

An example of the trace format is shown in Figure 1. This example shows one packet data call. The call originated from mobile station identifier (555) 249-0623 in region 7846 of the provider's network, using frequency 824 MHz. The call was placed at 10:05pm (79,525 seconds after midnight) on Tuesday March 9, 2004. The call lasted about 62 seconds. The intermediate reports (ACTIVE) show the active set of cell ids (CID) and sector ids (SID) with which the mobile station communicated during the call. Multiple

entries in this set indicate a soft handoff state between sites or sectors during the call.

```
MSID    5552490623
ESN     0xe3ce7469ace
SITE    7846
FREQ    824
START   2004 03 09 79525.080
ACTIVE  2004 03 09 79525.760
CID     287
SID     3
ACTIVE  2004 03 09 79526.240
CID     287
SID     3
CID     602
SID     2
          .
          .
          .
END     2004 03 09 79587.060
```

Figure 1. Example of Trace Format

We designed custom programs for analyzing this trace format and summarizing the data packet call behaviour on the cellular network. One C program (about 800 lines of code) processes the trace file in ASCII form. The program builds a state record for each packet call that includes user id, cell id, sector id, call start time, call end time, call duration, supplementary channel burst activity, and soft handoff events. As output, the program produces a file containing a time-ordered sequence of call records. In addition, a custom Java program (about 1500 lines of code) is used to produce the summary statistics and extract the data for plots from the traces. The statistics and plot data are produced per trace and overall for call and cell activity, user mobility, and per call characteristics. The results from our analysis follow in the next section.

3. Results

3.1. Overview

Figure 2 provides a graphical overview of the cellular data network traffic profile in our week-long data set. The horizontal axis of the graph shows time; our trace spans from 5:54am on Thursday March 25 to 4:21am on Thursday April 1. The vertical axis shows the number of packet data calls initiated in each consecutive (non-overlapping) 10-minute interval in the trace.

Several observations are evident from Figure 2. First, the packet data call arrival process is bursty. The jaggedness of the plot, even at the 10-minute sampling granularity,

Table 1. Summary Information for Trace Data Sets

Trace Name	Start Date	Start Time	Trace Duration	Total Pkt Calls	Average Pkt Calls/sec
Trace 10	Thu Mar 25	5:54am	24.8 hr	27,232	0.30
Trace 11	Fri Mar 26	6:36am	11.7 hr	9,707	0.23
Trace 12	Fri Mar 26	6:13pm	11.3 hr	12,225	0.30
Trace 13	Sat Mar 27	4:55am	25.2 hr	17,376	0.19
Trace 14	Sun Mar 28	5:18am	23.9 hr	15,661	0.18
Trace 15	Mon Mar 29	5:03am	24.7 hr	30,777	0.35
Trace 16	Tue Mar 30	5:43am	24.6 hr	29,576	0.33
Trace 17	Wed Mar 31	5:59am	22.4 hr	29,847	0.37
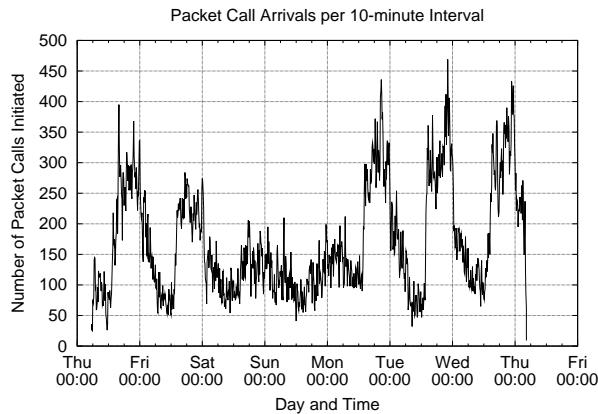
Figure 2. Packet Data Call Activity (1 Week)

shows pronounced fluctuations in the packet call data traffic. Second, the traffic pattern is non-stationary. This fact is obvious from Table 1, where the mean call arrival rates for the day-long traces (e.g., Trace 14 and Trace 15) differ by up to a factor of two. Furthermore, there is time-varying behaviour within each daily trace, as shown in Figure 2. There is a distinct diurnal pattern to the traffic, with heavier traffic on week days than on weekends, as observed by Kunz *et al.* [11]. Third, the most dominant traffic peaks appear in the late afternoon and evening hours. This behaviour is quite different from most Internet traffic studies, where the traffic pattern aligns more closely with the normal working hours (9am to 6pm). We attribute this difference in part to the pricing models offered by cellular network providers (i.e., discounted rates for air time on evenings and weekends). Another reason is the popularity of cellular data services with teenagers, who use their cell phones to browse the Internet and stay in touch with their friends outside of school hours.

Two implications follow from these observations. First, it is not possible to model packet data call arrivals using a single-rate Poisson process, as is traditionally used for voice traffic modeling. Rather, a time-varying rate process is re-

quired, as suggested by Paxson and Floyd [13]. Second, the "busy hour" for cellular data networks differs from wireline or cellular voice networks (e.g., 10pm versus 2pm). The latter observation is consistent with the results reported by Varga *et al.* [17].

3.2. Activity per Cell Site

Figure 3 shows the distribution of packet call activity observed at each cell site. In this week-long trace, there were 171,318 packet data calls observed from 139 different cell sites, for an average of 1,232 calls per site.

Figure 3(a) shows the distribution of packet calls by cell site. The horizontal axis shows the cell sites ranked according to level of activity, and the vertical axis shows the number of packet calls initiated from each cell site.

The distribution observed is highly non-uniform. A few cell sites have a high level of packet call activity, while many cell sites have a relatively low level of activity. The busiest cell site has 8,318 packet calls. We speculate that the busier cell sites are in densely populated areas, or have highly active users.

The skew to the distribution for per-cell activity is reminiscent of a Zipf-like distribution, suggesting power-law behaviour. However, more careful statistical analysis shows that the distribution is not heavy-tailed. In fact, the distribution, when discretized into bins, can be reasonably well approximated with a geometric or exponential distribution.

Figure 3(b) shows an example of this fitting with an exponential distribution. An exponential distribution is fitted to the number of calls per cell site, with a mean of 1,234.12. The Chi-square test and the Kolmogorov-Smirnov (KS) test produced p-values of 0.541 and 0.795 (two-tail), respectively, which indicate a good fit. We conclude from this that the distribution of call activity across cells for this data set is skewed, but not heavy-tailed. The distribution can be modeled well with a geometric or exponential distribution.

3.3. Activity per User

Figure 4 shows the distribution of packet call activity observed on a per user basis. Users are identified based on the mobile station identifier (MSID) in the trace. The 171,318 packet data calls in this trace came from 4,156 distinct users, for an average of 41 packet calls per user.

Figure 4 shows that the packet call activity per user is highly skewed. Figure 4(a) shows the packet calls per user on a linear scale. Users are ranked from highest to lowest according to level of activity. The busiest user generated 3,614 packet calls during the week. There were 1,053 users with only 1 packet call during the week.

Figure 4(b) shows the same distribution on a log-log scale. The "hump" in the middle of this distribution, as well as the piece-wise linear structure, suggests that there are several different categories of users. In particular, the top 40 users (less than 1% of the observed users) are heavy users of the network, generating about 33% of all the observed packet data calls. The next 140 busiest users (about 3.4% of all the users) account for about 25% of the packet data calls. The remaining users (several thousand) each contribute little to the overall packet data call activity on the network.

This skewed distribution of packet call activity per user exhibits heavy-tailed workload characteristics. A distribution is said to be heavy-tailed if the asymptotic shape of the distribution is hyperbolic. In mathematical terms:

$$Prob[X > x] \approx x^{-\alpha} \text{ as } x \to \infty \text{ where } 0 < \alpha < 2$$

The parameter α, referred to as the tail index, determines the heaviness of the tail of the distribution. Smaller values of α represent heavier tails (i.e., more of the "mass" is present in the tail of the distribution).

To illustrate the heavy-tailed property more clearly, Figure 5(a) presents an analysis using the aest tool developed by Crovella *et al.* [3]. The aest tool estimates the tail weight α for a heavy-tailed distribution. The graph shows a log-log complementary distribution (LLCD) plot of the calls per user, with probability on the vertical axis, and calls per user on the horizontal axis (each with log scale, and appropriately normalized [3]). The lowest curve in this plot shows the results for the raw data, while the successively higher curves show the results for the aggregated data, using a factor of 2 for each level of aggregation. The consistent slope of the plot over 5-6 levels of aggregation suggests the presence of a heavy-tailed distribution. The black dots on the curves indicate the points used to estimate the slope, which is $\alpha = 1.021$ in this example. This α value indicates a heavy-tailed distribution, since $\alpha < 2$.

3.4. Burst-Level Behaviour

In a CDMA2000 cellular network, data can be transmitted at the normal rate (1x) on the fundamental channel, or at higher rates (2x, 4x, 8x, or 16x) as *bursts* on a supplementary channel. The higher burst rates provide a more efficient means for transmitting large volumes of data. In most implementations, the supplementary channel bursts are triggered when the backlog of pending data exceeds a specified threshold, though some vendor equipment limits each supplementary channel burst to at most 5 seconds duration. The event records within our trace indicate when supplementary channel data bursts are used during a packet call. The trace indicates the time of the burst, the burst data rate, and the burst duration.

Figure 6 shows the results from burst-level analysis of packet call activity in our traces. Figure 6(a) shows the number of bursts generated per user, with the users sorted from highest to lowest based on burst activity. Figure 6(b) shows the same distribution on a log-log plot.

Figure 6 exhibits many of the same properties observed previously for user-level and cell-level activity. That is, the burst-level activity is highly skewed amongst the users, and the burst distribution suggests power-law behaviour.
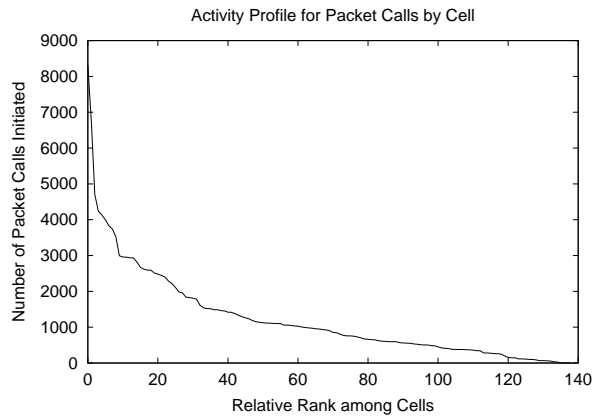
Figure 5(b) shows results confirming that the burst activity distribution is also heavy-tailed. This graph presents the results from the aest tool for the burst activity LLCD plot. There is a consistent slope to the plot over 4-5 levels of data aggregation, and the slope estimate is $\alpha = 1.042$. This α value is indicative of a heavy-tailed distribution.
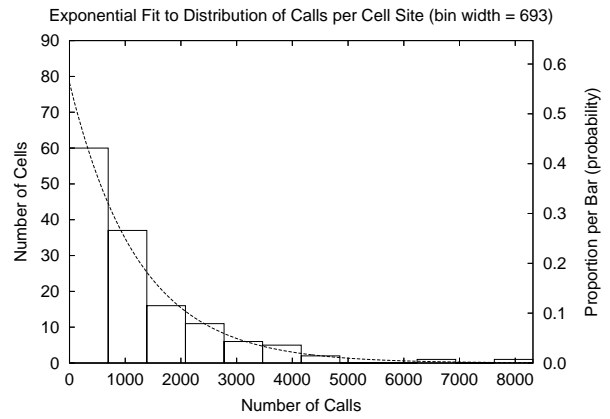
3.5. User Mobility

Our next analysis studies the mobility characteristics of users. By analyzing the trace file, we can identify the cell site at which each packet data call is originated. Extending this analysis on a per-user basis can identify user mobility characteristics, including the number of cell sites from which packet data calls are placed, and the dominant "home" cell site for a given user, if any.

Figure 7(a) shows the results from our analysis of user mobility for the top 500 users. The top line in the graph shows the number of packet data calls made by each user, with users sorted from the busiest user to the least busy user. The bottom line in the graph shows the number of cell site changes that occurred for that user during the trace (i.e., successive packet data calls were originated from different cell sites).

Two main observations with respect to user mobility are evident from Figure 7(a). First, the level of user mobility is generally quite low. Many users tend to be relatively stationary in their usage of the network, originating most or all of their successive calls from the same cell site, while
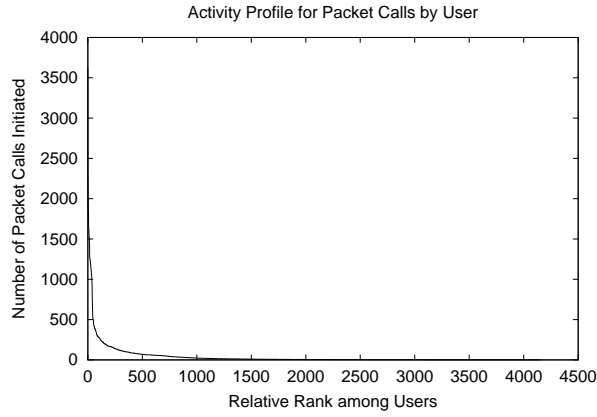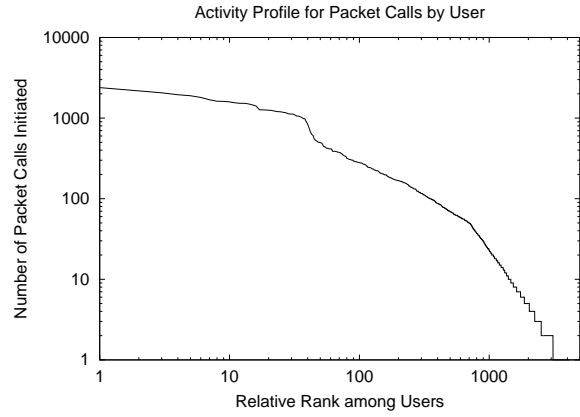
Activity Profile for Packet Calls by Cell

(a) Empirical Distribution

Exponential Fit to Distribution of Calls per Cell Site (bin width = 693)

(b) Fitted Distribution

Figure 3. Distribution of Packet Data Calls per Cell Site

Activity Profile for Packet Calls by User

(a) Linear Scale

Activity Profile for Packet Calls by User

(b) Log-Log Scale

Figure 4. Distribution of Packet Data Calls per User

Dataset Aggregation LLCD Plots with Points for Alpha Estimate of 1.021

(a) Calls per User

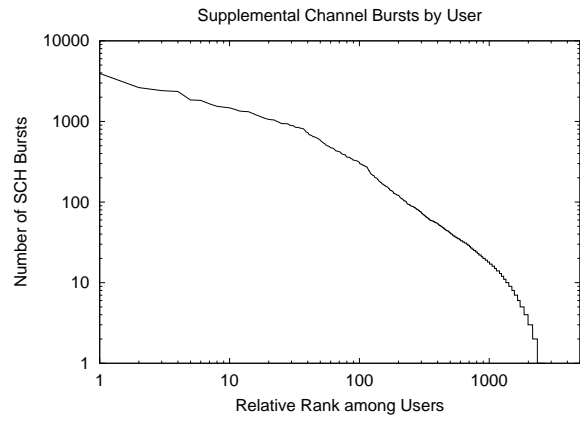Dataset Aggregation LLCD Plots with Points for Alpha Estimate of 1.042
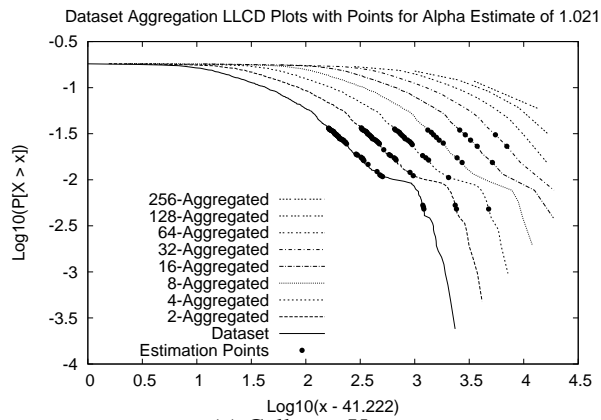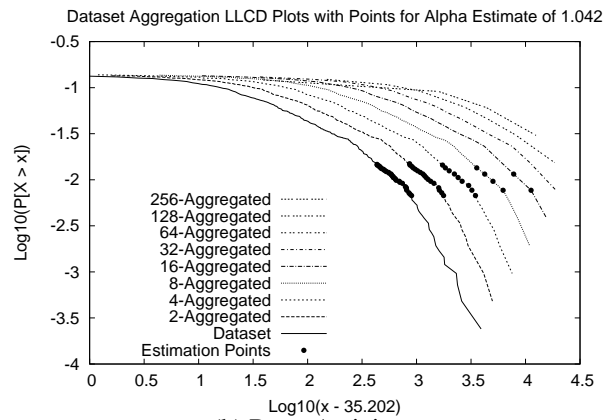
(b) Burst Activity

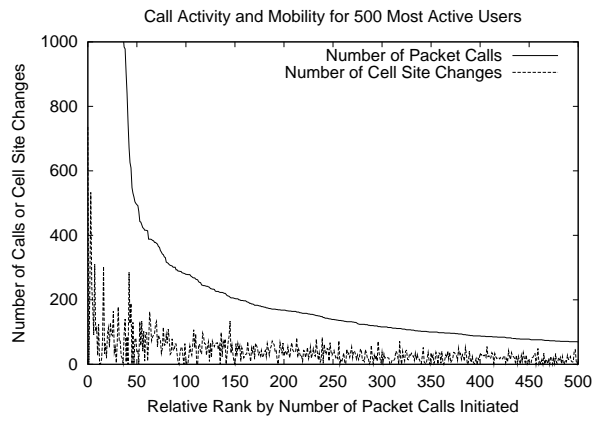**Figure 5. Analysis of Tail Behaviour using aest Tool**
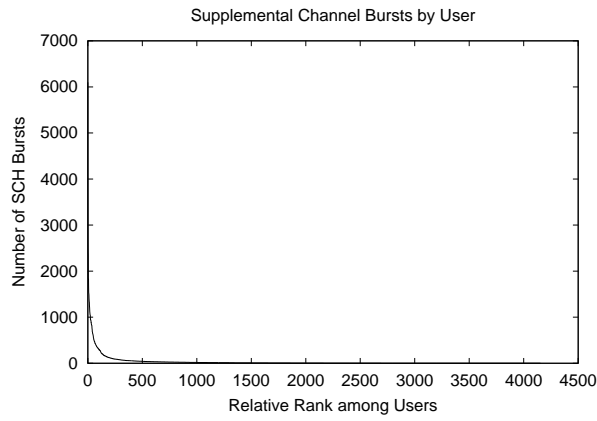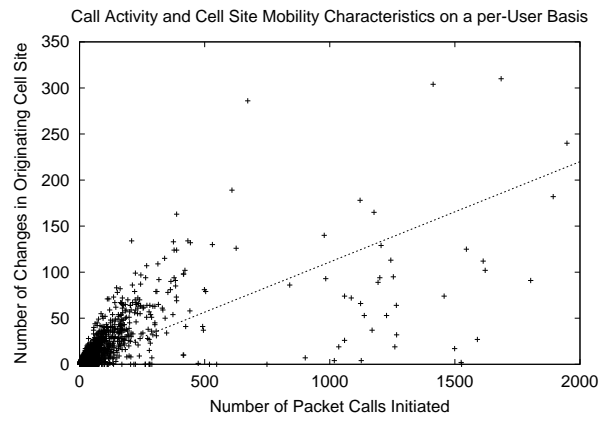
(a) Linear Scale

(b) Log-Log Scale

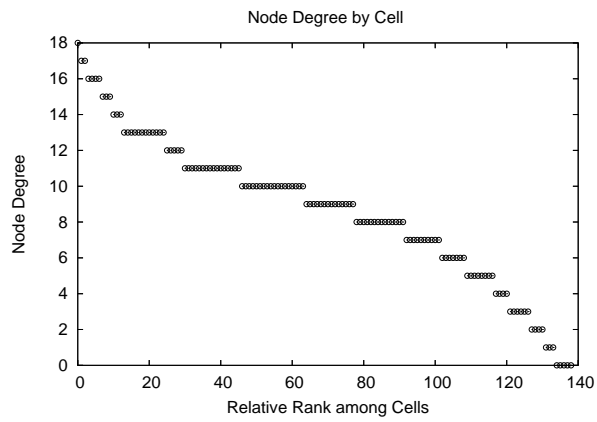**Figure 6. Supplementary Channel Burst Behaviour**
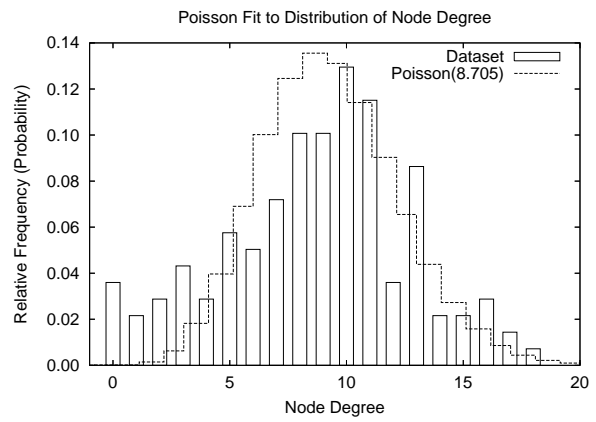

(a) Comparison of Activity and Mobility

(b) Scatter Plot and Regression Fit

Figure 7. Calls and Cell Site Changes per User

(a) Empirical Distribution

(b) Fitted Distribution

Figure 8. Distribution of Cellular Coverage Density

other users move occasionally. Second, the mobility characteristics (i.e., cell site changes) for a user are not strongly correlated to the level of packet call activity for that user. For example, the stationary users (i.e., those with 0 cell site changes) are scattered across the full range of call-level activity observed. The latter observation suggests that call activity level and cell site changes can be modeled as separate characteristics in a user mobility model.

Figure 7(b) provides a clearer representation of this information, by showing the relationship between the number of packet calls and the number of cell site changes observed on a per-user basis. Each point in the graph represents one user. Least-squares linear regression on this data produces a correlation coefficient of $R^2 = 0.55$, implying only modest correlation between call activity and cell site mobility. Additional results regarding user mobility are presented in a separate paper [5].

3.6. Soft Handoff Analysis

Our final analysis studies the richness of the coverage for the cellular network. We focus on the "soft handoff" states that are recorded in our traces. Recall that in a soft handoff, a mobile station is in simultaneous communication with multiple cell sites. We interpret this as an indication of physical proximity between the cell sites (at least in terms of wireless signal propagation), and use this information to construct an "adjacency matrix" representation between cell sites in the traces. We generate this "connectivity" information using the first two entries in each soft handoff report seen for each packet data call in the trace. This approach is applied throughout the entire trace to determine the number of "neighbours" observable by each cell site.

Our primary interest is in the "density" of the cellular network coverage. For the wired Internet, there is a known power-law structure to the topology [4]. Our goal is to see if there is similar structure in the cellular network coverage. Note that our approach to this analysis is not subject to the sampling bias issue identified by Lakhina *et al.* [12], since we are not doing random sampling of the topology. Rather, we are inferring connectivity information from *all* the calls and cell sites observed.

Figure 8 illustrates the density distribution results. Figure 8(a) shows the number of neighbouring cells observed for each cell site. Several cell sites appear as isolated islands of coverage, while some have as many as 18 neighbours. The sparse areas likely represent remote areas of coverage (or possibly cell sites with few mobile users). The denser sites likely represent metropolitan areas.

Figure 8(b) shows an approximate fit for the empirical data using a Poisson distribution. The Poisson distribution captures reasonably well the symmetry of the distribution, although it underestimates the tails as well as the overall variance of the distribution. A Chi-square test is applied to test the goodness of the fit. To verify the fit, the data are first arranged into an equiprobable histogram with variable cell sizes [6]. The Chi-square test statistic is then calculated at 8.168. The corresponding p-value is 0.417.

This model provides a first approximation for the empirical coverage density in an operational cellular data network. This model can be used in subsequent simulation models of the network topology.

4. Conclusions

This paper uses fine-grain measurements from an operational CDMA2000 cellular network to characterize wireless Internet data traffic. Our study uses one week of trace data from the network, representing 171,318 packet data calls from 4,156 users at 139 cell sites.

Our results illustrate several characteristics that differ from traditional Internet traffic assumptions. Our cellular network data traffic exhibits a diurnal pattern, with peak usage occurring in evening hours rather than during the business day. Statistical analysis of the results shows that the packet call arrival process is non-stationary, and burstier than a Poisson process.

Our results also identify power-law properties in several traffic characteristics. These include the distribution of packet call activity by user, as well as in the burst-level behaviour. These results are consistent with prior work demonstrating power-law characteristics in wired Internet traffic. Our results also provide some insight regarding user mobility characteristics. A novel analysis of soft handoff activity reveals the inherent structure in the cellular network coverage density.

Our ongoing work is pursuing two main directions. First, we are refining our analysis of user mobility, and developing user mobility models that are suitable for simulation studies [5]. Second, we are building application-layer traffic models (e.g., Web, WAP, email, SMS, video) to use in our capacity planning simulation studies. The knowledge and insights that we have gleaned from our traffic measurement study provide useful inputs to the traffic modeling process.

Acknowledgements

References

[1] A. Adya, P. Bahl, and L. Qiu, "Analyzing Browse Patterns of Mobile Clients", *Proceedings of ACM SIG-COMM Internet Measurement Workshop*, San Francisco, CA, pp. 189-194, November 2001.

[2] A. Adya, P. Bahl, and L. Qiu, "Characterizing Alert and Browse Services for Mobile Clients", *Proceedings of USENIX Technical Conference*, pp. 343-356, 2002.

[3] M. Crovella and M. Taqqu, "Estimating the Heavy Tail Index from Scaling Properties", *Methodology and Computing in Applied Probability*, Vol. 1, No. 1, 1999.

[4] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On Power-Law Relationships of the Internet Topology", *Proceedings of ACM SIGCOMM*, Cambridge, MA, pp. 251-262, September 1999.

[5] E. Halepovic and C. Williamson, "Characterizing and Modeling User Mobility in a Cellular Data Network", *Proceedings of 2nd ACM Workshop on Performance Evaluation of Wireless Ad Hoc, Sensor, and Ubiquitous Networks* (PE-WASUN 2005), Montreal, QC, October 2005. To appear.

[6] R. Jain, *The Art of Computer Systems Performance Analysis*, Wiley, 1991.

[7] W. Jeon and D. Jeong, "Call Admission Control for CDMA Mobile Communications Systems Supporting Multimedia Services", *IEEE Transactions on Wireless Communications*, Vol. 1, No. 4, pp. 649-659, October 2002.

[8] S. Keshav, "Why Cell Phones Will Dominate the Future Internet", *ACM Computer Communication Review*, Vol. 35, No. 2, pp. 83-86, April 2005.

[9] A. Klemm, C. Lindemann, and M. Lohmann, "Traffic Modeling and Characterization for UMTS Networks", *Proceedings of IEEE GLOBECOM*, Internet Performance Symposium, San Antonio, TX, November 2001.

[10] I. Koo, J. Ahn, J. Lee, and K. Kim, "Analysis of Erlang Capacity for the Multimedia DS-CDMA Systems", *IEICE Transactions on Fundamentals*, Vol. E82-A, No. 5, pp. 849-855, 1999.

[11] T. Kunz, T. Barry, X. Zhou, J. Black, and H. Mahoney, "WAP Traffic: Description and Comparison to WWW Traffic", *Proceedings of ACM MSWiM*, Boston, MA, pp. 11-19, August 2000.

[12] A. Lakhina, J. Byers, M. Crovella, and P. Xie, "Sampling Biases in IP Topology Measurements", *Proceedings of IEEE INFOCOM*, 2003.

[13] V. Paxson and S. Floyd, "Wide Area Traffic: The Failure of Poisson Modeling", *IEEE/ACM Transactions on Networking*, Vol. 3, No. 3, pp. 226-244, 1994.

[14] N. Shankaranarayanan, A. Rastogi, and Z. Jiang, "Performance of a Wireless Data Network with Mixed Interactive User Workloads", *Proceedings of IEEE International Conference on Communications (ICC)*, New York, NY, April 2002.

[15] D. Sharp, N. Cackov, N. Laskovic, Q. Shao, and L. Trajkovic, "Analysis of Public Safety Traffic on Trunked Land Mobile Radio Systems" *IEEE Journal on Selected Areas in Communication*, Vol. 22, No. 7, pp. 1197-1205, September 2004.

[16] D. Tang and M. Baker, "Analysis of a Metropolitan Area Wireless Network", *Proceedings of ACM MOBICOM*, Seattle, WA, pp. 13-23, August 1999.

[17] T. Varga, B. Haverkamp, and B. Sanders, "Analysis and Modeling of WAP Traffic in GPRS Networks", *Proceedings of ITCSS*, Antwerp, Belgium, 2004.

[18] B. Vujicic, N. Cackov, S. Vujicic, and L. Trajkovic, "Modeling and Characterization of Traffic in Public Safety Wireless Networks", *Proceedings of SCS International Symposium on the Performance Evaluation of Computer and Telecommunication Systems* (SPECTS), Philadelphia, PA, pp. 214-223, July 2005.