A COMPARATIVE MEASUREMENT STUDY OF THE WORKLOAD OF WIRELESS ACCESS POINTS IN CAMPUS NETWORKS

Félix Hernández-Campos [a] Maria Papadopouli [a,b]

a. Department of Computer Science, University of North Carolina at Chapel Hill, USA.
b. ICS-FORTH, Greece.
Emails: {fhernand,maria}@cs.unc.edu

Abstract— **Our goal is to perform a system-wide characterization of the workload of wireless access points (APs) in a production 802.11 infrastructure. The key issues of this study are the characterization of the traffic at each access point (AP), its modeling, and a comparison among APs of different wireless campus-wide infrastructures. Unlike most other studies, we compare two networks using similar data acquisition techniques and analysis methods. This makes the results more generally applicable. We analyzed the aggregate traffic load of APs and found that the log normality is prevalent. The distributions of the wireless received and sent traffic load for these infrastructures are similar. Furthermore, we discovered a dichotomy of APs: there are APs with the majority of clients that are uploaders and APs in which the majority of their clients are downloaders. Also, the number of non-unicast wireless packets and the percentage of roaming events is large. Finally, there is a correlation between the number of associations and traffic load in the log-log scale.**

I. Introduction

Wireless networks are increasingly being deployed and expanded in airports, universities, corporations, hospitals, residential, and other public areas to provide wireless Internet access. It is interesting to observe its evolution both in the spatial and temporal domain. While there is a rich literature characterizing traffic in wired networks, there are only a few studies available that examined and modeled wireless traffic load. Furthermore, there is no study that compares the different wireless infrastructures to generalize the models and characteristics of the traffic load. Access points (APs) are a critical element of the wireless infrastructure in campus network. The key issues of this study are the characterization of the traffic at each AP, its modeling, and a comparison among APs of two different wireless campus-wide infrastructures. Unlike most other studies, we compare two networks using similar data acquisition techniques and analysis methods. This makes the result more generally applicable.

In this paper, we study two large wireless infrastructures of the University of North Carolina at Chapel Hill (UNC) and Dartmouth College using a lightweight data acquisition methodology. The data was collected using the Simple Network Management Protocol (SNMP), the most widely available monitoring service in wireless platforms. Any AP in the market supports monitoring using SNMP, so it is important to understand how much operators and researchers can learn from SNMP data. Other types of data, such as packet or flow level data, are generally too detailed for this purpose, and their acquisition is much more resource-intensive. This paper makes use of SNMP data for analyzing traffic characteristics, such as the total number of bytes and packets that each access points sent and received during the monitoring period.

Our study considers three dimensions of the workload an AP: number of bytes sent and received, number of packets sent and received, and number of associations and roaming operations. In addition, we have also consider how the building types (*e.g.*, academic, residential, *etc.*) affect the characteristics of the AP workloads. While previous works have partially considered some of these aspects, our focus on access points is rather unique. Furthermore, we performed system-wide characterization, rather than focusing only the most utilized areas of the studied networks. We believe this type of analysis provides a useful view of the entire utilization of a wireless network, at least from the point of the access points that form the backbone of the wireless infrastructure.

In general, we found a surprising degree of similarity in the characteristics of the UNC and Dartmouth wireless networks. Our results therefore provide strong evidence in support of the development of parsimonious workload models of campus wireless networks. This type of modeling would make it possible to develop more realistic simulations and testbeds. The extended version of this paper [3] provides more details on our analysis, and a brief overview of the related work.

II. Data Acquisition

The data comes from the large campus wireless networks deployed at UNC and Dartmouth. UNC's wireless network uses 488 APs to provide coverage for 729-acre campus and a number of off-campus administrative offices. The university has 26,000 students, 3,000 faculty members, and 9,000 staff members. Dartmouth's network serves 190 buildings in a 200-acre campus. The university population includes 5,500 students and 1,200 faculty members. Personal laptops are

978-3-8007-2909-8/05/$20.00 ©2005 IEEE

required for undergraduates in both institutions, and almost all of them are equipped with a wireless interface.

The data in this paper was collected using SNMP for polling every AP on campus every five minute. We collected the UNC trace using a custom data collection system, being careful to avoid the pitfalls described in [2]. The system was implemented using a non-blocking SNMP library for polling each AP precisely every five minutes in an independent manner. This eliminates any extra delays due to the slow processing of SNMP polls by some of the slower APs. The UNC trace was collected between 9:09 AM, September 29th, 2004 and 12 AM, November 25th, 2004. The monitoring system did not suffer any problems during this period.

The Dartmouth trace corresponds to the most recent dataset studied in [2]. It was collected between November 1st, 2003, and February 28, 2004. The data was acquired using a similar approach, although the data collection system has some shortcomings that are described in section III-A. This trace includes 6,875 unique MAC addresses which were associated with one or more APs during the data collection period. This number is larger for the UNC trace, which reports on the activity of 14,712 unique MAC address. In summary, while the number of access points in both networks is similar, the population of clients is more than twice larger for UNC.

III. Analysis

A. *Traffic Load*

The first dimension of the workload of the wireless infrastructure that we examine in this study is the total traffic load in terms of bytes. These data were obtained from two cumulative counters, bytes received and bytes sent. These counters are encoded using only 32 bits, so wrap-arounds are frequent and they must be properly handled to reconstruct cumulative values above 2^{32} bytes. While this task seems straightforward, there are several pitfalls that can make the results completely bogus. Our careful analysis handles AP reboots, buggy firmware, irregular sampling for Dartmouth, and spurious resets (see [3] for further details).

Figures 1 and 2 provide an overview of the total traffic loads in the Dartmouth and UNC infrastructures. Since we are interested in studying the heterogeneity in the load of different APs due to their different uses, we consider here only those APs that remained operational for the majority of the tracing period. This means that the number of APs studied was 499 for Dartmouth (out of 557 present in the trace), and 447 for UNC (out of 488). The left scatter-plots in the two figures shows one symbol for each AP, comparing the total number of bytes that each access point received from its clients (x-axis) to the total number of bytes that it sent (y-axis). The plots illustrate the wide range of loads in the infrastructure. Some APs had extremely light loads (a few MBs during months of operations) while others were used much more heavily (hundreds of GBs). This is consistent for both campus networks. We can also observe in both plots a clear linear trend with a positive slope. This shows that while byte loads were generally asymmetric, no extreme cases were present.

The scatter-plots in Figures 1 and 2 use different symbols for the APs located in different types of buildings. The classification is based on main purpose of the building (*e.g.*, residential buildings are inhabited year-round by students). For both datasets, the majority of APs were located in academic or residential buildings. The left scatter-plots of Figure 1 and 2 show a wide range of traffic loads for each type of building. In both cases, the types of the most loaded APs were residential or library. Further analysis of this data using cumulative distribution functions (CDFs) revealed that the tails of the distributions of loads exhibit stochastic ordering.

The middle and right plots in Figure 1 and 2 show an interesting finding. These plots use the y-axis for the ratio of bytes sent to bytes received. This quantity characterizes the symmetry in the load of APs. The smaller the ratio, the more dominated the load of the APs was by data sent from its own clients. As the two middle plots illustrate, there is a clear downward trend when the ratio of sent to receive bytes is plotted against the total number of received bytes. This implies that APs with more bytes sent from their clients tend to send less data to them. We can say that these access points are dominates by *uploaders*, *i.e.,* clients that mostly serve data rather than download it from the Internet. We conjecture that this is due to peer-to-peer applications, which are fairly popular in both campus networks. The building type breakdown reveals some structure, although there is again a wide variety among the buildings of each type. Most social and dining buildings had a ratio above one, so uploading behavior was not very significant in them. Residential buildings account for a large fraction of the buildings with a ratio below 1.

The right plots in Figures 1 and 2 do not show the prominent linear trend found in the middle plots. This is most clear for the Dartmouth data. Rather than a positive trend, which would indicate that the APs with more bytes sent were dominated by downloaders, we find no trend (Dartmouth) or a slight downward trend (UNC). Since most clients are downloaders, this indicated that the total number of bytes sent by an AP increased as the number of clients increased, rather than as the clients become heavier downloaders. This is sharp contrast to the structure found for total received bytes, which increased as the APs became more dominated by uploaders.

The left plot of Figure 3 shows the CDFs, $Pr\{X \leq x\}$, for sent and received bytes. Dartmouth is heavier in both cases, as one would expect from a longer monitoring period. We also found that in general the distribution of AP aggregate loads can be well-approximated with a log-normal model. This is illustrated in the middle and left plots of Figure 3. These *quantile-quantile plots* compare the quantiles of an empirical distribution with the theoretical quantiles of a fitted model. Quantiles are represented using dots, with their theoretical value as the x-coordinate and their empirical value as their y-coordinate. In addition, the solid 45-degree line flanked by dashed confidence interval curves represents the area where

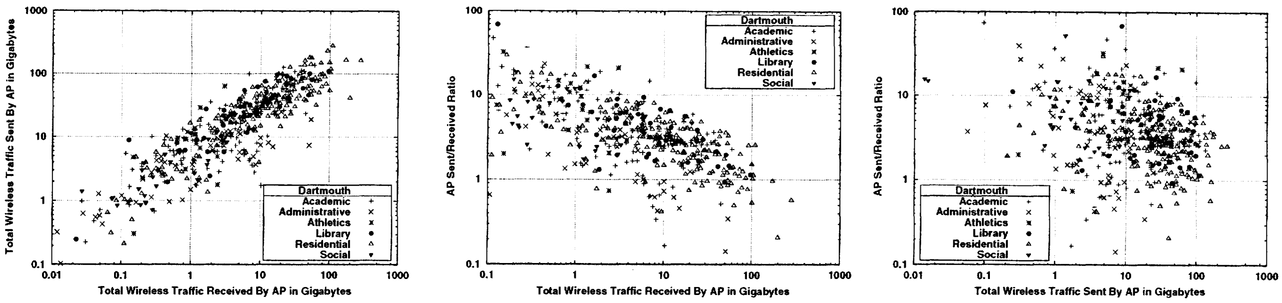

Fig. 1. Total traffic load in Dartmouth trace.

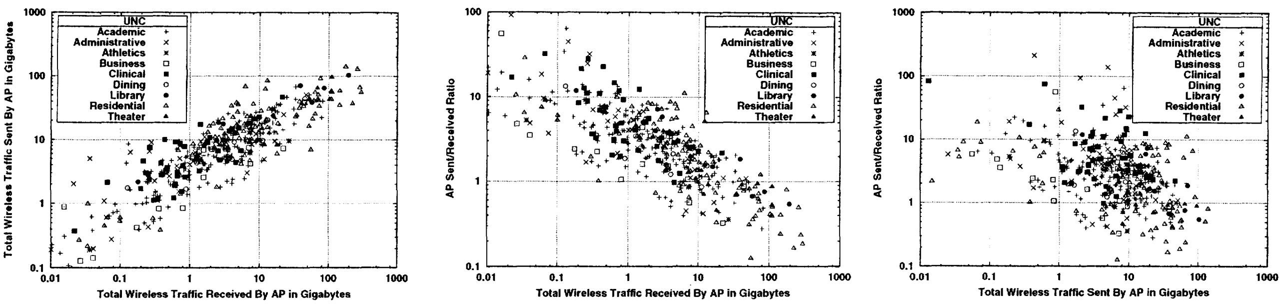

Fig. 2. Total traffic load in UNC trace.

the quantiles should appear in the case of a good fit. The middle plots, which correspond to the distribution of the total number of bytes sent by UNC APs, show an excellent match with the log-normal model. Note the theoretical quantiles correspond to those of a normal distribution, since the data has been transformed by taking the logarithm. Therefore, the system-wide distribution of sent bytes can be modeled accurately with a log-normal distribution. This is also true for the Dartmouth APs. On the contrary, the fit is not as good for received bytes. The highest quantiles show a systematic deviation above the upper confidence interval curve. This suggests that the empirical distribution has a tail that is significantly heavier than that of the fitted log normal model. Given the linearity of this deviation, we believe that a bimodal fit (*e.g.,* two log-normals with different parameters) would provide a good fit This is also true for the Dartmouth received bytes distribution. This finding is consistent with the results in Figures 1 and 2, which show two types of APs workloads (uploader vs. downloader dominated). Uploader dominated APs made the tails of the received bytes distributions heavier than captured by log-normal models. It is important to note that this conclusions cannot be extrapolated from the log-normality of flow sizes observed in previous studies [4] (except, of course, in extreme cases of APs with only one flow in total).

The previous plots examined the total load of the APs over the tracing period. It is also interesting to study the load during shorter intervals of time. Our SNMP data was sampled every five minutes, so it seemed natural to study AP loads using such interval of time. This analysis required to handle two types of irregularities in the SNMP polling rate, gaps and late polls, as described in [3].

The CDFs shown in the left plot of Figure 4 study the load of the APs during 5-minute intervals. This plot shows that the distributions of sent bytes are surprisingly similar for both networks. APs sent between 500 KB and 2 MB during most active intervals. In contrast, the distributions of bytes received from clients are quite different. The distribution for Dartmouth has a very light body, with less than 100 KB sent in 75% of the intervals. The UNC distribution shows two different regions: a very light one with 40% of the intervals; and a heavier one which is quite close to the distribution of sent bytes. This difference may seem at odds with the similarity between the two sites in Figures 1 and 2. However, the total bytes sent shown in those figures are dominated by the largest values. As the middle plot in the Figure shows, the distributions of sent bytes for Dartmouth and UNC are similar when we consider the probability per byte rather than per interval. While the number of intervals with very small loads was large, their impact was small in terms of the total bytes (*e.g.,* only 15-20% of the bytes came from the lower 95% of the intervals). Finally, the tails of these distributions is shown in the right plot of Figure 4 using a complementary cumulative distribution function, $Pr\{X > x\}$. The plot shows that the most utilized intervals follow similar distributions for both sent and received bytes. The sharp decrease around 1 GB is not surprising given the bandwidth limits in 802.11. The similarity, and linearity, of the curves between 1 and 100 MBs is more remarkable.

B. Packet Load

A scatter-plot of the total number of unicast packets observed in UNC APs are shown in the left plot of Figure 5. The plots for Dartmouth is very similar and is not included here. The range of packets loads is very variable, with APs

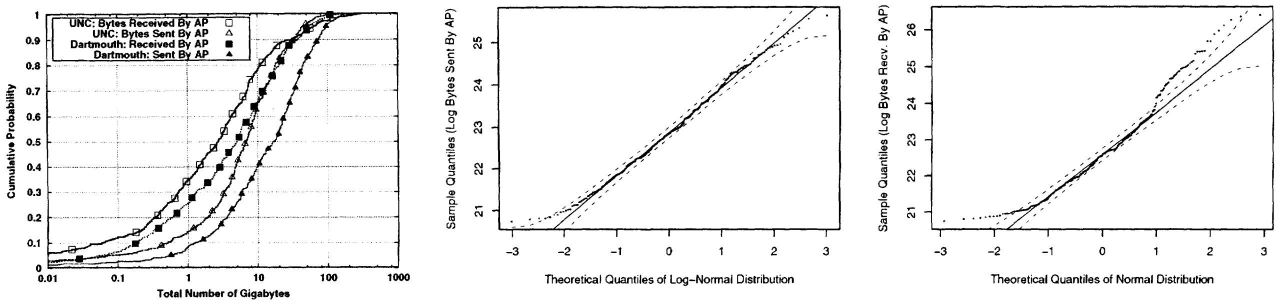

Fig. 3. CDFs for total traffic loads (left), and Q-Q plots for the distributions of total traffic sent (middle) and received (left) by UNC APs.

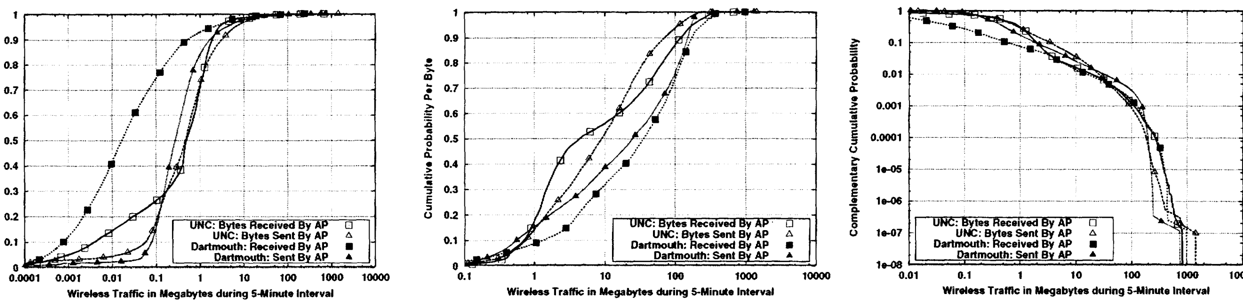

Fig. 4. Three different visualizations of the distributions 5-minute traffic loads for UNC and Dartmouth.

that handled only a few tends of thousands of packets during months of operation, while others handled tends of millions of packets. It is interesting to note that there is a far stronger correlation between the numbers of sent and received packets that between the corresponding numbers of bytes (see Figures 1 and 2). The byte plots showed a large variability around the linear trend which was due to the variable degree to which APs were dominated by more or less asymmetric communications (*e.g.*, lighter or heavier downloaders). On the contrary, when the total number of packets is examined, this heterogeneity is ameliorated by the ubiquitous use of TCP, which requires at least one acknowledgment packets for every two data packets. This severely limits the degree of asymmetry in packet load, and this observation can have implications for the design of APs and the improvement of the 802.11 protocols.

The middle plot in Figure 5 shows packets load for UNC access points due to non-unicast packets. The numbers are rather large, suggesting that this type of packet are quite significant in this type of infrastructures. Unfortunately, no similar data was collected for Dartmouth. Our preliminary analysis shows that these non-unicast packets mostly came from Net-Bios and ARP.

The right plot shows the average packet sizes for the APs in the UNC campus. Most average packets sizes are sizes are below 600 bytes, although we observe some degree of *axis-hugging*, where a small average received packet size corresponds to a large average sent packet size. This confirms the previous observation that some APs are dominated by highly asymmetric client behaviors (mostly uploading or mostly downloading), which sent mostly large (data) packets in one direction and mostly small (acknowledgment) packets in the other direction. Note that this plot was created by dividing the total number of bytes by the total number of packets (both unicast and non-unicast). A similar plot for Dartmouth using total bytes divided by total unicast packets results in average packet sizes well above the Maximum Transfer Unit of 802.11b, so it is clear that the number of non-unicast packets was also rather significant for Dartmouth.

C. Client Associations

Client association dynamics represent another important aspect of the workload of wireless APs. The SNMP counters in the Dartmouth and UNC trace include cumulative counts of the number of client that associated with each APs, number of clients that roamed into each access points, and number of clients that roamed away from each AP. The left plot in Figure 6 shows the CDF for these three parameters for UNC. As in previous cases, we find distributions that are quite heavy. One third of the APs received between 10,000 associations and 250,000 association during the monitored period. This represents daily averages between 170 and 4,310, which should be considered rather high. In addition, the distributions for roaming operations are also rather heavy. It is unlikely that this is purely due to client mobility, so we conjecture that associations instabilities are common, *i.e.*, clients that roamed to a nearby APs due to interferences and poor signal strength.

Previous studies [1] observed no correlation between the number of associations and the total traffic per AP. We have also studied this question and found only weak correlations for the Dartmouth and UNC traces (Pearson's correlation of 0.30 and 0.41 respectively). The middle plot of Figure 6 illustrates the absence of any clear trend in the Dartmouth

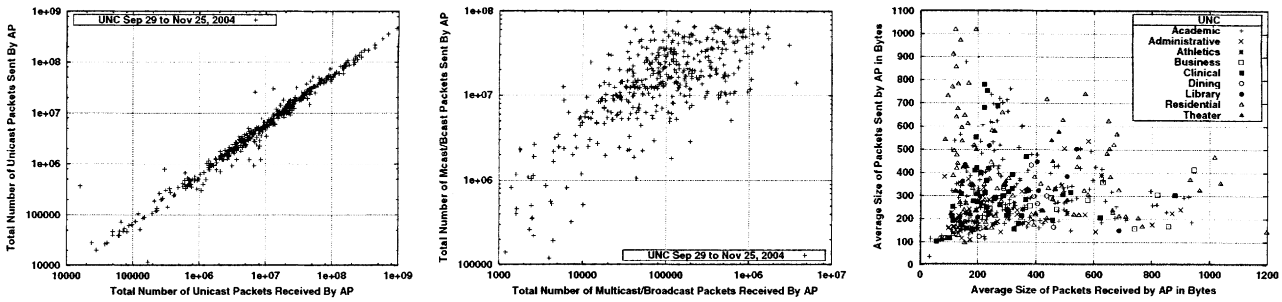

Fig. 5. Unicast (left) and non-unicast packets loads for UNC APs, and average packet sizes for UNC APs.

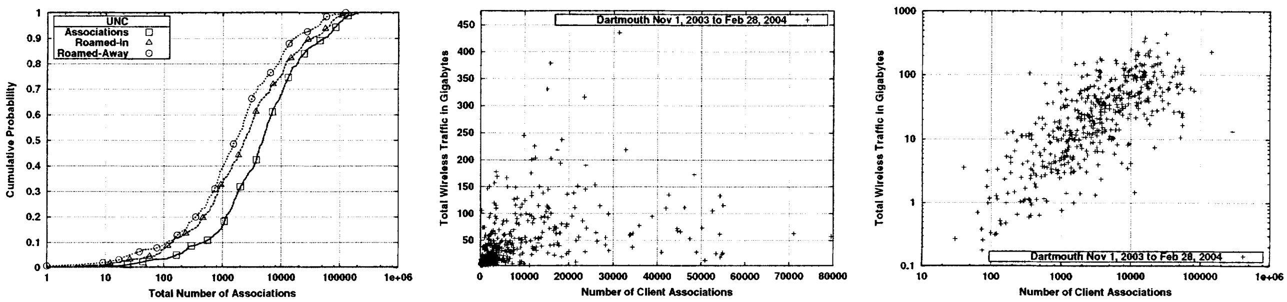

Fig. 6. Dist. of client associations for UNC APs (left) and correlation between load and number of associations for Dartmouth APs (middle and right).

data. However, when the data is examined in a log-log scale as in the right plot, it is clear that a linear upward trend is present. The correlation between the logarithm of the total bytes and the logarithm of the number of associations was 0.80 for Dartmouth and 0.74 for UNC. This means that $log_{10}t \approx a \times log_{10}n + b$ where t is the total traffic and n is the number of associations. Therefore $t \approx 10^b n^a$, and the total amount of traffic grows very quickly with the total number of associations. In our further analysis, we have plotted (not shown here) the total traffic against the average association size, and found a upward trend. This implies that higher loads in APs with more associations are not only due a greater number of associations but also to "bigger" associations, *i.e.*, associations that transfer a greater number of bytes in average.

IV. Conclusions and Future Work

We have studied the workload of APs in large campus networks. Our results show substantial similarities between these two environments, and open numerous avenues for further research. Our analysis of the load in bytes reveal interesting structure due to heavy uploading behavior. We intend to clarify the cause of this phenomenon by collecting and analyzing packet headers traces. This type of data would also help us to further analyze the causes of the shapes of the load distributions, and the origin of the large fraction multicast packets. Our finding of pervasive log-normality in the system-wide load of the two networks is intriguing, and we intend to study the generative process more carefully and propose formal parametric models for the different characteristics. We also intent to study its applicability to other periods (*e.g.*, daily loads).

We found surprisingly heavy distributions of total client associations and roaming operations. We are currently analyzing an alternative source of data, syslog event messages, which should help us to estimate the fraction of these events that comes from mobility rather than infrastructural problems. Additional SNMP data that reports client information should also prove useful.

We intend to study the spatial correlations of APs and classify APs based on various parameters (e.g., traffic characteristics, number of associations, and distinct clients). Furthermore, we aim to explore the topological properties of the wireless network infrastructure. Another ongoing effort focuses on forecasting of the traffic load at APs at various time scales.

V. Acknowledgments

We are grateful to David Kotz and his team for sharing the Dartmouth SNMP trace with us. This work was partially supported by the IBM Corporation under an IBM Faculty Award 2003/2004 grant.

References

[1] M. Balazinska and P. Castro. Characterizing mobility and network usage in a corporate wireless local-area network. In *Proc. of MobiSys*, May 2003.

[2] T. Henderson, D. Kotz, and I. Abyzov. The changing usage of a mature campuswide wireless network. In *Proc. of MobiCom*, Philadelphia, September 2004.

[3] F. Hernández-Campos and M. Papadopouli. A comparative measurement study of the workload of wireless access points in campus networks. Technical Report 353, ICS-FORTH, May 2005. ftp://ftp.ics.forth.gr/tech-reports/2005/2005.TR353_Workload_Wireless_Access_Points.pdf.

[4] X. Meng, S. H. Y. Wong, Y. Yuan, and S. Lu. Characterizing flows in large wireless data networks. In *Proc. of MobiCom*, pages 174–186, Philadelpia, 2004.