Application-based Characterization of the Traffic of a Campus-wide Wireless Network

Manolis Ploumidis[1,2], Maria Papadopouli[1,2], and Thomas Karagiannis[3]

[1] Institute of Computer Science, Foundation for Research and Technology-Hellas
[2] Department of Computer Science, University of Crete
[3] Microsoft Research

Abstract. With the increasing deployment of WLANs, network management and configuration of wireless Access Points (APs) has become one of the main concerns of network operators. While statistics and measurements regarding the overall usage of individual APs are readily available, the limited knowledge of the wireless traffic demand in terms of application types hinders efficient network provisioning. In this paper, we provide an extensive application characterization of a large scale campus-wide wireless LAN going beyond the port-number limitation across three dimensions. Specifically, we examine the application cross-section at the network, client and AP level. We find that the most popular application types both in terms of the number of flows, bytes and clients are web and peer-to-peer. Further, we show that while the majority of the APs is dominated by these two applications, APs that belong to the same building types appear to have large differences in their application traffic mix. Finally, we observe that clients use the wireless network mostly for one specific application, but the traffic share of each application is significantly affected as users move within the wireless network.

1 Introduction

Wireless local area networks (WLANs) are increasingly being deployed to address the growing demand for wireless access. As their user population increases, accurate traffic modeling and characterization of their workload are essential to facilitate efficient network management and better utilization of their scarce resources.

While there have been several studies looking at the application cross-section at wired networks (e.g., [11]), such attempts are limited in the case of wireless networks [3]. Besides the well-documented limitations of application identification [7], inherent additional complications in wireless networks, such as the increasing overheads of data collection due to the need of multiple monitoring points, cross-correlation of different type of traces, and transient phenomena due to the radio propagation and mobility, have led the community to assume that the expected workload of wireless networks follows the general trends of Internet applications.

This paper provides a multi-level application-based traffic characterization of a campus-wide wireless infrastructure from the perspective of the network, client, and AP. Such a characterization is crucial to our understanding of the application usage in order to build user profiles and develop better resource management and admission control mechanisms.

To avoid the "known-port limitation" (i.e., port-based classification into applications) [10] [6], we employed the *BLINC* tool [7] which performs classification of flows into applications based on the transport-layer footprint of the various application types. To identify AP- and user-specific characteristics, the flow-related statistics were coupled with syslog traces. Our findings include:

- The most popular applications are web browsing and peer-to-peer (P2P) accounting approximately for 60% of the total traffic. Most user profiles are also dominated by these two applications.
- While building-aggregated traffic application usage patterns appear similar, the application cross-section within the same building varies with the AP.
- Most wireless clients appear to use the wireless network for one specific application that dominates their traffic share.
- File transfer flows, such as FTP and P2P are heavier in the wired network than in the wireless one.
- The traffic share across applications is significantly affected when clients associate to new APs. This observation appears to be independent of the specific application type.

The rest of the paper is organized as follows. Section 2 presents related work while Section 3 describes the analyzed traces and the methodology followed. Sections 4,5 and 6 present our findings for the aggregate traffic, the AP and the client characterization. Finally, we conclude and discuss the implications of our findings in Section 7.

2 Related work

While there have been several measurement attempts in deployed wireless networks over the past few years, the focus has mostly been in identifying human behaviors in terms of mobility patterns [9] [2] [1] [12] with the goal of provisioning the wireless APs. These studies reflect a variety of wireless environments, such as campuses, auditoriums or enterprise networks. While our work touches upon the workload of APs, we are interested in studying the application cross-section across APs rather than the overall traffic.

Closer to our work, the authors in [3] [1] [12] briefly describe the breakdown of traffic into various application classes. These studies are however based on port-based application classification which has been shown to be highly inaccurate [10] [6] due to the usage of random port-numbers from the majority of emerging applications. Further, our work goes beyond simple application breakdown in our traces, by examining the variation of the application cross-section across APs and clients.

Set	Date	Start	Dur	Src.IP	Dst.IP	Packets	Bytes	Aver.Util.	Aver. Flows.
UNC (WLAN)	2005-04-13	12:00	178.2 h	2650 K	2454 K	1046 M	583 G	7.3 Mbps	25.8 K
UNC (WIRED)	2005-04-13	12:00	178.2 h	1998 K	1933 K	1022 M	521 G	6.5 Mbps	19.5 K

Table 1. General workload dimensions of our packet trace.

3 Data description

We analyzed traces from a large campus wireless network deployed at the University of North Carolina [13]. UNC's network provides 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.

Out of the total 488 APs that are part of the campus we monitor 382 of these in this study, belonging in 231 different buildings. These APs belong to three different series of the Cisco Aironet platform,namely, the state-of-the-art 1200 Series (269 APs), the widely deployed 350 Series (188 APs) and the older 340 Series (31 APs). The 1200s and 350s ran Cisco IOS while the 340s ran VxWorks.

Our traces consist of packet traces collected at one of the access routers at UNC and SYSLOG data from all APs. Syslog data were used to add MAC address and AP info to the flow-level data. In our study we observe approximately 9,125 distinct internal IPs which were mapped to approximately 2350 unique MAC addresses. Using our syslog traces, we were also able to map 74.6% of the packet level flows to specific MAC addresses and APs. Note that not all APs in the campus were monitored, hence the unmapped flows.

The total traffic that was correlated corresponds to 77.6% of the total traffic in the packet traces. Table 1 presents general dimensions of 7.5-day packet traces such as the number of source and destination IPs observed, the total number of packets and bytes as well as the average utilization observed for both the wired and the wireless sources found in our traces. We will occasionally use the latter data to compare application and client behavior across the wired and the WLAN.

4 Aggregate Traffic Characteristics

This section presents an overall application characterization of our traces at the wireless network level. First, we used the BLINC tool [7] to classify flows into applications. BLINC was able to classify approximately 87% of the flows. Then, we identified the dominant application types both in terms of the total number of bytes, flows, and the popularity among clients. To the best of our knowledge, this is the first studies looking at the application cross-section of a large campus-wide wireless network going beyond the "known-port" limitation. Finally, we compared our findings with previous studies both in wired and wireless environments.

4.1 Dominant application types

Although Web traffic has been traditionally prevalent, recent studies have highlighted an increasing interest in P2P applications. We identified the dominant applications in our traces by looking at byte, packet and flow statistics. The corresponding percentages are summarized in Table 2.

Application type	Flows (%)	Bytes (%)	Packets (%)	Clients (%)
NETWORK MANAGEMENT	9.84	0.35	1.40	34
CHAT	2.03	0.405	1.33	95
WEB	36.3	64.22	51.58	100
P2P	29.7	20.97	31.37	69
ON-LINE GAMES	1.10	0.01	0.068	8
FTP	0.90	1.32	1.56	19
MAIL	0.07	0.278	0.19	4
ADDRESS SCAN	6.33	0.107	0.53	95
PORT SCAN	0.39	0.270	0.26	3
STREAMING	0.10	0.149	0.178	1
UNKNOWN	13.0	11.89	11.50	97

Table 2. Application cross-section with respect to flows, bytes, packets and clients.

Table 2 shows that although less than half of the total flows are Web flows, it still accounts for nearly 65% of the traffic. P2P applications have a similar share in terms of flows that corresponds to roughly 21% of the total number of flows. While some applications may get a smaller share of the overall traffic, a significant percentage in terms of flows may have significant impact on APs (e.g., overhead in terms of state needed and delay in processing). For example, although the Network Management and scanning activity are responsible only for 0.35% and 0.1% of the bytes transfered, respectively, they amount collectively to approximately 17% of the total flows in the network! Thus, their effect in the overall network cannot be ignored.

4.2 Popularity of applications

Although byte and flow statistics reveal the dominant applications in terms of network traffic, they only indirectly hind on the popularity of each application. We define as *popularity* of an application the number of clients that had at least one flow for that specific application. Our findings are summarized at the last column of Table 2. Note that these percentages do not sum up to 100% since each client engages in more than one application.

Interestingly, all clients have at least one Web flow, while a significant percentage of them (68.7%) appears to be P2P users. Further, 8% of them use online games in our wireless network while almost each one had at least one unclassified flow. While mail traffic is only observed for a small number of our clients, this is the effect of our monitoring point at the edge of the network (i.e., internal traffic can not be monitored). Finally, note that almost all clients are scanned at least once by an attack flow!

4.3 Comparative study with other wired and wireless networks

We contrast our findings with three other application-based characterization studies in wireless and wired networks, namely, two wired campus networks (BLINC, and UNC) and two wireless campus networks (UNC, and Dartmouth). Note that although a direct comparison is not straightforward due to the differences in the monitored networks, time of collection as well as the varying definition of application classes across studies, we can still observe general application trends.

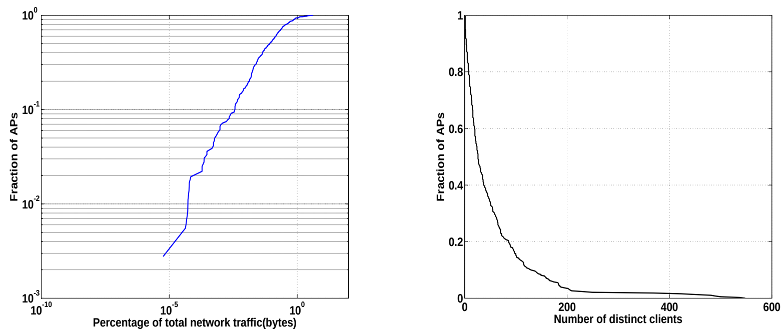

Fig. 1. LEFT: CDF of percentage of total network traffic across APs. RIGHT: CCDF of the number of distinct clients across APs.

To this end, we first compared the traffic share of the most dominant and popular applications (i.e., Web and P2P) of the UNC wireless network with the the share of the same applications captured at the wired component of the UNC network (i.e., traffic originating from wired clients) within the same time interval. Similarly, we contrast our findings to the BLINC campus trace studied in the original BLINC work [7], and the findings from the Dartmouth wireless network [3]. Table 3 summarizes the percentages of Web and P2P traffic for each one of these networks. Note that the BLINC campus trace, and the two UNC traces were all classified by BLINC, hence the findings can be directly comparable; port numbers were used in the case of the Dartmouth trace.

	BLINC wired	UNC Wired	UNC Wireless	Dartmouth Wireless
WEB	37.5%	48.68%	64.22%	28.6%
P2P	31.9%	34.85%	20.97%	19.3%

Table 3. Percentage of Web and P2P traffic(bytes) across four different networks

The results are remarkably similar for the two "wired" traces (BLINC and UNC campus networks), especially in the P2P case. On the contrary, the share of P2P traffic is significantly lower in both wireless traces amounting to approximately one fifth of the total traffic overall. The most significant difference across the wireless traces is the share of Web traffic which is roughly double in our data. While such a difference may simply reflect different usage patterns across the two wireless networks, it could also be attributed to the large percentage of unknown traffic in the Dartmouth trace.

5 AP characterization

This section focuses on application mix across APs and examined examined a general application-based characterization at the AP- and building- levels. As observed in previous studies, the overall distribution of traffic across APs is not uniform. Few APs are responsible for the largest amount of traffic. Figure 1 (left) presents the CDF of traffic across APs. Most APs transfer small amounts of traffic. For example, 3% of the APs are responsible for nearly 30% of the traffic.

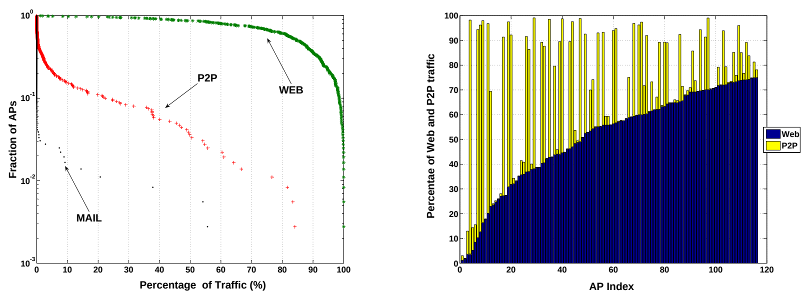

Fig. 2. LEFT: CCDF of the percentage of traffic across the various APs. Half of the APs are dominated by Web traffic (>70%), while approximately 10% of APs by P2P traffic. RIGHT: The share of P2P and Web across APs.

The reason for such a skewed distribution of traffic among APs is the varying popularity among APs. To validate this assumption, we plot the CCDF of the number of clients across APs in Figure 1 (right). This figure reveals that few APs are significantly more popular than others having been visited by more than 200 clients. As expected, examining these APs reveals that they are the ones with the highest traffic aggregation.

The distribution of traffic across APs varies with the application. We examined if the share of each application is similar across APs or whether specific applications dominate particular APs. Figure 2 (left) presents the CCDF of the traffic share of three distinct application types relative to the total traffic across APs. While high percentages of Web traffic appear in most of the APs, a small portion of them, roughly 10%, is dominated by P2P traffic. Other application types rarely dominate any of the APs with the exception of three APs being dominated by mail traffic and one by FTP. This unevenness is especially pronounced in the case of P2P. Figure 2 (right) shows the combined percentage of Web and P2P traffic across APs shorted by their Web traffic percentage. There are two modes. Specifically, either most APs are dominated by P2P traffic, or their share of P2Ptraffic is minimal. Note that while previous studies have found Web vastly dominating AP traffic, this is not the case for the majority of our monitored APs in our data.

p	WEB (%)	P2P (%)	FTP (%)	MAIL (%)	Unknown (%)
50	86.6	4.5	0.5	0.3	4.2

Table 4. Percentage of APs with a home application.

To study in more detail what application types dominate the traffic of APs, we define the *h*ome application of an AP, to be the application that represents more than p% of its traffic. Table 4 presents the breakdown of APs that have such a home application when $p = 50$, i.e., half of the traffic of the AP belongs to an application type. Note that approximately 4% of the APs are dominated by unknown traffic, while for roughly another 4%, a home application could not be defined since no application type is dominating this AP's traffic. Overall, there is an application preference towards specific APs in the wireless network.

Building Type	APs	WEB (%)	P2P(%)	FTP(%)	MAIL(%)	UNKN(%)
ACADEMIC	161	89	4	-	1.8	0.6
ADMINISTRATIVE	36	86	-	-	-	13
CLINICAL	22	77	14	-	-	9
ATHLETIC	12	50	-	-	-	33
RESIDENTIAL	46	96	-	-	-	2
BUSINESS	17	65	-	5	-	29

Table 5. APs per building category and the percentage of APs with a home application.

This is an important observation since it can direct traffic engineering decisions such as load balancing or filtering P2P traffic filtering at certain locations of the network.

5.1 Application usage patterns across buildings types

To further examine the spatial variation of the application cross-section, we grouped APs based on their building category. Working at building level circumvents several problems emerging when working at AP-level: non amenability to statistical processing, higher sensitivity of monitored traffic variables to the short-term propagation conditions, lack of scalability [8]. These categories reflect buildings with similar functionalities and allow us to examine whether the share of the application depends on these functionalities.

To this end, a similar analysis is performed using the notion of home application as defined in previous section. Table 5 presents the number of APs for 6 building categories, and the percentage of APs for which a home application existed. There is a weak correlation of the building category with the number of APs that have a home application(e.g., mail exists only in the academic buildings as a *home application*, while ftp is present only in the business category). This reinforces our intuition that distinct APs may require different configuration settings depending on the application or the type of building functionality.

AP ID	44	45	46	47	48	49	50	51	52
WEB (%)	89	59	88	88	**37**	80	**10**	88	**20**
P2P (%)	0.3	**36**	2	0	**54**	0	**84**	2	**76**

Table 6. Web vs. P2P traffic share for building ID 22 across APs.

The uneven traffic distribution in the application cross-section that was observed across buildings exists also across APs of the same building. Table 6 presents the percentage of Web and P2P traffic for all APs located in building 22. This building was chosen randomly among the buildings with the largest number of APs. While in most cases Web traffic dominates the overall traffic share, there are distinct APs (highlighted in the table) that show the exact opposite behavior with P2P traffic being the dominant application. Note that these statistics are not due to transient traffic phenomena, since our tracing period corresponds to several days.

6 Client characterization

The application-based characterization from the AP- and building- levels was extended to client level. This section examines the user behavior in terms of application usage to gain a better understanding of the underlying application trends. Characterizing the client behavior is essential in designing more efficient AP admission control and selection mechanisms based on user profiles.

We define the *home application* of a client as the application that is responsible for more than x% of that client's traffic. We observe that wireless clients have strong application preferences. For example, for x equal to 90%, more than half of the wireless clients have a home application. Such a strong preference holds for both bytes and number of flows.

x	NM	Chat	Web	P2P	Games	Ftp	Mail	Strm	Ascan	Pscan	Unkn
50	0.12%	0.30%	88.67%	3.82%	0%	0.15%	0.21%	0.09%	0.03%	0.03%	5.27%
75	0.03%	0.18%	75.12%	1.54%	0%	0.12%	0.12%	0.09%	0%	0%	2%
90	0.03%	0.15%	50.46%	0.89%	0%	0.12%	0.06%	0%	0%	0%	0.64%

Table 7. Percentage of clients per *home application*

Table 7 indicates the percentage of wireless clients that have a specific application types a a home application with various thresholds. The most prominent home application is the Web, while P2P appears to be the home application for only a minority of the clients. It is also interesting how the traffic mix varies for clients without a home application. Even in this case most clients are still dominated by Web. The second largest share of their data is accessed either through P2P or through an undefined application.

6.1 Client and application behavior over wireless

While user preferences with respect to applications over the wireless network appear to have similar trends as to wired networks [11] [6] (i.e., Web and P2P dominate), it is unclear whether client behavior is affected by the application performance over the wireless channel. To shed some light on client behavior over the wireless network, we compare the characteristics of the dominant applications over the wired and the wireless networks.

Flow sizes appear smaller for bulk file-transfer applications over the wireless network. Figure 3 (left) presents the CCDF of flow sizes in bytes for the FTP, Web and P2P applications as seen by wired and wireless clients. While in the case of Web traffic the two curves fall exactly on top of each other, P2P and FTP flows appear "lighter" in size in the wireless network, especially for larger flow sizes. There are two potential explanations for this observation, one is application-dependent while the second user-driven. First, especially in the case of P2P applications, loss or disconnected TCP flows severely affect performance; broken TCP connections will result in disconnecting from existing peers, which will further trigger peer discovery mechanisms and increase queue waiting times, hence ultimately decreasing overall the flow sizes. As we observed in an earlier study [4], the large number of retransmissions at the 802.11 MAC layer, increases both the packet delay and number of retransmitted or failed packets at the transport layer. This is consistent with Figure 3 (middle) which compares the number of flows per client between Web and P2P in the wireless network. Note that while overall the Web flows are heavier in terms of bytes and also the number of transfered bytes per client [4], the number of flows per client is larger in P2P.

[4] The figure is not presented due to space limitations.

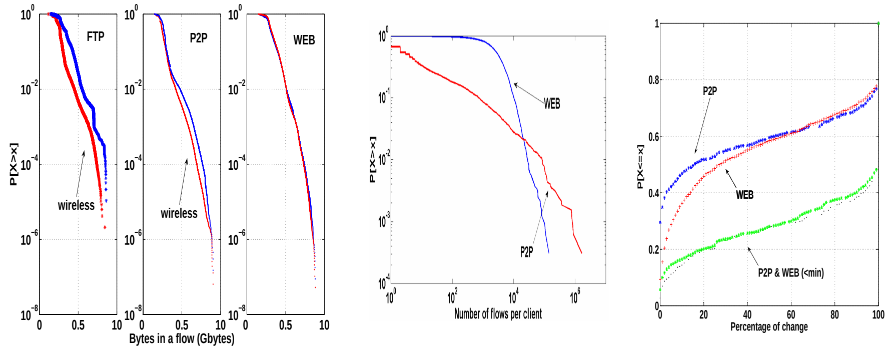

Fig. 3. LEFT: CCDF of flow sizes for FTP,P2P and Web. FTP and P2P flows appear heavier over the wired network. MIDDLE: Flows pers client for P2P and Web in the wireless network. RIGHT: CDF of the variability in the traffic share for P2P and Web when users associate to different APs.

Essentially, each P2P client appears to have a large overhead in the wireless network with numerous small flows, corresponding to control traffic. Second, users may avoid transferring large files over the wireless network because of the limited throughput, indicated also in Table 3 that compares the P2P traffic in wireless and wired. This observation holds for both FTP and P2P transfers.

Furthermore, mobility could provide another possible explanation for smaller flow sizes in P2P over the wireless network. As users associate to various APs, the performance of the more intensive applications such as P2P could be significantly affected. To test this hypothesis, we examined the differences in P2P and Web traffic shares across all clients as they move around the wireless network for a single weekday of our trace.

Figure 3 (right) presents the CDF of how the traffic share changes for P2P and Web as each client associates to multiple APs. The y-axis reflects the percentages of AP changes over all clients, while the x-axis represents the percentage of change in the overall traffic share for the specific application as a user associates to a new AP. The curves appear similar for Web and P2P applications and show for example that for 60% of the AP changes, the traffic share of the application will be affected by as much as 60%. Thus, it appears that associating to a new AP will significantly affect the overall traffic share of an application, and appears to be independent of the specific application. Note, that this effect may not be due to roaming but simply reflect user behavior or even traffic dynamics across APs. Yet, our results show an interesting observation that network operators should take into account.

To stress-test this hypothesis, we only looked at changes in the traffic share between APs where the overall association time in the previous AP was less than a minute. The two curves for Web and P2P are also presented in Figure 3 (right, bottom two lines), and fall on top of each other. Thus, even in very fast changes between APs, the effect appears not to depend on the application, but it is still significant.

7 Conclusions

This work provides a detailed three-level characterization of an operational campus-wide wireless network across APs, clusters of APs and clients. Our results can be employed to support better admission control and AP selection mechanisms, inform about the usage trends, and guide simulations with realistic traffic distributions for each application.

We found that while web traffic appears to dominate both the client and the AP share of the traffic, the substantial percentage of P2P applications bears a significant impact on the wireless network. P2P presents several undesirable side-effect for the wireless network, ultimately degrading its performance, such as: a) The always-active behavior of both the client machine and the AP (clients always have data to send and receive in such applications; b) the large number of flows as observed in the previous section that increase the contention at the AP which further leads to building up of AP queues and significant increase of the overall TCP RTT.

Furthermore, we showed that users have strong application preferences and mostly access the wireless network for one specific application. Finally, transitions from between APs impact the application usage. While it is unclear if such a behavior is user-driven, an application or a location side-effect, this observation is important for applications that are not delay and lag tolerant such as streaming, gaming and VoIP.

We intend to explore the impact of the wireless network and mobility on user-behavior and application performance. Furthermore, we will investigate how the underlying conditions of the MAC layer, such as packet delays and retransmission, affect application usage patterns. Extending our earlier work [5], we will explore the assymetry phenomena with respect to uploading and downloading behavior of APs and its application dependencies. We plan to use heuristics and statistical clustering techniques to profile clients based on their application characteristics and roaming patterns. Finally, we will perform a comparative study with traces from different wireless infrastructures and contrast the main user profiles and popularity. To encourage further experimentation along the lines drawn in this paper, we have made our datasets and tools available to the research community [13].

References

1. A. Balachandran, G. Voelker, P. Bahl, and P. Rangan. Characterizing User Behavior and Network Performance in a Public Wireless LAN, 2002.
2. Magdalena Balazinska and Paul Castro. Characterizing Mobility and Network Usage in a Corporate Wireless Local-Area Network. In *1st International Conference on Mobile Systems, Applications, and Services (MobiSys)*, San Francisco, CA, May 2003.
3. Tristan Henderson, David Kotz, and Ilya Abyzov. The changing usage of a mature campus-wide wireless network. In *MobiCom '04: Proceedings of the 10th annual international conference on Mobile computing and networking*, pages 187–201, 2004.
4. Felix Hernandez-Campos and Maria Papadopouli. Assessing the real impact of 802.11 wlans: A large-scale comparison of wired and wireless traffic. In *14th IEEE Workshop on Local and Metropolitan Area Networks*, Chania, Crete, Greece, 2005.

5. Felix Hernandez-Campos and Maria Papadopouli. A comparative measurement study of the workload of wireless access points in campus networks. In *16th Annual IEEE International Symposium on Personal Indoor and Mobile Radio Communications*, Berlin, Germany, 2005.
6. T. Karagiannis, A.Broido, M. Faloutsos, and kc claffy. Transport layer identification of P2P traffic. In *ACM/SIGCOMM IMC*, 2004.
7. T. Karagiannis, K. Papagiannaki, and M. Faloutsos. BLINC: Multi-level Traffic Classification in the Dark. In *ACM SIGCOMM*, August 2005.
8. Merkouris Karaliopoulos, Maria Papadopouli, Elias Raftopoulos, and Haipeng Shen. On scalable measurement-driven modeling of traffic demand in large wlans. Technical Report 383, ICS-FORTH, Heraklion, Crete, Greece, August 2006.
9. M. McNett and G. M. Voelker. Access and Mobility of Wireless PDA Users. In *Mobile Computing Communications Review*, 2005.
10. A. Moore and K. Papagiannaki. Toward the Accurate Identification of Network Applications. In *PAM*, March 2005.
11. Sprint Applied Research Group. http://ipmon.sprintlabs.com/packstat/packetoverview.php.
12. Diane Tang and Mary Baker. Analysis of a Local-Area Wireless Network. In *Mobile Computing and Networking (MOBICOM)*, 2000.
13. UNC/FORTH Archive of Wireless Traces. http://www.cs.unc.edu/Research/mobile/datatraces.htm.