Analysis of a Mixed-Use Urban WiFi Network: When Metropolitan becomes Neapolitan

Mikhail Afanasyev, Tsuwei Chen†, Geoffrey M. Voelker, and Alex C. Snoeren
University of California, San Diego and †Google Inc.
{mafanasyev,voelker,snoeren}@cs.ucsd.edu, tsuwei@google.com

ABSTRACT

While WiFi was initially designed as a local-area access network, mesh networking technologies have led to increasingly expansive deployments of WiFi networks. In urban environments, the WiFi mesh frequently supplements a number of existing access technologies, including wired broadband networks, 3G cellular, and commercial WiFi hotspots. It is an open question what role city-wide WiFi deployments play in the increasingly diverse access network spectrum. We study the usage of the Google WiFi network deployed in Mountain View, California, and find that usage naturally falls into three classes, based almost entirely on client device type. Moreover, each of these classes of use has significant geographic locality, following the distribution of residential, commercial, and transportation areas of the city. Finally, we find a diverse set of mobility patterns that map well to the archetypal use cases for traditional access technologies.

Categories and Subject Descriptors

C.2.3 [**Computer Communication Networks**]: Network Operations

General Terms

Measurement, Performance

1. INTRODUCTION

Municipal wireless networks have generated a great deal of excitement and controversy in recent years, as the promise of nearly ubiquitous Internet access for WiFi-capable devices has led many city governments and private entities to propose and deploy city-wide mesh networks. At the same time, the number and type of WiFi-capable devices have exploded due to the increasing popularity of laptops and WiFi-capable smartphones like the Apple iPhone. Yet mesh WiFi networks are far from the only networks on which such devices operate. In urban environments, the WiFi mesh frequently supplements a number of existing access technologies, including wired broadband networks, 3G cellular, and WiFi hotspots.

IMC'08, October 20–22, 2008, Vouliagmeni, Greece.

Given the plethora of alternative access technologies, the long-term economic feasibility of metropolitan mesh networks appears uncertain. In particular, it is unclear what role city-wide WiFi deployments play from a user's perspective, independent of any particular network agreement or charging policy. A great deal of academic research has focused on developing and improving wireless mesh protocols, and studies of deployed wireless networks have recently begun appearing in the literature [1, 5, 7, 8]. These studies focus almost exclusively on the operation and effectiveness of the mesh backbone, however; to the best of our knowledge, none have yet to report upon how clients actually use a metropolitan network.

We study the usage of the Google WiFi network, a freely available outdoor wireless Internet service deployed in Mountain View, California, consisting of over 500 Tropos MetroMesh pole-top access points. Due to its location in the heart of Silicon Valley and no-cost access policy, we expect usage in the Google network to represent an optimistic view of potential client demand in other urban networks. Using 28 days of overall network statistics in Spring 2008, we analyze the temporal activity of clients, their traffic demands on the network, the mobility of users as they roam through the city, and the diversity and coverage of users spread geographically in the network.

We find that network usage uniquely blends the characteristics of three distinctly different user populations into a single metropolitan wireless network; we call such a hybrid network *Neapolitan*.[1] Figure 1 shows one dramatic example of this usage variation: when plotting bytes transferred as a function of session length, three distinct clusters emerge: one cluster of short, light sessions at the left axis, another cluster of extremely long and heavy sessions at the far right, and a third that spans the full range of session lengths and sizes. If one classifies these sessions by device type as shown in the figure, three distinct user populations emerge: Local residents and businesses use it as a static WiFi mesh access network, a substitute for DSL or cable modem service. Laptop users have mobility and workload patterns reminiscent of campus and other public hotspot WiFi networks (labeled hotspot in the figure). Finally, smartphone users combine the ubiquitous coverage of cellular networks with the higher performance of wireless LANs.

Each of these classes has significant geographic locality in the Google WiFi network, following the distribution of residential, commercial, and transportation areas of the city. Additionally, we observe a diverse set of mobility patterns that map well to the archetypal use cases for traditional access technologies. Because the Google network is a production network—as opposed to a research prototype—user privacy is paramount. Hence, our study focuses exclusively on client aggregates; we make no attempt to

[1]Neapolitan ice cream consists of strawberry, chocolate, and vanilla ice cream all packaged side-by-side.

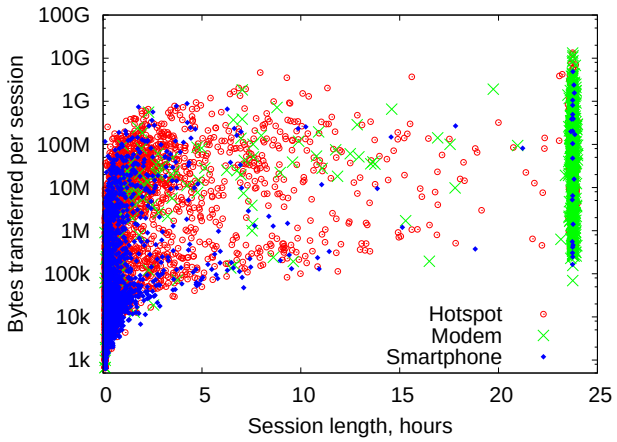


Figure 1: Bytes transferred as a function of session length during a typical 24-hour period.

isolate or analyze the traffic or mobility patterns of any particular clients. Moreover, we limit our traffic analysis to high-level application classification based upon protocol and port numbers. Finally, because we do not collect any client-side information, we report exclusively upon the behavior of clients that successfully connect to the Google network; potential clients that are either unable (due to, for example, a high rate of motion or poor signal strength at their location) or choose not to connect to the network are not represented in our data.

Due to the lack of comparable measurements for other networks, we are unable to comment on the generality of our conclusions. Instead, we hope that our study will encourage researchers and network operators to report upon the usage within other mesh deployments to inform the ongoing debate about the future of metropolitan WiFi. The remainder of this paper is organized as follows. We begin by surveying related work in Section 2 before describing the architecture of the Google WiFi network and our data collection methodology in Section 3. We analyze the disparate network usage patterns in Section 4 and then turn our attention to client mobility in Section 5. Finally, Section 6 considers the ramifications of observed usage on network coverage and deployment, and Section 7 summarizes our findings.

2. RELATED WORK

The Google WiFi network represents one of the latest in various community, commercial, and rural efforts to use commodity 802.11 hardware to construct mesh backbone networks. Since 802.11 was not originally tailored for use in a mesh, work in mesh network deployments has encompassed nearly all aspects of network design, including network architecture [5], MAC protocol development [21], routing protocol design [6], and network planning and provisioning [26].

Measurement studies of urban WiFi mesh networks inform such work in network design, implementation, and deployment. Aguayo et al. captured link-level measurements of the Roofnet community network in Cambridge, MA, to evaluate the network performance and reliability [1]. Camp et al. used the Technology For All (TFA) urban mesh network in Houston, TX, to characterize how control and management traffic degrade network performance [8], to develop models to correlate link characteristics with application performance, and to evaluate AP placement topologies to increase throughput [9]. Robinson et al. introduced low-overhead techniques for assessing mesh network geographic coverage for planning, evaluating the techniques on both TFA as well as the Google WiFi network we study in this paper [23]. Finally, Brik et al. combine both passive and active measurements of the MadMesh commercial mesh network in downtown Madison, WI, to evaluate mesh planning, deployment, routing, and user experience and workload [7]. In large part, these studies are orthogonal to our measurements of the Google WiFi network, and they therefore complement each other. The MadMesh study characterized some aspects of user activity, and we make comparisons when possible in Section 4. Points of comparison, such as the daily variation in the number of users, are in reasonable agreement, suggesting that at least some aspects of the Google WiFi network generalize beyond the uniqueness of being deployed in the heart of Silicon Valley. Otherwise, lacking comparable user workload and behavior studies, we hesitate to generalize beyond the one network we evaluated.

The "modem" users in our study are similar to users of community and commercial backbone mesh networks exemplified by Roofnet [1]. Community and commercial mesh networks often serve as multi-hop transit between homes, businesses, and public locales and the Internet. Mobility is possible, but not necessarily the primary goal; as such, network use tends to be similar to use with DSL or cable modem service. Their application workloads and network utilization are most useful as a point of comparison with the other two user populations in our study; they only exhibit mobility to the extent to which their AP associations flap over time.

The "hotspot" user base in our study most closely resembles user populations of single-hop access wireless LANs, such as university campus networks, both in the dominant applications used and the relatively limited user mobility. Numerous studies of indoor 802.11 networks have covered a variety of environments, including university departments [10, 11, 27], corporate enterprises [4], and conference and professional meetings [3, 14, 15, 18, 20, 24]. These studies have focused on network performance and reliability as well as user behavior from the perspectives of low-level network characteristics to high-level application use. With their more extensive geographic coverage, larger-scale studies of outdoor 802.11 networks on university campuses have provided further insight into mobility and other user behavior [12, 13, 16, 19, 25, 29].

The dominant presence of iPhone users represents perhaps the most interesting aspect of the Google WiFi user population. WiFi smartphones represent an emerging market early in its exponential adoption phase, yet it is the WiFi user population that is the least well understood. Tang and Baker's detailed study of the Metricom metropolitan wireless network [28] is most closely related to the smartphone population of the Google WiFi network. Metricom operated a Ricochet packet radio mesh network covering three major metropolitan areas. The study covers nearly two months of activity in the San Francisco Bay Area, and focuses on network utilization and user mobility within the network. Presumably cellular providers measure cellular data characteristics extensively, but these results are typically considered confidential.

Finally, we note that rural mesh networks in developing regions typically support targeted services [22], such as audio and video conferencing to provide remote medical treatment, and consequently have application characteristics specific to their intended use.

3. THE NETWORK

The Google WiFi network is a free, outdoor wireless Internet service deployed in Mountain View, California. The network has been continuously operational since August 16, 2006, and provides public access to anyone who signs up for an account. The network is accessible using either traditional (SSID GoogleWiFi) and secure

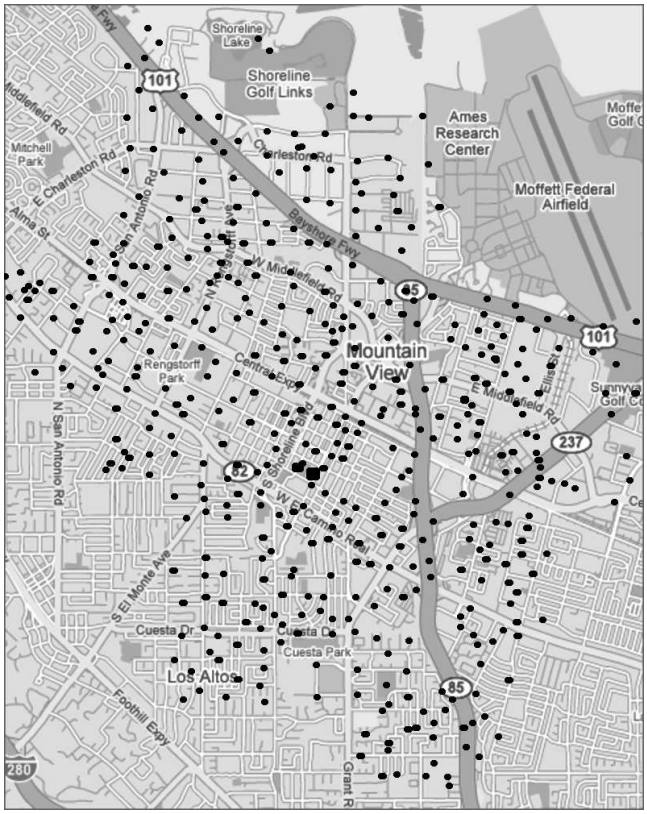


Figure 2: The Google WiFi coverage map.

(WPA/802.1x, SSID GoogleWiFiSecure) 802.11 clients. Aside from the standard prohibitions of SPAM, hacking, and other inappropriate activities, Google does not limit the types of traffic that can be transmitted over the network.[2] However, it does rate limit individual clients to 1 Mb/sec.

3.1 Network structure

The network consists of over 500 Tropos MetroMesh pole-top access points. Each Tropos node has a distinct identifier and a well-known geographic location; Figure 2 shows the approximate location of the Tropos nodes. Each Tropos node serves as an access point (AP) for client devices, as well as a relay node in a wide-area backhaul mesh that provides connectivity to the wired gateways. The topology of the Tropos mesh network is constructed dynamically through a proprietary Tropos routing algorithm. A pure mesh network of this scale exhibits significant traffic congestion at nodes close to the gateway router, however. To alleviate the congestion, the Google WiFi network is hierarchically clustered around approximately 70 point-to-point radio uplinks that serve as a fixed long-haul backbone for the mesh network.

Traffic is eventually routed to one of three distinct wired gateways spread across the city, which then forward the traffic to the main Google campus where it is routed to a centralized authorization and authentication gateway. Google provides single sign-on authentication and authorization service, but, at the link layer, 802.11 client devices continue to associate with each Tropos AP individually. All Tropos nodes support the RADIUS accounting standard [17] and provide periodic updates to the central server.

[2]The Google WiFi Terms of Service are available at `http://wifi.google.com/terms.html`.

Field	Units
Acct-Status-Type	Start/Interim-Update/Stop
NAS-Identifier	Tropos ID string
Calling-Station-Id	client MAC address
Acct-Session-Time	seconds
Tropos-Layer2-Input-Octets (TLIO)	bytes
Tropos-Layer2-Output-Octets (TLOO)	bytes
Tropos-Layer2-Input-Frames (TLIF)	frames
Tropos-Layer2-Output-Frames (TLOF)	frames
Acct-Input-Octets (AIO)	bytes
Acct-Output-Octets (AOO)	bytes
Acct-Input-Packets (AIP)	packets
Acct-Output-Packets (AOP)	packets

Table 1: Partial contents of a RADIUS log record.

3.1.1 Mesh topology

While not the main focus of our study, we collected basic information about the mesh topology through an administrative interface exported by the Tropos nodes. The relatively dense deployment of APs provides significant path diversity. On average, only 5% of APs have a unique neighbor (with a signal quality of at least 14 dBm [26]); the median AP can communicate with at least four neighboring APs, and the most well-connected 10% have more than eight potential next hops. The mesh topology continuously evolves during our study (the median node changes parents once every two days), but remains relatively stable in the short term (the most dynamic node changes parents slightly more than once an hour over the course of the trace). The hierarchical structure ensures most clients have short paths to the gateway: the majority of clients active during our study are within two hops of the gateway, and less than 10% are more than three hops away.

3.1.2 Access devices

To extend the network coverage indoors, Google recommends the use of WiFi modems, or bridges, which are typically outfitted with more capable antennas than a standard 802.11 client. WiFi modems often provide a wired Ethernet connection or serve as an in-home wireless AP, allowing the connection of multiple physical machines. While Google does not manufacture or sell WiFi modems, it has recommended two particular WiFi modems to users of the Mountain View network. In particular, Google suggests the PePLink Surf and the Ruckus MetroFlex. Additionally, in certain portions of the city, Google has deployed Meraki Mini mesh repeaters to extend the reach of the Tropos mesh.

3.2 Data collection

We analyze a trace of 28 days of accounting information collected by the central Google WiFi RADIUS server during the Spring of 2008. Periodic updates are generated by all Tropos nodes for each associated client every fifteen minutes. Tropos nodes issue additional updates when clients first associate or disassociate (either explicitly—which is rare—or through a 15-minute timeout). Table 1 shows the portion of the RADIUS log records that we use for our study. For the purposes of this paper, we focus almost exclusively on layer-three information: we do not consider the link layer behavior of the network. (Although we do make occasional use of layer-two accounting information as described below.)

Additionally, to facilitate our study of the types of application traffic in the network (Section 4.2.2), we obtained five-days worth of packet-header traces collected at the central Internet gateway of the Google WiFi network. The header trace contains only (a prefix of) the first packet of each flow for the first fifteen minutes of each hour. Because the trace was collected at the gateway—as op-

posed to inside the wireless mesh itself—we do not observe layer-two protocol traffic such as ARP, nor many DHCP requests handled by the Tropos nodes themselves. Moreover, we only observe layer-three traffic entering or leaving the Google WiFi network; our traces do not contain traffic whose source and destination both reside inside the WiFi network. A comparison of the trace content to the statistics reported by the RADIUS logs (which do include traffic internal to the network) indicates the volume of such traffic is negligible, however.

3.2.1 Data correction

During the course of our analysis, we discovered several bugs in the Tropos accounting mechanism. In particular, a number of fields are susceptible to roll-over, but such events are readily detectable. More significantly, the Acct-Output-Octets (AOO) field is occasionally corrupt, leading to spurious traffic reports for roughly 30% of all client sessions. Tropos confirmed the bug, and informed us that the latest version of the Tropos software fixes it. Unfortunately, our traces were collected before the software update was applied.

Luckily, the layer-two traffic information reported by the Tropos nodes appears accurate, so we are able to both detect and correct for corrupt layer-three traffic information. We detect invalid log records by comparing the number of layer-two output octets (TLOO) to the layer-three count (AOO); there should always be more layer-two octets than layer-three due to link-layer headers and retransmissions. If we discover instances where the layer-three value is larger than layer two, we deem the layer-three information corrupt and estimate it using layer-two information:

$$\widehat{AOO} = \begin{cases} AOO & \text{if } AOO \geq TLOO, \\ TLOO \cdot (AOP/TLOF) & \\ \qquad - (32 \cdot AOP) & \text{otherwise.} \end{cases}$$

In other words, we scale the layer-two octet field based upon the ratio of layer-two frames to layer-three packets to account for link-layer loss, and subtract 32 bytes per packet for link-layer headers.

3.2.2 Client identification

To preserve user privacy, we make no attempt to correlate individual users with their identity through the Google authentication service. Instead, we focus entirely on the client access device and use MAC addresses to identify users. Obviously, this approximation is not without its pitfalls—we will incorrectly classify shared devices as being one user, and are unable to correlate an individual user's activity across devices. While we speculate that a number of users may access the Google WiFi network with multiple distinct devices (a laptop and smartphone, for example), we consider this a small concession in the name of privacy.

We have aggregated clients into groups based upon the class of device they use to access the network. We classify devices based upon their manufacturer, which we determine based upon their MAC addresses. In particular, we use the first three octets, commonly known as the Organizationally Unique Identifier (OUI). Because many companies manufacture devices using several OUIs, we have manually grouped OUIs from similar organizations (e.g., "Intel" and "Intel Corp.") into larger aggregates. Table 2 shows some of the most popular OUI aggregates in our trace.

Apple bears particular note. While we have attempted to determine which OUIs are used for iPhones as opposed to other Apple devices (PowerBooks, MacBooks, iPod Touch, etc.), we have observed several OUIs that are in use by both laptops and iPhones. Hence, accurately de-aliasing these OUI blocks would require tedious manual verification. For the purposes of this paper we have lumped all Apple devices together, and consider them all to be iPhones. Somewhat surprisingly, this appears to be a reasonable approximation. In particular, we estimate that 88% of all Apple devices in our trace are iPhones.[3]

Class	Manufacturers	Count
Smartphone (45%)	Apple	15,450
	Nokia	138
	Research in Motion (RIM)	107
Modem (3%)	Ruckus	525
	PePLink	297
	Ambit	224
Hotspot (52%)	Intel	9,825
	Hon Hai	1,931
	Gemtek	1,735
	Askey Computer Corp.	667
	Asus	385

Table 2: A selection of manufacturers in the trace and distinct client devices seen, grouped by device class. The fraction of total devices in each class is in parentheses.

To estimate the population of iPhone devices, we observed that Apple products periodically check for software updates by polling a central server, `wu.apple.com`. iPhones in particular, however, poll `iphone-wu.apple.com`, which is a CNAME for `wu.apple.com`. Hence, if one considers the DNS responses destined to an iPhone device polling for software updates, it will receive responses corresponding to both `iphone-wu.apple.com` and `wu.apple.com` (either because the DNS server proactively sent the A record of `wu.apple.com`, or the client subsequently requested it). Other Apple devices, on the other hand, will only receive an A record for `wu.apple.com`. We compare the total number of DNS responses destined to clients with Apple OUIs for `iphone-wu.apple.com` to those for `wu.apple.com` present in our packet header traces, and determine that the Gateway sees 1.13 times as many responses for `wu.apple.com`. We therefore conclude that 88% of the `wu.apple.com` responses actually resulted from queries for `iphone-wu.apple.com`.

iPhones constitute the vast majority of all devices we have classified into the *smartphone* group, although we see several other manufacturers, including Research in Motion—makers of the Blackberry family of devices—and Nokia in the trace. As discussed previously, Ruckus and PePLink are two brands of WiFi modems that Google recommends for use in their network. Moreover, neither company appears to manufacture other classes of WiFi devices in any large number. Hence, for the remainder of the paper we have combined Ruckus and PePLink OUIs into a larger class that we term *modem*. (We also include Ambit, whose only WiFi-capable devices appear to be cable modems.) Finally, for lack of a better term, we classify the remaining devices as *hotspot* users. While it is extremely likely that some portion of these devices are misclassified (i.e., some modem and smartphone devices are likely lumped in with hotspot devices) the general trends displayed by the hotspot users are dominated by Intel, Hon Hai, and Gemtek, manufactures well known to produce a significant fraction of the integrated laptop WiFi chip-sets. (Notably, Hon Hai manufactures WiFi chip-sets used in the Thinkpad line of laptops.)

[3]Apple released the 3G version of the iPhone after the completion of this study. A comparison of the number of iPhone devices present on the network in late July 2008 shows that while a significant fraction of the iPhone user population upgraded to a new device (or at least the new software release), the total number of iPhone devices did not increase significantly. The traffic patterns, however, have changed (see Section 4.2).

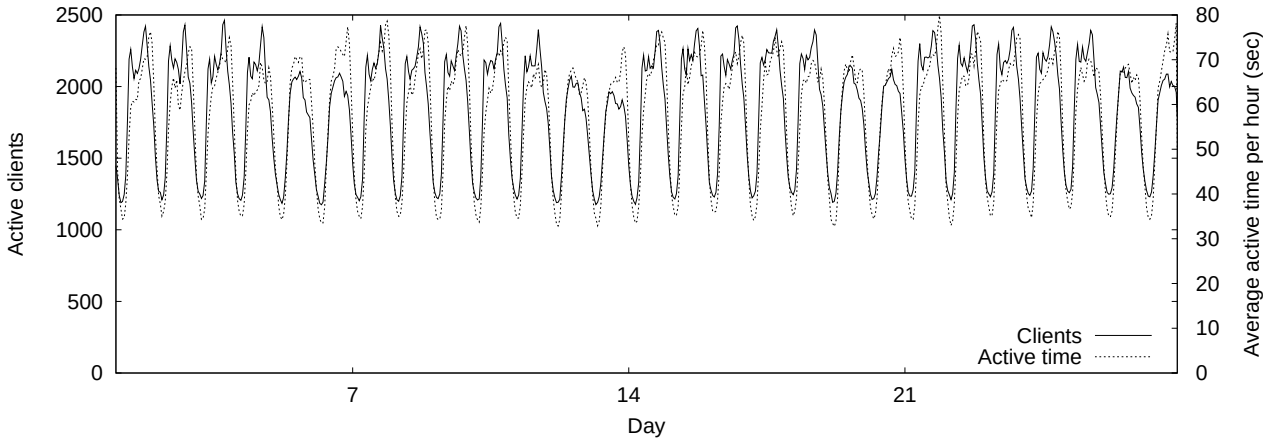


Figure 3: Usage of the Google WiFi network for the duration of the trace, measured in 15-minute intervals.

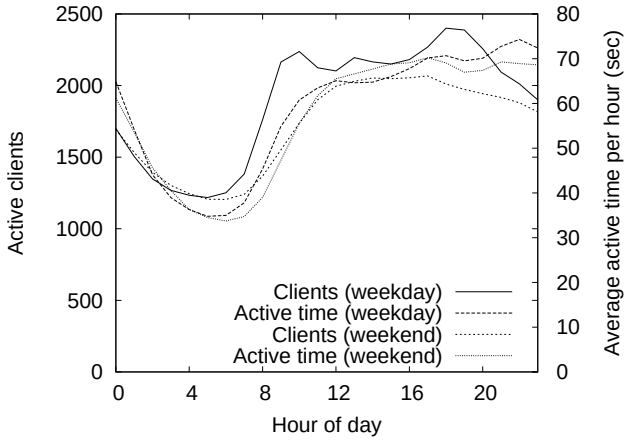


Figure 4: Average daily use of the Google WiFi network.

4. USAGE

In this section we analyze when various classes of clients are active in the Google WiFi network, and then characterize the application workload these clients place on the network.

4.1 Activity

We begin by looking at overall aggregate network activity. Figure 3 shows the number of active clients using the network (left y-axis) and their average activity time (right y-axis) per 15-minute interval for the entire trace. In our analyses, we consider a client to be *active* for a 15-minute reporting interval if it sends at least one packet per second during the interval. If a client sends fewer packets, we deem it to be active for a prorated portion of the interval—i.e., a client that sends at least 54,000 packets is deemed active for the entire interval, while a client that sends 18,000 packets is said to be active for 5 of the 15 minutes. We choose this metric in an attempt to reduce the contribution of devices that are simply on but likely not being used, as such devices still tend to engage in a moderate rate of chatter [2]. We calculate activity time as the average number of seconds each client was active during the hour.

The results show that the Google WiFi network has a substantial daily user population, peaking around 2,500 simultaneous users in any 15-minute interval. The curves also show the typical daily variation seen in network client traces, with peaks in both users and activity during the day roughly twice the troughs early in the morning. Weekend use is lower than on weekdays, with roughly 15% fewer users during peak times on the weekends. When users are connected, they are active for only a small fraction of time. On an hourly basis, users are active only between 40–80 seconds (1–2%) on average.

Figure 4 shows the daily variation of aggregate network behavior in more detail. The figure has four curves, two showing the number of clients (left y-axis) and two showing average hourly client activity (right y-axis) on a typical weekday and weekend day. At the scale of a single day, variations over time in the number of clients and their activity become much more apparent. For example, there are multiple distinct peaks in clients on the weekday during morning rush hour (9 am), lunch time (12:30 pm), and the end of evening rush hour (6 pm); weekends, however, are much smoother. Further, the largest peaks for the number of clients and activity are offset by four hours. The number of clients peaks at 6 pm at the end of rush hour, but activity peaks at 10 pm late in the evening. We note that the diurnal characteristics in the number of clients of the Google network match those of the MadMesh network [7], suggesting at least one high-level similarity in user populations in two widely separated locales.

This behavior reflects the kinds of clients who are using the network and how they use it. Figure 5(a) similarly shows the daily variation of client usage on weekdays as in Figure 4, but separates clients by the type of device they use to access the network. The graph shows three curves corresponding to the number of active modem, smartphone, and hotspot clients each hour. Separated by device type, we see that the different types of clients have dramatically different usage profiles. The number of modem clients is constant throughout the day. This usage suggests homes and businesses with potentially several computers powered on all day, with "chatty" operating systems and applications providing sufficient network traffic to keep the wireless access devices constantly active (analyses of network traffic in Section 4.2 shows that these users do have substantial variation in traffic over time). Hotspot users show more typical diurnal activity, with peak usage in late afternoon twice the trough early in the morning. Hotspot user activity is also high for more than half the day, from 9 am until 11 pm at night.

Smartphone users show the most interesting variation over time. The curve shows three distinct peaks during the day (9 am, 1 pm, and 6 pm), suggesting that smartphone usage is highly correlated with commute and travel times and that the devices are active while

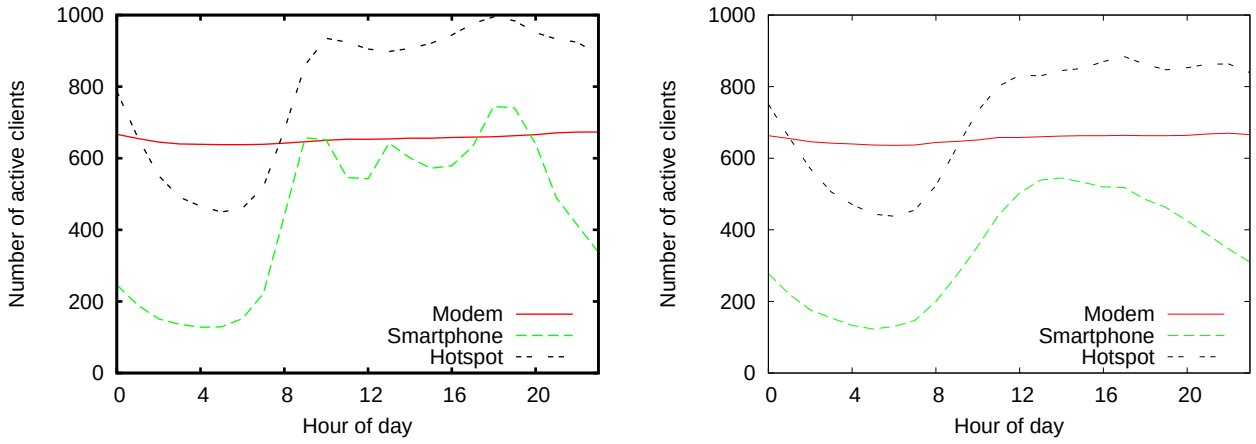


Figure 5: Hourly usage of the Google WiFi network divided between weekdays and the weekend.

users are mobile (Section 5 explores mobility behavior further). Further, smartphone usage is much more heavily concentrated during the day. Peak client usage at 6 pm is four times the trough at 5 am in the morning. There are a number of possible explanations for this behavior. One is that the majority of smartphone users are commuters, and therefore are only within range of the network during the day. Another is that, although they may make voice calls, users do not access WiFi during the evening, perhaps preferring to access the Internet with laptops or desktops when at home.

Figure 5(b) similarly shows the number of active clients by device type as Figure 4, but for a typical day on the weekend. Comparing weekdays with the weekend, we see little difference for modem and hotspot users. Modem users remain constant, and, although there are approximately 10% fewer hotspot users during the highly active period than on the weekday, the period of high activity remains similar. Smartphone users again exhibit the most notable differences. Smartphone peak usage no longer correlates with commute times, peaking at midday (1 pm) and diminishing steadily both before and after.

4.2 Traffic

The results above show how many and when clients are active. Next we characterize the amount of traffic active clients generate.

Figure 6 plots a CDF of the total amount of data transferred by clients of each class per day. Only active clients are included; if a client did not connect at all during a day, that data point was not included in the graph. For ease of presentation, we combine upload and download traffic as opposed to reporting each individually. The daily ratio of download to upload traffic remains relatively constant across our trace at approximately 3.15:1, although there are interesting distinctions between device classes. Hotspot and modem users are roughly equivalent, at 2.9 and 3.2 to one, respectively, while smartphone usage was noticeably more skewed at 5.9:1.

Figure 7 shows the distribution of transfer rates for 15-minute intervals when the clients were active for the entire trace period. In other words, if a client sends less than one packet per second during an interval, that interval is not included. The graph shows curves for each of the three user populations. Recall that Google limits transfer rates to 1 Mb/sec per client, or approximately 128 KB/sec. Very few active periods approach this limit, though, so it has little impact on sustained traffic demands by users.

The transfer rates vary substantially among the different populations. The median rates in active periods are 3 KB/sec for modem users, 512 bytes/sec for hotspot users, and 128 bytes/sec for smartphone users. Note that the very low transfer rates in bytes/sec are an artifact of the measurement infrastructure. The trace records have a granularity of 15 minutes, so low transfer rates reflect short activity averaged over a relatively long time interval. Modem activity has the overall highest transmission rates: the bulk of of active periods (80%) transmit at 256 bytes/sec or higher, and 20% at 8 KB/sec. Hotspot activity is roughly uniformly distributed across the range: over 80% of hotspot transfer rates fall between 64 bytes/sec and 8 KB/sec, with tails at either extreme. Smartphone activity falls into three regions. Much of smartphone activity exhibit very low rates (40% less than 96 bytes/sec), the next 40% of activity is linear between 96 bytes/sec and 768 bytes/sec, while the remaining 20% have higher rates.[4]

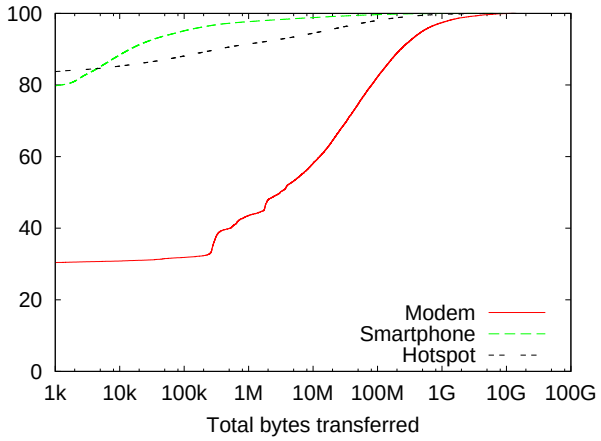


Figure 6: Total bytes transferred (in and out) by each type of client per day (log-scale x axis).

4.2.1 Sessions

Next we characterize how long clients are active when associated with the network. We observed up to 379 distinct sessions

[4]A short follow-up study after the release of the 3G iPhone (July 2008) indicates a noticeable uptick in the amount of data transferred by the smartphone class, perhaps due to the enhanced functionality of the new software version.

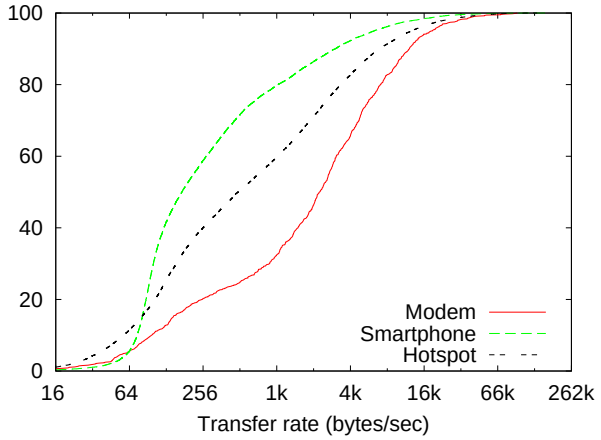


Figure 7: Instantaneous transmission rates during activity periods for each type of client.

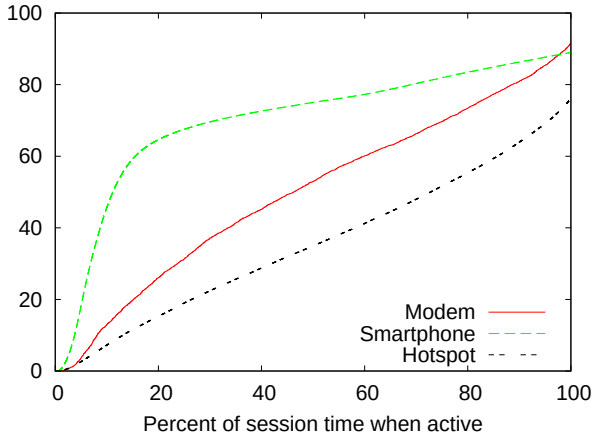


Figure 9: Percentage of the session during which the client was active.

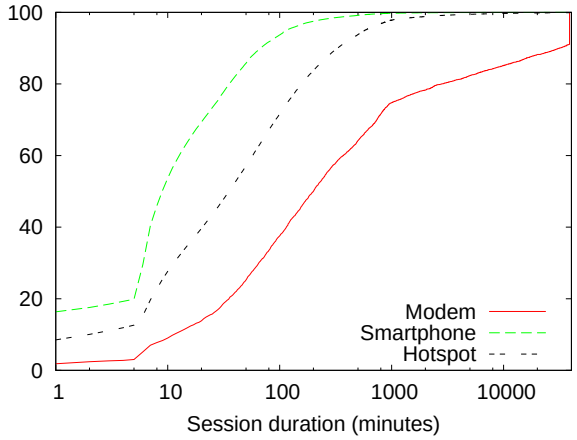


Figure 8: CDF of session lengths, in minutes (x axis in log scale). The entire trace is only 40,320 minutes long.

per client, with the median client connecting only twice and a full 35% appearing only once. At the high end, almost 7% of clients connected at least once per day, on average, and more than 10% connected at least once per weekday (20 times).

Figure 8 shows the distribution of session lengths during our trace for the different client populations. We define a client session as the period of time between 802.11 association and disassociation with an access point. Clients in the different user populations exhibit different session length distributions. A significant fraction of modem clients have sessions that span the entire trace; although 65% of modem sessions are shorter than a day, these shorter sessions are due to oscillations between access points (see Section 5). Many hotspot clients have sessions shorter than an hour: the median hotspot session length is 30 minutes, while 30% of hotspot sessions longer than two hours. Smartphone clients have the shortest session lengths: over half the sessions are less than 10 minutes, and only 10% are longer than an hour.

Just because clients are associated with the network does not necessarily mean that they are active during the entire session. Figure 9 shows what fraction of their sessions the clients were actually active. Not only do smartphone users have short sessions, their session activity is quite low. For over half of smartphone sessions, clients are active for less than 10% of the time. This low activity suggests that users have their phones and WiFi turned on when in the network, but use Internet applications only infrequently. Modem clients are much more active during their sessions. Over 40% of their sessions are active at least half the time. Finally, hotspot clients are the most active when connected to the network; the median session is active almost 75% of the time. This activity suggests that hotspot users connect to the network with the intention to use it, and disconnect when finished.

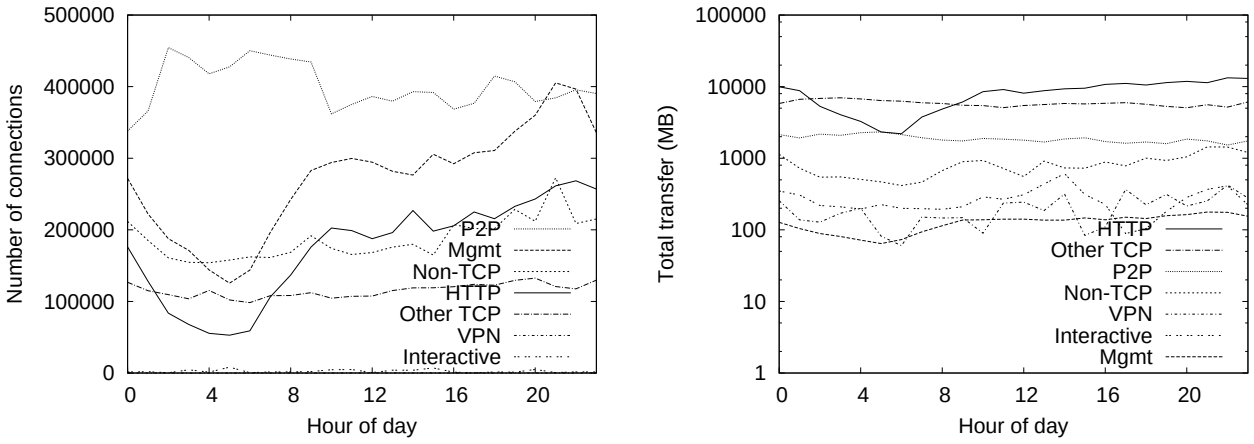
4.2.2 Application classes

It is natural to ask what types of traffic the Google WiFi network carries. Using a five-day packet header trace spanning a weekend during our larger trace, we classify the first packet of each flow based on protocol and port numbers. Figure 10(a) plots the number of connections for each traffic class as a function of the time of day. While our port-based traffic classification mechanism is imperfect, it is clear that peer-to-peer connections constitute a significant fraction of the network use. (While most of the traffic is BitTorrent, we see a remarkable amount of "Thunder" traffic, a Chinese peer-to-peer protocol also known as Xunlei, which operates on UDP port 15000.) Interestingly, peer-to-peer usage appears to be relatively time insensitive, which is consistent with users that leave their file sharing clients on almost all the time.

Web traffic is significantly more diurnal, seeing a significant dip in the early morning hours, and peaking in the evenings. Perhaps the most unusual feature is the dramatic variation in the frequency of management (ICMP, DHCP, and DNS) connections. The vast majority of this traffic is actually mDNS "dnsbugtest" traffic, however. In fact, Figure 11(a) shows that almost all of it stems from a few particular modem devices.

The other two main connection contributors, other TCP and non-TCP show less significant—but still apparent—diurnal trends. We group SSH, telnet, X windows, and similar remote log-in protocols into an interactive class; perhaps not surprisingly they represent a consistently negligible fraction of the total connections. Finally, we observe very few VPN connections, despite the fact that Google promotes Google Secure Access, a free VPN provided by Google for use on the Google WiFi network—although the VPN connections that do exist turn out to be relatively heavy.

The picture for bytes is similar. Figure 10(b) plots the total amount of data transferred in the network as a function of hour



(a) Number of connections (b) Bytes transferred

Figure 10: Hourly usage of the network per application class.

of the day. HTTP and other TCP traffic clearly represent the lion's share of the traffic. We suspect that other TCP is largely peer-to-peer traffic that we failed to properly classify. Identified peer-to-peer traffic forms the next tier of usage, along with non-TCP traffic which we suspect represents VoIP and other multimedia transfers. The log-scale y-axis provides a better view on the interactive and VPN traffic, which shows a subtle diurnal trend. Finally, we see that management flows, while frequent, constitute a very small fraction of the total traffic in terms of total bytes transferred.

Figure 11 breaks down each of the two preceding graphs by client type. To do so, we build a mapping between the client MAC addresses and assigned IP addresses in the RADIUS logs, and then classify the traffic logs by IP address. Not surprisingly, the three device types show markedly different application usage. Smartphones, in particular, generate very few connections, and almost all their bytes are Web or other TCP applications. We surmise that the bulk of the other traffic is made up by streaming media (e.g., UPnP-based iPhone video players like Mooncat) and VoIP traffic, but further analysis is required.[5]

The distinctions between modem and hotspot users are far more subtle. It is worth noting, however, that there are an order of magnitude more hotspot users than modem users, yet the modem users place similar aggregate traffic usage demands on the network. Both modem and hotspot users show a significant amount of peer-to-peer, Web, and non-TCP traffic. Of note, the modem P2P users appear to receive much higher per-connection bandwidth than the Hotspot users, which is consistent with our observations about the instantaneous bandwidth achieved by each client type (c.f., Figure 7). Hotspot users are significantly more likely to use interactive remote login applications than modem users, but we have not attempted to determine why that may be.

Finally, we observe that almost all the connection volume in the management class stems from modem clients—Ruckus devices in particular. While many devices in our trace periodically issue "dns-bugtest" mDNS requests, some Ruckus devices issue thousands of queries during each 15-minute interval. The precise cause of this behavior deserves further investigation.

[5]Assuming iPhones are extremely unlikely to be using BitTorrent clients (although at least one exists), we use significant BT activity (more than 1 MB) as a filter to pull three presumably misclassified Apple laptops out of the Smartphone grouping.

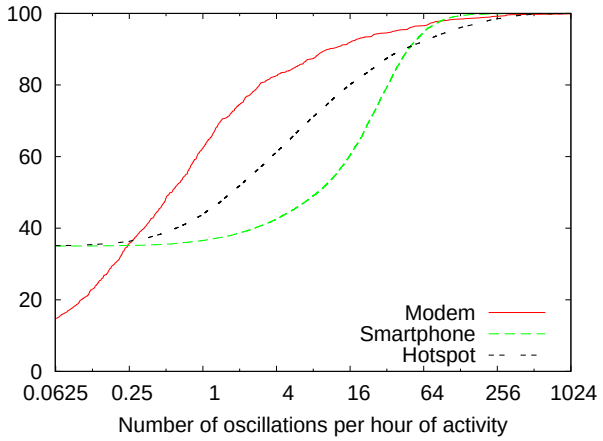


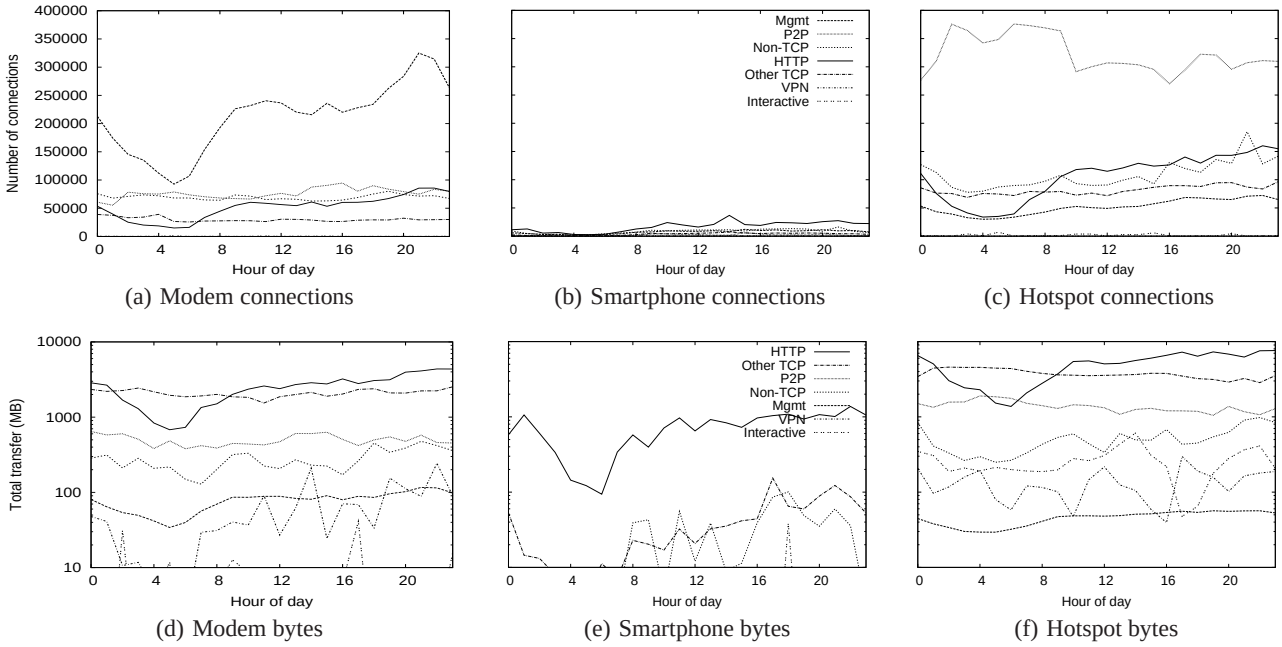
Figure 12: CDF of the number of oscillations per hour (x axis is log scale).

5. MOBILITY

We now turn to questions of client mobility; in particular, we study how frequently, fast, and far hosts move. Because clients do not report their geographical location, we use the location of the AP to which they associate as a proxy for their current location. The Google WiFi network has varying density, but APs are approximately 100 meters apart on average. While that estimate provides an effective upper bound on the resolution of our location data, it is possible that clients may associate to APs other than the physically closest one due to variations in signal propagation.

5.1 Oscillations

Moreover, signal strength is a time-varying process, even for fixed clients. To gain an appreciation for the degree of fluctuation in the network, we consider the number of oscillations in AP associations. To do so, we record the last three distinct APs to which a client has associated within the last hour. If a new association is to one of the previous three most recent APs, we consider it an oscillation. (While it is possible that our definition captures some instances of physical movement, only five oscillation occurrences include APs physically separated by distances of 1500 meters or more, so we believe it to be a reasonably accurate approximation.)



(a) Modem connections (b) Smartphone connections (c) Hotspot connections

(d) Modem bytes (e) Smartphone bytes (f) Hotspot bytes

Figure 11: Number of connections (a–c) and bytes (c–e) per hour for each device type.

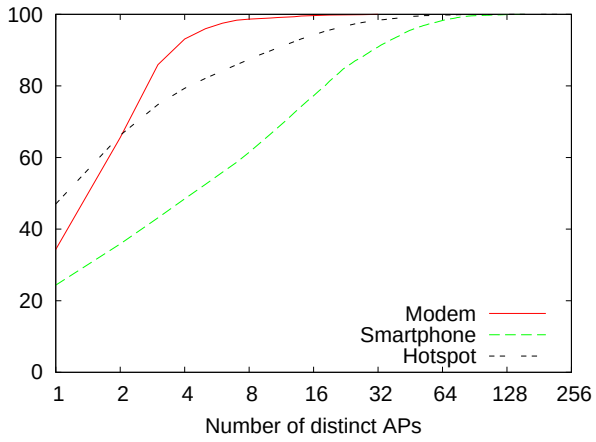


Figure 13: CDF of the number of distinct APs a client associates with over the course of the trace.

Using this definition, we detect a high frequency of oscillations in the data. Figure 12 plots the number of oscillations per hour for each client type. Overall, we see that 50% of clients oscillate at least once an hour, and individual clients oscillate as frequently as 2,900 times an hour (almost once a second). The rate of oscillation varies between client types, with modems exhibiting the lowest rate of oscillation—likely because they are physically fixed, and oscillate only due to environmentally induced signal strength variation—and smartphones the highest. We eliminate oscillations from the association data used in the remainder of this section in an attempt to more accurately capture physical movement—as distinct from RF movement due to changes in signal strength.

5.2 Movement

We plot the number of distinct APs to which a client associates during the course of our trace in Figure 13. Roughly 35% of all devices associate with only one AP; this corresponds well to the fraction of clients that appear only once in the trace (c.f., Section 4.2.1). As one might expect, each client class exhibits markedly different association behavior. Modems tend to associate with a very few number of APs—likely nearby to a single physical location. Smartphones, on the other hand, frequently associate with a large number of APs; 50% of smartphones associate with at least six distinct APs, and the most wide-ranging of 10% smartphones associate with over 32 APs. Hotspot clients, on the other hand, are significantly less mobile—the 90% percentile associates with less than 16 APs during the four-week trace. We observe, however, that both the smartphone and hotspot populations are skewed by a significant number of clients that appear only once in the entire trace.

If we restrict the time window to a day—as opposed to 28 days as above—the distribution shifts considerably (not shown): 90% of all clients connect to at most eight APs per day on average, with only a handful of clients connecting to more than 16 APs. Fully 90% of modems, 70% of hotspot users, and 40% of smartphones connect to only one AP per day on average.

Next, we consider how geographically disperse these APs are. In particular, we study the distance traveled between consecutive associations by a single client. Figure 14 plots the average distance in meters between non-oscillatory client associations. Not surprisingly, very few devices associate with APs less than 100 meters apart, as there are few locations in the city with closely spaced APs (the library is a notable exception). At the other extreme, we see devices that travel over six miles between associations—roughly the maximum distance between APs in the network.

It is frequently possible to connect to a number of different APs from one physical location. If we assume that modem devices move infrequently (most are likely installed in users' homes), we can infer that the Google WiFi signal travels at most 500 meters from an AP. Moreover, by considering the number of APs with which modems associate in Figure 13, we conclude that most locations in the city (where WiFi modems are installed) can reach at most four APs. While this contrasts with the reported connectivity of Tropos

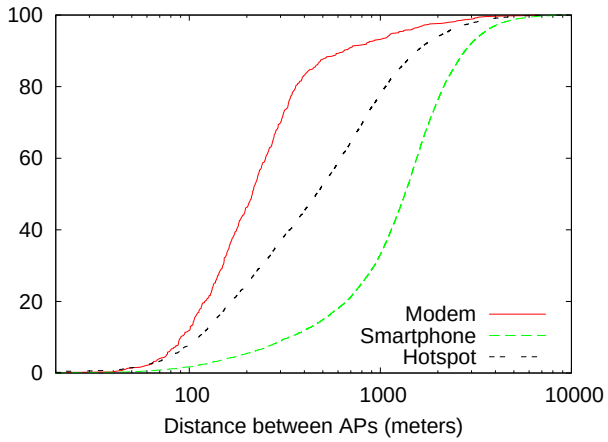


Figure 14: CDF of the average distance between consecutive client associations.

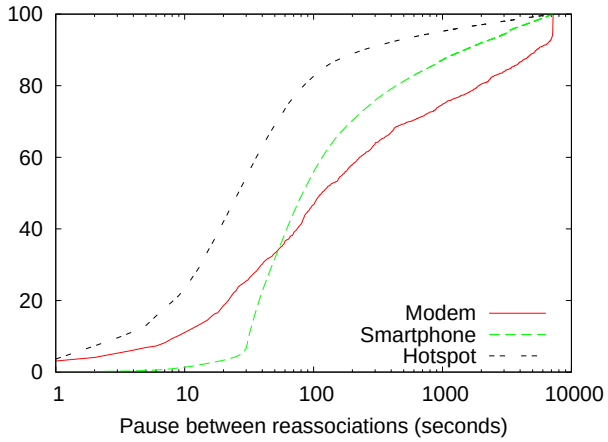


Figure 15: CDF of pause time for each class of client.

nodes (c.f., Section 3.1.1), APs are outfitted with commercial-grade antennae and located on top of light poles, which frequently provides line-of-sight signal propagation to nearby APs.

While smartphones appear to travel further than hotspot clients on average, both show significant range. The median smartphone travels well over half a mile (approximately 1050 meters) between associations, compared to a quarter mile for hotspot clients. The 90-th percentile smartphone travels just slightly farther—1200 meters—than the median, while hotspot usage is more varied: the 90-th percentile user travels almost three times as far as the median.

Finally, to understand how fast clients are moving, we plot the pause time between associations in Figure 15. Interestingly, we note that smartphones rarely re-associate in less than thirty seconds, but usually within two minutes. In contrast, a significant fraction of modems go very long periods without re-associating (likely because the remain constantly attached to the same AP). The majority of hotspot users, on the other hand, re-associate between ten seconds and one minute after their last association.

If one considers a scatter plot of AP distance as a function of pause time (not shown), there is high density along the y axis (instantaneous reassociation) until about 750 meters, with a (comforting) void delineated by roughly the 75 mph line. Symmetrically, we see a significant portion of users that reassociate roughly 200 meters away over all time scales, indicating varying rates of travel between adjacent APs. The graph is significantly less dense in the regions slower than five minutes and further than 500 meters, however.

6. COVERAGE

So far, we have considered characteristics of the users of the network. In this section, we turn our attention to the network itself and ask two distinct questions. First, we consider whether the network is utilized differently in different parts of the city. Secondly, we ask to what extent the full coverage of the network is necessary—in other words, is it possible to deactivate certain APs from time to time and preserve the overall user experience.

6.1 Diversity

The usage of the Google WiFi network varies based on physical location. Table 3 considers three disjoint regions of the city—one residential, one commercial, and one simply a thruway (Highway 101) at four distinct periods throughout the day: 5–6 am, 9–10 am, 3–4 pm, and 6–7 pm corresponding to the peaks and valleys of Figure 4. For each time period and region, we show the number of clients, activity time across those users, and total bytes transferred. To facilitate comparison across time periods and areas, yet preserve the privacy of users in these select geographic areas, we normalize the histograms for each particular value (bytes, activity, and users) to the average for that value over all classes of clients and time periods—in other words, the sum of all the histograms for a particular value is thirty six.

We see significant differences between the network use across the geographic areas. Not only does the proportion of modem, smartphone, and hotspot users vary across locations, but the usage patterns within these user classes also differs substantially. For example, we see far more smartphones in the transit area surrounding Highway 101 than any other type of device, but the smartphones show almost no usage. Indeed, the few hotspot users we do see transfer more data cumulatively than the smartphones. In contrast, smartphones are far less prevalent in the residential area, appearing in similar numbers to hotspot users. However, those we do observe are substantially more active than those in the transit area. Not surprisingly, modem users represent a significant fraction of the residential usage, at least in terms of traffic and activity if not in total number. Moreover, their usage appears less time dependent than the other devices.

The commercial area is the most active, with significant usage across all three classes of clients. Modem activity is similar to that in residential areas, but the absolute number of both smartphones and hotspot users is significantly higher. Mobile (i.e., smartphone and hotspot) usage peaks in the commercial area in the middle of the afternoon (hotspot usage is off scale, with a normalized byte count of 6.2 and user count of 5.4), yet remains strong across all periods, unlike the other two, which show far less usage in the early morning hours. Unsurprisingly, the number of clients in the transit area peaks during rush hours, while residential usage is highest during the evening (not shown).

6.2 Concentration

For a metropolitan network covering an entire city, an interesting deployment question is to what extent the full set of nodes in the network are actively being used. Figure 16 shows the average number of simultaneous clients supported by each AP over the course of a day. Clients are distributed widely: the busiest AP supports just over 14 simultaneous clients and all but the least-utilized 5%

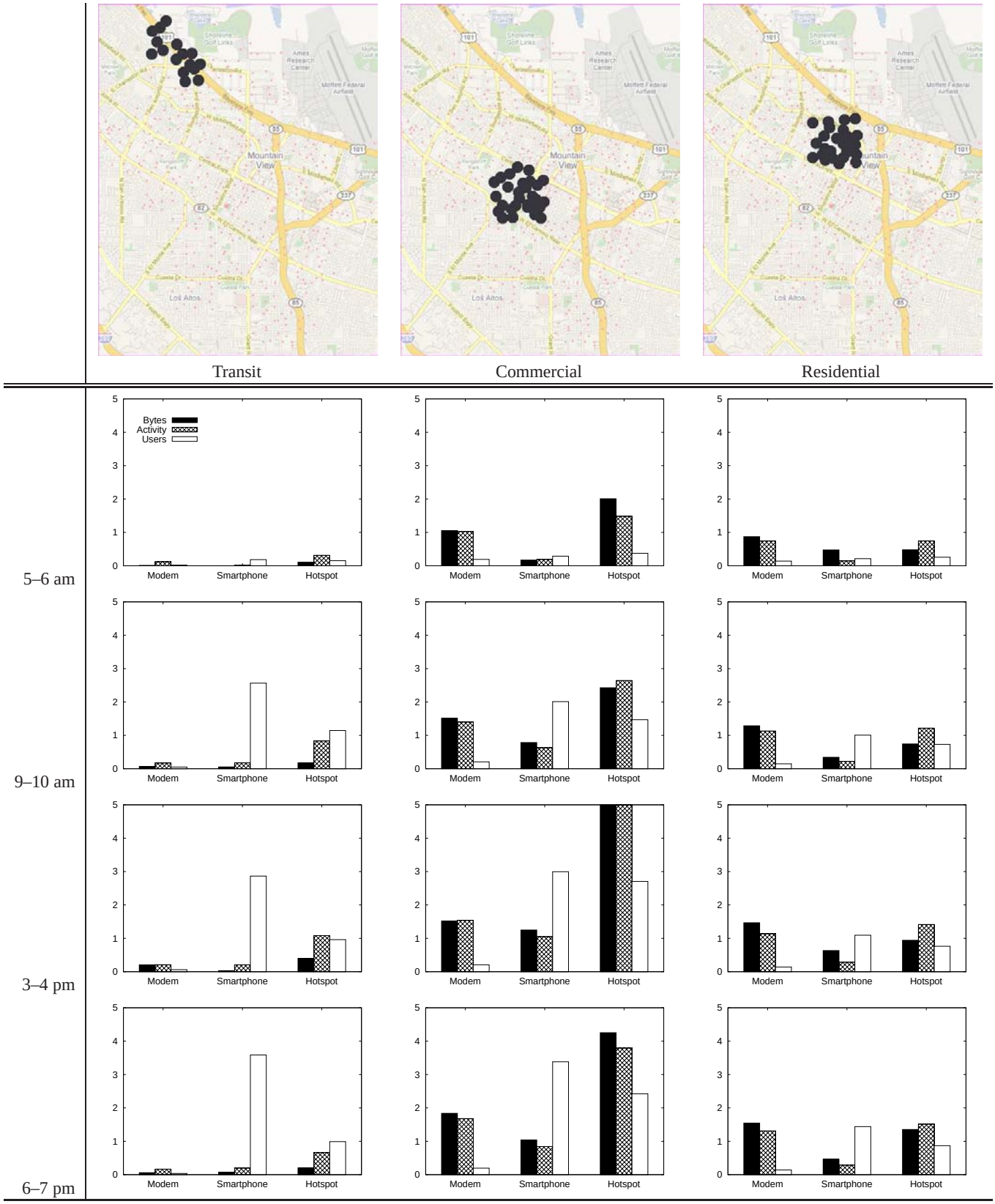


Table 3: Network usage for representative time periods across different parts of the city.

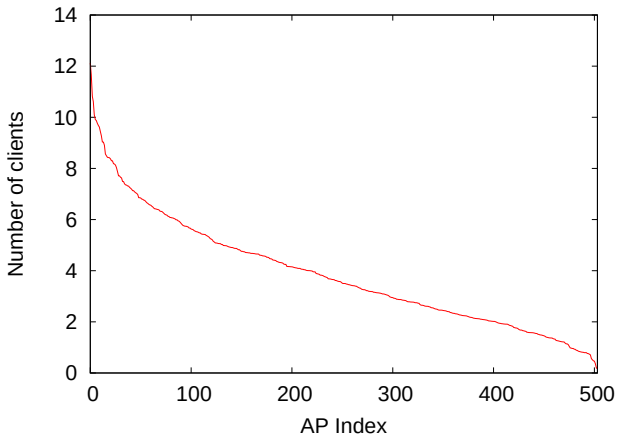


Figure 16: The average number of clients per day for each AP.

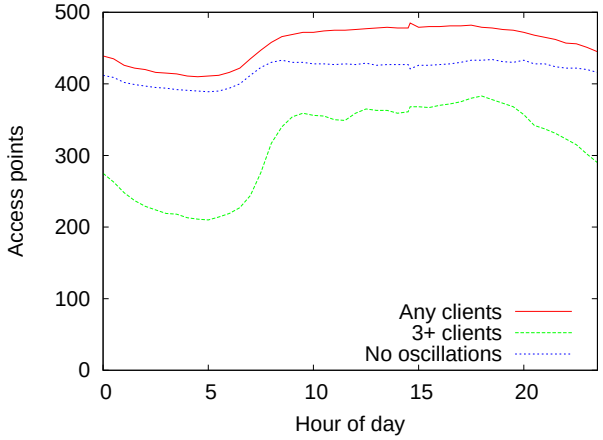


Figure 17: The number of access points in use as a function of time of day, based upon clients served.

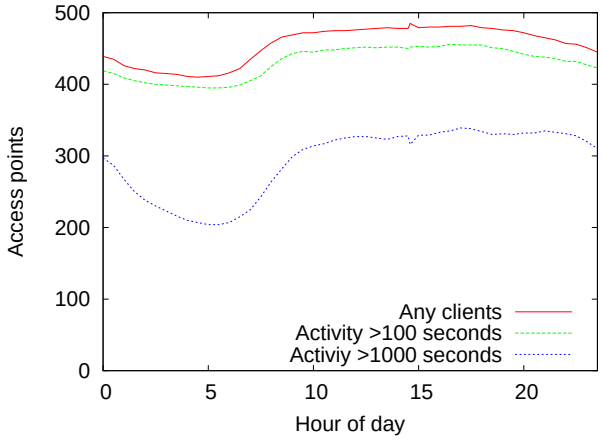


Figure 18: The number of access points in use as a function of time of day for various activity thresholds.

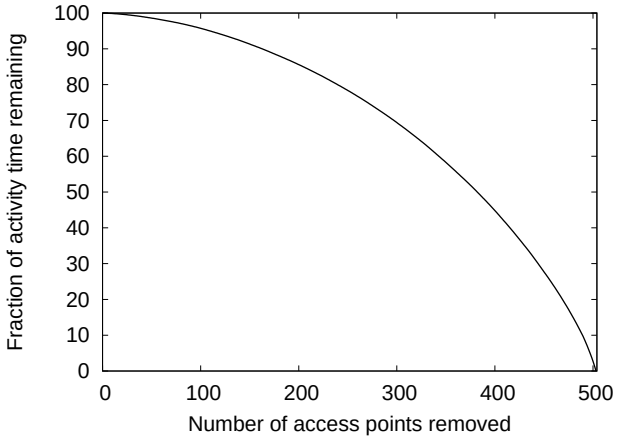


Figure 19: Effect of removing Tropos access points on total network activity time.

are serving at least one client on average. Obviously, by this accounting, all APs contribute substantially to the network coverage.

The number of clients using the network varies by a factor of two over the course of the day, however. Hence, one might expect a similar variation in the number of APs in active use. Figure 17 plots the number of access points in use throughout the day for several definitions of "in use." The "any clients" line shows that even in the dead of night over 80% of the APs are servicing at least one client. The diurnal usage pattern is much more apparent if we consider only heavily used APs, e.g., those with three or more simultaneous clients. Of course, simply removing "lightly used" APs might leave some clients without access. Hence, we plot a final line, "no oscillations," which counts only APs that are servicing one or more clients that have no alternative. Because we do not have access to client-side 802.11 information, we have no way to know definitively if a client has more than one accessible AP at its current location. Here, we consider a client to have an alternative AP if it is currently associated to an AP that has been (or will be) involved in oscillatory behavior at some point that day. In other words, if there exists some client, C, that oscillates between APs X and Y at any point in the day, we consider any clients (not just C) associated with either X or Y to have alternatives. The "no oscillations" line plots the number of APs required to cover the set of clients currently active using our definition of alternatives.

An alternative way to view network coverage is not in terms of client connectivity but rather in terms of aggregate network activity as in Figure 3. Figure 18 replots the "any clients" line from Figure 17, but compares it against the APs that supported at least 100 and 1000 seconds of activity in aggregate per 15-minute interval, respectively. We calculate the total activity time at each AP and sort them in order of increasing activity time, with the least active node first. Figure 19 shows the results of successively removing nodes in sorted order. The x-axis shows the number of access points removed (in sorted order of increasing activity time). The y-axis shows the fraction of all activity time a given set of nodes contribute. At each step, we calculate the fraction of activity time contributed by all of the remaining nodes—the first step corresponds to the activity of all of the nodes, the second to all nodes minus the least active node, etc. Interestingly, we do not find a heavy tail to the curve, indicating that all nodes are relatively active and contribute to useful network coverage throughout Mountain View.

7. CONCLUSION

In this paper, we study the usage of the Google WiFi network, a freely available outdoor wireless Internet service deployed in Mountain View, California. We find that the aggregate usage of the Google WiFi network is composed of three distinct user populations, characterized by distinct traffic, mobility, and usage patterns that are characteristic of traditional wireline, wide-area, and localized wireless access networks. Modem users are static and always connected, and place the highest demand on the network. Hotspot users are concentrated in commercial and public areas, and have moderate mobility. Smartphone users are surprisingly numerous, have peak activity strongly correlated with commute times and are concentrated along travel corridors, yet place very low demands on the network.

8. ACKNOWLEDGMENTS

This work was completed while Mikhail Afanasyev was at Google Inc. The authors thank Chris Uhlik and Bill Coughran at Google Inc. for their continuous support of this study. They are further indebted to Rick Dean at Tropos for assistance with the RADIUS log information and to Brandon Enright, Justin Ma, Stefan Savage, and the anonymous reviewers for comments on earlier versions of this manuscript. This work is funded in part by the UCSD Center for Network Systems (CNS), Ericsson, NSF CAREER grant CNS-0347949, and Qualcomm through the UC Discovery program.

9. REFERENCES

[1] D. Aguayo, J. Bicket, S. Biswas, G. Judd, and R. Morris. Link-level measurements from an 802.11b mesh network. In *Proceedings of ACM SIGCOMM*, Sept. 2004.

[2] M. Allman, K. Christensen, B. Nordman, and V. Paxson. Enabling an energy-efficient future internet. In *Proceedings of HotNets*, Nov. 2007.

[3] A. Balachandran, G. M. Voelker, P. Bahl, and P. V. Rangan. Characterizing User Behavior and Network Performance in a Public Wireless LAN. In *Proceedings of ACM SIGMETRICS*, June 2002.

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

[5] J. Bicket, D. Aguayo, S. Biswas, and R. Morris. Architecture and evaluation of an unplanned 802.11b mesh network. In *Proceedings of ACM Mobicom*, Aug. 2005.

[6] S. Biswas and R. Morris. Opportunistic routing in multi-hop wireless networks. In *Proceedings of SIGCOMM*, Aug. 2005.

[7] V. Brik, S. Rayanchu, S. Saha, S. Sen, V. Shrivastava, and S. Banerjee. A measurement study of a commercial-grade urban WiFi mesh. In *Proceedings of ACM Internet Measurement Conference*, Oct. 2008.

[8] J. Camp, V. Mancuso, O. Gurewitz, and E. Knightly. A measurement study of multiplicative overhead effects in wireless networks. In *Proceedings of IEEE INFOCOM*, Apr. 2008.

[9] J. Camp, J. Robinson, C. Steger, and E. Knightly. Measurement Driven Deployment of a Two-Tier Urban Mesh Access Network. In *Proceedings of ACM MobiSys*, June 2006.

[10] Y.-C. Cheng, M. Afanasyev, P. Verkaik, P. Benkö, J. Chiang, A. C. Snoeren, S. Savage, and G. M. Voelker. Automating cross-layer diagnosis of enterprise wireless networks. In *Proceedings of ACM SIGCOMM*, Aug. 2007.

[11] Y.-C. Cheng, J. Bellardo, P. Benkö, A. C. Snoeren, G. M. Voelker, and S. Savage. Jigsaw: Solving the puzzle of enterprise 802.11 analysis. In *Proceedings of ACM SIGCOMM*, Sept. 2006.

[12] T. Henderson, D. Kotz, and I. Abyzov. The Changing Usage of a Mature Campus-wide Wireless Network. In *Proceedings of ACM Mobicom*, Sept. 2004.

[13] F. Hernández-Campos and M. Papadopouli. A Comparative Measurement Study of the Workload of Wireless Access Points in Campus Networks. In *Proceedings of IEEE PIMRC*, Sept. 2005.

[14] A. P. Jardosh, K. N. Ramachandran, K. C. Almeroth, and E. M. Belding-Royer. Understanding Congestion in IEEE 802.11b Wireless Networks. In *Proceedings of ACM Internet Measurement Conference*, Oct. 2005.

[15] A. P. Jardosh, K. N. Ramachandran, K. C. Almeroth, and E. M. Belding-Royer. Understanding Link-Layer Behavior in Highly Congested IEEE 802.11b Wireless Networks. In *Proceedings of ACM E-WIND*, Aug. 2005.

[16] D. Kotz and K. Essien. Analysis of a Campus-wide Wireless Network. In *Proceedings of ACM Mobicom*, Sept. 2002.

[17] C. R. Livingston. Radius accounting. RFC 2866, IETF, June 2000.

[18] R. Mahajan, M. Rodrig, D. Wetherall, and J. Zahorjan. Analyzing the MAC-level Behavior of Wireless Networks in the Wild. In *Proceedings of ACM SIGCOMM*, Sept. 2006.

[19] M. McNett and G. M. Voelker. Access and Mobility of Wireless PDA Users. *Mobile Computing and Communications Review*, 9(2), 2005.

[20] K. N. Ramachandran, E. M. Belding-Royer, and K. C. Almeroth. DAMON: A Distributed Architecture for Monitoring Multi-hop Mobile Networks. In *Proceedings of IEEE SECON*, Oct. 2004.

[21] B. Raman and K. Chebrolu. Design and evaluation of a new MAC protocol for long-distance 802.11 mesh networks. In *Proceedings of ACM Mobicom*, Aug. 2005.

[22] B. Raman and K. Chebrolu. Experiences in using WiFi for rural internet in India. *IEEE Communications Magazine, Special Issue on New Directions In Networking Technologies In Emerging Economies*, 45(1):104–110, Jan. 2007.

[23] J. Robinson, R. Swaminathan, and E. Knightly. Assessment of urban-scale wireless networks, with a small number of measurements. In *Proceedings of ACM Mobicom*, Sept. 2008.

[24] M. Rodrig, C. Reis, R. Mahajan, D. Wetherall, and J. Zahorjan. Measurement-based Characterization of 802.11 in a Hotspot Setting. In *Proceedings of ACM E-WIND*, Aug. 2005.

[25] D. Schwab and R. Bunt. Characterising the Use of a Campus Wireless Network. In *Proceedings of IEEE Infocom*, 2004.

[26] S. Sen and B. Raman. Long Distance Wireless Mesh Network Planning: Problem Formulation and Solution. In *Proceedings of World Wide Web Conference*, May 2007.

[27] D. Tang and M. Baker. Analysis of a local-area wireless network. In *Proceedings of ACM Mobicom*, Aug. 2000.

[28] D. Tang and M. Baker. Analysis of a metropolitan-area wireless network. *Wireless Networks*, 8:107–120, 2002.

[29] S. Thajchayapong and J. M. Peha. Mobility Patterns in Microcellular Wireless Networks. In *Proceedings of IEEE Wireless Communications and Networking*, Mar. 2003.