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. In this paper, we study the usage of the Google WiFi network deployed in Mountain View, California. We 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.

1. INTRODUCTION

While WiFi was initially designed as a local-area access network, mesh networking technologies have led to increasingly expansive deployments of WiFi networks. Indeed, a number of municipalities have deployed city-wide WiFi networks over recent years. 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 such devices operate on. In urban environments, the WiFi mesh frequently supplements a number of existing access technologies, including wired broadband networks (cable, DSL, etc.), 3G cellular (EVDO, EDGE, etc.), and commercial WiFi hotspots. It is an open question what role city-wide WiFi deployments play in the increasingly diverse access network spectrum.

In this paper, we study the usage of the Google WiFi network, a freely available outdoor wireless Internet service deployed in Mountain View, California, and operational since August 2006. The network consists of over 500 Tropos MetroMesh pole-top access points and serves up to 2,500 simultaneous clients at a time. Using 27 days of overall network statistics in Spring 2008, we analyze the temporal activity of clients, the applications they use and 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] These user populations naturally fall into three classes based almost entirely on client device type. 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. And with a metropolitan WiFi network, smartphone users combine the ubiquitous coverage of cellular networks with the higher performance of wireless LANs. 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.

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 patters 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 focused on network architecture [5], MAC protocol development [18], routing protocol design [6], and network planning and provisioning [22], with measurement targeted to evaluating the

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

performance and reliability of the network itself [1]. Community and commercial mesh networks typically serve as multi-hop transit between homes, businesses, and public locales and the Internet. Mobility is possible, but not necessarily an intended feature; as such, network use tends to be similar to use with DSL or cable modem service. Rural networks in developing regions typically support *targeted application services* [19], such as audio and video conferencing to provide remote medical treatment, and consequently have application characteristics specific to their intended use. The static access users in our study are similar to users of community and commercial backbone mesh networks. 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 "campus" wireless LAN has been measured most extensively by the research community. Numerous studies of indoor 802.11 networks have covered a variety of environments, including university departments [7, 8, 23], corporate enterprises [4], and conference and professional meetings [3, 11, 12, 15, 17, 20]. 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 [9, 10, 13, 16, 21, 25]. The laptop user base in our study most closely resembles these outdoor campus user populations, both in the dominant applications used and the relatively limited user mobility.

The dominant presence of iPhone users represents 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 [24] 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.

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 (WPA/802.1x, SSID GoogleWiFiSecure) 802.11 clients. Aside from the standard

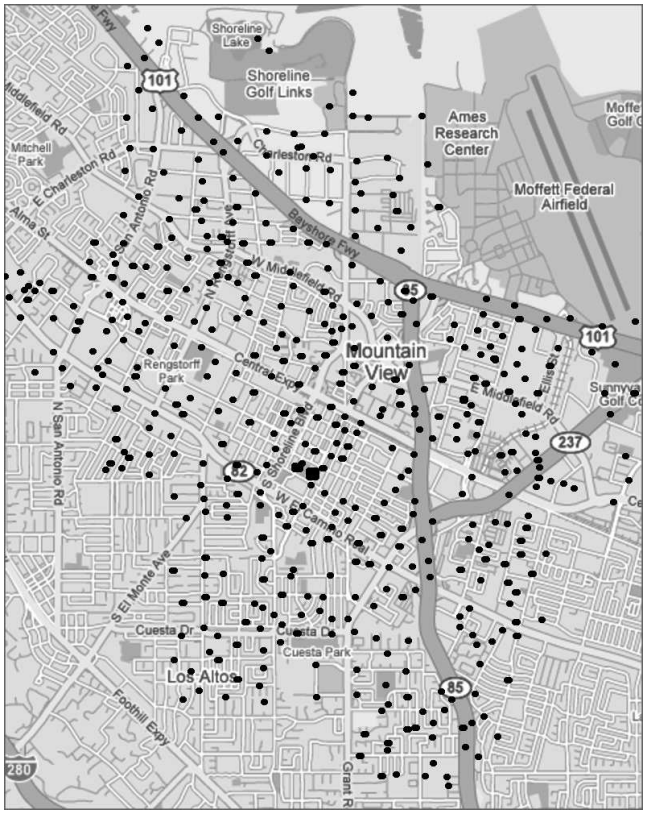


Figure 1: The Google WiFi coverage map.

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 1 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 cen-

[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.

tralized authorization and authentication gateway. Google provides single sign-on authentication and authorization service, but, at the link layer, 802.11 client devices continue associate with each Tropos AP individually. All Tropos nodes support the RADIUS accounting standard [14] and provide periodic updates of client activity to the central server.

3.1.1 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

In this study, we analyze a trace of 27 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 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 the first packets of each flow for the first fifteen minutes of each hour. Because the trace was collected at the gateway—as opposed to inside the wireless mesh itself—we do not observe layer-two protocol traffic such as ARP, nor many DHCP requests which are handled by by the Tropos nodes themselves.

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 easily 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 confirms the bug, and informs 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 off 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 different 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.

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.

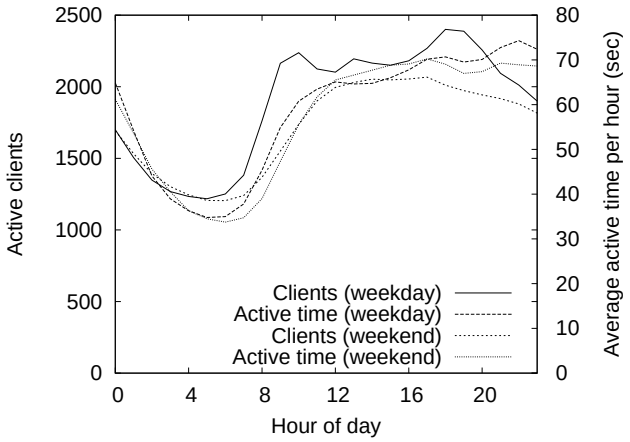


Figure 3: Average daily use of the Google WiFi network.

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.

In order to estimate the population of iPhone devices, we leverage the fact 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 `wu.apple.com` A record, 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 that conclude 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 recommended 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 mis-classified (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.)

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 2 shows the number of active clients using the network (left y-axis) and their average activity time (right y-axis) per fifteen-minute interval for the entire trace. In our analyses, we consider a client to be *active* for a fifteen 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

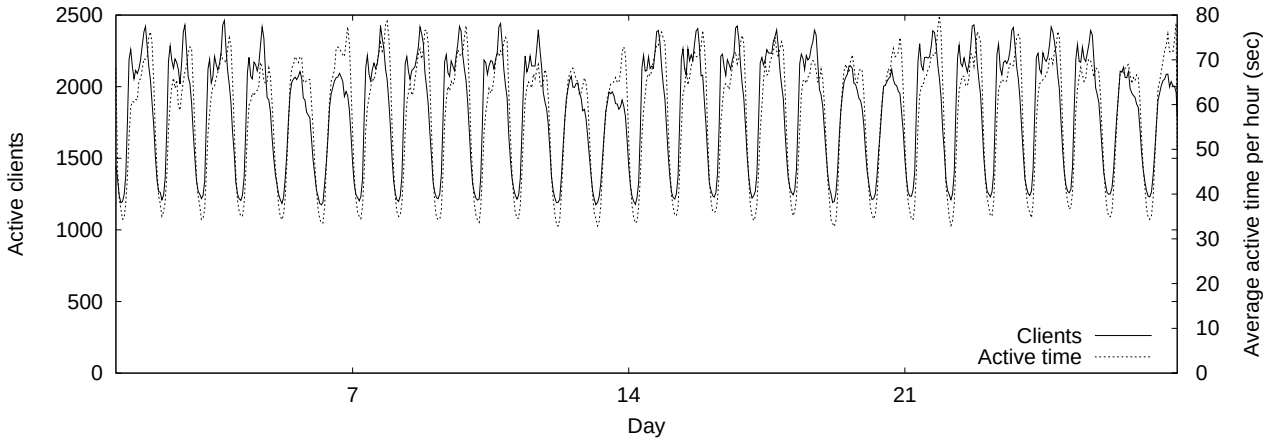


Figure 2: Usage of the Google WiFi network for the duration of the trace, measured in 15-minute intervals.

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 3 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 (6pm); 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. This behavior is due to a combination of the kinds of clients who are using the network and how they use it.

Figure 4(a) similarly shows the daily variation of client usage on weekdays as in Figure 3, 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 9am until 11pm at night.

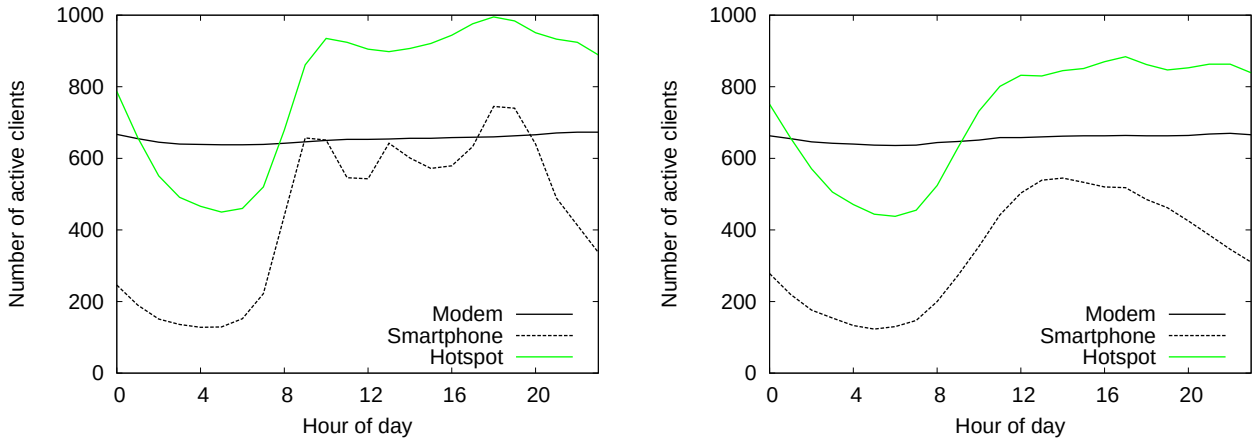
Smartphone users show the most interesting variation over time. The curve shows three distinct peaks during the day (9 am, 1 pm, and 6 pm), suggesting that smartphone usage is highly correlated with commute and travel times and that the devices are active while users are mobile (Section 5 explores mobility 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. The period of high activity is slightly shorter than hotspot users (9 am to 8 pm).

Figure 4(b) similarly shows the number of active clients by device type as Figure 3, 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 fewer hotspots users during the highly active period than on the weekday, the period of high activity remains similar. Smartphone users, however, again exhibit the most notable differences. Smartphone peak usage no longer correlates with commute times, peaking at midday (1pm) 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 5 plots the CDF of total amount of data transferred (the sum of upload and download) by clients of each class



(a) Weekdays (b) Weekend

Figure 4: Hourly usage of the Google WiFi network, broken down by day of the week.

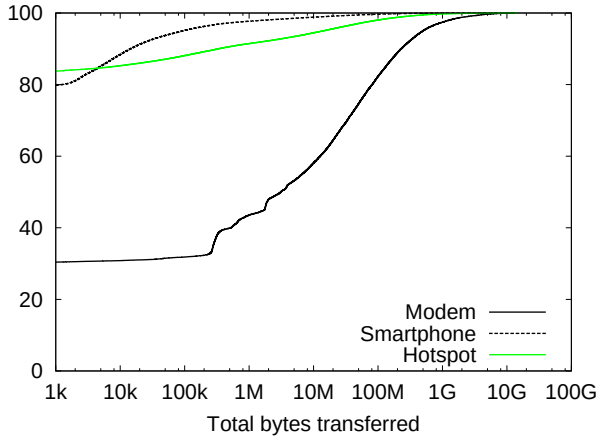


Figure 5: Total bytes transferred (in and out) by each client per day. (Note the log-scale x axis.)

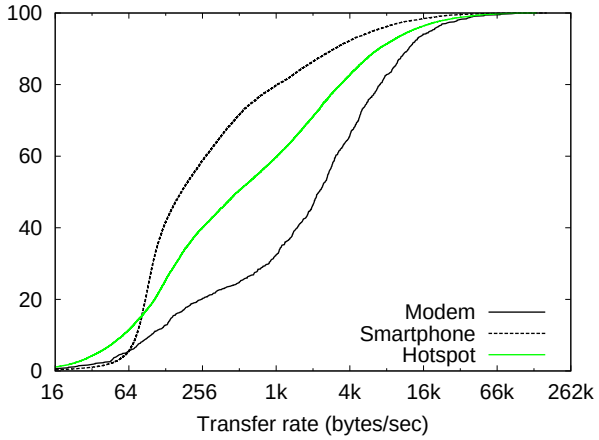


Figure 6: Instantaneous transmission rates during activity periods, broken up by client type.

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.

Figure 6 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 three 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 very little impact on extended 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 are uniformly 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, and the remaining 20% have higher rates.

4.2.1 Sessions

Next we characterize how long clients are active when associated with the network. We observed up to 379 distinct sessions per client, with the median client connecting only twice and a full 35% appearing only once. Almost 7% of

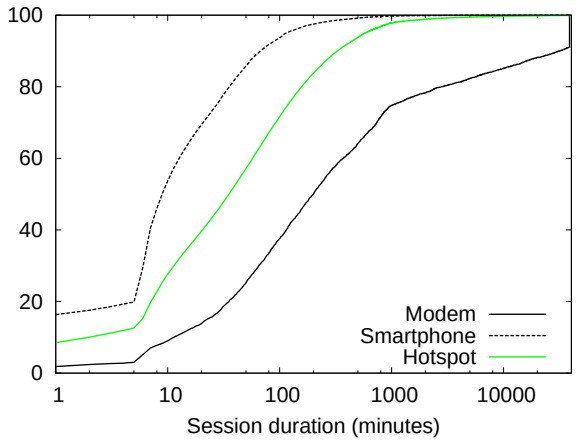


Figure 7: CDF of session lengths, in minutes (x axis in log scale). The entire trace is only 40,320 minutes long.

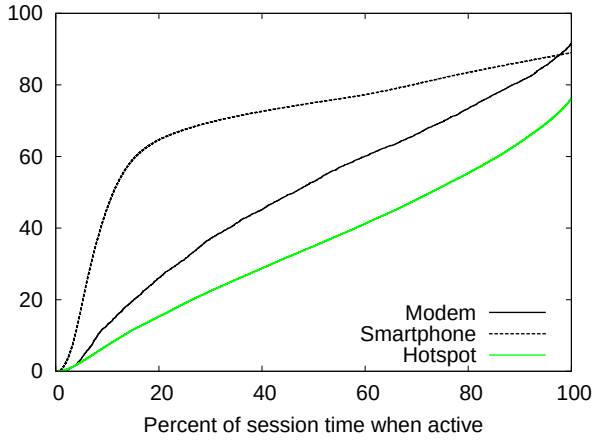


Figure 8: Percentage of the session during which the client was active.

clients connected at least once per day, on average, and more than 10% connected at least once per weekday (20 times).

Figure 7 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 also exhibit different session length distributions. A significant fraction of modem clients have sessions that span the entire trace; although 65% of those 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. But a substantial fraction are rather long, with 30% of the sessions longer than two hours. Smartphone clients have the shortest session lengths. Over half of 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 8 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.

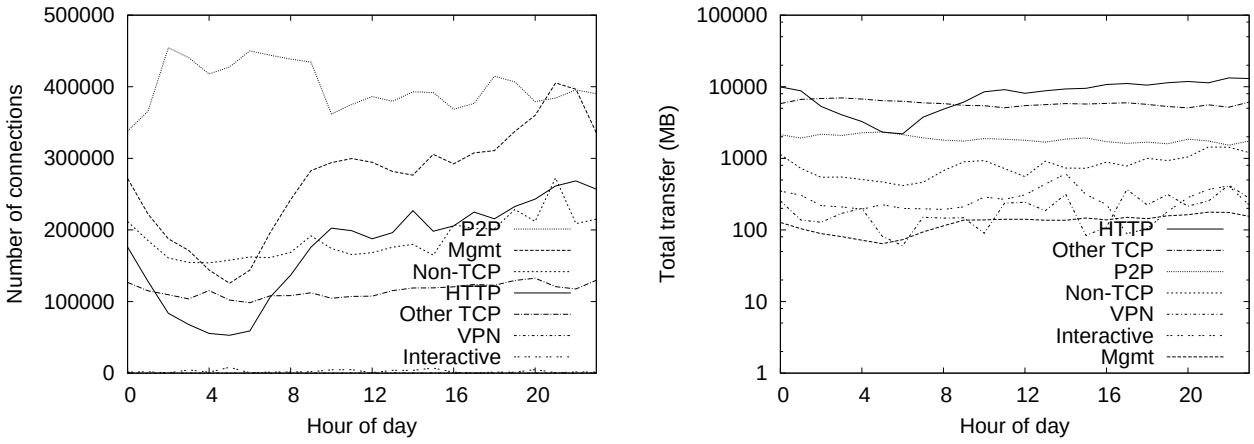
4.2.2 Application classes

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 packets of each flow based on protocol and port numbers. Figure 9(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 most unusual feature is the dramatic variation in the frequency of management (ICMP, DHCP, and DNS) connections. It turns out that the vast majority of this traffic is is actually mDNS "dnsbugtest" traffic. In fact, Figure 10(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 they turn out to be relatively heavy.

The picture for bytes is similar. Figure 9(b) plots the total amount of data transferred in the network as a function of hour 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



(a) Number of connections

(b) Bytes transferred

Figure 9: Hourly usage of the network per application class.

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 10 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) and VoIP traffic, but further analysis is required.

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 6). 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 "dnsbugtest" mDNS requests, some Ruckus devices issue thousands of queries during each 15-minute interval. The precise cause of this behavior deserves further investigation.

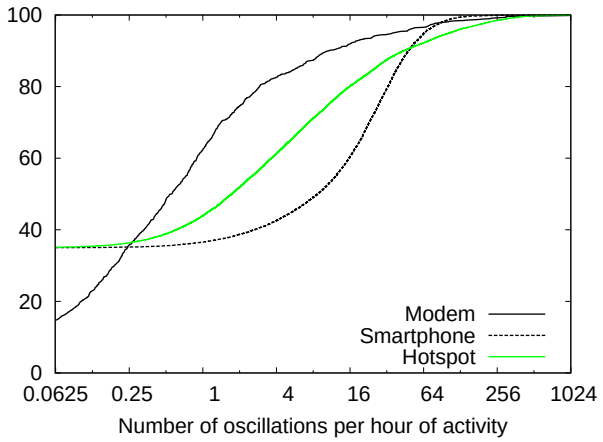


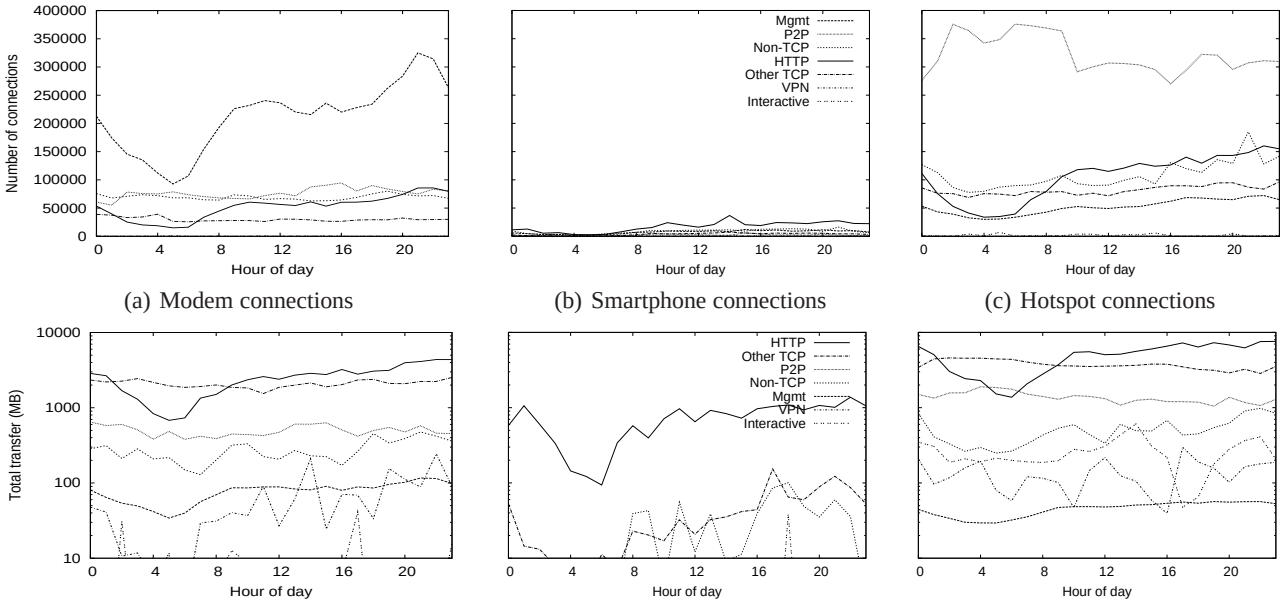
Figure 11: CDF of the number of oscillations per hour (x axis is log scale).

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 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. If a new asso-



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

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

Figure 10: Number of connections (a–c) and bytes (c–e) per hour for each device type.

ciation is to one of the previous three most recent APs, we consider it an oscillation. Using this definition, we detect a high frequency of oscillations in the data.

Figure 11 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 2900 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, although the extreme tail is heaviest for hotspot users. To more accurately capture physical movement—as distinct from RF movement due to changes in signal strength—we eliminate oscillations from the association data used in the remainder of this section.

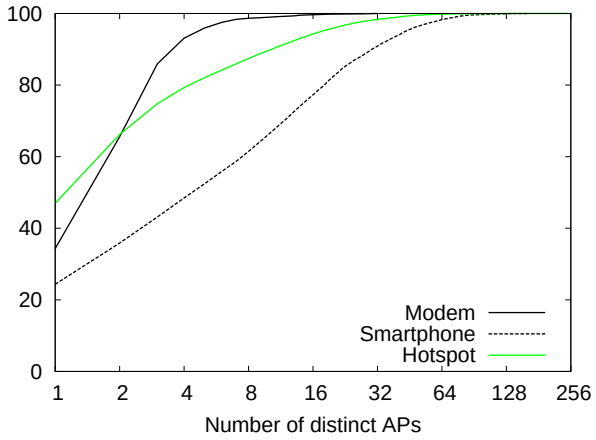


Figure 12: CDF of the number of distinct APs a client associates with over the course of the trace.

5.2 Movement

We plot the number of distinct APs to which a client associates during the course of our trace in Figure 12. 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—he 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. A 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 13 plots the average distance in meters between non-oscillatory client

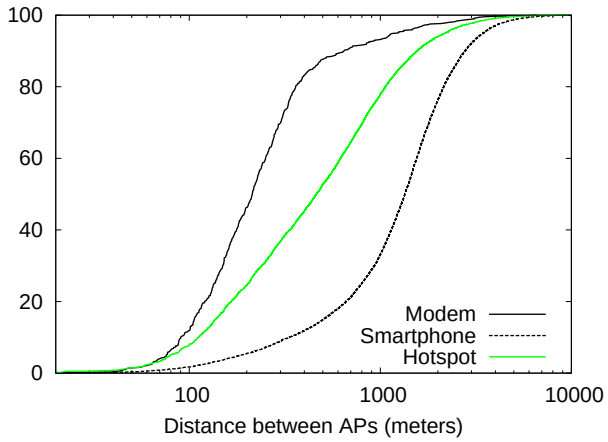


Figure 13: CDF of the average distance between consecutive client associations.

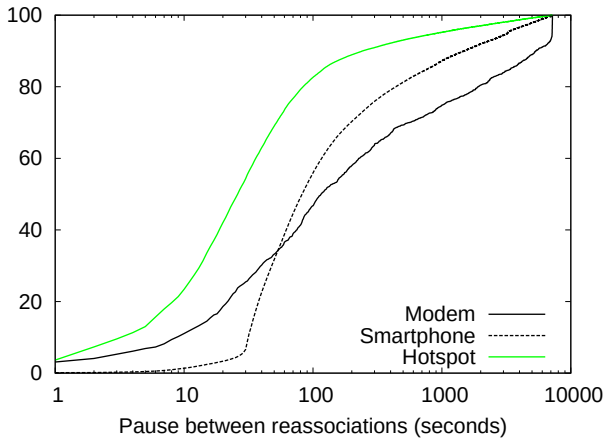


Figure 14: CDF of pause time for each class of client.

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 modems associate to in Figure 12, we conclude that most locations in the city (where WiFi modems are installed) can reach at most four 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, in order to understand how fast clients are moving, we plot the pause time between associations in Figure 14. 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 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, 9–10am, 3–4pm, and 6–7 pm corresponding to the peaks and valleys of Figure 3. 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

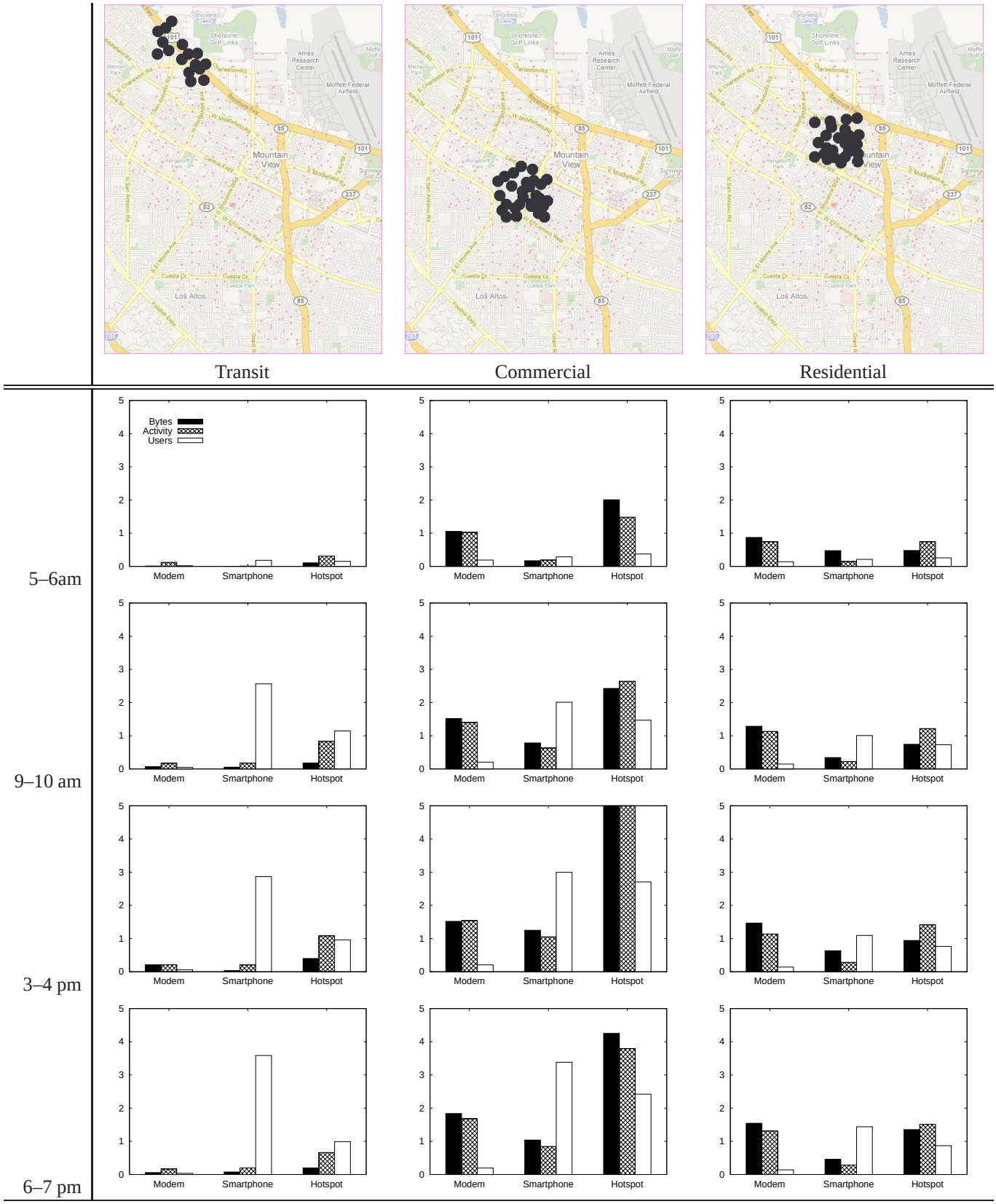


Table 3: Network usage for representative time periods across different parts of the city.

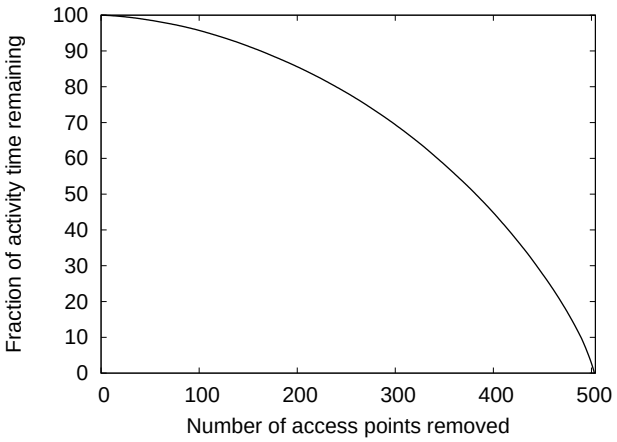


Figure 15: Effect of removing Tropos access points on total network activity time.

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. As a final experiment, we calculated the total activity time for each pole top Tropos node. We then sorted the nodes in increasing activity time, with the least active node first. Starting with all of the nodes, we then successively removed nodes in sorted order. At each step, we calculated the fraction of activity time contributed by all of the nodes together — the first step corresponds to the activity of all of the nodes, the second to all nodes minus the least active node, etc.

Figure 15 shows the distribution of activity time for this experiment. 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. Somewhat surprisingly, 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. 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*, 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*, 2003.
[5] J. Bicket, D. Aguayo, S. Biswas, and R. Morris. Architecture and Evaluation of an Unplanned 802.11b Mesh Network . In *Proceedings of Mobicom*, August 2005.
[6] S. Biswas and R. Morris. Opportunistic Routing in Multi-Hop Wireless Networks. In *Proceedings of SIGCOMM*, August 2005.
[7] 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 the ACM SIGCOMM Conference*, Kyoto, Japan, Aug. 2007.
[8] 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 the ACM SIGCOMM Conference*, pages 39–50, Pisa, Italy, Sept. 2006.
[9] T. Henderson, D. Kotz, and I. Abyzov. The Changing Usage of a Mature Campus-wide Wireless Network. In *Proceedings of ACM Mobicom*, Sept. 2004.
[10] 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*, 2005.
[11] 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 IMC*, Oct. 2005.
[12] 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.
[13] D. Kotz and K. Essien. Analysis of a Campus-wide Wireless Network. In *Proceedings of ACM Mobicom*, Sept. 2002.
[14] C. R. Livingston. Radius accounting. RFC 2866, IETF, June 2000.

[15] 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.
[16] M. McNett and G. M. Voelker. Access and Mobility of Wireless PDA Users. *Mobile Computing and Communications Review*, 9(2), 2005.
[17] 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*, 2004.
[18] B. Raman and K. Chebrolu. Design and Evaluation of a new MAC Protocol for Long-Distance 802.11 Mesh Networks. In *Proceedings of the 11th Annual International Conference on Mobile Computing and Networking (MobiCom'05)*, August 2005.
[19] 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, January 2007.
[20] 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*, 2005.
[21] D. Schwab and R. Bunt. Characterising the Use of a Campus Wireless Network. In *Proceedings of IEEE Infocom*, 2004.
[22] S. Sen and B. Raman. Long Distance Wireless Mesh Network Planning: Problem Formulation and Solution. In *Proceedings of the 16th Annual International World Wide Web Conference (WWW 2007)*.
[23] D. Tang and M. Baker. Analysis of a Local-Area Wireless Network. In *Proceedings of ACM Mobicom*, 2000.
[24] D. Tang and M. Baker. Analysis of a metropolitan-area wireless network. *Wireless Networks*, 8:107–120, 2002.
[25] S. Thajchayapong and J. M. Peha. Mobility Patterns in Microcellular Wireless Networks. In *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC)*, March 2003.