How to Tell an Airport from a Home: Techniques and Applications

Andreas Pitsillidis[⋆∗] Yinglian Xie[‡] Fang Yu[‡]
Martin Abadi[‡†] Geoffrey M. Voelker[⋆] Stefan Savage[⋆]

[⋆] University of California, San Diego
{apitsill,voelker,savage}@cs.ucsd.edu

[‡] Microsoft Research Silicon Valley
{yxie,fangyu,abadi}@microsoft.com

ABSTRACT

Today's Internet services increasingly use IP-based geolocation to specialize the content and service provisioning for each user. However, these systems focus almost exclusively on the current position of users and do not attempt to infer or exploit any qualitative context about the location's relationship with the user (e.g., is the user at home? on a business trip?). This paper develops such a context by profiling the usage patterns of IP address ranges, relying on known user and machine identifiers to track accesses over time. Our preliminary results suggest that rough location categories such as residences, workplaces, and travel venues can be accurately inferred, enabling a range of potential applications from demographic analyses to ad specialization and security improvements.

Categories and Subject Descriptors

C.2.0 [**Computer Communication Networks**]: General—*Security and protection*; C.2.3 [**Computer Communication Networks**]: Network Operations—*Network monitoring*

General Terms

Algorithms, Measurement, Security

1. INTRODUCTION

Designers of Internet services increasingly seek to specialize their behavior and content to reflect differences among their users (e.g., location, bandwidth, reputation, etc.). Location is perhaps the most widely examined of these features and has been used to help address system problems ranging from load balancing [5] to email reputation [10], as well as to enable entirely new classes of capabilities, such as geographically-specialized advertising and proximity-oriented social networking.

However, much of the technical effort in this area has focused on accurately inferring location from available identifiers (either from IP addresses [3, 9, 11, 13] or WiFi mapping [2, 4, 7]) and comparatively little thought has been devoted to the "meaning" of these locations.[1] For example, a given user might access an Internet service from four different locations: from their home, from their business, from a hotel on a business trip, or from a cafe. While today we can use commercial IP geolocation tools to map each of these accesses to city-level positions, the salient context surrounding these positions is lost.

We argue that such "location context" has a broad range of uses and can add value for researchers, service providers, and application developers alike. For example, restaurants will likely find value in providing geolocalized advertising to all users, while plumbers might limit such specialization to users accessing the Internet from their "home" city. Similarly, data center operators might consider migrating user account state for users whose residence has changed, but a user accessing the same account from afar while traveling might not merit the overhead of such migration.

In this paper we explore the proposition that "location context" can be inferred based on the dynamics of user mobility. In particular, we argue that rough location characterizations such as "residence", "workplace", or "travel venue" (e.g., hotel, airport, cafe, etc.) can be automatically derived for many IP address ranges. Moreover, we show that this classification can be achieved without significant infrastructure or cooperation from network operators. Instead, we rely purely on the analysis of Internet service logs (in particular logs of email logins and automated software updates) combined with existing IP geolocation tools.

[∗]The work was done at Microsoft Research Silicon Valley.

[†]Martín is also affiliated with the University of California, Santa Cruz, and the Collège de France.

[1]Some previous efforts have sought to extract additional context from IP addresses, including proxy status [6], dynamic address assignment [15], accurate mapping to hosts or users [14], and identifying offered services [12]; however, none of these projects have concerned location or indeed any issues at a level of abstraction visible to typical users.

Hotnets '10, October 20–21, 2010, Monterey, CA, USA.

The core idea behind our approach is simple: that the "kind" of location at a particular IP address is implicitly revealed by the mobility patterns of the users who operate there. For example, address ranges from which the same users appear consistently are more likely to be residences or workplaces, while address ranges that host a large number of distinct users who are not seen again are more likely to represent Internet infrastructure for transient users (e.g., airports or cafes). Conversely, a user who normally operates out of a "residence" IP block, but is later seen to send requests from a "travel" IP range in a major city hundreds of miles away can be inferred to be traveling. Despite confounding factors in such analyses (e.g., inaccuracies of IP geolocations, VPNs, or Satellite VSAT links), we find that mining across a large number of users can potentially provide fairly robust results.

Finally, we describe a range of use cases for location context, from pure demographic research—tracking aggregate population migration over time—to concrete applications for data migration, security, and specialized search/advertising.

2. CATEGORIZING IP ADDRESSES

Our method for deriving home and travel IP addresses relies on mining user online activity logs. In this section, we describe our datasets, our current methods, and initial results with validations.

2.1 Dataset Description

We use three large data sets for our study: a set of update events for a large, ubiquitously deployed software package (referred to as *Software Update*), collected from Aug 2009 to Jan 2010, and two sets of user login events to a popular Web mail provider, collected from Aug to Sep 2008 (*Email-2008*) and Aug to Sep 2009 (*Email-2009*), respectively. The data sets contain sampled records with anonymized host or user IDs, IP addresses, and activity timestamps. For simplicity, we refer to both types of IDs as *users* hereafter. Thus the mobility results in the software update and email datasets reflect host mobility and user mobility, respectively.

As a basis for categorizing IP addresses, we use data from Quova [3], which provides the mapping of IP addresses to geolocation. We also determine distances between events by calculating the geographic distance of the corresponding IP addresses using the Quova data. Since we found that Quova's geolocation results were either unavailable or inaccurate for IP addresses outside of the United States, we restrict our analysis to events within the US. Further, to ensure that we have a sufficient number of events to reason about the mobility patterns of each user, we focus on only the set of users for which we have at least 15 events from IP addresses within the US. Table 1 shows the number of events and unique users in the traces after this filtering.

2.2 Classification Methods

Without prior knowledge of user travel activities, we begin with inferring IP addresses of home locations, defined as *home IP addresses*. We use this information for the subsequent inference of IP addresses used at work places (*work IP addresses*) and those used at travel locations (*travel IP addresses*). During the process, we take steps to minimize errors introduced by the inaccuracies of the geolocation data, which is unavoidable in some cases. Since our main purpose is to illustrate the type of classification one can derive, we choose to be conservative in our current methods, rather than obtaining an exhaustive list of IP addresses of each type.

	Total events	**Unique users**
Software Update	0.5×10^9	4.0×10^6
Email-2009	8.4×10^9	16.7×10^6
Email-2008	7.5×10^9	29.4×10^6

Table 1: Events and users after filtering.

(1) Identifying Home IP Addresses. We begin our classification by first deriving home IP addresses. For now, we broadly refer to home IP addresses as those that are associated with users in their general home city and metropolitan area. Later we perform further fine-grained classification of these IP addresses into work IP addresses as well.

In each trace we process each user's events individually to identify home IP addresses. Because of DHCP effects and local travel, home IP addresses could change frequently. However, they should remain in the same location both topologically and geographically. So for each IP address, we first identify its BGP prefix as well as its geographic location (city and state names), and store them in a <BGP, Location> pair. We then identify the frequent locations accessed by the user by counting the number of days a given <BGP, Location> pair appears in the trace. We use a sliding window of size 30 days, advanced by an increment of one week. If we observe at least 15 active days within such a window, we record the particular pair as a home location candidate. At the end of our processing, we tag all IP addresses associated with home location candidates as *candidate home IP addresses* for that particular user.

(2) Identifying Travel IP Addresses. After having obtained the home IP address information, we proceed to infer travel IP addresses. For each event that occurred more than 250 miles away from its user's home location, we consider the IP address associated with the event a *candidate travel IP address* for that user. The 2009 NHTS report [1] estimates that the largest average traveling distance per person and per transportation mode, for work reasons, is roughly 100 miles. Thus we conservatively set our threshold to 250 miles, which is also bigger than the diameter of any large city in US.

(3) Filtering Proxy/VPN IP Addresses. We found that a small number of proxy or VPN IP addresses introduce false IP geolocation information, confusing further analysis. For instance, social networking sites may automatically log users into user-provided email accounts. Such login events correspond to the IP addresses of the social networking sites, not the users. To identify these IP addresses, we calculate a "miles per hour" (mph) metric between two IP addresses that

were used consecutively in time by a user. The intuition is that a user's physical travel speed has an upper bound (set to 500 mph, a typical commercial airplane speed). If a transition between a travel and a home IP is faster than the upper bound, then no physical travel is actually involved and we remove the IP address from consideration.

(4) Pruning Candidate Sets. We then resolve the inconsistency between candidate home and travel IP sets, where different users may place the same IP address into different categories. Such inconsistencies could arise for a variety of reasons. For instance, IP addresses can have inaccurate geolocation information and it may be inherently difficult to pinpoint their locations (e.g., satellite IP addresses). Also, inferring based on individual user activities may create inaccuracies. For example, we may misclassify the home and travel addresses for a user who travels to a location for an extended period of time.

To increase the confidence level of our classification, for each IP address we look at its user population and examine the degree of consensus on its classification. If (1) at least two users consistently tag this IP address as a home (or travel) IP *and* (2) more than half of its population tags this IP address in the same way, then we conservatively regard our initial classification as accurate. Otherwise, we prune the IP address and the corresponding ranges (from the BGP table) from the final home (or travel) IP address set.

(5) Identifying Work IP Addresses. Finally, we differentiate IP addresses used at workplaces from those used at residences, as a first look into the possibility of performing such fine-grained classification. This differentiation provides the benefit of reflecting user commute patterns.

We find that 7–8% of users have more than one initial home IP range. For these users, some of the ranges may represent actual residences, while others may correspond to work locations. To distinguish these two possibilities, we perform day-of-week analysis, as work IP addresses tend to be used only during work days, while home addresses are mostly used at night and during weekends. Because users are from different time zones, we analyze only the weekday and weekend patterns. We first select home IP address pairs that (1) are from different BGP ranges and (2) are both accessed on a single day. For the corresponding IP range pairs, if such daily access patterns repeat for more than one day, then we tag any range that has been active for at least 6 days out of a 7-day window as a residence IP range; we also use the consensus heuristic above to further increase confidence. We classify any other range as a work IP range.

2.3 Analysis and Validation

We apply our method to all three data sets. Table 2 shows the number of home and travel IP addresses identified in each data set. Our method can conservatively identify millions of home IP addresses across all three datasets.[2] Not surprisingly, the numbers of travel IPs are one order of magnitude smaller. Tables 3 and 4 show the top home and travel cities by number of users. We observe that the top home cities all have large populations. The top travel cities, however, also include cities with major airports (e.g., Dallas and Denver) and high-tech companies (e.g., San Jose).

	Software Update	Email-2009	Email-2008
Home	4,419,532	1,509,617	4,088,555
Travel	64,963	126,244	284,818

Table 2: Number of inferred home and travel IP addresses in each data set.

		Software Update	Email-2009
Home	Desktop	51.9%	48.9%
	Laptop	48.1%	51.1%
Travel	Desktop	7.9%	6.6%
	Laptop	92.1%	93.4%

Table 5: Desktop and laptop breakdown.

	Software Update		Email-2009	
	Home	Travel	Home	Travel
/24s	882	16	1,107	107
Full IPs	11,050	20	5,797	127

Table 6: Overlap with known residential broadband network IP addresses.

Since the three input data sets have similar properties, we use only the Email-2009 dataset for deriving work IP addresses. In total we identified 12,431 such addresses.

Next we validate our results. To our knowledge, there are no publicly available data sets we can use to verify our classification of IP addresses based on usage. Instead, we perform two tests on our results as sanity checks.

Machine Types. Our software update data set contains information on the brands and models of the computers that performed updates. Given that travel IP addresses are more likely associated with mobile devices such as laptops, we first compare the types of machines observed on home and travel IP addresses. Since the software update data set is collected in 2009, we perform this check on the results derived from the Software Update and Email-2009 data sets only.

Table 5 confirms our expectation that the majority (92–93%) of the machines appearing at travel IPs are indeed laptops. In contrast, for home IP addresses we do not see significant differences between desktops and laptops. Note that it is plausible to see a small percentage of desktops on travel IPs because travel locations can contain desktop machines, e.g., a hotel lobby machine shared amongst many guests.

Known Residential Broadband IPs. We also obtained

[2]The Email-2008 and Email-2009 data sets were collected using different sampling rates, which resulted in capturing data for fewer users in the latter. Consequently, we identified fewer home IP addresses from the Email-2009 dataset. The big increase of events per user shown in Table 1 was due to activity from a number of popular online services present only in the Email-2009 dataset. These services act as aggregators and leverage email credentials for authentication, thus do not represent actual Web mail login attempts.

Software Update			Email-2009			Email-2008		
Location	# Users	% All Users	Location	# Users	% All Users	Location	# Users	% All Users
New York, NY	91740	3.1%	New York, NY	120732	3.2%	New York, NY	295441	3.7%
Chicago, IL	68879	2.4%	Los Angeles, CA	98625	2.6%	Los Angeles, CA	207864	2.6%
Los Angeles, CA	62623	2.1%	Chicago, IL	76005	2.0%	Phoenix, AZ	148192	1.8%
Atlanta, GA	53146	1.8%	Phoenix, AZ	74671	2.0%	Tampa, FL	137344	1.7%
Phoenix, AZ	44174	1.5%	Tampa, FL	69755	1.8%	Washington, DC	125269	1.6%
Washington, DC	43890	1.5%	Washington, DC	61205	1.6%	Chicago, IL	121614	1.5%
Tampa, FL	42657	1.5%	San Diego, CA	48901	1.3%	San Diego, CA	106694	1.3%
Houston, TX	40493	1.4%	Atlanta, GA	42709	1.1%	Philadelphia, PA	89093	1.1%
Philadelphia, PA	33548	1.2%	Philadelphia, PA	40754	1.1%	Atlanta, GA	82715	1.0%
San Diego, CA	32007	1.1%	Salt Lake City, UT	38245	1.0%	Seattle, WA	82383	1.0%

Table 3: Top home cities by number of users.

Software Update			Email-2009			Email-2008		
Location	# Users	% All Users	Location	# Users	% All Users	Location	# Users	% All Users
Atlanta, GA	8919	7.2%	Atlanta, GA	14501	5.9%	Chicago, IL	30501	6.2%
Dallas, TX	7124	5.7%	Chicago, IL	12450	5.1%	Atlanta, GA	28956	5.9%
Chicago, IL	6939	5.6%	Dallas, TX	11768	4.8%	Dallas, TX	24629	5.0%
New York, NY	6376	5.1%	Sacramento, CA	11488	4.7%	Sacramento, CA	24176	4.9%
Denver, CO	5089	4.1%	Newark, NJ	10549	4.3%	San Jose, CA	23794	4.9%
Newark, NJ	3875	3.1%	San Jose, CA	10255	4.2%	Newark, NJ	18534	3.8%
San Jose, CA	3819	3.1%	New York, NY	9508	3.9%	Ashburn, VA	16324	3.3%
Sacramento, CA	3780	3.0%	Ashburn, VA	8232	3.3%	New York, NY	13021	2.7%
Ashburn, VA	3506	2.8%	Denver, CO	7130	2.9%	Fort Worth, TX	12484	2.6%
Orlando, FL	3137	2.5%	Orlando, FL	5240	2.1%	Phoenix, AZ	12026	2.5%

Table 4: Top travel cities by number of users.

a small set of verified residential broadband network IP addresses provided by [8]. This dataset consists of 3,613 IP addresses anonymized by removing the last 8 bits of each address, spanning 2,099 distinct /24 subnets.

We evaluate the overlap of our results with this known set of addresses in the following two ways. First, since the provided data is essentially /24 IP blocks, we convert our result sets into the same format and determine their overlap at this granularity. Second, we expand the broadband residential IP set to include all individual addresses in the /24 range and then examine the overlap. Again, we compare only the IP sets derived using the 2009 data sets. Table 6 shows the comparison results. As expected, there is a significant difference between the home and travel IP sets: only a small number travel IP addresses overlap with the broadband residential IP set, yet the number of overlapping home IP addresses is orders of magnitude more. This difference suggests that we correctly identified home IP addresses.

While we strive to maximize our confidence in the validity of the current results, we acknowledge that finer-grained classifications are likely to be possible.

3. USER MOBILITY

In this section, we perform an initial study on user mobility to illustrate the type of mobility patterns one might derive from our classification.

3.1 Long-term Mobility

The concept of home locations for users enables us to examine long-term mobility trends as users change their home locations—literally move home—over long time scales. Understanding such long-term mobility can provide a global view of the user population for tasks such as resource provisioning and location-based feature planning. Although ISPs may derive their own customer moving trends with user profiles or network traces, their data points are often limited to a single administrative domain. Government agencies also collect census data to study relocation trends. However, a census requires great manual effort and is only feasible infrequently.

Origin	Destination	# Users	% All Users
New York	Illinois	561	4.3%
Texas	West Virginia	551	4.2%
Colorado	Arizona	539	4.1%
California	Illinois	499	3.8%
Washington	Texas	440	3.4%
Montana	Colorado	348	2.7%
Virginia	Illinois	286	2.2%
Texas	Florida	224	1.7%
New York	Florida	201	1.5%
Wyoming	Montana	156	1.2%

Table 7: Top location pairs for long-term migration.

We use the email data sets to identify users whose home locations changed from one year to the next. Referring back to Table 3, we see that New York and Los Angeles remain the top two cities in 2008 to 2009 with the largest number of user home IP addresses, while Chicago's number increased significantly. Further examining the home IP relocation patterns for individual users, Table 7 shows the top pairs of location changes over the year in terms of movement between states. To conclude that the change was a relocation, we use a conservative threshold of at least 500 miles in the dis-

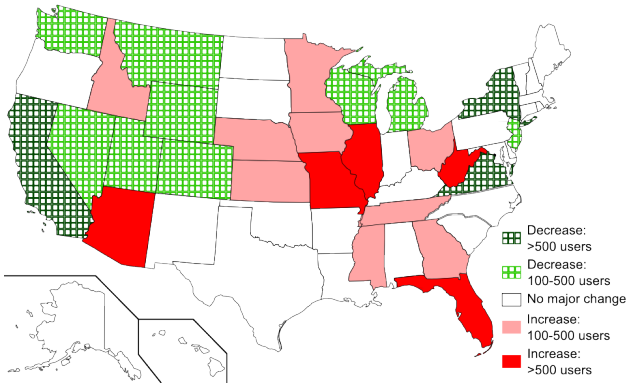

Figure 1: Long-term relocation patterns across states.

Origin	Destination	# Users	% All Users
Los Angeles, CA	San Jose, CA	491	0.8%
Atlanta, GA	Tampa, FL	432	0.7%
Chicago, IL	Minneapolis, MN	399	0.6%
Chicago, IL	New York, NY	381	0.6%
Phoenix, AZ	San Jose, CA	336	0.5%
Ashburn, VA	Los Angeles, CA	322	0.5%
Atlanta, GA	Dallas, TX	314	0.5%
Atlanta, GA	Orlando, FL	300	0.5%
San Diego, CA	San Jose, CA	287	0.5%
Chicago, IL	Saint Louis, MO	282	0.4%

Table 8: Top short-term travel pairs.

tance home locations changed. We see that Illinois (mostly Chicago) indeed is a popular relocation destination. More generally, Figure 1 shows that the user population we study exhibits a coast-to-inland moving trend from 2008 to 2009, perhaps because of reduced job opportunities in locations such as New York and California.

3.2 Short-term Mobility

Next we examine short-term user mobility, where users travel temporarily and return to their home locations later. Such short-term travel patterns are often useful for applications that would benefit from user population profiles, e.g., targeted advertisement. Figure 2 shows that close to 40% of users travel no further than 500 miles, suggesting strong locality of short-term mobility.[3] Interestingly, Table 8 shows that the top travel route is from Los Angeles to San Jose. On the other hand, Chicago and Atlanta are popular travel sources, possibly because of a combination of large populations and major travel hubs.

4. APPLICATIONS

Since we perform our inference at the IP-address level, many services can easily take advantage of this information for processing requests from users or remote hosts. In this section, we outline a few applications, focusing on search and advertisement and on security. Many other applications

[3]The results obtained by using different input data sets are very similar, so we present only the measurements from the Email-2009 data set here.

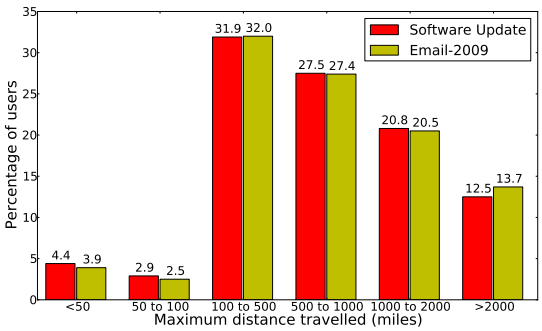

Figure 2: Maximum distance traveled.

may be viable. For instance, strategies for data migration and caching might benefit from the categorization: users typically spend just a few hours at airports, while they could spend days at hotels and years at home locations, and these different time scales may inform the decision of whether to migrate user data.

4.1 Search and Advertisement

The classification of home and travel IP addresses provides new opportunities for online applications to understand user profiles. Based on the location categories of a user, we could infer their interests or intentions to customize search results and to target advertisement. For example, travel-related links may have higher rankings when the query comes from a travel IP address.

We have verified, as an example, that search patterns from known airport IP ranges are different from those in nearby residential IP ranges. We examined the search log from a large search engine, computed the percentage of users that make queries containing ten travel-related keywords including "flight", "airline", and "travel", and observed three times higher percentages at the airports.

Advertising can also be more specialized based on the location of the user; it is for example less likely that users are interested in home repair services, if they are utilizing travel IPs. In a similar manner, local attractions can be of more interest to such users.

Using two weeks of query and ad-click logs from a large search engine, we extracted two sets of queries that resulted in online-ad conversions from travel and home IP address sets, respectively. For each set of queries, we examined the vocabularies and identified the top 100 most frequent words. In total about 30% of these query words are disjoint (i.e., appeared in one set but not in the other). Table 9 shows the top five query words from the disjoint sets. Not surprisingly, home users are more interested in advertisement about job opportunities (with keyword "employment" and "applications"), while travel users tend to search for information related with entertainment, accommodations, and transportation. Further, for the remaining 70% words that are popular across both set of IP addresses, their rankings are also signif-

	Disjoint query words (top 5 only)
Travel IPs	moving, universal, movies, inn, cars
Home IPs	employment, applications, cruises, food-stamps, college

Table 9: Top disjoint query words.

	Home			Travel		
	IPs	Events	IDs	IPs	Events	IDs
Total (10^6)	1.5	432.2	3.1	0.1	5.8	1.4
Malicious (10^3)	6.5	595.8	190.9	1.1	119.7	101.1
Percentage	0.4%	0.1%	6.2%	0.9%	2.0%	7.2%

Table 10: Prevalence of malicious activity events.

	Home	Travel
Events per IP	286.3	46.3
User IDs per IP	2.0	11.1
Events per user ID	140.8	4.2

Table 11: Home and travel IP event statistics.

icantly different. For example, "hotel" is among the 20 most popular words for travel users, but is used less frequently by home users (with rank 56).

4.2 Mobility and Security

Next, we explore whether home and travel IP addresses show different security properties. The answers to this question could have implications for defense strategies.

We evaluate the security-related characteristics of the home and travel IP sets derived from the Email-2009 data. For this application, we rely on labels from the data provider distinguishing malicious spammer accounts from legitimate user accounts.[4] To our surprise, Table 10 shows that the percentage of malicious login events is 20 times higher at travel IP addresses than at home IP addresses, suggesting access from travel IP addresses should operate under more stringent security policies. Moreover, the larger percentage of malicious login events at travel locations also suggests there might exist other types of malicious activities on those hosts, such as spreading attacks to compromise more computers. Thus machines at travel IPs may potentially be more vulnerable than home computers.

The overall event statistics in Table 11 also show interesting characteristics. For the number of user IDs per IP address, the travel set is an order of magnitude higher than the home IP set (i.e., a larger concentration of users in a smaller travel IP space). However, the number of events per user in the travel set is two orders of magnitude smaller, as users often spend less time in travel locations and so access their accounts less frequently.

Because travel IP addresses are shared much more frequently, directly blacklisting them is likely to generate higher false positive rates. On the other hand, attackers may prefer choosing travel locations to launch their attacks because of the denser user population. An interesting direction of future work is thus to explore different defense strategies at travel locations.

Furthermore, the different security properties of travel and home IP addresses suggest new possibilities for detecting the use of compromised accounts by attackers and other similar fraudulent activity. For example, user-login events from several new IP addresses within a few hours seem more likely to be benign if these addresses are at travel locations than if they are in residences. In the former case, the user may simply be traveling; in the latter case, the events may be from bots that employ the user's account for sending spam.

[4]These labels were marked based on detection performed by the data provider. They typically have a low false positive rate, but may have false negatives.

5. SUMMARY

This paper explores the notion of "location context". Specifically, we develop techniques to derive context about the locations where users access the network (home, work, travel) by analyzing user mobility patterns from large service logs. We show that these techniques can scale to the Internet despite a number of practical challenges. Using concrete examples, we show that "location context" can provide interesting characteristics of users and their activities for a variety of applications.

6. REFERENCES

[1] National Household Travel Survey. http://nhts.ornl.gov/.
[2] Navizon virtual GPS service. http://www.navizon.com/.
[3] Quova. http://www.quova.com/.
[4] Skyhook. http://www.skyhookwireless.com/.
[5] S. Agarwal, J. Dunagan, N. Jain, S. Saroiu, and A. Wolman. Volley: Automated data placement for geo-distributed cloud services. In *Proceedings of NSDI*, Apr. 2010.
[6] M. Casado and M. J. Freedman. Peering through the shroud: The effect of edge opacity on IP-based client identification. In *Proceedings of NSDI*, Apr. 2007.
[7] J. Cheng, S. H. Y. Wong, H. Yang, and S. Lu. SmartSiren: Virus detection and alert for smartphones. In *MobiSys*, June 2007.
[8] M. Dischinger, A. Haeberlen, K. P. Gummadi, and S. Saroiu. Characterizing residential broadband networks. In *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement (IMC'07)*, Oct 2007.
[9] B. Eriksson, P. Bardford, J. Sommers, and R. Nowak. A learning-based approach for IP geolocation. In *Proceedings of the Passive and Active Measurement Conference (PAM)*, Apr. 2010.
[10] S. Hao, N. A. Syed, N. Feamster, A. G. Gray, and S. Krasser. Detecting spammers with SNARE: Spatio-temporal network-level automatic reputation engine. In *Proceedings of USENIX*, Aug. 2009.
[11] V. N. Padmanabhan and L. Subramanian. An investigation of geographic mapping techniques for Internet hosts. In *Proceedings of ACM SIGCOMM*, Aug. 2001.
[12] I. Trestian, S. Ranjan, A. Kuzmanovic, and A. Nucci. Unconstrained endpoint profiling (Googling the internet). In *Proceedings of ACM SIGCOMM*, Aug. 2008.
[13] B. Wong, I. Stoyanov, and E. G. Sirer. Octant: A comprehensive framework for the geolocalization of Internet hosts. In *Proceedings of NSDI*, Apr. 2007.
[14] Y. Xie, F. Yu, and M. Abadi. De-anonymizing the Internet Using Unreliable IDs. In *Proceedings of ACM SIGCOMM*, Aug. 2009.
[15] Y. Xie, F. Yu, K. Achan, E. Gillum, M. Goldszmidt, and T. Wobber. How dynamic are IP addresses? In *Proceedings of ACM SIGCOMM*, Aug. 2007.