ISSN 1744-1986

Technical Report N° 2009/ 11

I Know What You Did Last Summer: risks of location data leakage in mobile and social computing.

L. Jedrzejczyk and B. A. Price and A. K. Bandara and B. Nuseibeh

13 November, 2009

Department of Computing
Faculty of Mathematics, Computing and Technology
The Open University

Walton Hall, Milton Keynes, MK7 6AA
United Kingdom

http://computing.open.ac.uk

I Know What You Did Last Summer: risks of location data leakage in mobile and social computing.

Lukasz Jedrzejczyk*, Blaine A. Price*, Arosha K. Bandara*, Bashar Nuseibeh*
Department of Computing, The Open University, UK*
L.Jedrzejczyk; B.A.Price; A.K.Bandara; B.Nuseibeh@open.ac.uk

ABSTRACT

Advances in mobile and web technologies have brought unwanted access to often sensitive data, ranging from our personal details, where we work, where we live and even behavioral patterns. The increasing use of social networks and location-aware mobile applications raise a number of concerns, including the issue of ensuring users' privacy. In order to explore those concerns we conducted an exploratory study in re-identifying people based on their movements and publicly available information. We observed anonymous users of a location-based social networking application in their natural environment and demonstrated how to re-identify them based on that data. In addition to discovering location-based private data, we were also able to find quite a number of facts from their private lives. This article reports on the methodology we used, ethical issues related to informed consent and user's reaction to the fact of being re-identified.

INTRODUCTION

One of the biggest changes in mobile computing in recent years has been the ability of devices to self-locate. Hundreds of millions of cameras, PDAs and smart phones are now running thousands of different location-aware applications which make use of WiFi base station proximity/triangulation, GPS, or in some cases manual geo-tagging. Last year, one third of new mobile phones launched had GPS capability. According to the latest report from Research and Markets [1], this ratio will likely rise to one half of all handsets by 2013. Statistics from the leading mobile applications platforms such as Apple's iPhone AppStore, Android Market or Nokia's Ovi shows that location-aware applications are being downloaded at an incredible rate. The current undisputed leader, Apple's iPhone AppStore, has seen more than 2 billion downloads of applications, with a half billion programs in the last quarter alone [2] and over 80,000 individual applications. Of these almost 10,000 are location-aware [3] and the numbers continue to grow.

Cheap ubiquitous mobile broadband access and the widespread use of social networking has led to endemic active and passive data sharing, including location data. Although access to a single location update might not seem very interesting or invasive to an individual's privacy, we were interested in discovering how much could be discovered by analyzing a regularly updated dataset of location records over a period of time. To avoid observation bias, we wanted to look at publicly available data provided by people who didn't know we were studying them. Some of the key questions driving our research were:

- *Is it possible to re-identify anonymous users based on the publicly accessible movement data combined with other, freely available, data ?*
- *How well do users understand the risks of location data sharing?*
- *What choices make users more vulnerable to location data leakage?*

The research presented in this article demonstrates the main vulnerabilities in commercial applications that can lead to inadvertent data leakage and retention. We show how we used publicly accessible data and data-mining tools to re-identify anonymous users of location based service and how we cross-referenced users' data against their online presence in order to reconstruct some private facts from their life. We also discuss some techniques and principles that allow location data sharing and location based services (LBSs) while minimizing data leakage.

RELATED WORK

Privacy is researched intensively from many different perspectives. In this paper, we consider privacy from the perspective of re-identification risk. Re-identification is matching a user in two datasets by using some linking information such as name, date of birth or location, in order to disclose one's identity or an unknown attribute [4].

There are a vast number of re-identification examples in the literature. The well known example of re-identification [5] shows how the triple (date of birth, gender and zip code) can be used to successful re-identification of 87% of American population. Matching common columns (DOB, gender, zip code) between two datasets (anonymous medical data and voters list) enabled researchers to establish a link between particular records. In results researchers were not only able to identify users' identities but also their health problems, address, ethnicity and the political party they supported.

Meaningful information about different aspects of our life can be inferred from social networking sites or online

forums. One of the recent, though not yet published, work in this area is "Gaydar", a project initiated by two MIT students, Jernigan and Behram. [6]. They applied statistical analysis to users' contact lists and correctly revealed sexual orientation of 10 Facebook users, including those, who decided to keep that information private. In another Facebook study, Lindamood et. al [7] performed a scalable inference attack in order to reveal undisclosed information. They predicted Facebook users' political views based on profile information and connections.

In order to re-identify users of popular movie ratings portal, Frankowski et al. developed algorithms that match data from a publicly available dataset with information from the portal's forum. They showed that it is possible to re-identify a large number of users in a sparse relation space, in which mere presence of particular attribute can result in anonymous information disclosure [4].

The work described so far was strongly focused on personal attributes re-identification, which is only partial interest for work presented in this paper. Another piece of research we used in our work was focused on location data mining. Several pioneering efforts in this area have explored whether an algorithm can discover users' meaningful places from spatial data. Ashbrook [9] used a k-means clustering algorithm for identifying important places. In the 'Commotion System', Marmasse [10] proposed using ad hoc location learning algorithms based on arrival and departure times. Sander and Ester developed a density-based algorithm, gdbscan [11] while Zhou et al. [12] proposed another density-based algorithm DJ-CLUSTER. Zhou et al. also found that a very useful element in the location data mining process is the shape of movements and proposed a set of four visualizations that can represent particular groups of different places [13].

Though the above algorithms can identify significant location from movements, they cannot automatically discover the relation between the user and place. Assigning a meaningful label to a location was one of our objectives. An attempt to automatically tag places was made by Liao et al [14]. They used machine learning techniques to successfully identify home and work places of five subjects. Hoh et al. examined movement data of 65 drivers and were able to automatically detect home locations of 85% of them [15]. Krumm [16], whose work most closely matches our study, inferred the home address of 172 known subjects by using four different algorithms: Last Destination, Weighted Median, Largest Cluster and Best Time. A particularly interesting aspect of this work is the attempt to disclose subjects' identity based on inferred address, which was also one of our objectives. In order to achieve this goal, Krumm used the Windows Live Search API's white pages lookup feature and reverse geo-coding service. The number of successful re-identifications was quite low, varying from 1.2% to 5.2%. This means that he could identify the forename and surname of maximum 9 people from the dataset of 172, assuming that reverse geo-coding service contains information about all inferred home addresses. In our study we were interested in inferring significant places and identities of users based on their time-stamped movements (similarly to Krumm) and web pseudonym, which gives us more options for re-identification.

THE STUDY

In order to ensure that we were studying natural behaviour, one of our requirements was that the participants not know that we were going to view their data as they were generating it. This presented an obvious ethical problem of informed consent. In discussions with our institution's ethics panel, we noted that the social location sharing system we were using only associated pseudonyms (usernames) with the timestamp, latitude and longitude. The panel agreed to approve the study provided we did not analyze or store data of identified individuals without consent. We resolved this problem by dividing our data collection and analysis into three parts. Firstly we collected location data from all the users in our identified group but they were identified only by their self-chosen login name. We then performed an initial analysis on all these users to isolate a group we believed we could discover the identity of and we discarded the remaining data. Once we found a method of contacting a user, we did so, in order to identify ourselves, explain what we wanted to do with their data, and ask for permission to perform further analysis. Those who agreed were further analyzed in the second phase while those who didn't reply, or refused, had their data deleted. Those who we re-identified successfully were further interviewed in phase three.

Phase one – data collection, quantitative data analysis and informed consent

Our first task was identifying a suitable dataset. In order to start collecting the data we had to identify a popular location sharing service that permitted bulk access to groups of user data. Once we identified the service and tested our tools we began collecting data for a period of 76 days starting from the 12th of May 2009. We collected more than 3,000,000 movement records of 1,700 users worldwide. Due to the demonstrative nature of our study and the workload requirement we did not want to re-identify all of them, therefore we selected an interesting subset to study. The main criteria for selecting data for further analysis were the number of location readings in total, the number of different locations and number of active days. By active days we mean the number of different days when users updated their location. Another important criterion was availability of contact information on a user's public profile, which gave a possibility of contacting subjects before we started detailed analysis. It also suggested a method of communication in a case of partial re-identification. From the original pool of 1,700 users we chose 54 that met the contact criterion. We reduced this to 17 after applying our data quality criterion.

We then tried to contact the partially-re-identified users to seek permission to analyze the historical movement data that we had collected. We explained that we would not identify them or publish any private information about them, and that they could withdraw at any time. After one week, 3 out of 17 subjects respond to our request, 2 of them were happy to participate without further questions (U2, U3). One subject (U1) was more suspicious and checked our project website first, he also contacted the location based service provider he was using. After we provided him some additional information about our project and our ethical protocol he gave us permission to use his location data. The data about users that did not reply to our request within one week were discarded and removed from the database. We also deleted the re-identification reports about these users and deleted the backup files leaving only the data relating to the consenting participants. We then completed the analysis of the data (phase two) and presented participants with all the information we found (phase three).

Phase two – data analysis

Once we obtained consent we began a detailed analysis. The three subjects that consented to participate in our study generated 748 locations in total and updated their location 993 times. The least active user (U1) was active only for 27 days, whether the most active user (U2) updated his position every day during data collection period. A statistical summary of our dataset is presented in Table 1.

	U1	**U2**	**U3**
Input data summary			
Different locations	81	484	183
Updates	238	565	190
Active days	27	76	32
Mean Updates per day	8.8	7.4	5.9
Mean Updates per location	2.9	1.2	1.03

Table 1. Statistical summary of dataset used in our study. Results in table are based on latitude/longitude coordinates rounded up to 3 decimal points.

First difficulty – challenges in analyzing location data

We did not want to be very intrusive therefore we decided not to analyze all data about the users in detail. Our main objective was to demonstrate how location data can be misused, and what information can be inferred from movement records. The key facts we wanted to learn about users in this phase were: *Where does the user live? Where does he/she work? What is the user's daily movement pattern?* In addition to that we also identified some fact from user's life i.e. *What user did on particular day?*

In order to cope with the size of the dataset, we developed a mash-up analysis and visualization tool made up of 8 parts:

1. **Locations** – a list of movements showing the real address of the location, time, date, name of the day and time user spent there. Stops were marked by different colors.
2. **Visits per location** (weighted median method [16]) – a bar chart presenting the mostly visited locations.
3. **Distance by date** – a bar chart presenting the daily movements distance in km for each day. By using this tool we could easily spot if the user's daily movement had any patterns.
4. **End and start of the day** (Last destination algorithm [16]) – a table showing first and last location of the day.
5. **Results on map: markers** – a map presenting visited locations, different colors of markers represent number of visits at particular location, which helped us identify clusters of users' significant locations.
6. **Results on map: trajectories** – a map presenting all daily movements as lines. This helped us identify daily movement patterns and significant locations.
7. **Street View** – the Google Maps Street View tool enabled us to look at the physical location from street level as opposed to the map overview. This tool is linked with markers and trajectories maps and can be activated by clicking on markers.
8. **Day by day movement trajectories** – this tool presented a visualization of a user's movement' trajectory for each day.

Initial study results

In our attempt to re-identify participants, we started our analysis by finding significant locations for each user, such as home and work addresses. We identified two methods for finding the home location. The first method is similar to the Largest Cluster (LC) method [16], is based on identifying center point within the cluster with most number of points as the origin of a trajectory. The center, star-shaped, point indicates the user's home (Figure 1).

The second method is based on pattern recognition in daily movement records. We observed that some of our subjects live according to a very predictable pattern: they usually left and come back home at certain time. When looking at daily distance chart we could observe that pattern. The average daily distance during working days was almost exactly the same for each day, but during the weekend we observed longer distance, which suggests a trip (Figure 2). For example, when we looked at the daily trajectories map for each of selected days we noticed that our user went out of the city for a weekend (Figure 3). When we looked at

smaller variation in daily distance chart and compared it to the actual movements on the map we found that user went to Costco (presumably shopping). One shortcoming of this method is that it is sensitive to the noise.

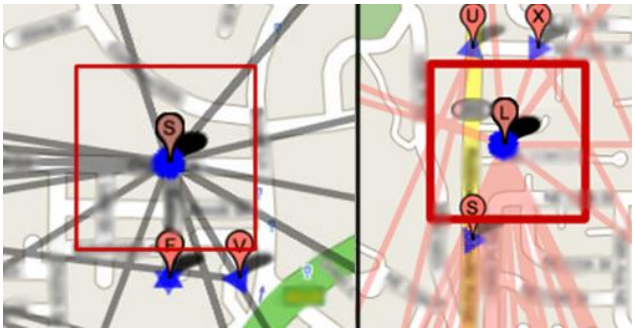

Figure 1. Finding home using star visualization method.

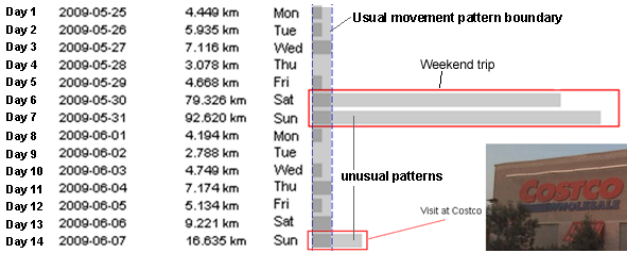

Figure 2. Using 'daily distance' and 'daily movements' modules in analyzing movement behavior. The longest distance suggests unusual behavior, i.e. days 6 and 7 (Figure 3).

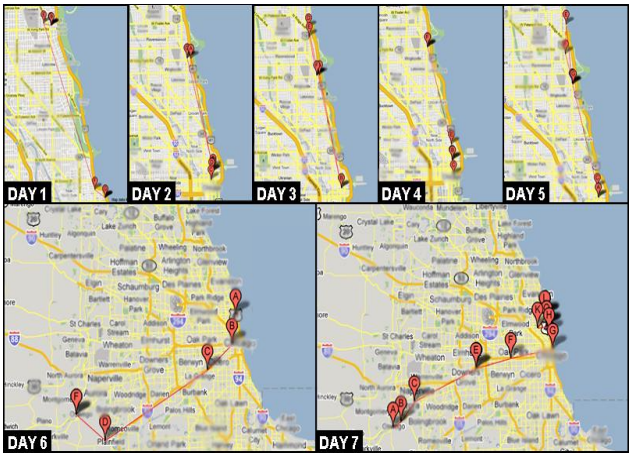

Figure 3. Movements trajectories of one week (days 1-7 from Figure 2) from U2's life.

In addition to the raw location data analysis, we also cross-referenced the users' movement data and login names with information available through search engines and often identified their real names. We ran different queries on popular search engines in order to find out more about users' personal details and interests. We also used people search engines such as pipl.com. Our subjects very often used the same login name for different websites, which enabled us to validate some of our findings and find more details about them. For example, when one of our participants (U3) went to San Francisco, he stayed at Hotel X and posted information on his Twitter account that *he is going to go to the gym at X*. We compared the date of this post with the user location at this time. The results showed the same username was in the Hotel X at the time of posting that tweet, which confirmed the Twitter user as our user (Figure 4). We found that social networking sites, such as Twitter and LinkedIn play a significant role in user re-identification and help validate findings based on location data mining.

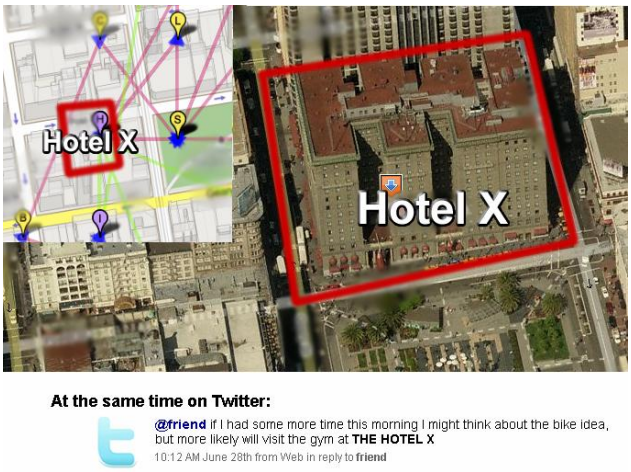

Figure 4. Using entries on social networking sites to validate our findings. The figure shows the location where U3 was at a particular time and an entry from his Twitter account posted on the same time.

Based on the collected data we managed to identify the following about users' locations: home and work address, shopping activities, single trips, places they stopped for a meal, daily movement patterns and visited hotels. The detailed results of our analysis are presented in Table 2.

Phase three – results presentation and interviews

After the re-identification process we created a report for each individual. The reports contained two types of information: the results of our analysis (note: table 2 only shows whether or not we found information in order to preserve privacy, the individualized report contained the actual data) and the "how-to" information, presenting how we tied different data together to produce each result.

In order to be fully transparent we showed our participants all the data we collected, including raw movement records that were not used in the detailed analysis. We also gave our participants access to the analysis tools we used and explained how we used them in the re-identification process.

First reactions to results

In the reports we explained our methodology to participants. We also showed them what choices made them vulnerable. Their first reactions to the results were quite positive and did not indicate strong privacy concerns. They

were more focused on the accuracy and complexity of reports, rather than privacy issues. Participants raised more concerns about our "how to" report, and reported their concerns about the simplicity of the re-identification process. U1, which we found was most concerned about privacy, said "*... you've done something anyone could do, which is frightening really*".

	U1	**U2**	**U3**
Name and surname	Yes	Yes	Yes
Email	Yes	Yes	Yes
Date of birth	Yes	Yes	Yes
Phone no	No	Yes (work)	No
Occupation	Yes	Yes	Yes
Family information	No	Yes	Yes
Picture	Yes	No	Yes
Home address	Yes	Yes	Yes
Work address	Yes	Yes	Yes
Identified facts from live	Several shopping trips, weekend trip, stop for a meal	Weekend trip and shopping, daily pattern identified (time + movements)	Trip to San Francisco
Web presence	Very low web presence (mostly forums)	Medium web presence, mostly forums or not fully updated profiles	Rich web presence, user available on forums, social networking websites. Runs own blog.
Other	Picture, mobile device used	Picture of a house	Picture, mobile devices used, picture of a house, personal interests information, car make.

Table 2. Final results of our analysis.

In next stage we tried to understand what their main motivation for using a LBS. The motivations varied across our subjects. U1 used the service to allow his girlfriend track him. She felt more comfortable about knowing where he was when he was driving up and down the country. Another motivator for using this LBS was tracking past movements and the ability to locate on the map. Our participants did not mention that they used the service to share their location with other, anonymous service users. U1 was surprised when we told him that his data are publicly available and can be accessed by strangers. One important statement he made was that "*it has to be made clearly by service providers that if you become a member of a user group your information is publicly available*". This particular comment highlights two problems with LBSs that relate to both privacy and usability: (1) explaining the where possibly unanticipated data leakage may occur; and (2) making privacy mechanisms, which are already there, more visible and easier to use.

Accuracy

In the report we sent to our participants, we tried to identify a number of places from their life, including their home address and work place. In order to validate our methodology we asked them how accurate our findings are. The answers showed that, though we were able to locate precisely their homes on the map, the addresses provided by the reverse geo-lookup were not always right: U1 "*address is nearly right, it's actually 27 Street name, knowing what the GPS image of my house looks like it's easy to assume it's 25. I live in a back to back terrace which is behind number 27*". We also made a mistake in the address of U2's work place, we found the correct building, but the name we provided in the user's report was wrong.. In case of U3, we were able to provide very precise home and work address: "*You identified not only the part of the building where my office is, my home and hotel I stayed in (...)*". Clearly the numbers we chose for detailed analysis do not allow us to generalize our results, however the answers suggest that the methodological direction we have chosen for location data analysis is effective. The two methods we used (described in the initial study results section) enabled us to precisely find significant locations for all participants. The discrepancies between addresses we provided and real locations are caused by incorrect results returned to us from Google's Reverse Geo-coding Service, which we used to transform coordinates into human-readable addresses.

Location sharing vs. privacy perception

Interviews showed that we did not find sensitive details about our participants; there was nothing in our results that made our participants feel very uncomfortable in terms of activities or places they visited. Indeed, in such a small sample of random strangers it would be surprising if we stumbled upon an affair or a visit to an adult entertainment establishment. The interviews showed that disclosing a location as an anonymous person does not have a strong impact on their privacy perception. One participant (U2) said that though we found his home address, work place and identified his daily pattern he did not found it "*damaging*". Alternatively U1 highlighted a security risk of living according to a pattern: "*I wouldn't like that information fallen into darker hands, especially if I'm a 9-5 worker.*". U2, when asked about the potential security risks of sharing location data, said: "*People who do want to find me can probably find me easily, if somebody wants to find me personally by my name they would have to know that I*

use "the service" and what my login is. Otherwise they would have to decrypt all of the public data to find who might have what … I didn't see that much of privacy problem". It is important to mention that U2 was the most privacy unconcerned from all 3 participants.

Two of our participants (U1, U3) said that taking part in our study was a real "*eye opener*" for them. U1 reported that he was very surprised we could get access to his location data and extract it. Asked about future use of the service, he said that he will still use it, but first he needs to opt out off all public groups he joined. What we found really interesting is that location is not the information that people want to hide. What our participants really want to control, is the correlation between location, their identity and activities. U1: "*I don't mind anyone knowing that I went for shopping to London. What is scary is that you don't know me and you could extrapolate my home address, work address and then get personal information by other sources of the Internet.*"

Although U3 was not very concerned about privacy and still contributes to several LBSs, he reported that we highlighted "*one or two leakages"* that he was not aware of. The biggest privacy concern we noticed during interviews is related to publishing non-anonymous location information and linking it to activity information posted on social networking sites. We asked U3 about possible changes in his online behavior triggered by our report said he will not change, and he recently started contribute to new LBS. However, he will be more careful about publishing personal information. He also added that his real privacy concern comes from publishing "*I'm on the holiday with the family for a week*" with his home address being so visible, which increases the risk that his home will be burgled.

Views on privacy control

One of the final questions we asked our participants was about privacy controls in LBSs. U2, who we found was less concerned about privacy, highlighted a need for easy on/off option to hide his location. At the moment to achieve this he has to switch the entire service on or off. Surprisingly he was the only participant with disclosure control practice, which we believe has something to do with his level of expertise. U1 reported that he would like to control location broadcasting based in relation to people or organization and time: "*If I broke down in the car, knowing them* (sic) *automatically where I am – brilliant. Ordering pizza. In that situation yes, because I have relation with them. I wouldn't like my information be known by people/organization that I have no relationship with.*"

Similarly to U2, U3 highlighted a need for a coarse grained control mechanism. He reported that there is a risk in sharing aggregated location data and said that he would only like to share contemporary data with authorized users. U3 was the only user that suggested a detailed mechanism for sharing location at different levels of granularity: "*X is asking for your current location. Return [a] Exact, [b] Town, [c] State, [d] Country, [e] Pre-defined 'fake', [f] Reject*". What is interesting here is that he also suggested use of deception in managing privacy which was explored by Adam [17].

DISCUSSION

The main purpose of our study was to examine the risks of anonymous location sharing and how this information can be cross-referenced with one's online presence in order to re-identify a real person. Based on three individuals we showed that by using only time-stamped movements we can not only find significant locations such as home or work but also identify regular patterns in movement behaviour or even identify the part of the building where one works.

Our participants did not express strong privacy concerns about sharing their location information. Instead the factor that we found to have a real impact on their privacy is cross-referencing location data with their identity and activities. U1 and U3 expressed that concern during interviews. U2, did not find our results "*damaging*" and has no privacy concerns at all. We believe that it is related with his level of expertise and technological-awareness. The perceived unlinkability between his movements and real identity was a factor that made him feel comfortable about sharing location.

Although the detailed study was small, it highlighted some important risks in the combination of location data leakage and how individuals present themselves on the Internet. Our study confirmed that people using the same pseudonyms or posting activity information on several social services made them more vulnerable to potential attacks than users that had no visible web presence at all.

Another issue we observed was that there is a relation between usability, privacy and security. One of our participants reported a lack of feedback about the future use of user's data in the LBSs. In fact, the user was not notified about any consequences of joining the group and all of his actions within the system were based purely on his (erroneous) assumptions. This is a good example of the privacy and security implications of a bad design, which is illustrated in Figure 5. The user's data was publicly available due to a misunderstanding of how the service works. It violated his privacy because information he did not want to share with strangers was publicly accessible. Consequently, he lost control over his data, which led to re-identification of his personal attributes by using location data mining and simple search activities for user's web pseudonym. Cross-referencing both, personal attributes and meaningful location information enabled us to re-identify his real identity. Had our re-identification procedure been performed with malicious intent, the information gathered could have been used to cause serious (physical or financial) harm to the user.

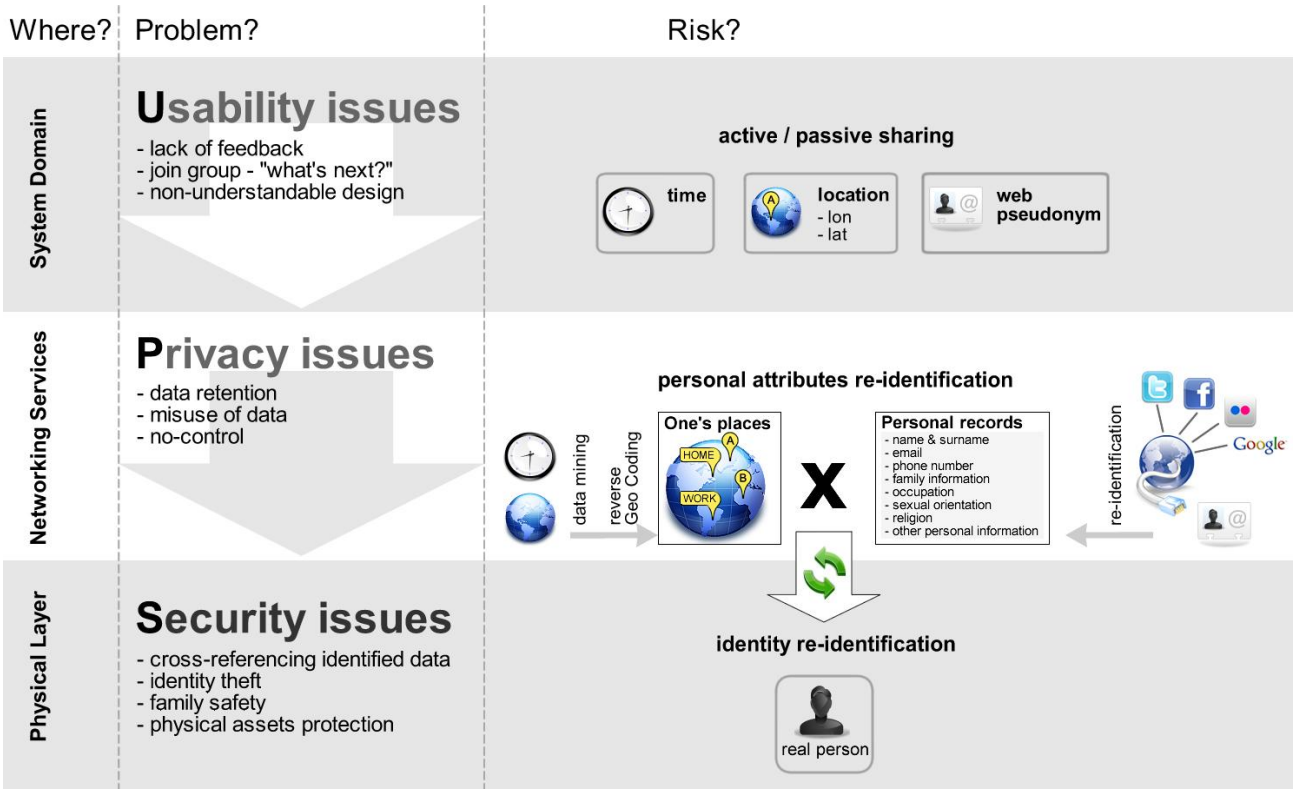

Figure 5. Relation between usability, privacy and security.

CONCLUSION

Being re-identified using location data and publicly available information can violate not only one's privacy but also security. Our "fake attack" on a real location based service exploited several potential vulnerabilities both in the service we used, and in people's everyday privacy choices. We highlighted significant security vulnerability within this LBS: the ability for the anonymous collection of large amounts of location data. We also highlighted usability problem related to the lack of feedback in the system. It resulted in a discrepancy between people's assumptions of privacy and real situations. We showed that cross-referencing location data with information publicly available on the web can lead to full re-identification of data owners and raise both privacy and security concerns. Our study was only exploratory but it suggests opportunities for further research:

- Repeating the study with a larger number of participants in order to verify to what extent our current findings can be generalized;
- Assuring communication between the system and user, and exploring how it meets the user's expectations;
- Developing tools that help people understand actual privacy settings;
- Incorporating meaningful places discovery algorithms into the location data analysis tool.

In order to understand people's perception of location privacy and elicit requirements for future privacy aware location based services we need to further explore the above ideas. Future work on the re-identification problem of privacy should expand our experimental study and focus on tools that can help users understand their privacy choices. In particular, we need to focus on mechanisms that could identify behavioral and privacy choices making users vulnerable to potential attack.

ACKNOWLEDGMENTS

This research is part of the PRiMMA project (Privacy Rights Management for Mobile Applications, http://primma.open.ac.uk) and was funded by the EPSRC (EP/F024037/1 and EP/F023294/1). We thank the participants of our studies for their commitment and willingness to be interviewed. Additionally, we would like to thank Prof. Marian Petre for her help in improving the presentation of our results, and members of the PRiMMA project team for their feedback on the drafts of the paper.

REFERENCES

[1] "United Kingdom Location Based Services (LBS) Market Forecast, 2009 - 2013: Total spend in the LBS market in the UK to rise to $406 million in 2013 - Market Research Reports - Research and Markets." http://www.researchandmarkets.com/reportinfo.asp?report_id=1080767, Last accessed 11th November 2008

[2] "Apple Shares App Store Stats: 85k Apps Available, 2 Billion Downloads So Far."

http://www.techcrunch.com/2009/09/28/apple-shares-app-store-stats-85k-apps-available-2-billion-downloads-so-far/, Last accessed 11th November 2008

[3] "AppStore Search Engine." http://www.uquery.com, Last accessed 11th November 2008

[4] D. Frankowski, D. Cosley, S. Sen, L. Terveen, and J. Riedl, "You are what you say: privacy risks of public mentions," *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, ACM, 2006, p. 572.

[5] P. Samarati and L. Sweeney, "Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression," *Proceedings of the IEEE Symposium on Research in Security and Privacy*, 1998, pp. 384–393.

[6] "MIT's Facebook "Gaydar" - Is it Homophobic?" http://www.fastcompany.com/blog/dan-macsai/popwise/mits-facebook-gaydar-it-homophibic, Last accessed 11th November 2008

[7] J. Lindamood, R. Heatherly, M. Kantarcioglu, and B. Thuraisingham, "Inferring private information using social network data," *Proceedings of the 18th international conference on World wide web*, ACM New York, NY, USA, 2009, pp. 1145-1146.

[8] T.F. Gieryn, "A space for place in sociology," *Annual Review of Sociology*, 2000, pp. 463-496.

[9] D. Ashbrook and T. Starner, "Learning significant locations and predicting user movement with GPS," *Wearable Computers, 2002.(ISWC 2002). Proceedings. Sixth International Symposium on*, 2002, pp. 101-108.

[10] N. Marmasse and C. Schmandt, "Location-aware information delivery with commotion," *Lecture Notes in Computer Science*, 2000, pp. 157-171.

[11] J. Sander, M. Ester, H.P. Kriegel, and X. Xu, "A Density-Based Algorithm in Spatial Databases: The Algorithm DBSCAN and Its Applications," *Data Mining and Knowledge Discovery*, 1998, pp. 169-194.

[12] C. Zhou, D. Frankowski, P. Ludford, S. Shekhar, and L. Terveen, "Discovering personal gazetteers: an interactive clustering approach," *Proceedings of the 12th annual ACM international workshop on Geographic information systems*, ACM New York, NY, USA, 2004, pp. 266-273.

[13] C. Zhou, P. Ludford, D. Frankowski, and L. Terveen, "How Do People's Concepts of Place Relate to Physical Locations?," *Lecture notes in computer science*, vol. 3585, 2005, p. 886.

[14] L. Liao, D. Fox, and H. Kautz, "Location-based activity recognition using relational Markov networks," *Proc. of the International Joint Conference on Artificial Intelligence (IJCAI)*, Citeseer, 2005.

[15] B. Hoh, M. Gruteser, H. Xiong, and A. Alrabady, "Enhancing security and privacy in traffic-monitoring systems," *IEEE Pervasive Computing*, vol. 5, 2006, p. 38.

[16] J. Krumm, "Inference attacks on location tracks," *Lecture Notes in Computer Science*, vol. 4480, 2007, p. 127.

[17] K.A. Adam, Balancing Privacy Needs With Location Sharing in Mobile Computing. PhD Dissertation, The Open University, 2009.