Who's Your Best Friend? Targeted Privacy Attacks In Location-sharing Social Networks

Vassilis Kostakos, Jayant Venkatanathan, Bernardo Reynolds
Madeira Interactive Technologies Institute
University of Madeira
{vk, vjayant, bernardo.reynolds}@m-iti.org

Norman Sadeh, Eran Toch
School of Computer Science
Carnegie Mellon University
{sadeh, eran}@cs.cmu.edu

Siraj A. Shaikh
Faculty of Engineering and Computing
Coventry University
s.shaikh@coventry.ac.uk

Simon Jones
Department of Computer Science
University of Bath
s.jones2@bath.ac.uk

ABSTRACT
This paper presents a study that aims to answer two important questions related to targeted location-sharing privacy attacks: (1) given a group of users and their social graph, is it possible to predict which among them is likely to reveal most about their whereabouts, and (2) given a user, is it possible to predict which among her friends knows most about her whereabouts. To answer these questions we analyse the privacy policies of users of a real-time location sharing application, in which users actively shared their location with their contacts. The results show that users who are central to their network are more likely to reveal most about their whereabouts. Furthermore, we show that the friend most likely to know the whereabouts of a specific individual is the one with most common contacts and/or greatest number of contacts.

Author Keywords
Location sharing, privacy, privacy attacks.

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms
Experimentation, Human Factors.

INTRODUCTION
The study tries to answer two important questions relating to targeted location-sharing privacy attacks. First, given a group of users and the social ties amongst them, is it possible to predict which of these users is likely to reveal the most about their whereabouts? Second, given an individual user within a particular social network, is it possible to predict which of her friends knows most about her whereabouts? To answer these, the paper presents a longitudinal study of real-time location sharing whereby the patterns of information exchange and privacy policies of a large group of users are analysed and modelled.

Real-time location sharing applications are gaining wide adoption, with a number of commercial systems now available on the market, including Foursquare, Facebook Places, and Google Latitude. Such services are frequently used in the context of online social networks (OSN), whereby one's real-time location becomes yet another sharable aspect of one's online profile. With the increasing adoption of online location sharing services, understanding the privacy implications and potential targeted attacks enabled by this new technology, becomes crucial.

A conventional approach for engineering a privacy attack is to attempt to gain ongoing access to the target's whereabouts, thereby building up a profile of that user's behaviour. In this paper, we assume that location sharing practices are likely to follow the trend of other OSN profile properties and propagate through the network of friends. The key assumption, therefore, is that a target's location can be visible to friends of friends. From the attacker's perspective, this has the benefit that they do not get "too close" to the target while still they are able to collect information about the target's location on an ongoing basis.

The two questions that this paper addresses are key in instrumenting a targeted attack against users. Such an attack would first identify a suitable target amongst a set of users. Once this has been achieved, the attacker then identifies a "weak link" in the target's list of friends. The "weak link" is a friend of the target whom the attacker will attempt to befriend in order to gain direct access to the target's whereabouts by becoming a friend of a friend. Therefore, the attacker is likely to seek for weak links who are most likely to have full access to the whereabouts of the target.

UbiComp'11, September 17–21, 2011, Beijing, China.

RELATED WORK

Sharing perceptions and strategies

Substantial research shows that people approach for developing rules and ultimately policies for sharing information with others are strongly related to the presentation of self [13], and also relate to the formulation of dialectic and dynamic behavioural mechanisms depending on circumstantial context [2] or the conjugation of disclosure, identity and temporal boundaries [24]. What was once achieved with walls, doors and other physical or architectural constraints is still to be adapted to today's communication means [34]. Privacy management is an intricate process and is further augmented in a computer mediated environment. On social networking sites, privacy regulation is a socio-technical activity involving interaction with the technological system and the group context. Individuals' privacy behaviour in such systems involves a mixture of technical and mental strategies. For instance, a technical strategy may involve the use of privacy settings to regulate content distribution to select audiences [30], while research has also shown that considering tie strength is another strategy for developing rules for disclosure [36].

Despite the evidence suggesting that users adopt objective strategies for controlling their privacy online, previous work has identified a discrepancy between people's privacy attitudes towards sharing information and their actual sharing patters [1,23]. This behaviour has been termed the "privacy paradox". For instance, a study revealed a high discrepancy between stated concerns and actual behaviour towards sharing static profile information on Facebook [1]. Other studies have further established the privacy paradox on social networking sites [34].

While it is not clear whether the privacy paradox also applies to people's perceptions towards location sharing, it certainly highlights the needs for collecting quantitative data on people's location-sharing behaviour rather than relying purely on subjective data.

Location-sharing privacy

There is an increasing amount of work on understanding users' location-privacy needs in ubiquitous and location-aware systems relying on techniques such as diary studies [4], interviews [14], surveys [17], scenarios [16, 35] and lab and field observations [5]. Research suggests that users may start with relatively coarse and conservative preferences [28]. Over time, they take advantage of controls exposed to them and exhibit more sophisticated sharing behaviours, controlling the availability of the data through mechanisms such as disabling the service [4] or obtaining feedback about which users can see or have seen their information [14,16,28,33]. Users are also sceptical about the usefulness of location sharing in day-to-day activities, suggesting that current practices (such as calling somebody up) are sufficient [4]. However, the usefulness of such services was acknowledged in more stressful situations involving unfamiliar environments or in crisis and safety scenarios in general [14]. In such situations, information usefulness outweighs privacy concerns. Furthermore, prior work has shown that people's presence in different physical environments is likely to affect their willingness to trust and actually engage in interaction with location-based services [18].

Research investigating sophisticated privacy mechanisms, such as customizable privacy policies, has indicated that, without new interface technologies, they can present significant challenges for users. One recent study reports participants failing to implement their desired policies with a high degree of accuracy [28]. Furthermore, it also noted that although participants varied considerably in the time they spent defining their policies (between 5 and 8 minutes), the duration of this period was not strongly correlated to final policy accuracy.

It has also been observed that the recipients of the location data are typically more significant to users than the locations being shared. Perhaps unsurprisingly, users are more willing to share information with friends than acquaintances or strangers [5,33]. While recipient identity seems to be the strongest factor influencing one's willingness to share her location [10,20] time and location restrictions have been shown to also be important in capturing people's preferences [5]. Research has also shown that users are sensitive to the reactions of recipients if location information is denied or not made available [14, 28], suggesting that systems need to incorporate an element of plausible deniability. However, users do make distinctions in sharing particular locations: additional privacy is required at home when compared to work [31].

Privacy attacks on OSNs

Targeted privacy attacks on OSNs have been demonstrated in the past. Attempts to construct social graphs for individuals from available public listings are already shown to be feasible [6]. Once achieved, social graphs can be clustered for segregating groups into sub-groups in terms of different spheres of activity for an individual [37,15]. Further results show that even hidden communities can be detected with reasonable effort [22]. This work shows that given an individual of interest, it is possible to identify a close group around that person, which may potentially be used in order to get closer to the target. To some extent, this is already done by authorities targeting criminals coordinating their activities using OSNs [7,9], and it usually involves some level of active probing [29] which in the context of OSNs may mean striking friendships with individuals close to the target so as to avoid detection.

The characteristics of privacy attacks in the context of location sharing differ from privacy attacks online social networks because of two reasons. Location-sharing applications include information about users' physical

whereabouts, which can lead to access to one's physical self. Empirical evidence show that users fear that revealing their location to people they do not trust may lead to physical and property harm [32]. Furthermore, users' decisions on location sharing may differ considerably than decisions taken in the context of social networks, making this subject worthwhile of investigation.

Identifying "weak links"

Prior work on social networks may be used to derive some hypotheses about who is likely to share information with whom on a social network. For instance, Petronio's theory of Communications Privacy Management (CPM) describes an iterative process of rule development, boundary coordination and boundary turbulence [25]. Rule development can be defined as the process of developing regulations about who to tell what. These regulations guide our everyday disclosures, and are a function of our context and disclosure goals. Ties of differing strength have varying disclosure norms, thus Stutzman theorizes rule development is a function of network composition [30]. For example, a network that is more heavily focused on strong ties may require higher levels of privacy, as disclosures among strong ties are more personal in nature [36]. This suggests that network structure may be used as a basis for attempting to predict disclosures amongst individuals.

Recent work on sharing ephemeral information shows that rule development is a function of tie strength [27]. In order to test CPM's rule development process on the context of posting content to Facebook, users were presented various scenarios of information disclosure and were prompted to decide how and with whom to share that information with. Results show users are more prone to share with stronger ties as opposed to weak ties. These findings where uniform across the various scenarios of information disclosure presented to participants. Intended and expected audience for both profile and ephemeral information was a function of tie strength [27,30]. Both authors report that users' perceived audience for the information they share is mostly composed of strong ties.

STUDY

Definitions

The following are definitions of metrics used in the study that follows.

- *Social Graph*: A set of individuals and the explicit friendship ties amongst them.
- *Degree Centrality* : The degree centrality of a user is the number of direct connections (or "friends") that the user has in the social graph. These were the friends of the user on Facebook that were also users of Locaccino.
- *Betweenness Centrality* : The betweenness centrality of a user is the number of shortest paths between all pairs of nodes in the social graph that pass through the node representing the user. For a more thorough description of the betweenness centrality, the reader is directed to [11].
- *Openness* : The *openness* of the ordered pair (A, B) of users is the percentage of simulated location requests made to A by B that were granted by A's policies.
- *Trust* : The *trust* of a user A is the mean of the openness values (A, B) where B ranges over all of A's friend. i.e. it is the average openness of user A towards all her friends.
- *Trustworthiness* : The trustworthiness of a user A is the mean of the openness values (B, A) where B ranges over all of A's friends. i.e. it is the average openness of A's friends towards A.
- *Trust Rank :* Given a user A and a user B who is a friend of A, the *trust rank of B with respect to A* is *i* if there are precisely *i-1* friends $C_1, C_2 \ldots C_{i-1}$ of A such that the openness of (A, C_j) , $1 \leq j < i$, is greater than the openness of (A, B). i.e. the trust rank is obtained by ranking A's friends in terms of how much they are trusted by A.
- *Degree Rank* : Given a user A and a user B who is a friend of A, the *degree rank of B with respect to A* is *i* if there are precisely *i-1* friends $C_1, C_2 \ldots C_{i-1}$ of A such that the degree centralities of $C_1, C_2 \ldots C_{i-1}$ are greater than that of B. i.e. the degree rank is obtained by ranking A's friends in terms of their degree centralities.
- *Mutual Rank* : Given a user A and a user B who is a friend of A, the *mutual rank of B with respect to A* is *i* if there are precisely *i-1* friends $C_1, C_2 \ldots C_{i-1}$ of A such that the number of common friends A has with each of $C_1, C_2 \ldots C_{i-1}$ is greater than the number of common friends that A has with B. i.e. the mutual rank is obtained by ranking A's friends in terms of how many mutual friends they have with A.

Hypotheses

Previous work suggests a relationship between social network structure, tie strength and the patterns of disclosure amongst individuals (e.g. [30]). In attempting to identify which individual is more likely to reveal information about their whereabouts, one may hypothesise that individuals who are more central to the network are more likely to do so. A possible explanation would be that such individuals are more likely to engage in collaboration and coordination activities, therefore it may be more likely that they are willing to share their real-time location with others. This reasoning provides ground for the first experimental hypothesis:

- H1: Individuals who are more central to the social graph are likely to reveal the most about their location.

Upon determining a suitable person to target, the next step in a potential attack would be to befriend someone from the target's social network. Considering that previous literature

suggests that reciprocity is an important driving force in social networks [26], one can expect that the target is likely to share their location with someone in the social network out of their desire to reciprocate. Hence, the friend of the target with the most number friends, who by means of H1 is likely to share their own location, is someone with whom the target may wish to share their location in order to reciprocate. That person is therefore a potential "weak link" whom the attacker might befriend in order to get closer to the target. This leads to the second experimental hypothesis:

- H2: The target's friend with the highest degree has higher probability of knowing more about the target.

Finally, it can be argued that shared membership and being part of the same community would be suggestive of two individuals who may be possibly involved in joint activities requiring coordination. In addition, literature on homophily has shown that individuals who share mutual friends are more likely to be alike, thus likely to engage in joint activities [21,8]. It is therefore plausible to hypothesise that individuals who belong to the same group are more likely to share their real-time location with each other, thus becoming candidate "weak links". This leads to the third experimental hypothesis:

- H3: The target's friend with most common ties with the target knows most about the target.

To test these hypotheses, the location-sharing system described next was deployed and used by a large group of users longitudinally.

System

The study was conducted by deploying Locaccino, a real-time location sharing application integrated in the OSN Facebook. The application consists of two components: a Web application component and a mobile component. Various version of the mobile component were developed to run on multiple mobile platforms: windows and apple laptops, and Symbian smartphones. The purpose of this component is to collect in real-time a user's location and then upload it to a central server.

The Web application component of Locaccino (Figure 1) allows users to set preferences regarding how their location is shared with their Facebook contacts. Users are given the option to create policies in order to manage their location sharing. Policies specify the conditions under which the location should be revealed to another user. These conditions include the identity of the recipient of the information, the time and day, and the actual location where the user is. For instance, one may specify a policy to allow work colleagues to obtain one's location only during work hours and when they are in town.

Participants were recruited on campus using advertisements on-line and via email, as well as through national press covering the features of our system.

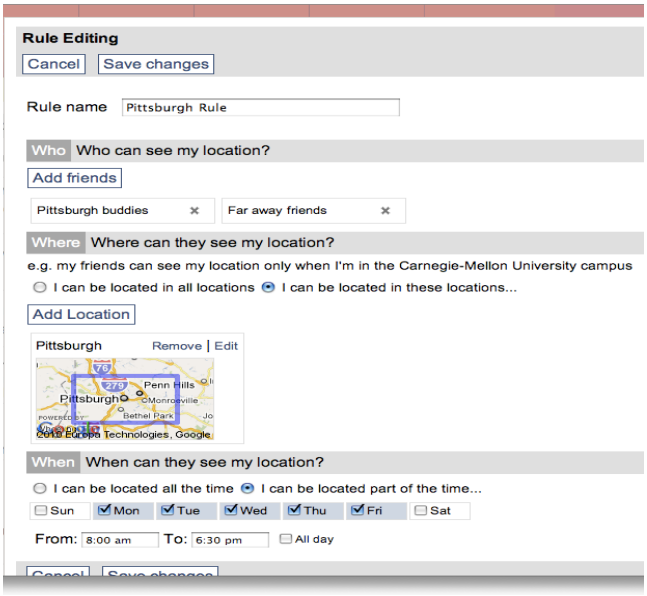

Figure 1. Screenshot of Locaccino's functionality that allows users to construct their location sharing policy rules.

The system was used longitudinally and more than 300 users installed the application and actively begun using it to share their location with colleagues. For the purposes of the study presented here, the following information was collected about users:

- Social graph: An undirected unweighted graph describing the friendship between all the participants. In this graph, a node represents a user, and two nodes are connected if they are friends on Facebook.
- Policy graph: A directed weighted graph describing the privacy policies between the users. In this graph each node represents a user, and user A is connected to user B if user if these two users are friends on Facebook. In addition, the weight of the edge from user A to user B is a value between 0 and 1 based on the "openness" of user A towards user B. The weight of the opposite edge, i.e. the openness of user B towards user A, is independent and may be a different.

The openness value of (A,B) was calculated as the percentage of B's possible requests that were granted by A's policies. The openness value from one user towards another represents the extent to which a user is willing to share their location with another user. In our case we rely on users' policies to capture and quantify this feature. Specifically, to generate a value representative of the openness between two users we conducted the following procedure. For each pair of users (A,B) in the dataset we ran a simulation whereby user B repeatedly requested the location of user A. These simulated requested were processed by the policies of user A, and the result was either positive or negative, thereby either showing or hiding user A's location respectively. During this analysis the

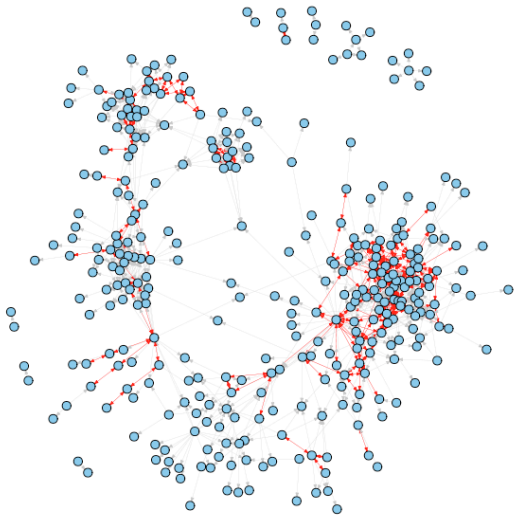

Figure 2. The graph representing the participants (nodes) and their trust relationships as directed edges. Mutually open relationships are highlighted in red.

movement of user A was the same as recorded during the study.

RESULTS

The study ran for a month with 340 users who were already users of Facebook. The derived policy graph contained 1778 policy rules, two for each of the 889 friendship ties within the user population (Figure 2).

Each policy described the openness of one user towards other users, ranging from 0 to 1. For each user the average openness that they show towards their friends was calculated (referred to as "trust" towards others) and is summarised in Figure 3, while the average openness that a user was shown by his friends (i.e. their "trustworthiness") is shown in Figure 4.

Hypothesis testing

H1: Individuals who are more central to the social graph are likely to reveal the most about their location.

A Kruskal_Wallis non-parametric test of independent samples [e.g. 12] showed that there was a significant effect of a node's betweenness on that node's trust towards its direct connections ($H(45)=82.111$, $p<0.001$) but not on that node's trustworthiness ($H(45)=56.168$, $p=0.123$). Furthermore, there was a significant effect of degree centrality on node trust ($H(23)=82.076$, $p<0.0001$) and also on node trustworthiness ($H(23)=35.276$, $p<0.05$).

H2: The target's friend with the highest degree has higher probability of knowing more about the target.

To test this hypothesis all users with less than 2 friends in the dataset were discarded from the analysis, leaving 247 users. This data was discarded because no comparison can

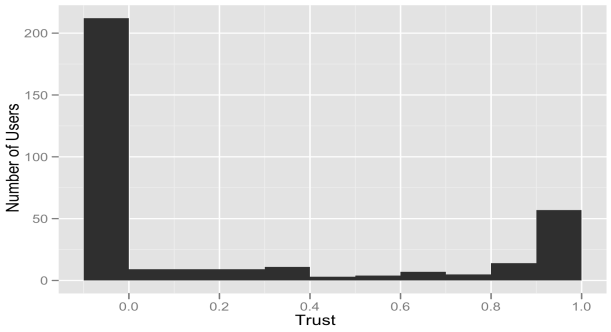

Figure 3. Histogram of distribution of nodes' average openness (i.e. the average of all outgoing ties for each node)

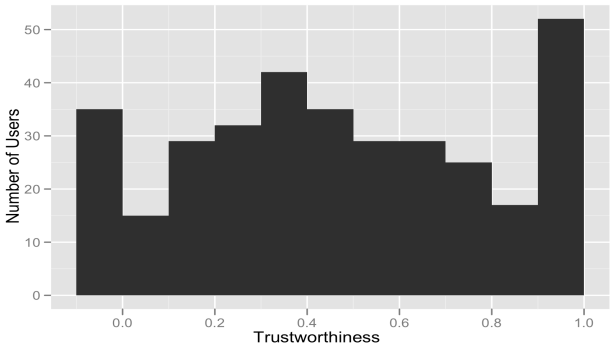

Figure 4. Histogram of nodes' average trustworthiness (i.e. the average of all incoming ties for each node).

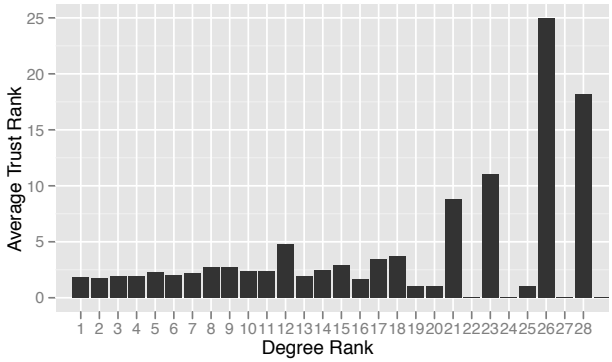

Figure 5: Degree rank of nodes (x-axis) versus the average trust rank (y-axis) for all nodes of a specific degree rank

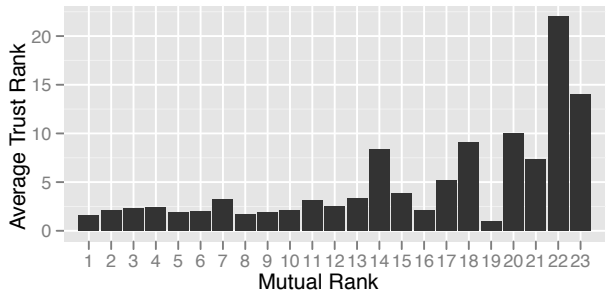

Figure 6. Histogram of Mutual rank (x-axis) vs. average trust rank (y-axis) for all nodes of a specific CommonFriends rank.

be carried out for users with a single friend. For each user A, all of A's friends were ranked in terms of how much they are trusted by A (Trust Rank), and in terms of how many friends they actually have (Degree Rank). This gave for each friendship relationship in the data two values: Trust Rank and Degree Rank respectively (Figure 5). A chi-square test showed a significant relationship between Degree Rank and Trust Rank ($x^2=3981.723$, $dF=744$, $p<0.001$) while there was a positive correlation between the two variables (0.239, $p<0.01$).

H3: The target's friend with most common ties with the target knows most about the target.

To test this hypothesis all users with less than 2 friends were discarded from the analysis, leaving 247 users. For each user A, all of A's friends were ranked in terms of how much they know about A (Trust Rank), and in terms of how many mutual friends they have with A (Mutual Rank). This gave us for each friendship relationship in the data two values: Trust Rank and Mutual Rank respectively (Figure 6). A chi-square test showed a significant relationship between Mutual Rank and Trust Rank ($x^2=3210.841$, $dF=682$, $p<0.001$), while there was a positive correlation between the two variables (0.252, $p<0.01$).

Structural analysis

Finally, an analysis was conducted to assess the extent to which friends with the highest degree are the same as friends with a large number of common friends. A chi-square test showed a significant relationship between Degree Rank and Mutual Rank ($x^2=6548.051$, $dF=528$, $p<0.001$), and a positive correlation between Degree Rank and Mutual Rank of 0.81 ($p<0.01$) as shown in Figure 7.

Furthermore, a triad analysis was conducted, to assess the extent to which there exists a bias in how trust and trustworthiness was distributed across the network. The analysis was conducted by first classifying each bi-directional edge in one of three possible states: balanced-high (meaning both people are sharing in full or partially), balanced-low (meaning that both people are not sharing), and unbalanced (meaning that one person is sharing while the other is not). Given the three possible labels for each edge, there exist 10 possible "templates" for triads, depending on the combination of its bidirectional edges (see Table 1). Each triad in the graph was labelled appropriately, and the frequency of occurrence of each template was calculated.

In addition, for each of the 10 templates the theoretical expected frequency of occurrence was calculated, as described in [19], by assuming that the same edges were randomly distributed on a graph with identical topography. The relationship between the observed and expected frequency for each of the 10 templates is shown in Figure 8. The figure shows a modest correlation ($R^2=0.75$), with the exception of the data point at (119,225) corresponding to

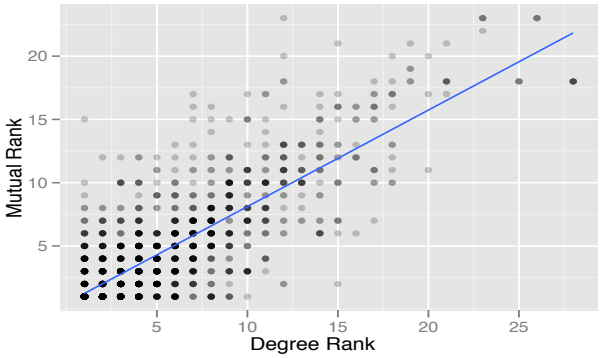

Figure 7. X-axis: degree rank of a neighbour wrt another node. Y-axis: common friends rank of a neighbour wrt another node. Darker dots indicate overlapping points.

Template		Expected Frequency	Observed frequency
1		292	209
2		142	119
3		119	225
4		23	22
5		39	10
6		1	19
7		16	5
8		3	1
9		3	0
10		0.7	28

Unbalanced Balanced-Low Balanced-High

Table 1. Expected frequency (given a random model) and observed frequencies for each of the 10 possible triad templates.

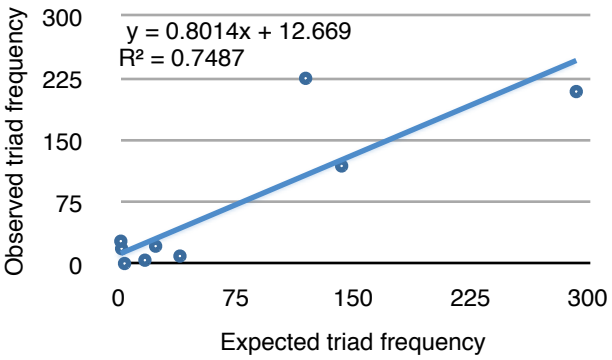

Figure 8. Correlation between the expected and observed frequencies for each of the 10 possible triad templates.

the template "Balanced-high, Unbalanced, Unbalanced". Removal of this point would substantially improve the correlation (R^2=0.96).

DISCUSSION

Targeted location-sharing privacy attacks

This paper proposes a threat model related to location-sharing privacy, whereby the attacker attempts to collect data about a target's longitudinal movements. To do this, the attacker first needs to identify suitable targets such that his chances of success are maximised. Once a suitable target is identified, then the attacker attempts to gain access to the target in order to collect data about the target's location, but not "too close" to avoid detection. Therefore, even though a strategy for collecting data on a target's whereabouts would be to attempt to befriend the target directly, that increases the chances of the attacker being noticed. This paper assumes that the attacker can attempt to collect data about the target by befriending one of the target's friends, i.e. a "weak link". This will make the attacker a "friend of a friend" of the target, which is arguably adequate to gain access to the target's location. In order to achieve this, the attacker needs to figure out which of the target's friend are more likely to have access to the target's location data, and are therefore a more suitable person to befriend. The results from Figure 3 show that, on average, nodes exhibit a bimodal distribution of trust which is weighted towards not sharing at all. Hence, if the attacker picks a target's friend at random, they are about four times more likely to not gain access at all to the target's data.

To assess the extent to which such an attack can be engineered, the study presented here answers two important questions relating to this kind of targeted location-sharing privacy attacks. First, given a group of users and the friendship ties amongst them, is it possible to predict which of these users is likely to reveal the most about their whereabouts? Second, given an individual user within a particular social network, is it possible to predict which of her friends know most about her whereabouts?

The study presented here captured a measure of "openness" between individuals, which reflects the probability that a request for someone's real-time location is likely to be satisfied. An advantage of using a generic measure, which we refer to us trust (when a person of interest is open towards someone else) and trustworthiness (when someone else is open towards a person of interest), is that it can be applied across multiple features of online social networks. Therefore, while commercial location-sharing systems vary in features and their capabilities evolve over time, the measure of trust and trustworthiness is likely to remain an underlying driver in guiding users' decision to share their location with others.

Identifying a suitable target

The motivation for H1 was to suggest a way in which the attacker can identify users who are more likely to share their location with friends. The hypothesis was that individuals who are more central to the social network reveal the most about themselves, motivated by the observation that such individuals are more likely to engage in collaboration and coordination activities. Our results suggest that a user's network centrality as measured by betweenness and degree centrality had a significant effect on the amount of trust that user was willing to show towards their friends, thus supporting hypothesis.

The results suggest that individuals who are more central to their network are more likely to be willing to share their location with others, and therefore they are good targets for a potential attacker. Hence, an attacker can conduct a basic analysis of the network structure to identify central nodes, and then attempt to target more central nodes since they are more likely to share their location. It can be argued that individuals who are more central to the network are more socially active, and maintain more social relationships. This is likely to require them to take part in more social activities, and therefore it can be argued that these conditions require more coordination on their part. This offers one explanation as to why the findings in this study suggest that more central users did in fact share their location more often.

How to target individuals

Once the attacker has identified a target who is likely to be open and share their location, the next step is to develop a strategy for targeting that individual. The threat model discussed in this paper assumes that the attacker will not attempt to befriend the target directly, since that bears a high risk of being detected. Instead, the attacker can attempt to befriend someone from the target's friends since that can give them access to the target's location data without bringing them "too close" to the target. Therefore, the next step for the attacker is to identify a "weak link" in the target's list of friends, or a person who is likely to be granted access to the target's location data. The study tested two possible strategies for identifying weak links: based on the number of friends that a weak link may have (H2), and based on the number of common friends that the weak link may have with the target (H3).

Prior studies have shown the importance of reciprocity in social interactions, thus providing the motivation for H2. More specifically, studies have shown that when an individual performs a favour or act that bestows trust upon another individual, that individual is likely to feel obliged to reciprocate the favour or act. The motivation for H2 comes from this perceived obligation and from H1. The results show a positive correlation between the amount of trust that an individual bestows on each of his friends and the number of friends of those friends. This suggests that the attacker

can identify a suitable "weak link" of the target by considering the target's list of friends and identifying those individuals with the highest number of friends of their own. Such individuals are more likely to be social active, and are therefore more likely to choose to share their location with the target (see H1). The target, by virtue of reciprocity, is therefore more likely to share their location with such individuals.

A competing, and possibly complementary hypothesis for identifying weak links is H3, which states that the target's friend with most common ties with the target knows most about the target's whereabouts. This can be due to the fact that the existence of common friends can indicate shared membership in a community or organisation. The results show that there is a significant positive correlation between the trust of a target towards each of his friends and the rank of that friend in terms of the number of mutual friends he has with the target. The results provide a clear strategy for how an attacker can identify a "weak link", which entails identifying who from the target's list of friends has the highest number of common friends with the target. One explanation for these findings is that individuals who share many friends, and are thus likely to belong to the same social groups, are more likely to share their location in order to coordinate their activities better, as well as to maintain an increased awareness of each other's ongoing activities.

Finally, it should be pointed out that the analysis provides evidence that H2 and H3 are directly related. Since both hypotheses were supported by the analysis, this is not surprising. The results show a strong positive correlation between H2 and H3 in that a target's friends who have many friends are also likely to have a lot of common friends with the target. One explanation for this relationship may be that individuals who have many friends of their own are more likely to be extroverts who socialise and engage in multiple social interactions activities. Their behaviour therefore increases the likelihood of them being friends with mutual friends with the target simply because they have a lot of friends.

Triads and small group privacy

The results of the structural analysis presented here offer insights into how, in the context of location-sharing, triads of users distribute and balance trust and trustworthiness. In addition to being useful in understanding the behaviour of our participants, these results are also useful in situations where only partial information may be known about the network.

The structural analysis shows that even though under a completely random model we expected to observe only one triad where all three members trust each other (template 10 in Table 1), we actually observed 28 such triads. Furthermore, the correlation analysis highlights triad template 3 as being substantially different from the overall correlation pattern between expected and observed frequencies. In this case, this result shows that we observed quite often situations where two people trust each other but both maintain unbalanced relationships with a third individual. This is a balanced situation and expected to be more frequent in a realistic setting than in a purely random environment [e.g. 11].

The results from this analysis coincide with prior work in that people tend to avoid unbalanced situations and prefer the comfort of balanced triads . Furthermore, these results can be used to make predictions in situations where incomplete information has been collected about individuals. This is possible since given three individuals and 2 of the 3 relationships between them, we may be able to predict the third relationships. For example, given a triad with two balanced-high relationships, the chances of the third relationship being balanced-low is very close to zero, unbalanced is 15%, and balanced-high is 85%.

Protection against such privacy attacks

The attack described here assumes that the attacker is trying to gain longitudinal access to the target's whereabouts, and does so by avoiding detection since they do not need to befriend the target directly, but only one of their friends. Assuming that on average users have about 150 friends in a social network, then the attacker's strategy ensures that he is one of about 22000 people who are friends-of-friends of the target, making detection much harder.

One strategy that the platform could follow in case of a pull-based location-sharing model would be to ensure that individuals are notified if anyone is making too many location-sharing requests. This could be implemented in the form of a user-defined threshold or as a nudging mechanism intended to help people refine their sharing preferences [3]. In the case of a push-based model, the users can ensure that their information is visible only to their friends directly, and to no-one beyond that. Similarly, limits could be imposed on how often a user can update their location, hence offering an upper bound on how much users can reveal about their whereabouts. However, such solutions seem to contradict the needs of commercial systems which appear to strive for increasing the amount of shared information.

Making useful predictions

While the work described here was framed in the context of a privacy attack, the hypotheses that were tested may be useful in developing user-friendly features that can automatically provide useful suggestions to users. For instance, the hypotheses discussed earlier provide an indication on how to identify a person who is likely to know the whereabouts of an individual of interest. It may be the case that the individual of interest has not logged into the location-sharing system to update their location, due to technical difficulties, time constraints, or any other plausible difficulty. Under such circumstances, the system

may be able to make automated suggestions about who to ask regarding the whereabouts of the person of interest based on a simplistic network-structure analysis. Therefore, in cases of high urgency it is possible to offer such recommendations as a fall-back strategy.

Limitations of the study

In a realistic environment there may be multiple factors affecting the sharing of information, many of which are inadvertently manipulated by users. For instance, battery life and group norms may be important factors that urge users to hide or share their location. These were not taken into account in this study.

Furthermore, this study presents and tests a generic strategy for engineering such an attack. Clearly, the fine details of the social platform where this information is recorded and shared are important, and may facilitate or hinder the success of such an attack. For instance, being friend of a friend may be "too close" or "too far" to obtain location information, while some auditing mechanism may allow users to see who is viewing their location information repeatedly.

Finally, it is important to take into consideration here the fact that users of this location sharing application start with a default privacy policy of not sharing their location information with anybody in the network. We cannot rule this out a contributing factor to our result that more central nodes trust more, as they are also likely to be seasoned users of the system and hence have invested more time to articulate their location sharing preferences.

CONCLUSION

This paper presents a study that aims to answer two important questions related to targeted location-sharing privacy attacks: (1) given a group of users and their social graph, is it possible to predict which among them is likely to reveal most about their whereabouts, and (2) given a user, is it possible to predict which among her friends knows most about her whereabouts.

The results show that users who are more central to their social network (both locally and globally) are more likely to share information about their location, and hence are more "vocal". In addition, the findings show that given a target, that target's friend who either has many friends or many common friends with the target is more likely to be trusted by the target.

The findings of this study are important both in understanding how privacy attacks can be engineered and how they can be prevented. An important next step for this work is the application of these insights for the development of automated protection and suggestion mechanisms that will make the sharing of real-time location safer and more useful.

ACKNOWLEDGEMENTS

This work is funded by NSF grants CNS-0627513, CNS-0905562, CNS-1012763, by CyLab at Carnegie Mellon under grants DAAD19-02-1-0389 and W911NF-09-1-0273 from the Army Research Office. Additional support has been provided by Google and the CMU/Portugal Information and Communication Technologies Institute and the Portuguese Foundation for Science and Technology (FCT) grant CMU-PT/SE/0028/2008 (Web Security and Privacy). The authors would also like to acknowledge the entire Locaccino team, including Jason Hong, Lorrie Cranor, Paul Hankes Drielsma, Justin Cranshaw, Patrick Gage Kelley, Jialiu Lin and Michael Benisch for their contributions.

REFERENCES

1. Acquisti, A. and Gross, R. (2006). Imagined communities: awareness, information sharing, and privacy on the Facebook. Proc. PET 2006, Springer, 36-56.
2. Altman, I. (1977). Privacy Regulation: Culturally Universal or Culturally Specific? Journal of Social Issues, 33 (3), 66-84.
3. Balebako, R., Leon, P.G., Mugan, J., Acquisti, A., Cranor, L.F., Sadeh, N. (2011) Nudging Users Towards Privacy on Mobile Devices, CHI 2011 workshop article, May 2011
4. Barkhuus, L. (2004). Privacy in location-based services, concern vs. coolness. Mobile HCI 2004 workshop: Location System Privacy and Control.
5. Benisch, M., Kelley, P.G., Sadeh, N., Cranor, L.F., (2010). Capturing Location-Privacy Preferences: Quantifying Accuracy and User-Burden Tradeoffs. Personal and Ubiquitous Computing (PUC). Forthcoming.
6. Bonneau, J., Anderson, J., Anderson, R. and Stajano, F. (2009) Eight Friends Are Enough: Social Graph Approximation via Public Listings. SNS'09.
7. Choo, K-K. R. and Smith, R. G. (2008). Criminal Exploitation of Online Systems by Organised Crime Groups. Asian Criminology (2008) 3:37–59
8. Christakis, N., Fowler, J.H. (2008). The collective dynamics of smoking in a large social network. The New England journal of medicine 358(21): 2249-58.
9. CISC (2010) Report on Organized Crime. Criminal Intelligence Service Canada. Available At: http://www.cisc.gc.ca [Last Accessed 25th October 2010]. ISBN 978-1-100-51931-9.
10. Consolvo, S., Smith, I. E., Matthews, T., LaMarca, A., Tabert, J., and Powledge, P. (2005). Location disclosure to social relations: why, when, & what people want to share. CHI 2005, 81-90.

11. Easley, D., and Kleinberg, J. (2010). Networks, Crowds, and Markets: Reasoning About a Highly Connected World. Cambridge University Press.
12. Gibbons, J.D. (1993). Nonparametric statistics: An introduction. Sage University Paper series on Quantitative Applications in the Social Sciences, 07-090.
13. Goffman, E. (1959). The presentation of self in everyday life. Garden City, NY: Doubleday Anchor.
14. Hong, J. I. and Landay, J. A. (2004) An architecture for privacy-sensitive ubiquitous computing. MobiSys '04, 177-189.
15. Jones, S., and O'Neill, E. (2010). Feasibility of structural network clustering for group-based privacy control in social networks. In Proceedings of the Sixth Symposium on Usable Privacy and Security (SOUPS '10). ACM, New York, NY, USA, , Article 9 , 13 pages
16. Kelley, P. G., Hankes Drielsma, P., Sadeh, N., and Cranor, L. F. (2008). User-controllable learning of security and privacy policies. AISec 2008, 11-18.
17. Khalil, A. and Connelly, K. (2006). Context-aware telephony: privacy preferences and sharing patterns. CSCW '06, 469-478.
18. Kostakos, V. and Oakley, I. (2009). Designing Trustworthy Situated Services: an Implicit and Explicit Assessment of Locative Images' Effect on Trust. CHI, Boston, USA, pp. 329-332.
19. Kostakos, V. and Venkatanathan, J. (2010). Making friends in life and online: Equivalence, micro-correlation and value in spatial and transpatial social networks. IEEE SocialCom, Minneapolis, USA, pp. 587-594.
20. Lederer, S., Mankoff, J., and Dey, A. K. (2003). Who wants to know what when? privacy preference determinants in ubiquitous computing. CHI 2003, ACM Press, 724-725.
21. McPherson, Miller, Lynn Smith-Lovin, and James M Cook. "Birds of a Feather: Homophily in Social Networks." Annual Review of Sociology 27(1):415-444.
22. Nagaraja, S. (2008) The economics of covert community detection and hiding. WEIS: Workshop on the Economics of Information Security.
23. Norberg, Patricia A, Daniel R. Horne, and David A. Horne (2007). The Privacy Paradox: Personal Information Disclosure Intentions versus Behaviors. Journal of Consumer Affairs, 41(1):100-126.
24. Palen, L., Dourish, P. (2003). Unpacking Privacy for a Networked World. In Proc. of the Conference on Human Factors and Computing Systems: CHI 2003, ACM Press 129–136.
25. Petronio, S. (2002). Boundaries of Privacy: Dialectics of Disclosure. SUNY, Albany, NY
26. Regan, D.T. (1971). Effects of a favor and liking on compliance. Journal of Experimental Social Psychology, 7(6):627-639.
27. Reynolds, B., Venkatanathan, J., Goncalves, J., and Kostakos, V. (2011). Sharing Ephemeral Information in Online Social Networks: privacy perceptions and behaviours. In proceedings of INTERACT.
28. Sadeh, N., Hong, J., Cranor, L., Fette, I., Kelley, P., Prabaker, M., and Rao, J. (2008). Understanding and capturing people's privacy policies in a mobile social networking application. Personal and Ubiquitous Computing 13, 6, 401-412.
29. Shaikh, S. A., Chivers, H., Nobles, P., Clark, J. A. and Chen, H. (2008). Network reconnaissance. Network Security, 2008(11):12-16.
30. Stutzman, F., Kramer-Duffield, J. (2010) Friends only: examining a privacy-enhancing behavior in Facebook In Proc. of the Conference on Human Factors and Computing Systems: CHI 2010, ACM Press 1553-1562.
31. Toch, E., Cranshaw, J., Drielsma, P.H., Tsai, J. Y., Kelley, P. G., Cranor, L., Hong, J., Sadeh, N. (2010) Empirical Models of Privacy in Location Sharing, in Proceedings of the Twelfth International Conference on Ubiquitous Computing. Ubicomp 2010
32. Tsai, J., Kelley, P.G., Cranor, L.F., and Sadeh N. (2010). Location- Sharing Technologies: Privacy Risks and Controls. Journal of Law and Policy for the Information Society, 2010.
33. Tsai, J.Y., Kelley, P., Drielsma, P., Cranor, L.F., Hong, J., and Sadeh, N. (2009). Who's viewed you?: the impact of feedback in a mobile location-sharing application. CHI '09, 2003–2012.
34. Tufekci Z. (2008). Can You See Me Now? Audience and Disclosure Management in Online Social Network Sites. Bulletin of Science and Technology Studies. Volume 11, Number 4, June 2008 , pp. 544-564(21).
35. Wagner, D., Lopez, M., Doria, A., Pavlyshak, I., Kostakos, V., Oakley, I., Spiliotopoulos, T. (2010). Hide And Seek: Location Sharing Practices With Social Media. MobileHCI '10, 55-58.
36. Wellman, B. and Wortley, S. (1990). Different Strokes from Different Folks: Community Ties and Social Support. American Journal of Sociology 96, 3, 558-588.
37. Xu, X., Yuruk, N., Feng, Z. and Schwieger, T. A. J. (2007) SCAN: a structural clustering algorithm for networks. Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, 824-833.