Privacy-Preserving Friendship Relations for Mobile Social Networking

W3C Workshop on the Future of Social Networking – Position Paper

Sören Preibusch – Alastair R. Beresford
University of Cambridge · Computer Laboratory
William Gates Building, 15 JJ Thomson Avenue
Cambridge CB3 0FD, United Kingdom
Firstname.Lastname@cl.cam.ac.uk

ABSTRACT

This position paper explores the dual nature of friendship relations as an enabler but also as a pitfall for privacy in social networks. The privacy-deteriorating consequences of unilateral friendship disclosure are quantified for an existing medium-sized social network site. A lightweight implementation of hidden friendship using existing standards is proposed and assessed for practicable and secure deployment in a mobile networking scenario.

1. INTRODUCTION

Social Network Sites (SNS) are a popular platform for interacting and communicating amongst individuals and groups. Some sites have a professional focus whilst others promote "networking" as a leisure activity. The relationships built, used and eventually discontinued among members of an online social network no longer replicate the offline world but constitute a new form of socialising in its own right – including online relationships for which no offline counterpart exists. The cultivation of existing acquaintances and the browsing for new contacts induces a level of active participation unrivalled by other Web 2.0 phenomena. Whilst Wikipedia and video or photo platforms such as YouTube or flickr are mainly used passively, more than 80% of SNS users become active network members [2]: they explore and browse the profiles of other users, search for contacts, write text contributions or comment on existing texts in the community and modify their own profile. In 2008 nearly 30% of all people connected to the Internet are members of at least one SNS and three out of four teenagers participate in online communities [2].

Centralised social networks are predominant today and rely on a central authority the users trust with their data. The SNS operator is therefore in possession of the entire profile and networking data of all network members. Even if they trust the operator to observe correct data handling practices, the databases of the operator may be compromised by criminals or confiscated by law enforcement authorities. Indeed, file sharing networks – the first form of mainstream social networking on the Internet – now use peer-to-peer protocols because centralised architectures, such as Napster and Audiogalaxy, were brought down by litigation. With emerging decentralised authentication schemes such as OpenID, the distributed storage of profile information is a good next step towards a privacy-enhanced social network.

Socialising online is now a common part of daily Web usage and this habitualisation drives a demand for increasingly ubiquitous access to a social network site. SNS members experience social pressure to participate regularly in the network and being deprived from access creates a feeling of unease [3]. Consequently, social networking increasingly occurs via the mobile Internet. This thirst for mobile social networking is satisfied by companies such as T-Mobile who advertise "Internet on the move" and encourage Internet connectivity from mobile phones with commentary such as "Check the BBC website or look at Facebook – browse the whole internet while you're out and about."

Mobile usage of social networking platforms increases the inherent privacy risks because rich spatio-temporal data may be added to the profile and networking data [4]. This same ubiquitous usage imposes limits on the operability of a privacy-enhancing infrastructure because a centralised authorisation server may be unreachable under intermittent connectivity. Near-field communication technologies such as Bluetooth or WiFi should therefore be used for both the content channel and for determining and enforcing the privacy settings of SN members in a peer-to-peer fashion. An SN member can then determine locally whether another member shall be, for instance, granted access to her profile.

In this position paper we motivate the need for both public and private friend relationships in social networks (Section 2) and explain why maintaining public and private friend relationships in a centralised architecture is easier than in a peer-to-peer one. We identify and quantify the privacy risks of unilaterally disclosed symmetric relations in a real social network (Section 3). We then go on to provide a lightweight technical implementation using standard FOAF files (Section 4), and review this approach with regard to computational overhead and security considerations, focusing on deployment in a mobile social networking scenario (Section 5).

We have motivated the benefits of a peer-to-peer architecture for mobile social networking and make two further contributions: (1) we demonstrate the need for hidden friendship relations in a social network and explain why this is non-trivial in a distributed architecture; and (2) we provide an example peer-to-peer architecture for implementing hidden friendship relations using the existing FOAF standard.

Figure 1: Friends are privileged in getting access to a social network member's profile data (detail of the Facebook privacy settings page). As browsing other members' profiles is the favourite activity on social network sites, there is a keen interest in not divulging its entire contents to the public.

2. PRIVACY AND FRIENDSHIP IN SOCIAL NETWORKS

Friendship is the most fundamental relation in a social network. It is a relation between two members of the network and carries the understanding of friendship from the offline world. Whilst the formal establishment of a friendship relation in a social network is only a matter of a mouse click, the preceding activities for agreeing to engage in a friendship resemble the offline world: messages are exchanged, notes are left, and gifts are offered. In particular, these precursors of friendship justify a symmetry that is enforced by the social network: user A is a friend of user B if and only if user B is a friend of user A.

In SNS, friendship conveys the ability to perform privileged operations, such as reading otherwise private content: a friend may see more details of a member profile than an ordinary user (Figure 1). The symmetry of a friendship relation matches the strongly reciprocal nature of self-disclosure [1] and sustains the reciprocity of the data disclosure; a unilateral revocation of the friendship entails a bilateral discontinuation of friendship and therefore also disclosure.

An existing friendship relation may also be the premise for executing privileged actions such as sending private messages, initiating a private chat session, forwarding messages via email or SMS, or adding content to the user profile. As such, a friendship relation is an *enabler for privacy* because privacy-intruding access is restricted to a set of friends.

However, the symmetry of friendship relations is also a *source of privacy breaches*. To cope with the heterogeneous privacy preferences amongst the members of a social network, a user can herself determine which information to disclose and to whom. The distinction between private and public can be fine-grained or coarse-grained: on Facebook, for instance, a member can set access control for virtually every data entry of her profile separately. In contrast, members of the professional social network LinkedIn can only differentiate between a public view and an insider view of their profile pages. These access control mechanisms can typically also be applied to a list of friends as well as individuals.

Public lists of friends power the social network as they provide a visible and codified manifestation of a member's social network. These lists are a source for finding opportunities for communication and interaction. The user herself can quickly access her favourite interaction partners; a stranger can explore the network and hop from profile to profile by exploring friendship links. Indeed, such browsing is the most popular activity in SNs [2]. Service providers profit from undirected profile browsing because it makes visitors stay longer. Openly visible lists of friends also promote network growth: potential newcomers may use the size of existing friendship connections as an indicator of the overall "buzz" and hence assess the network's attractiveness before joining.

On the other hand there is also a genuine interest in keeping one's friendship lists private because it is in itself personal information. The formal manifestation of a friendship relation makes it the method of choice for inferring social closeness. Friendship relations are subsequently used in academia and in industry to extract the social connections between users and apply inference mechanisms over these relations with the underlying idea being that personality traits, interests, and socio-demographic characteristics propagate along the lines of a friendship. If user A is a middle-aged liberal professional than one may infer that her friend B is also a middle-aged liberal professional with higher confidence than for an unrelated user C. If A is interested in buying hybrid fuel cars than targeting eco-car advertisements to her friend B seems more sensible than to an unknown user C.

In the above example, the friendship relation is the source of a privacy invasion for B, and A may wish to protect her friends from such abuse. In addition, user A may not wish to reveal her friends because of other social, legal or professional reasons. For example, investigating journalists have a professional interest in not revealing their sources; executive professionals may wish to maintain secret ties with friends working at competitor companies; and teenagers may feel peer group pressure such as "why are you friends with this loser but not with me?".

The establishment and revocation of online friendships has already been shown to have consequences in the offline world. In the US, "a woman is divorcing her husband after she claims he had a 'virtual' affair with a computer-generated female character" [6] in Second Life. The woman reported that "the solicitor wasn't at all surprised – she said it was her second divorce case involving Second Life that week." She now has a new man in her life, someone she met while playing World of Warcraft. In Japan, a woman virtually killed her former husband by misusing his login credentials after he had dissolved their online marriage [5].

3. SYMMETRY AS A PRIVACY CHALLENGE

Having acknowledged the need to keep at least some friendship relations private, there is a design challenge that stems from the symmetry of friendship relations. Even if a member of a SNS decides to hide her list of friends, her friends may themselves independently publish their lists of friends publicly. Such a unilateral friendship disclosure results in a privacy breach because the entire friendship relation becomes inferable, as detailed in Figure 2. The privacy of a friendship relation is only maintained if both parties do not disclose the existence of the relation.

Using the disclosure behaviour of members in an existing SNS, we can quantitatively assess the privacy-breaching con-

sequences of unilateral friendship disclosure. We have explored a medium-sized German social network and built a graph of the friendship relations that exist between the 120,000 members using publically-accessible data. (In our analysis all figures are rounded to two significant digits). The site lets users decide whether their list of friends should be publicly visible or not. Originally, all friendship relations were hidden; when public lists were introduced, disclosure was declared to be the default setting. The requirement for explicit opt-out solicited harsh criticism from the active users in the community. Users who ceased to log into the site prior to the introduction of public friendships have therefore not had a chance to change their disclosure settings.

Table 1 quantifies the friendship revelation behaviour on the analysed social network site. The analysis differentiates between all users and the sub-group who are currently active, which is identified using a simple heuristic over the users' profile pictures: we assume that members of the social network who are or have been active participants would have uploaded a personal profile picture to replace the default image. This heuristic does not assume recent active participation but tests for activity at some point in the past.

	total	active
users	120,000	26,000
- hiding friends	1,900	1,700
- hiding friends [%]	1.6%	6.4%

Table 1: The share of users hiding their friends is higher among the active users.

We found 1,900 users hide their friends of which at least one friend could be inferred for 1,300 users; consequently more than two thirds of the users are subject to a privacy breach because of unilateral friendship disclosure. In total, 47,000 directed friendship links were identified on the SNS and used for the friendship inference (Table 2). For every other friendship publicised by both parties, there is a friendship that was discovered because one of the involved members of the social network made her list of friends public. Only friendship relations hidden by both parties could not be detected.

bi-directionally public friendships	19,000
one-directionally public friendships	9,000
public self-friendships	100
directed friendship links (total)	47,000

Table 2: The share of users hiding their friends is higher among the active users.

The data from those users who publish their list of friends can be used to ascertain the inference ratio for friendships: the number of friends made public via the user's list of friends (that is outgoing friendship links) is compared to her number of friends inferred by incoming friendship links. For instance, if a user A is known to have five friends and four other members of the SN were found to list A as their friend, then the friendship inference ratio is $4/5 = 80\%$. In a social network with fully public friendship relations, both numbers are equal and the inference ratio is 100%. For the actual social network we analysed, the average friend discovery ratio was found to be 89% for active users (97% for the entire sample). That is, on average, almost all of a user's friends could be correctly inferred by just analysing the friendship links published by other members of the social network.

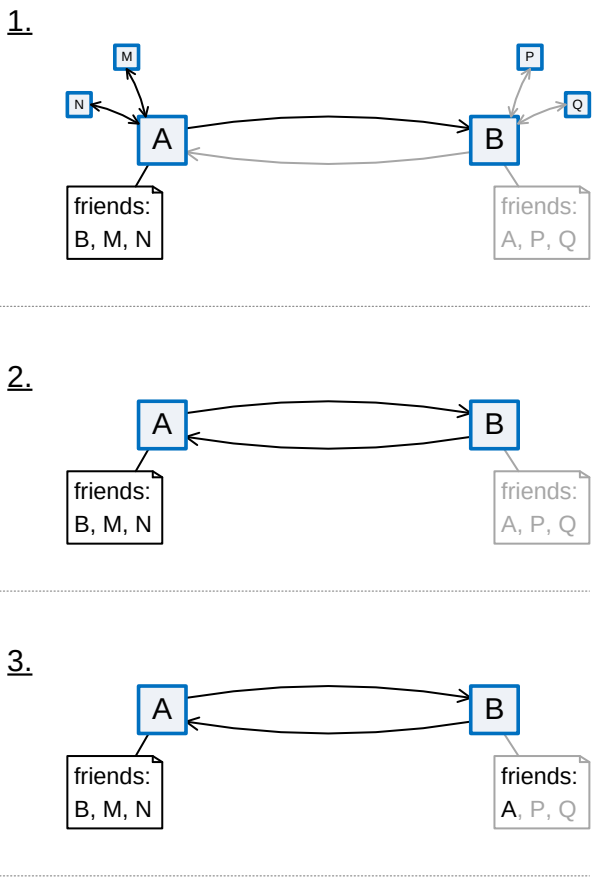

Figure 2: The friendship inference triple jump: the symmetry of friendship enables a privacy breach upon the uni-lateral disclosure of a friendship relation. (1) Bob wants to keep confidential his list of friends (greyed out), but Alice reveals her list in which Bob is included. The enforced symmetry of friendship relations enables an inference of friends between B and A (2). In consequence the contents of Bob's list become partly inferable (3). This inference affects only the friendship between A and B; B's friendship relations with P and Q are not (yet) unveiled.

4. IMPLEMENTING HIDDEN FRIENDSHIP RELATIONS

In a social network with a centralised authority and clients being constantly able to establish a connection to this authority, hidden friendships can easily be implemented by views on a user's list of friends. A customised view on the outgoing friendship relations of a SN member is generated for each request. Links for which the friends have agreed not to make them public will purged from the view unless the hidden friend herself is requesting the list. Potentially, this procedure can be expanded to incorporate access con-

trol based on group membership. In a large social network enough, the number of hidden friends could be displayed in addition without a substantial loss in privacy.

In a social network which is distributed in nature by either a peer-to-peer architecture or by the pragmatic constraints of mobile ad-hoc networking, the approach sketched above is not applicable. Friendship relations are no longer stored centrally but on each of the members' clients. Even if a central authority still exists, connectivity to it may be intermittent or too costly to establish.

The FOAF (Friend-of-a-friend) standard has emerged from the Semantic Web initiatives to encode personal information and relationships in a machine-readable format. Its design goals include the codification of a local view of the social network built on friendship and trust one situates oneself in. Built on top of RDF, FOAF files can include varying amounts of personal information – such as name and e-mail address – and links to other persons to be discovered by URIs. The author of a FOAF file provides an encoding of her own community, makes it available at a URI and discoverable by creating links to it. Because everybody can publish FOAF files autonomously, it is the method of choice for storing profile and relationship information in a distributed social network.

Using the `foaf:knows` property, the author of a FOAF file can manifest a link that exists with other persons, whose own FOAF files can in turn be referenced using the `seeAlso` property of RDF schema. A `knows` relationship indicates some level of reciprocated interaction between the parties, without however, having a central authority that could enforce this symmetry. The reciprocity can, however, be checked by searching for a corresponding back-link in the FOAF file of the referenced acquaintance with the understanding that a friendship only exists if the reciprocity holds. FOAF is therefore a well-suited technology for a distributed social network infrastructure. Yet, the `foaf:knows` properties still establish public friendship relations.

We propose to encrypt the identifiers of referenced persons to whom a hidden friendship shall be established. There is a *public friendship* between user A and user B if and only if A `foaf:knows` B and B `foaf:knows` A. There is a *hidden friendship* between user A and user B if and only if A `foaf:knows` $H_B(\mathrm{A})$ and B `foaf:knows` $H_A(\mathrm{B})$ with H_U being a secure hash function with a key only known to U.

As an example, consider the small social network depicted in Figure 3. There is a public friendship between A and C and both of them have listed each other in their FOAF files. These FOAF files are public and anybody can validate the reciprocity of the friendship. In particular, A and C themselves can check the symmetry of the relation prior to allowing a friend-restricted action. A and B want to be hidden friends. They do not reference one another; one cannot identify a relationship between them. The entry $H_B(\mathrm{A})$ in A's FOAF file can only be decrypted by B, similarly only A can decrypt the entry $H_A(\mathrm{B})$ in B's FOAF file. If A approaches B to perform a privileged action such as reading B's online diary, A can present $H_B(\mathrm{A})$ as a credential. Figure 4 shows the corresponding FOAF file fragments of A and B. The

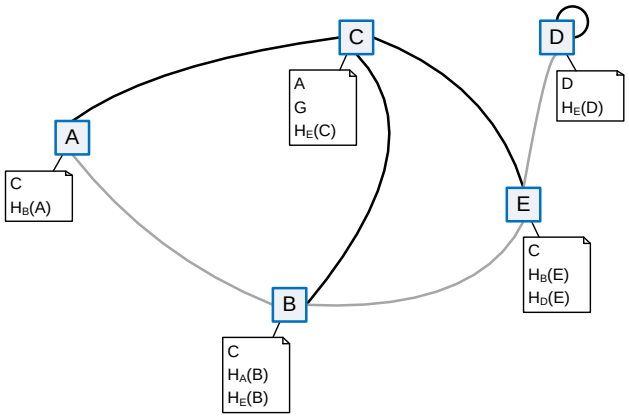

Figure 3: Detail of a social network with the users A, ..., E. Black (grey) arcs indicate public (hidden) friendship. Each entity has attached a public FOAF file with the entities she `foaf:knows`.

creation of a hidden friendship between two users equals a securely hashed identifier being added to each other's FOAF files. Standard mechanisms such as SSL or certificates can be applied transparently on top of our approach to ensure secure transmission and strong authentication.

5. DISCUSSION

Our approach of securely hashing identifiers for hidden friendship relationships presents several theoretical and technical advantages which are particularly valuable for deployment in mobile social networking:

Accuracy. Public friendship relations remain public and openly queryable. Private friendships cannot be identified by outsiders.

Syntactically seamless. Stating `foaf:knows` relations with encrypted identifiers in a FOAF file is compatible with the FOAF vocabulary and the underlying RDF standard. FOAF files stating hidden friendships are valid.

Cacheability. Because the coding of a hidden friendship does not rely on secrecy but instead on the secure hash of the friend's identifier, the members of a social network can continue to publish their FOAF without restriction. In particular, an access restriction would not add any further confidentiality to a hidden friendship. This public character of a FOAF file allows for its caching on proxies or on a SN member's device itself to cope with intermittent connectivity. Moreover, FOAF files can be aggregated and served by a third-party directory service. An expiry date of the FOAF files indicates the timespan after which a fresh retrieval must be made. Whilst neither RDF or FOAF provide for expiration, delivering FOAF files over HTTP enables cache control (for instance `Cache-Control: public, max-age=86400, must-revalidate` indicates a 24h lifetime). The expiry date is consequently user-defined, but global to the FOAF file implying the absence of socially divisive friendship discrimination.

Versioning and archivability. Archiving of personal FOAF

```
<foaf:Person rdf:ID="me">

  <foaf:name>A@sns.example.com</foaf:name>

  <foaf:knows> <foaf:Person>
      <foaf:name>C@sns.example.com</foaf:name>
  </foaf:Person> </foaf:knows>

  <foaf:knows> <foaf:Person>
      <foaf:name>U?LhJM1F?([y7QO,D</foaf:name>
  </foaf:Person> </foaf:knows>

</foaf:Person>
```

```
<foaf:Person rdf:ID="me">

  <foaf:name>B@sns.example.com</foaf:name>

  <foaf:knows> <foaf:Person>
      <foaf:name>C@sns.example.com</foaf:name>
  </foaf:Person> </foaf:knows>

  <foaf:knows> <foaf:Person>
      <foaf:name>,kZ#H8h)k^Le~M;RD</foaf:name>
  </foaf:Person> </foaf:knows>

  <foaf:knows> <foaf:Group>
      <foaf:name>WHr|7htwHj^.L1d5?</foaf:name>
  </foaf:Group> </foaf:knows>

</foaf:Person>
```

Figure 4: Excerpts of the FOAF files of users A (top) and B (bottom). One cannot tell that A and B are friends, yet each of them may test for one another still listing the other as a friend.

files – as durable caching – and versioning within a public or personal archive is possible and can be used to share portions of a personal database such as MyLifeBits.

Disclosure breaks the hidden friendship. If A and B are hidden friends an A unveils this friendship by replacing the entry "H_B(A)" in her FOAF file by "B", then the friendship is broken because B will no longer be able to confirm the presumed friendship by querying for "H_B(A)". If A does not replace, but adds an entry "B", then this will not create a public friendship, because there will be no matching entry "A" in B's FOAF file. As a result, friendship relations cannot be disclosed uni-laterally.

Immunity to replay attacks. Wireless communication is susceptible to interception and unencrypted data transmissions over WiFi is still common. The entry "H_B(A)" that B transmits to A when they engage in a friendship may be overheard and placed in another FOAF file by an attacker. However, the validity of "H_B(A)" is contingent upon this entry being located in A's FOAF file. The attacker, C, would need to produce "H_B(C)", which is impossible because H_B is only known to B. Therefore, B can identify such a forgery.

Negligible computational overhead. Establishing, querying, and revoking a hidden friendship relation requires small overhead. When two members A and B in a social network become hidden friends, they need to compute the encrypted identifiers $H_A(B)$ and $H_B(A)$. The same token has to be computed when checking for the existence of a friendship relation: if approached by A, B computes $H_B(A)$ and tries to find this entry in A's FOAF file. This querying can be optimised to $O(1)$. The revocation of friendship is achieved by the deletion of the corresponding entry in one's own FOAF file.

Obfuscation and deniability. Members can obfuscate their number of hidden friends by adding `foaf:knows` entries with arbitrary identifiers from the domain of the secure hash function. If the hash function's domain is a subset of its co-domain, one cannot tell apart hidden and uni-lateral public friendship relations. A member can then deny having hidden friends.

Non-exclusiveness. Public and hidden friendship relations do not interfere and both can co-exist, even with the same parties.

To the same extent to which friendship relations sustain higher level connections among the members of a centralised social network (see Section 2), the proposed approach for coding public and hidden friends in distributed FOAF files empowers privileged connectivity among social network members. Applications include the consumption of user-generated content to which access is restricted, and real-time communication among friends. A user may publish semi-static content (a blog, a diary, a photo journal) and restrict access to her friends without them knowing who exactly is entitled to read an entry. If a user's mobile device senses the proximity of a fellow member, it may check for an existing friendship link before sending status updates to the discovered user. The latter may in turn check herself for an established friendship prior to accepting the transmission. Similarly, geographically distant users may engage in an instant messaging conversation that new chatters can join if they are friends with all existing interlocutors.

In conclusion, hidden friendships overcome the privacy-deteriorating consequences of uni-lateral friendship disclosure which were proven to be substantial. Securely hashed identifiers in FOAF files are a lightweight, yet secure implementation of hidden friendship particularly suited for deployment in a mobile networking scenario.

6. REFERENCES

[1] Derlega, VJ; Metts, S; Petronio, S; Margulis, ST *Self-disclosure*, 1993

[2] Fisch, M, Gscheidle, C. Mitmachnetz Web 2.0: Rege Beteiligung nur in Communitys, in: *Media Perspektiven* **7**, (2008) pp. 356-364.

[3] Katz, JE; Aakhus, M. (eds.) *Perpetual Contact. Mobile Communication, Private Talk, Public Performance*, 2002

[4] Preibusch, S; Hoser, B; Gürses, S; Berendt, B. Ubiquitous social networks – opportunities and challenges for privacy-aware user modelling, in: *Workshop on Data Mining for User Modelling at UM 2007*, pp. 50-62

[5] sueddeutsche.de. *Japanerin nach 'Mord' an virtuellem Ehemann festgenommen*, October 27, 2008

[6] Western Morning News. *Virtual world 'affair' ends with real-life divorce*, November 14, 2008