Privacy-preserving concepts for supporting recommendations in decentralized OSNs

Marcel Heupel
Chair for IT Security Management
Information Systems Institute,
University of Siegen, Germany
heupel@wiwi.uni-siegen.de

Mohamed Bourimi
Chair for IT Security Management
Information Systems Institute,
University of Siegen, Germany
bourimi@acm.org

Simon Scerri
Digital Enterprise Research Institute
National University of Ireland,
Galway, Ireland
simon.scerri@deri.org

Dogan Kesdogan
Chair for IT Security Management
Information Systems Institute,
University of Siegen, Germany
kesdogan@uni-siegen.de

ABSTRACT

Recommender systems depend on the amount of available and processable information for a given purpose. Trends towards decentralized online social networks (OSNs), promising more user control by means of privacy preserving mechanisms, lead to new challenges for (social) recommender systems. Information, recommender algorithms rely on, is no longer available, (i.e. central user registries, friends of friends), thus shared data is reduced and centralized processing becomes difficult.In this paper we address such drawbacks based on identified needs in the decentralized OSN di.me and present concepts overcoming those for selected functionalities. Besides this, we tackle the support of privacy advisory, warning the user of risks when sharing data.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; K.6.m [**Security**]: Privacy; K.8.m [**Personal Computing**]: Miscellaneous

General Terms

Security, Design, Ontologies, Semantic Web, Social networking

Keywords

privacy and security, linking data, online social networks, decentralized social networks, di.me

MSM '13 Paris, France

1. INTRODUCTION

Nowadays, recommender systems are omnipresent, especially in online social networks (OSNs). They are recommending friends, events, locations and are used to optimize the placement of advertisements. The powerfulness and meaningfulness of recommender systems primarily depends on the amount of available and processable information for a given purpose. Recent trends towards decentralized OSNs (DOSNs) and enhanced privacy preserving mechanisms create new challenges for (social) recommender systems. DOSNs promise more user-control with respect to information disclosure compared to server-centric approaches. From the privacy point of view, server-centric solutions mostly allow for building fully-fledged user-profiles and have many other linkability and security issues. Nevertheless, centralized solutions offer great possibilities for recommendation, especially for providers who could access and process users' data (by ensuring end-users' consent in the terms and conditions of the followed usage policies). However, many information the recommender algorithms rely on, is no longer available in DOSNs such as friends of friends lists or user registries, thus having a negative effect on the support of recommendations and their quality.

In this paper we address selected challenges we were confronted with in our work in the European research project di.me. We present several privacy-preserving concepts in order to cope with those challenges and help using traditional (centralized) recommendation techniques in decentralized (social) environments, i.e. for selected cases related to (i) recommending contacts without disclosing contact lists, (ii) building lists of potential friends of contacts, for informing privacy advisory, and (iii) processing location information with non-linkable anonymous transactions. Besides this, we present the feasibility of those concepts in form of an approach providing user supportive privacy advisory, warning the user of potential risks when sharing data. We

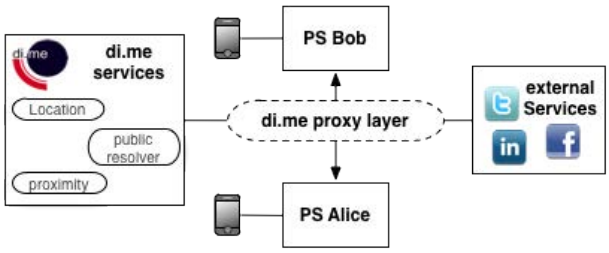

Figure 1: di.me architecture

thereby try to fulfill *multilateral security needs*[1] in order to maximize users' privacy by simultaneously providing recommendations as good as possible.

The remainder of this paper is structured as follows: In Section 2 we provide minimal background information about di.me, followed by the elucidation of identified challenges. In Section 3 we present our concepts to overcome those challenges. Section 4 compares our contribution to related work and Section 5 concludes by also addressing future directions.

2. PROBLEM ANALYSIS

The di.me project aims at integrating all personal data in a personal information sphere, guarded by a single, user-controlled point of access, located on a user-controlled server and realizing decentralized communication, avoiding external data storage and undesired data disclosure.[2]

The decentralized and user-control oriented di.me design brings some implications with it. DOSNs usually have a different view on the data than centralized systems. They operate only on a limited amount of data, in extreme cases only on information that was explicitly shared or is publicly available. This leads to the problem, that there is no central entity knowing who is connected to whom as in some OSNs like Facebook or LinkedIn. Consequently there is also no information about contacts of my contacts (e.g. friend-of-a-friend; FOAF). In order to give valuable recommendations in this respect (e.g., recommend new contacts) or privacy advisory (e.g., warn when sharing private information to untrusted persons) it is of major importance to find out to whom my contacts are connected to. However, contact information is very valuable, and should not be shared to the public. This is especially the case in situations where the user is interested in being anonymous. It has been proven (cf. [7]) that it is quite easy to re-identify people in different OSNs just by analyzing their social graph. Many studies state that people often have a tendency to share information to a larger audience than intended [1]. This might come from a lack of awareness or design of the application (intended or unintended by the service provider). Besides a considerable damage of personal reputation, this could also lead to violations of law, if sharing information of others [13]. **Thus, we identify the need of supporting contact related functionality in multilaterally secure and privacy-preserving way (Need 1; N1).**

[1]Considering security needs and requirements of all involved parties in a balanced way.

[2]The interested reader is referred to e.g., [11] or the project website (`http://www.dime-project.eu`), for more information about the project background.

Another key functionality of the di.me userware is not just giving recommendations, e.g., interesting locations nearby, but also to give advisory concerning trust and privacy. Therefore di.me's semantic core uses ontologies and information extraction techniques to provide various privacy-related features. Computational techniques extract personal information, to raise personal and social awareness and support the users' privacy. This opens the possibility to give intelligent advisory when the user is sharing information (profile information, files, microposts etc.) and might breach the user's or any other implicated person's privacy. In the case of microposts, we are employing Natural Language Processing (NLP) techniques in order to identify contained context and activity information. Another way to overcome FOAF functionality can be reached by extending this functionality of analyzing micro blogs, check-ins, status messages etc. in order to build social graphs and identify common contacts (see also [2]). **Thus, we identified the need to extend di.me for building such potential social contact graphs (N2).**

As di.me supports ubiquity, capabilities of modern smartphones are being leveraged for context-based recommendations in selected scenarios. Such devices and mobile operating systems (such as iOS or Android) offer native support (e.g. GPS sensor information) to be used in mobile and location-based scenarios. Usually, these use-cases are only dealing with the local processing of the own location, because there are a lot of privacy concerns and legal obligations, when processing location data. When it comes to recommender systems, also when considering the previously mentioned context-based recommendation, accurate location information becomes quite valuable. This is also the case for di.me, where we have the functionality to recommend the creation of ad-hoc groups, which means the formation of a group of contacts, based on their proximity to the user (e.g., in the same room for a certain time). Obtaining the proximity to one another, requires a certain degree of location disclosure (more exact disclosure will allow more exact proximity determination). In privacy-oriented environments like di.me, it is also needed, to have the possibility to share the current location without disclosing identity information, for example to support anonymous location information check-ins, or location recommendations. **Thus, we identify the need of privacy-respecting processing and sharing of context information**[3] **(N3).**

3. APPROACH

In the following, we present privacy-preserving concepts for our identified needs (**N1**-**N3**). We focus thereby on the di.me use-cases which are in most cases also applicable to other systems with similar, decentralized architecture.

For meeting **N1**, we suggest here a concept for privacy-respecting, pairwise comparison of contact lists, revealing only common contacts and nothing else. Even if people do not wish to share their complete contact list to the public, or even friends, it might still be of interest to them to know about common friends with a contact. To preserve the privacy of users, a suitable approach is to use special cryptographic mechanisms, allowing to share an encrypted version of the contact list. There are several algorithms fitting the problem (like e.g., Secure-Two-Party computa-

[3]In our example location and proximity information.

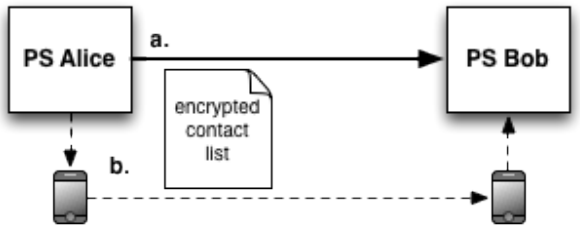

Figure 2: Exchange of encrypted contact list

tion, Private Set Intersection, Anonymous Credentials and others), but when considering performance and security the best approach is the one described by De Cristofaro et al. in [6]. They proposed a protocol for *Private Discovery of Common Social Contacts* allowing to compare contact lists and reveal only matching entries. In contrast to other approaches it uses *certified contacts*, in order to avoid impostors to add arbitrary contacts to their list, just in order to reveal all contacts of the other party. The protocol is carried out without the need of trusted third parties and also allows to leave selected contacts out of the processing, either because of context (e.g., leave out family and friends when just intend to reveal common business contacts) or other reasons. In di.me the protocol will be carried out directly between the PS's of two users, both sending an encrypted list of contacts to each other (see arrow a in Figure 2). Since the exchange does not depend on a specific network or communication protocol, it would also be possible to perform it on the users' smartphones using bluetooth for communication, if the users are in close physical proximity (arrow b. in Figure 2).

To meet **N2**, we describe the extension of the semantic core for building social graphs of potential contacts. In di.me we use the term *LivePosting* to refer to microposting practices (e.g. tweeting) on the Social Web. Microposts are retrieved from an integrated service API (including Facebook, Twitter and Google+). They can be re-shared with contacts in di.me, or pushed to other social networks, either as individual posts or within a broadcast *LiveStream* (which includes future posts). These so-called LiveStreams offer an opportunity to construct Social Graphs based on the identification of people with whom a person has regular contact. In order to construct Social Graphs for both a person and his/her contacts, references to people in microposts belonging to the visible network need to be i) discovered in the text and ii) reconciled with known person representations. The first stage of this process is enabled by an Information Extraction (IE) pipeline that decomposes each post to identify references to locations, activities and people [9]. The IE pipeline, implemented on top of GATE[4] and the ANNIE IE system [5], considers both semi-structured data (e.g. tagged persons) as well as unstructured text. To address the latter, NLP techniques are employed to parse textual microposts into typed named entities. Once typed entities have been discovered in a LivePost, we target their reconciliation with existing entities in local and external knowledge repositories. For this purpose, the IE pipeline also employs the GATE Large-Knowledge Base(KB) Gazetteer. In the context of this paper, this component is useful to dynamically extend the underlying gazetteers with person names extracted from online profiles, in order to improve their future matching to person-type entities, discovered in microposts. One limitation of this approach, and also a possible privacy-threat, is the possibility for a person to have multiple online profiles, including anonymous profiles and profiles where the owner adopts an alias. For our semantic reconciliation technique to be reliable, it is necessary to integrate multiple online profiles belonging to the same person into one general representation. In [4] we describe a technique for detecting semantic equivalence between one or more contacts' profiles (e.g., as retrieved from different social networks), in order to merge them under a unique person representation. Furthermore, the same technique is also used to attempt the discovery of possibly anonymous or fake profiles for a known contact, based on a high coincidence of attributes (especially inverse functional ones). This is not only useful to avoid privacy threats when sharing your own data with untrusted people, but also to get warnings about unintended linkability between your own profiles. The hybrid technique performs part syntactic-, and part semantic-based similarity measures on profile attributes (as supported by the NCO Ontology[5]), in order to produce a similarity score between any two profiles that do not necessarily have the same level of completion.

In order to satisfy **N3**, we elaborate concepts concerned with location storage and processing. Sharing dynamic information like the current location can be difficult in DSONs as it requires a lot of synchronization between peers in order to be up to date. Therefore, we envisioned in the di.me project, to tackle a hybrid approach. Location information can be stored (and updated respectively) on a centralized proximity service, but the sharing (meaning "allowing access") will happen in a decentralized way, by decoupling the link between location information and a person. One of the first basic constraints of the proximity service is that it may not reveal the location of persons directly. In order to get proximity information, i.e. information who is nearby, the user has to send his/her own location to the service and will get a list of nearby contacts as a response. Since the proximity service does not know the contacts of the requesting person, and the requesting user should also not be able to identify all people nearby (especially the ones unknown), the list of nearby contacts is anonymized. This means, the proximity service will only disclose a list of specially created identifiers (e.g. random UUIDs[6]). In order to be re-identifiable to selected contacts, the user needs to share this identifier once to contacts he/she is willing to share his location to. Further updates need only be communicated to the proximity service. It is possible to revoke the location sharing for selected contacts, by just using a new UUID and sending this to all the contacts that should still receive the location information. This approach still leaves room for future improvements. So it is envisioned to integrate an anonymous credential system, like. e.g., IBM's *idemix* [3] into the proximity service authentication mechanism. idemix is already being used in the di.me user registry[7] to guarantee only authorized access to the service (only registered di.me users

[4] `http://gate.ac.uk/`

[5] `http://www.semanticdesktop.org/ontologies/nco/`

[6] Universally Unique Identifier

[7] The di.me user registry is a service, where the users can optionally publish a minimal profile, in order to be found by other users.

can query for other users) but also allows anonymous, unlinkable transactions, making it impossible to link queries to a single user and would satisfy the most concerned users.

4. COMPARISON TO RELATED WORK

In [11], we analyzed existing social networks adopting the basic approach of decentralized privacy-ensuring systems. We restricted the set of compared systems to those being in a rather stable state (e.g., Diaspora, Friendica, Jappix, Kune and StatusNet), since in cases of many alpha releases it's almost impossible to get reliable results for an evaluation. All systems considered in [11] were analyzed with respect to several requirements, of whose the most relevant in the context of this paper is *Intelligent User Support with Context Sensitive Recommendations and Trust Advisory* with was not tackled by any of the other systems [2]. Recent work provides ways for recommending movies, products or collect opinions [10][12][8]. However, all presented prototypes or solutions follow a centralized way of processing data and do not address challenges which could arise in DOSNs as discussed above in this paper with use-cases from the di.me project. To our best knowledge, we are not informed about related work addressing functionalities related to requirements **N1** and **N2** for DOSNs. With respect to **N3**, there are many works in the security community that describe similar approaches. Our contribution in this respect is the application in the field of Semantic Web and recommender systems.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we analyzed selected drawbacks of privacy enhancing DSONs compared to OSNs with respect to effective recommendation of things (i.e. related to contact and context based information). We presented three privacy preserving concepts for meeting needs of selected use-cases in the di.me project with respect to (i) recommending contacts without disclosing contact lists, (ii) building potential friends lists for informing privacy advisory, and (iii) processing location information. All these concepts target to overcome identified drawbacks which arise in DOSNs and could hinder recommendation support. In fact, the concepts base on leveraging established security and Semantic Web functionalities by re-using them in the DOSN context for preserving privacy of end-users, while also increasing the available information for recommendation mechanisms. To our best knowledge, the presented concepts are not addressed in related work, at least not in this form and combination, namely, by basing on described security and semantic techniques. Future work will focus on implementing and integrating these concepts in di.me. One of the main challenges thereby, is to optimize the building of social graphs with respect to recognized entities (i.e. contacts) in live posts etc. This task could be made less ambiguous by involving other techniques for more accurate contacts' matching.

6. ACKNOWLEDGMENTS

This work was supported by the EU FP7 (FP7/2007-2013) project digital.me, under grant agreement no. 257787.

7. REFERENCES

[1] M. Beye, A. Jeckmans, Z. Erkin, P. Hartel, R. Lagendijk, and Q. Tang. Literature Overview - Privacy in Online Social Networks. 2010.

[2] M. Bourimi, I. Rivera, S. Scerri, M. Heupel, K. Cortis, and S. Thiel. Integrating multi-source user data to enhance privacy in social interaction. In *Proceedings of the 13th International Conference on Interaccion Persona-Ordenador*, INTERACCION '12, pages 51:1–51:7, New York, NY, USA, 2012. ACM.

[3] J. Camenisch and E. Van Herreweghen. Design and implementation of the idemix anonymous credential system. In *CCS '02: Proceedings of the 9th ACM conference on Computer and communications security*, pages 21–30, New York, NY, USA, 2002. ACM.

[4] K. Cortis, S. Scerri, I. Rivera, and S. Handschuh. Discovering semantic equivalence of people behind online profiles. In *Proceedings of the 5th International Workshop on Resource Discovery (RED 2012)*, 2012.

[5] H. Cunningham, D. Maynard, K. Bontcheva, and V. Tablan. GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications. In *Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL'02)*, 2002.

[6] E. De Cristofaro, M. Manulis, and B. Poettering. Private discovery of common social contacts. *International Journal of Information Security*, 12:49–65, 2013.

[7] S. Labitzke and I. Taranu. What your friends tell others about you: Low cost linkability of social network profiles. *Proceedings of the 5th International ACM Workshop on Social Network Mining and Analysis*, 2011.

[8] A. Papangelis, G. Galatas, and F. Makedon. A recommender system for assistive environments. In *Proceedings of the 4th International Conference on PErvasive Technologies Related to Assistive Environments*, PETRA '11, pages 6:1–6:4, New York, NY, USA, 2011. ACM.

[9] S. Scerri, K. Cortis, I. Rivera, and S. Handschuh. Knowledge discovery in distributed social web sharing activities. *Making Sense of Microposts (# MSM2012)*, pages 26–33, 2012.

[10] E. Shen, H. Lieberman, and F. Lam. What am i gonna wear?: scenario-oriented recommendation. In *Proceedings of the 12th international conference on Intelligent user interfaces*, IUI '07, pages 365–368, New York, NY, USA, 2007. ACM.

[11] S. Thiel, M. Bourimi, R. Giménez, S. Scerri, A. Schuller, M. Valla, S. Wrobel, C. Frà, and F. Herman. A requirements-driven approach towards decentralized social networks. In *Proceedings of the International Workshop on Social Computing, Network, and Services*, 2012.

[12] C. Wartena, W. Slakhorst, and M. Wibbels. Selecting keywords for content based recommendation. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, CIKM'10, pages 1533–1536, New York, NY, USA, 2010. ACM.

[13] S. Wrobel, M. Bourimi, M. Heupel, F. Herrmann, and M. Valla. Towards a minimal framework considering privacy and data protection goals for social networking platform providers. In *The Power of Information Conference '13, Extended Abstracts, Brussels*, 2013.