Enforcing Privacy in Decentralized Mobile Social Networks

Hiep H. Nguyen[1], Abdessamad Imine[1], and Michaël Rusinowitch[1]

LORIA/INRIA Nancy-Grand Est
France
{huu-hiep.nguyen,michael.rusinowitch}@inria.fr,abdessamad.imine@loria.fr

Abstract. This position paper first summarizes work done by the first author on location privacy and differential privacy. These techniques will help to solve privacy problems in *decentralized mobile social networks*, which is the main theme of his PhD research. The paper then briefly reviews the state-of-the-art in privacy-preservation of social graphs and clarifies the lack of attention to graph sharing in decentralized setting. Finally, some initial ideas on how to realize such *soft* decentralized access controls are described.

1 Motivation

The prevalent architectures of social networks mainly rely on the central repository of user data, so users must trade their privacy for free service. The centralized setting makes offline data processing operations (e.g. publishing a anonymized version of a social graph) easier and attracts a lot of interest in the security community. Parallel to this line of research, there also exists another line dedicated to decentralized alternatives [1] in which users manage their own data and access policies. Shifting from the free-service (centralized) paradigm to the user-centric (decentralized) one should, in principle, increase user confidence or trust, but understanding how to properly share information in this architecture remains a huge challenge in user privacy preservation. For example, early efforts such as Diaspora [2] and Persona [3] protect user privacy at the cost of less utility.

As a specific sub-problem, anonymizing social network data is more challenging [13] than anonymizing relational data. First, it is harder to model the adversarial background knowledge. Vertex/edge labels, neighborhood graphs and the like can be used to re-identify individuals, not only the conventional quasi-identifiers. Second, to measure information loss in perturbed network data is not easy because nodes and edges are not independent as items in tabular data. Third, anonymization techniques in social network data usually ask for deeper understanding of affects caused by a single graph modification. In decentralized settings, those anonymization challenges may be exacerbated as we need the collaboration of users to reach a global (noisy) view of the social graph.

[1] http://en.wikipedia.org/wiki/Comparison_of_software_and_protocols_for_distributed_social_networking

[2] https://diasporafoundation.org/

1.1 Thesis's objectives

Privacy in social networks involves privacy requirements for identities, links, contents, activities disclosure-resistance. The early vision of the PhD thesis is on understanding intrinsic difficulties of decentralized mobile social networks and to propose mechanisms for privacy-preserving data sharing (e.g. partial friend list sharing). We aim to start with simple models with several assumptions about the attacker's knowledge and dynamics of networks. While there are many structural attacks in centralized setting [2], we witness little similar research on those attacks in the decentralized one. One important reason may be due to the lack of the global view of the social graph; hence we conjecture that there will be different attack scenarios. We seek for an appropriate combination of cryptography and anonymization techniques to address the problem. Our ultimate objective is to make significant contributions both theoretically and empirically to the development of decentralized alternatives to the current online social networks (OSN).

The two following subsections summarize two main contributions of the first author in his master course. In each subsection, an anticipated application of the corresponding technique will be presented.

1.2 Location Privacy

Anonymization methods (e.g. k-anonymity) are extensively used in location privacy with a typical configuration of a anonymizer between the LBS server and mobile users. By cloaking user locations into bigger sets satisfying certain thresholds, users' true locations are concealed. Privacy in this setting means *"privacy by blending yourself in a crowd"*. *MeshCloak* [8] brings novelties in the cloaking problem by leveraging the real street maps for cloaking purpose. It tries to solve a harder problem of *continuous queries* which guarantees the privacy even in consecutive queries. Compared to previous work on the same problem, it produces much smaller cloaking areas, so reducing the processing burden at the LBS server. To formally quantify the privacy level of the new model, *MeshCloak* takes into account the stronger attackers knowledge by modelling user moving patterns based on time-homogeneous semi-Markov process with dwelling time at states. Specifically, each user with historical traces and a moving speed has a mobility profile. Under the assumption that the attacker may construct such mobility profiles and mount maximum likelihood attacks, *MeshCloak* still achieves high level of privacy (measured as high incorrectness of attacks).

Our technique and the anonymization methods applied to the publication of whole social graphs [11] are similar in the sense of hiding by blending protected objects into bigger sets. Meanwhile, existing techniques in [8] may help us in problems related to dynamics in social networks (arrival/leave of nodes as well as creation/deletion of links).

1.3 Differential Privacy

The second work [9] is on *differential privacy* [5], specifically applied to *linear counting queries* [7]. Differential privacy is an in-focus paradigm for publishing useful statistical information over sensitive data with rigorous privacy guarantees. It requires that the outcome of any analysis on database is not influenced substantially by the *existence* of any individual (so the name *differential*). Differential privacy has been successfully applied to a wide range of data analysis tasks. The comparative study [9] juxtaposes the state-of-the-art schemes and identifies advantages/disadvantages of each one as well as proposes improvements for better scalability and lower injected noise.

With strong theoretical foundations, differential privacy is a promising tool for formulating and solving certain privacy issues in both centralized/decentralized and offline/online settings for data publication within social networks [11].

2 Privacy Preservation of Social Graphs

Social graphs considered in the state-of-the-art vary in a wide range, from unlabeled, simple, undirected ones to data-rich ones. This section reviews a few typical privacy-preserving techniques in centralized and decentralized settings.

2.1 Centralized setting

There is a large body of work focusing on how to share social graphs owned by organizations without revealing the identities or links between users involved. The existing anonymization methods fall into four main categories. The methods of first category provide k-anonymity by *deterministic* edge additions or deletions, assuming attacker's background knowledge regarding some property of its target node. The second category includes *random* additions, deletions and switches of edges to prevent the reidentification of nodes or edges. The methods falling in the third category assign probabilities to edges to add uncertainty to the original graph. Finally, the fourth class of techniques cluster together nodes into supernodes of size at least k. Note that the last two classes of schemes induce *possible world* models, i.e., we can retrieve graph samples that are consistent with the anonymized output graph.

2.2 Decentralized setting

Decentralized social networks consist of *federated* and *distributed* models. In distributed setting, the social graph is split between many holders, and even between users in which each node only knows his friends. To privately aggregate the global graph, we need collaborative privacy-preserving sharing mechanisms. Unfortunately, not much attention has been given to this topic except in a few works, for example [10]. In tabular data aggregation, the distributed sharing problem seems easier and exposes a lot of efficient mechanisms, e.g. [1].

As we aim at mechanisms for decentralized social networks, the access control on network data is of importance. One of our concerns is how to devise “soft” access policies via collaborative anonymization. An anticipated scenario, presented in the next section, is that given different trust levels between a user and his friends, the user adds different noises to his true friend list before transmitting them to his neighbors.

3 Ongoing Work

At the early stage of the thesis, we survey the state-of-the-art of privacy-preserving methods in decentralized social networks as well as in the centralized ones. Privacy protection techniques can be classified roughly as cryptography-based and anonymization-based. Cryptographic primitives support a wide range of OSN-related operations like encryption to a group, group key revocation with attribute-based encryption (ABE) [3], searching on encrypted data, secure multiparty computation. However, this set of solutions incur unavoidable trade-off between secrecy, performance and complexity. Anonymization techniques while avoiding costly encryption may have other intricacies in designing online protocols, e.g. protocol for establishing a trust relationship, controlling the boundary of shared data (even in noisy forms). In addition, new protocols in our models require deliberate choice of replication schemes and consistency models.

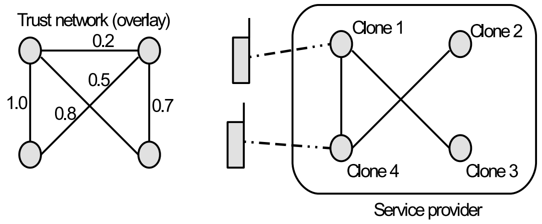

Fig. 1: Autonomous sharing of friend lists with trust-based *soft* access control

We propose initial ideas of autonomous relationship sharing (highly sensitive data) via trust-based access control. Fig. 1 depicts the sharing model under the assumption of Clone2Clone-like [6] distributed social networks. Each mobile user possesses one clone in the cloud for high availability and may allow the clone to perform actions on behalf of him. We assume an overlay trust network [4] between users to guide the sharing activity. Intuitively, each user determines the amount of noise added to his true friend list based on the trust levels (*fairness*) to a friend and then sends out the noisy friend list. The friend list should get noisier (*monotonicity*) as propagating within the network. Our goal is to allow each user to get friend lists from many sources, directly or indirectly, and to properly reconstruct the (noisy) global graph for his own purposes. An interesting question is how to guarantee the convergence of information in aggregated graph. We may start with several metrics such as randomness [12] and apply mining techniques for uncertain graphs to quantify the remaining utility in the graph.

4 Conclusion

Parallel to the success of centralized social network services based on the freemium model, decentralized alternatives have also been proposed as a response from the research community in order to better address privacy concerns of users. Following this direction, our methodology is to see first the big picture of decentralized social networks by studying their challenges and opportunities. We then aim to formalize problems of moderate size that can be stated and solved efficiently using both tools in cryptography and anonymization. One such in-focus problem is the collaborative and autonomous relationship sharing in cloud-based peer-to-peer architectures.

References

1. D. Alhadidi, N. Mohammed, B. C. Fung, and M. Debbabi. Secure distributed framework for achieving ε-differential privacy. In *Privacy Enhancing Technologies*, pages 120–139. Springer, 2012.
2. L. Backstrom, C. Dwork, and J. Kleinberg. Wherefore art thou r3579x?: anonymized social networks, hidden patterns, and structural steganography. In *Proceedings of the 16th international conference on World Wide Web*, pages 181–190. ACM, 2007.
3. R. Baden, A. Bender, N. Spring, B. Bhattacharjee, and D. Starin. Persona: an online social network with user-defined privacy. In *ACM SIGCOMM Computer Communication Review*, volume 39, pages 135–146. ACM, 2009.
4. B. Carminati, E. Ferrari, and A. Perego. Enforcing access control in web-based social networks. *ACM Transactions on Information and System Security (TISSEC)*, 13(1):6, 2009.
5. C. Dwork. A firm foundation for private data analysis. *Communications of the ACM*, 54(1):86–95, 2011.
6. S. Kosta, V. C. Perta, J. Stefa, P. Hui, and A. Mei. Clone2clone (c2c): Peer-to-peer networking of smartphones on the cloud. In *5th USENIX Workshop on Hot Topics in Cloud Computing (HotCloud13)*, 2013.
7. C. Li, M. Hay, V. Rastogi, G. Miklau, and A. McGregor. Optimizing linear counting queries under differential privacy. In *PODS*, pages 123–134. ACM, 2010.
8. H. Nguyen, D. Chae, J. Kim, and J. Kim. Meshcloak: A map-based location privacy model for continuous query in mobile services. Manuscript.
9. H. Nguyen and J. Kim. A comparison of differentially private linear counting queries. Manuscript.
10. T. Tassa and D. J. Cohen. Anonymization of centralized and distributed social networks by sequential clustering. *Knowledge and Data Engineering, IEEE Transactions on*, 25(2):311–324, 2013.
11. X. Wu, X. Ying, K. Liu, and L. Chen. A survey of privacy-preservation of graphs and social networks. In *Managing and mining graph data*, pages 421–453. Springer, 2010.
12. X. Ying and X. Wu. On randomness measures for social networks. In *SDM*, volume 9, pages 709–720. SIAM, 2009.
13. B. Zhou and J. Pei. Preserving privacy in social networks against neighborhood attacks. In *Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference on*, pages 506–515. IEEE, 2008.