PhD 2017

Title:
Sensitive Data Protection in Online Social Networks

- **Location:** LORIA /INRIA --- Nancy, France

- **Project-team:**
PESTO

- **Scientific Context:**

The majority of social networks (like Facebook, LinkedIn, etc) provide control functions to limit the visibility of certain data (such as friend list, wall posts and images) to a specific user group. However, most of users are unaware of the risks associated with the publication and exchange of personal data on social networks. For example, publishing and sharing location information on Facebook could easily lead to a burglary. Privacy risks may arise from explicit and/or implicit information that we can learn from online data. These risks due to data sharing are constantly increasing, allowing privacy attacks with unfortunate consequences, and making people very reticent to remain socially active (e.g. staying connected with friends and expanding friendship circle). To practice online social activities with greater confidence and less risk, it is imperative to devise tools that allow users to control themselves the usage that their data can be destined to. These tools assist users to detect and minimize the dissemination of personal information.

- **Missions:**

The main objective of this PhD project is to design software tools based on original solutions for self-controlled social data privacy. To do that, we will investigate anonymization and obfuscation techniques for preventing unintentional sensitive information disclosure from social network users and/or from their friends. We plan therefore to address the following issues:

1. Detection of privacy vulnerabilities

Each user has a profile containing some personal attributes (such as gender, age, location and religious and political affiliations) and describing relationships and interactions with other users. From these attributes, we have to state precisely the meaning of sensitive data/information. To do this, we will make use of qualitative and quantitative studies based on the preferences of social networks' users on the data/information that they consider sensitive.

Privacy risks may appear either directly after online data publication (e.g. finding a user's phone number within a wall post) or indirectly through an inference of private information (e.g. deducing sexual orientation from some friendship relations). In this part, we will propose a methodology for characterizing and building direct and indirect attacks. For direct attacks, given a user target, we will provide efficient algorithms for analysing the user vicinity in order to detect a subset of missing attributes. Since indirect attacks allow extracting information not explicitly stated in user profiles, we will combine algorithmic and statistical approaches to infer data with high probability. Extensive experiments will be conducted on real social network datasets to demonstrate the effectiveness of both attacks.

2. Definition and enforcement of privacy policies

When privacy vulnerability is detected, it may arise from one or several users linked by friendship relations. Any policy for privacy protection should be fair (i.e. it does not affect the privacy of other users) and optimal (i.e. it does not isolate completely user from the friendship circle). To eliminate or minimize privacy vulnerabilities, we plan to explore two trade-off techniques. The first one must combine optimally two possible actions: (a) hiding sensitive attributes (such as home address, email address and phone number) and (b) not disclosing friends to others. Besides a binary logic (publish or hide), the second technique enables us to change the semantics of the published information in such a way it becomes less accurate (or noised). This technique has to adapt some anonymization methods [1,2,3] (used for offline publication) for online user interactions.

Starting from differential privacy [4,5] and k-anonymity [6] tools, known to be quite successful for preserving individual privacy when querying databases, we will have to adapt theses concepts to social graph structures and user-centered applications where each node of the graph may have different privacy requirements. Moreover linked users share partial information about each other and this information may lead to privacy issues. Hence when a user needs to install a particular privacy policy this cannot rely solely on her/his personal setting but a collective responsibility. Therefore a trust negotiation protocol has to be designed so that each node in the social graph may inherit non-disclosure or anonymity constraints from her/his relatives. These constraints may require to be balanced with utility, and a game-theoretic solution will be investigated to find a trade-off.

Key words: Social networks, Privacy, Graph algorithms, Anonymization, Differential Privacy, Game Theory.

- **Bibliography:**

[1] H. H. Nguyen, A. Imine and M. Rusinowitch. “A Maximum Variance Approach for Graph Anonymization”. In Symposium on Foundations and Practice of Security (FPS), Best paper, 2014, pp. 49-64.

[2] H. H. Nguyen, A. Imine and M. Rusinowitch. “Anonymizing Social Graphs via Uncertainty Semantics”. In ACM Symposium on Information, Computer and Communications Security (ASIACCS), 2015, pp. 495-506.

[3] H. H. Nguyen, A. Imine and M. Rusinowitch. “Differentially Private Publication of Social Graphs at Linear Cost”. In IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 2015.

[4] C. Dwork. “Differential Privacy”. In International Colloquium on Automata, Languages and Programming (ICALP), 2006.

[5] H. Xi, M. Ashwin and D. Bolin. “Blowfish Privacy: Tuning Privacy-utility Trade-offs Using Policies”. In ACM SIGMOD International Conference on Management of Data, 2014.

[6] L. Sweeney. k-anonymity: a model for protecting privacy. International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 10 (5), 2002; 557-570.

- **Skills and profile**: Applications are invited from those with a Master qualification in computer science and good knowledge in a directly relevant area (graph algorithmic, probability and statistics, security, privacy or data-mining). The applicants should have a passion for collaborating in inter-disciplinary research and the academic skills to be part of a research team aiming to make a high impact.

- **Additional information:**

Supervision and contact:
Abdessamad Imine
Pesto Team
Université de Lorraine & LORIA-INRIA, Nancy, France
Tel: +33 (0)354958535
email: abdessamad.imine@loria.fr

or

Michaël Rusinowitch
Pesto Team
LORIA-INRIA, Nancy, France
Tel: +33 (0)383593020
email: michael.rusinowitch@loria.fr

Additional links: *(if any)*

Duration: 3 years
Starting date: between Oct. 1st 2017 and Jan. 1st 2018
Salary: 1 958 euros gross monthly (about 1 580 euros net) during the first and the second years. 2 059 euros the last year (about 1 661 euros net). Medical insurance is included.

The required documents for applying are the following :
- CV;
- a motivation letter;
- your degree certificates and transcripts for Bachelor and Master (or the last 5 years if not applicable).
- Master thesis (or equivalent) if it is already completed, or a description of the work in progress, otherwise;
- all your publications, if any (it is not expected that you have any).
- At least one recommendation letter from the person who supervises(d) your Master thesis (or research project or internship); you can also send at most two other recommendation letters.

The recommendation letter(s) should be sent directly by their author to the prospective PhD advisor.

All the documents should be sent in at most 2 pdf files; one file should contain the publications, if any, the other file should contain all the other documents. These two files should be sent to your prospective PhD advisor (in addition to the application on the web).

Help and benefits :

- Possibility of free French courses
- Help for finding housing
- Help for the resident card procedure and for husband/wife visa
- Lunch cost at Inria is 2,72 €