Privacy Management in Agent-Based Social Networks (Doctoral Consortium)

Nadin Kökciyan
Supervisor: Pınar Yolum
Department of Computer Engineering, Bogazici University
Istanbul, Turkey
nadin.kokciyan@boun.edu.tr

ABSTRACT

In online social networks (OSNs), users are allowed to create and share content about themselves and others. When multiple entities start distributing content, information can reach unintended individuals and inference can reveal more information about the user. Existing applications do not focus on detecting privacy violations before they occur in the system. This thesis proposes an agent-based representation of a social network, where the agents manage users' privacy requirements and create privacy agreements with agents. The privacy context, such as the relations among users, various content types in the system, and so on are represented with a formal language. By reasoning with this formal language, an agent checks the current state of the system to resolve privacy violations before they occur. We argue that commonsense reasoning could be useful to solve some of privacy examples reported in the literature. We will develop new methods to automatically identify private information using commonsense reasoning, which has never been applied to privacy context. Moreover, agents may have conflicting privacy requirements. We will study how to use agreement technologies in privacy settings for agents to resolve conflicts automatically.

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*Multiagent systems*

General Terms

Algorithms, Management, Design

Keywords

Privacy, Commitment, Ontology

1. INTRODUCTION

Typical examples of privacy violations on social networks resemble violations of access control. In typical access control scenarios, there is a single authority (i.e., administrator) that can grant accesses as required. However, in social networks, there are multiple sources of control. That is, each user can contribute to the sharing of content by putting up posts about herself as well as others. Further, audience of a post can reshare the content, making it accessible for others. These interactions lead to privacy violations, some of which are difficult to detect by users.

Our review of privacy violations reveal two important axis for understanding privacy violations. The first axis is the main contributor to the situation. This could be the user herself putting up a content that reveals unwanted information or it could be other people sharing content that reveals information about the user. The second axis is how the information is revealed; if the information was itself unwanted or the information led to new information being revealed (i.e., through inferences). According to these two axis, we identified four types of privacy violations. In type (i), a user shares some content with some privacy settings, the system acts against these settings and shares the content with people that it was not supposed to. In type (ii), an information about a user is shared by another person. In online social networks, information about a user can easily propagate in the system, without a user's consent. In type (iii), a user puts up a content on the social network without realizing that more information can be inferred from her post; e.g., giving away location information through a landmark. In type (iv), a friend's action leads to a privacy leakage but the leakage can only be understood with some inferences in place; e.g., a friend's tag revealing friendship status. Moreover, a content may lead to privacy violations because of its semantics. A post may annoy or insult the user, or it may include private information; e.g., sharing a post that reveals the user's politic affiliation. This thesis develops an approach for managing users' privacy constraints in online social networks for detecting privacy violations and guide the user to protect her privacy as well.

2. APPROACH

We would like to detect privacy violations and warn the user before privacy violations occur in the system. In another words, we envision an intelligent digital assistant, which would automatically interact with the user and other users to protect the user's privacy. However, we focus on intelligent reasoning over the content and the privacy requirements of users. Our work proposes an agent-based representation of a social network, where each agent (e.g., digital assistant) represents a user. The agents manage users' privacy requirements, which are represented via privacy agreements. A privacy agreement should be structured so that agents can process it automatically and reason about it. A logic-based

Appears in: *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2015), Bordini, Elkind, Weiss, Yolum (eds.), May 4–8, 2015, Istanbul, Turkey.*

representation would be appropriate since agents can infer new information from the existing knowledge. Hence, the privacy context, such as the relations among users, various content types in the system, and so on are represented with a formal language. By reasoning with this formal language, an agent checks the current state of the system to resolve privacy violations before they occur. An agent notifies its user to take an action according to detection results.

We have an initial framework, PriGuard, that detects privacy violations in the OSN [1]. We have developed a PriGuard ontology to represent the details of a social network such as users, relationships between users and content being shared. Semantic rules are used for making further inferences using developed ontology. Our approach consists of four steps. (i) A user of the OSN specifies her privacy constraints where she declares her privacy constraints using PriGuard[1] interface. The user specifies who can or cannot see some specific content such as media, location, people that the user is together with. (ii) Then, agents create privacy agreements between the users and the system through commitments [3]. The contents of a commitment are represented using PriGuard ontology as well. OSN commits to the user to act according to the generated commitments. (iii) Following this, an agent generates the statements wherein these commitments would be violated. For this, we model violation statements as Prolog rules. (iv) Finally, the system checks whether these statements hold in the current state, which would mean a violation of privacy. For detection, PriGuard uses the ontology, semantic rules that are used by the OSN for semantic operations, the current state of the social network and the violation statements. If PriGuard can prove a statement, then the corresponding commitment is violated and the user is notified to take an appropriate action. Our approach can detect privacy breaches that result from commitment violations, commitment conflicts as well as inferences that cannot be detected in mainstream social networks. In the extension of our work, we evaluate the scalability of our approach on generated as well as Facebook data. For this, we replicate a privacy example where a user shares a content with everybody and tags his friend. However, his friend does not want to show this content to anyone. In a real-network of 4093 users with 88234 relations, it takes approximately 4 minutes to detect privacy violations with a memory consumption of 45MB on an ordinary computer. Our preliminary results are encouraging and show that our approach can scale to real-world networks.

Privacy violations are taking place because of different privacy concerns, based on context, audience, or content that cannot be enumerated by a user up front. Accordingly, privacy should be handled per post and on demand among all that might be affected. Hence, a post should be compatible with the user's privacy constraints, and other users' privacy constraints as well. In a recent work, we propose an agent-based social network where agents negotiate on the privacy concerns that will govern the content. We employ a negotiation protocol and use it to settle differences in privacy expectations. An agent reasons on its user's privacy constraints and decides on whether a post is compatible with its user's privacy constraints. Then, an agent can approve an offer made by a negotiator agent or reject it by providing structured reasons. The negotiator agent collects reasons from other agents, it can revise its post if necessary so that it can satisfy privacy constraints of others, or it can publish the post as it is. Privacy violations are minimized since agents negotiate before sharing a post.

[1]A demonstration is available at `http://mas.cmpe.boun.edu.tr/nadin/priguard`

3. FUTURE DIRECTIONS

So far we have an initial implementation that can represent privacy agreements (between an agent and the OSN) through commitments and the corresponding violations statements, and detect privacy violations in a centralized way. First, we want to extend this implementation to cover privacy agreements between agents. Second, we want to improve our initial model to detect privacy violations before they would occur in the system. For this, we will improve our model by enabling agents to reach agreements automatically. Agents may have conflicting privacy constraints and we want them to be able to resolve such conflicts. In our preliminary work, we use a negotiation protocol so that agents can agree on a mutually acceptable privacy agreement. In the next step, we will investigate further how to use agreement technologies in privacy context; e.g., formulation of an offer or a counter-offer in negotiation; formulation of an argument and an attack in argumentation.

Various privacy examples are reported in the literature; we argue that commonsense reasoning [2] could be useful to solve some of them. For this, some information can be extracted from a post; e.g., the context of a post, entities in a post text or a picture and so on. Such information can be processed by an agent to decide whether a post is private. For example, if an agent knows that a diamond ring is in a picture, and the context is an engagement; then, it can proactively notify its user that the picture might be private and suggest him to not show the picture to his girlfriend. Hence, the user can take an action (revise, publish or delete post) to protect his privacy. We will study logic-based (e.g., Cyc) and language-based (e.g., ConceptNet) commonsense reasoning tools, which have never been applied to privacy context. Finally, we will improve our model so that it can be used for detecting privacy violations in a distributed way since users cannot trust a central entity (e.g., the OSN operator) to protect their privacy. This requires collecting evidence from other agents in the social network. We will develop new methods for agents to collect such evidence and process it for detecting privacy violations. We will evaluate our approaches on scenarios from the literature.

Acknowledgments

This research has been supported by The Scientific and Technological Research Council of Turkey (TUBITAK) under grant 113E543.

REFERENCES

[1] N. Kökciyan and P. Yolum. Commitment-based privacy management in online social networks. In *First International Workshop on Multiagent Foundations of Social Computing at AAMAS*, 2014.

[2] J. McCarthy. Artificial intelligence, logic and formalizing common sense. In *Philosophical Logic and Artificial Intelligence*, pages 161–190. Springer, 1989.

[3] M. P. Singh. An ontology for commitments in multiagent systems. *Artificial Intelligence and Law*, 7(1):97–113, 1999.