Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16)

Privacy Management in Agent-Based Social Networks

Nadin Kökciyan
Department of Computer Engineering
Bogazici University, 34342 Bebek, 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.

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.

Approach

We envision an intelligent digital assistant, which would automatically interact with the user and other users to protect the user's privacy in the social network. For this, 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 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 accordingly.

In our recent work, there are three contributions that we make. (i) A meta-model for agent-based online social networks is developed. Hence, agent-based social networks, privacy requirements, and privacy violations can be formally

defined. (ii) We develop a semantic approach called PRIGUARD that conforms to this meta-model. This approach is based on description logic to represent social network information and multiagent commitments (Singh 1999) to represent users' privacy requirements. We propose a detection algorithm that performs reasoning using the description logic and commitments on a varying depths of social networks. (iii) We build an open-source software tool PRIGUARDTOOL that implements the approach using ontologies (Kökciyan and Yolum 2014). First, a user of the OSN specifies her privacy constraints where she declares her privacy constraints using PRIGUARDTOOL[1] interface. The user specifies who can or cannot see some specific content such as media, location, people that the user is together with. Second, agents create privacy agreements between the users and the system through commitments. OSN commits to the user to act according to the generated commitments. Third, an agent generates the statements wherein these commitments would be violated. Finally, the system checks whether these statements hold in the current state, which would mean a violation of privacy. For detection, PRIGUARDTOOL 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 PRIGUARDTOOL can prove a statement, then the corresponding commitment is violated and the user is notified to take an appropriate action. We show that our approach can detect different types of privacy violations. We evaluate the scalability of our approach on generated as well as Facebook data. In a real-network of 4039 users with 88234 relations, it takes approximately 0.4 seconds to detect privacy violations 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 (Mester, Kökciyan, and Yolum 2015). 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. In a recent work, we propose an agent-based social network where agents negotiate with each other through arguments. When a user decides to share some content, the user's agent first contacts agents that are involved in that content. These agents evaluate the content using their own ontology. If the content is acceptable, then the content can be shared as it is. However, if one of the agents rejects the content, then the agent provides *arguments* to defend why the content should not be shared as it is. The agent's ontology is used to generate arguments. The requesting agent can continue the dispute by providing further arguments. This continues in a turn-taking fashion and eventually when there are no more arguments to be provided, the existing arguments are evaluated in an argumentation system, which decides on sharing the content or not.

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

Future Directions

Before AAAI 2016 conference, I will focus on improving our PRIGUARD approach. First, I will work on a formal proof of soundness and completeness of our proposed detection algorithm. Second, I will use large social networks to see the scalability of our approach. Third, I will compare the use of negotiation and argumentation techniques in the privacy context. After the conference, my plan is to work on commonsense reasoning to solve some of the privacy examples reported in the literature. 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. I will use existing commonsense reasoning tools such as Cyc and ConceptNet. Finally, I 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 approach on scenarios from the literature and real-world social networks.

Acknowledgments

This research has been supported by The Scientific and Technological Research Council of Turkey (TUBITAK) under grant 113E543 and by the Turkish State Planning Organization (DPT) under the TAM Project, number 2007K120610.

References

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

Mester, Y.; Kökciyan, N.; and Yolum, P. 2015. Negotiating privacy constraints in online social networks. In Koch, F.; Guttmann, C.; and Busquets, D., eds., *Advances in Social Computing and Multiagent Systems*, volume 541 of *Communications in Computer and Information Science*. Springer International Publishing. 112–129.

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