Linked Data and Mobile Application Privacy

Evan W. Patton[1] and Ilaria Liccardi[2,3]

[1] Dept. of Computer Science, Rensselaer Polytechnic Institute, Troy, NY USA
[2] Oxford e-Research Centre, University of Oxford, Oxford, UK
[3] MIT CSAIL, Cambridge, MA USA

Abstract. The smartphone applications ecosystem lacks significant controls and accountability mechanisms to protect users' privacy from snooping applications, especially when context may significantly impact the decision to share sensitive information. In this paper, we discuss some of the existing privacy concerns of mobile applications. In particular we highlight how semantic web and linked data technologies have the potential to help improve privacy controls through rules, queries, and reasoning. These technologies have the potential to obtain and maintain the integrity of private data, but there are privacy challenges that may result from widespread deployment of mobile semantic technologies.

Keywords: privacy, mobile applications, linked data, accountability

1 Introduction

Mobile devices are a ubiquitous part of modern life, with over 2 billion estimated cellular connections in the world today [6]. The market for mobile applications, or 'apps', is significant. The increase in the number of apps has been accompanied by an increase in the amount of tracking and other forms of data collection on users, primarily for the purpose of generating advertising revenue to support application developers. A recent report by the United States Federal Trade Commission [4] shows that users are not always aware of the amount nor the extent to which companies track their behavior. We propose that Resource Description Framework (RDF) based technologies, such as rule engines, ontologies, and reasoners, can be used to assist mobile users in preserving their privacy by constructing a more robust platform for maintaining provenance about privacy-oriented data. Such tools will help users understand where their information is stored, how it is used, and create meaningful and informed decisions of how their information are shared. The paper is structured as follows:

- **People's reactions and attitudes towards current app access**: We highlighted related research that explores people's reaction when made aware of the kind of access apps have to their personal data. We highlighted possible reasons for developers wanting to access to these data (section 2).
- **Smartphone privacy scenarios**: We compared two hypothetical scenarios of apps accessing one's personal data. We described an implementation using semantic technologies that could grant or deny access to personal data based on context (section 3).

- **Possible solutions using semantic web technologies**: We presented possible implementations using semantic web technologies which are aimed at helping users' preserve their privacy (section 4).

2 Related Research

Personal information is collected at a large scale from smartphone devices. Different operating systems have their own ways and mechanisms of presenting the access requested by different apps and of gaining consent from users.

Android and Windows devices present users a set of permissions requested by each app prior to installation. On these devices, users must accept all permissions before the app can be installed. Often users are not aware of the type of access, and even if they are able to understand they often do not know the reasons, motivations and purposes for accessing this information.

Apple devices use a different method for informing and gathering consent. An app that requires access to users' personal information alerts the user to obtain permission the first time it is required. Users can decide whether to grant permission or not. While it is not always the case, denying access can result in the app failing to work properly or stopping working all together.

In all cases, apps can request access to personal data for monetization rather than core functionality. Mobile advertising is one of the most common and lucrative ways for app developers to monetize their apps [13]. Research on Android apps found that almost half of the apps (49% (473) of 964 apps) in the Google Play store requested multiple permissions only needed for advertising libraries [14]. Often users have no way of understanding differences in permissions, especially if personal information is legitimately needed by the app. For example, weather apps need the user's location to gather and display location-based weather. However they can also send that information to advertisers to serve targeted ads. Users are often unaware of this access, particularly because only a small percentage of apps (fewer than 10%) have privacy policies specified on their description page [10]. Even when present, users often ignore privacy policies because they are often too vague or incomprehensible due to technical and legal jargon [11]. While Apple's permission model allows users to have more fine-grained control over what data are accessed, users often grant access to all the information the app requires because they are unaware that the app uses their personal information for purposes other than the app's functionality.

People are often unaware that apps may collect their personal data [3] as permission mechanisms are often difficult to understand [11] and part of this collection can happen silently in the background [18]. When users are made aware of this collection, they feel much less willing to share data which they perceive be extremely sensitive [15]. Some express shock and a desire to remove the app [1,12] or experience a sense of "creepiness" that results in a loss of trust [16]. The perceived sensitivity of data is subjective and can also vary within the individual's context [15]. For example, a user might be willing to share when he or she is at a certain location or while engaging in a certain activity (e.g., relaxing), but not when performing another (e.g., working). Users are also less willing to disclose personal data when more specific usage details are provided [15], which suggests that user awareness of how data are used is critical to preserving privacy. It is impossible to consent to data collection for every foreseeable purpose, given the incomplete, missing or difficult to understand information users receive when making the decision about whether to install an app [11,15].

3 Apps and Personal data: Access, sharing and consent

This section outlines a hypothetical, yet realistic scenario that depicts what happens when we carry and/or use smartphone devices. We then propose a possible improvement to the scenario using semantic technologies to grant or deny access to personal data using contextual information.

3.1 How currently is people's data shared and access?

Alice needs to drive to work, but before she starts driving, she checks *routeInfo*, an app for real-time traffic updates (many such apps are present in app markets). To provide this information, *routeInfo* requests access to GPS. Alice is not aware, however, that other apps on her phone (e.g., a game she installed for her child) are also passively listening to these GPS updates to identify nearby businesses. These apps (like the majority of apps) use targeted advertising to generate revenue.

Alice arrives at work; her smartphone recognizes her work WiFi "Work-WiFi". Alice's company provides all employees with an app *openSesame* to navigate the building which grants access to rooms and personal offices.

Alice has another app, *remindME*, installed on her phone to help her with her errands, such as groceries and dry-cleaning. The app alerts Alice as she leaves work of her errands. She proceeds with all her to-dos while other apps are silently listening (e.g., apps gathering GPS updates for advertisement, *openSesame* knowing she is not longer at work). She runs her errands and drives back home. Her smartphone recognizes her home wifi "Home-WiFi".

Developers who have collected Alice's location can estimate that Alice lives at this address (Alice is there from evening to morning, everyday). Using public (possibly linked) data repositories, this information can be used to determine the average household income, home value, debt, number and length of car ownership, demographics, likely political affiliation, etc. All this information can be used by the developers for revenue (e.g. selling this information to data brokers), advertisement, etc. Alice never wanted or intended to supply such specific information, however. The lack of control over passive applications and types of information accessed has caused her device to leak significant private information about her without her **express** permission or intention.

3.2 How can semantic technologies improve smartphone privacy?

Alice needs to drive to work, but before she starts driving, she checks *routeInfo*, an app for real-traffic updates. To provide this information, *routeInfo* requests

GPS information and her context is set {DRIVING}. Alice only wants the foreground app to have access to GPS updates while driving. Since she has launched the app it receives GPS but other apps do not.

Alice arrives at work; her smartphone recognizes her work WiFi "Work-WiFi" and sets her context to *at*.{WORK}. Her company provides employees with an app, *openSesame*, that grants access to rooms and personal offices. Only *openSesame* is allowed to access information in the WORK context. She attends a meeting, her smartphone updates her context to *at*.{WORK} ∩ *located_in*.{BUILDING_A#ROOM_304B}, *openSesame* knows she is in the room.

Alice has another app, *remindME*, installed on her phone that helps her with her errands, such as groceries and dry-cleaning. The apps alerts Alice of her errands after she leaves work. As she proceeds with all her to-dos, her smartphone updates her context to *intent*.{RUNNING_ERRANDS}. *openSesame* can not collect the location in this context, nor can other apps unless Alice has allowed it.

She runs her errands and drives home. Her smartphone recognizes her home wifi "Home-WiFi" and her smartphone updates her context to *at*.{HOME}; all apps are prevented from collecting location.

Developers have collected Alice's location, but only for specific contexts. While Alice's personal information can still be deduced, the information is more sparse than before. Alice is also aware of the possibility for developers to access her location when using an app. While her device may leak personal information, it is dependent on the context information set by Alice.

4 Employing semantic technologies for privacy

Given limitations of current app permission models, we propose that semantic technologies will improve the experience for users by enabling systems to:

1. Capture **provenance** of sensitive data to explain possible privacy violations.
2. Model user's **intent** to share data (e.g., via description logic class expressions or rules).

Since many semantic technologies assume an open world, privacy preservation benefits from the fact that information can be *unknown*. We briefly explore some ways in which semantic technologies have been used to model privacy and how they might be adapted for mobile deployments.

4.1 Privacy information modeling

The Accountability in RDF (AIR) language was developed as a means of generating and tracking explanations on the Semantic Web [7]. Similar rule efforts include Rule Interchange Format (RIF) [8] and Semantic Web Rule Language (SWRL) [5]. PROV-O [9] provides an ontology for modeling provenance information in a linked way. Using a PROV middleware, one could track how private information are manipulated by different applications and shared with third parties. Users could also specify a custom privacy policy using intensional (i.e., described) contexts and/or rules to describe additional limits on applications.

4.2 Controlling access to private information

Tao et al. demonstrate query answering with privacy preservation in the $\mathcal{EL}$ description logic [17]. They show that a set of secrets can be protected by inverting the application of $\mathcal{EL}$ expansion rules to compute an "envelope of secrets." Queries are answered using secrets so long as the system will not reveal sufficient information in the envelope that could infer secrets. Further, the computational complexity of both $\mathcal{EL}$ and the envelope creation is polynomial, which makes mobile deployments feasible. It is an open question whether secrets expressible in a more expressive logic (e.g., $\mathcal{ALC}$) will face tractability issues.

The WebID[4] and WebACL[5] community groups at the W3C are iterating on frameworks for identification, authentication, and authorization built on RDF-based representations. While these technologies are primarily aimed at web-based deployments, one can see how their approach to access control using constraints structured in RDF could be repurposed for a mobile environment.

Chen et al. explore use of semantic technologies in a Context Brokering Architecture for mobile environments [2]. They use the architecture to prototype an intelligent meeting room that provides services by inferring when users enter the room from sensor data streams. While context can be used to customize an environment to the user's needs, we note that devices might be unintentionally informing untrusted external systems about user context as well.

5 Conclusion

The contextual features provided by mobile devices are a double-edged sword for end users. They provide significant information to third-party applications to tailor behavior to one's situation, but can also provide others, e.g. advertisers, a significant look at one's behavior. Worse, the permissions models on mobile phones lack transparency and fine control over how context is used and disseminated. We have presented a scenario where semantic technologies could be used to capture provenance of private data and enforce user privacy policies.

Other techniques may be necessary to supplement linked data approaches to modeling dynamic user context. Markov models, for example, may provide a more compact, statistical representation of past context in a space and time complexity amenable to mobile devices compared with tableaux reasoning about context provenance modeled in PROV-O. Due to the error present in the various sensors used to infer a user's context, probabilistic models and logics will also be relevant to enabling privacy-preserving technologies for mobile devices.

There are a number of open research questions for the community. Do we need to develop novel representation languages for context and privacy, especially in the face of probabilistic contextual features, or is RDF sufficient? What is the maximal feature space for context and privacy in description logics and is their evaluation on mobile devices sufficient? How can reasoning about context be made efficient on mobile devices to reduce energy drain?

[4] `http://www.w3.org/wiki/WebID`

[5] `http://www.w3.org/wiki/WebAccessControl`

References

1. Balebako, R., Jung, J., Lu, W., Cranor, L.F., Nguyen, C.: Little brothers watching you: Raising awareness of data leaks on smartphones. In: Proc. 9th SOUPS. pp. 12:1–12:11 (2013)
2. Chen, H., Finin, T., Joshi, A.: An intelligent broker for context-aware systems. In: Adjunct proceedings of Ubicomp. vol. 3, pp. 183–184 (2003)
3. Felt, A.P., Egelman, S., Wagner, D.: I've got 99 problems, but vibration ain't one: A survey of smartphone users' concerns. In: Proc. 2nd SPSM. pp. 33–44 (2012)
4. FTC Staff: Mobile privacy disclosures: Building trust through transparency. Tech. rep., Federal Trade Commission (2013)
5. Horrocks, I., Patel-Schneider, P.F., Boley, H., Tabet, S., Grosof, B., Dean, M.: SWRL: A semantic web rule language combining OWL and RuleML. Tech. rep., W3C (2004)
6. ICT Data and Statistics Division: The world in 2014: ICT facts and figures. Tech. rep., International Telecommunications Union (2014)
7. Kagal, L., Jacobi, I., Khandelwal, A.: Gasping for air why we need linked rules and justifications on the semantic web. Tech. rep., MIT (2011)
8. Kifer, M., Boley, H.: RIF overview (2nd edition). Tech. rep., W3C (2013)
9. Lebo, T., Sahoo, S., McGuinness, D.L.: PROV-O: The PROV ontology. Tech. rep., W3C (Apr 2013)
10. Liccardi, I., Pato, J., Weitzner, D.J.: Improving Mobile App selection through Transparency and Better Permission Analysis. J. Privacy and Confidentiality 5(2), 1–55 (2014)
11. Liccardi, I., Pato, J., Weitzner, D.J., Abelson, H., De Roure, D.: No technical understanding required: Helping users make informed choices about access to their personal data. In: Proc. 11th MOBIQUITOUS. pp. 140–150 (2014)
12. Lin, J., Amini, S., Hong, J.I., Sadeh, N., Lindqvist, J., Zhang, J.: Expectation and purpose: Understanding users' mental models of mobile app privacy through crowdsourcing. In: Proc. 14th Ubicomp. pp. 501–510 (2012)
13. Merisavo, M., Vesanen, J., Arponen, A., Kajalo, S., Raulas, M.: The effectiveness of targeted mobile advertising in selling mobile services: an empirical study. Int. J. Mobile Commun. 4(2), 119–127 (2006)
14. Pearce, P., Felt, A.P., Nunez, G., Wagner, D.: Addroid: privilege separation for applications and advertisers in android. In: Proc. 7th ASIACCS. pp. 71–82 (2012)
15. Shih, F., Liccardi, I.: Privacy tipping points in smartphones privacy preferences. In: Proc. 33rd CHI. pp. 807–816 (2015)
16. Shklovski, I., Mainwaring, S.D., Skúladóttir, H.H., Borgthorsson, H.: Leakiness and creepiness in app space: Perceptions of privacy and mobile app use. In: Proc. 32nd CHI. pp. 2347–2356 (2014)
17. Tao, J., Slutzki, G., Honavar, V.: Secrecy-preserving query answering for instance checking in EL. Tech. rep., Iowa State University (2010)
18. Zhang, F.: Assessing intrusiveness of smartphone apps. Tech. rep., Masters thesis, MIT (2012)