Privacy Engineering? Why?

***Abstract*—**

I. Introduction

Use of software applications in daily lives of users is growing [1]. With the pervasiveness of technology users today disclose a huge amount of personal information, such as their address, credit card and banking details and even their health conditions, into online applications. When users disclose more data into software applications, the ways through which user privacy could be compromised through these software systems increase. For example, if a user disclose his/her location through his mobile phone to an untrusted application continuously, it becomes easy for a stranger to predict the location of that user on a given time on a given day. Therefore, the concerns for user privacy in software systems have become indespensable .

In order to enable users to protect their privacy within commonly used software systems (mobile applications, social networking platforms) it is important to understand how users make their decisions when they disclose their personal information into these software systems. When disclosing their personal information into software systems users are faced with many conditions. For example, how important is this information for this particular system? How sensitive is this particular information is to me? How visibile would this information be once I disclose them?. However, how these conditions relate to each other, and how users perceive these conditions and the impact of these conditions are not readily observable. This not only makes users uncomfortable when interacting with software systems, but also makes it difficult for privacy researchers and software developers to make data collection decisions when designing software systems.

In order to ensure user privacy preferences are met, developers need to be able to measure and quantify privacy of certain data items are for users. For such a measurement to be done, first one needs to identify the factors that affect privacy of data items for a user and the relationship among these factors. For example, would users be more comfortable sharing their credit card number with their banking application than their social networking account? Would users be equally comfortable sharing their blood group with their banking application? Would users be more comfortable sharing their age than their birthdate with their social networking account? Answering these questions are difficult yet important. A privacy risk metric that could determine the privacy of different data items in a given setting would be beneficial for users, software developers and also privacy researchers. From a user perspective, knowing how private their data are would help them to decide whether or not to disclose certain data items into software systems. From a developer's perspective they could decide which data to collect and which data to avoid when designing software systems in order to protect user privacy within the systems. For privacy researchers this information could be used when making measurements on user privacy among different software systems and provide guidance for law makers and authorities to enforce data protection regulations.

In this work, we attempt to determine a privacy measurement for data elements using three parameters, sensitivity of the data element to the user, visibility of the data element within the software system once disclosed and the relatedness of the data element to the software system. We observe how these three parameters affect data dicslosure decisions of users when they interact with software systems through a user survey. We conducted an online study with 151 participants. We incorporate the relationship among the three parameters above to determine a metric to measure the privacy associated with different data items in a given context of a software system. For example, what is the privacy risk component associated with disclosing the age of a user into a social networking application?

The contribution of this paper is determining how visibility, relatedness and sensitivity of data elements affect the discluse decisions of users when they interact with software systems. The paper is structured as follows. We first discuss prevoius work that has identified the parameters (sensitivity, visibility and relatedness) in measuring privacy in the background section. Then we describe our user study followed by the results. Next, we discuss the implicatins of our findings followed by the conclusion and future work.

References

[1] F. T. Commission *et al.*, "Fair information practice principles," *last modified June*, vol. 25, 2007.

Appendix A
Appendix A Survey Questionnaire