Secure Portable Tokens for Sensitive Questionnaires Surveys

Athanasia Katsouraki[1,2], Luc Bouganim[1,2], Benjamin Nguyen[3], Paul Tran-Van[1,4]

[1]INRIA Paris-Rocquencourt France Fname.Lname@inria.fr
[2]U. Versailles St-Quentin-En-Yvelines France Fname.Lname@uvsq.fr
[3]INSA – Centre Val de Loire France benjamin.nguyen@insa-cvl.fr
[4]Cozy Cloud France paul@cozycloud.cc

ABSTRACT

Sensitive Questionnaire Surveys are becoming popular, with a wider spectrum of researchers who conduct this kind of study. However, the more sensitive the questions are, the most skeptical the participants become, leading to either untruthful or lack of response. The arising question here is whether individuals would rather disclose their sensitive information by answering a questionnaire survey in a system that provides more tangible security than simply using an online survey web site. We describe an experimentation using both a secure and a vulnerable approach of a system dedicated to Sensitive Questionnaire Surveys. The first approach uses Secure Portable Tokens (SPT), a device that stores securely both the questionnaire survey and the participants' answers. The second approach involves a central server (MySQL server) in which participants will be connected and respond to the survey. The results will be used to carry out skills' evaluation, quantified and statistical analysis. The objective of this experimentation is to show to which extent the secure hardware is able to cover privacy deficiencies, as well as to encourage people to deliver more information, in the context of Sensitive Questionnaire Surveys.

I. MOTIVATION

Nowadays, surveys and questionnaires are frequently used by researchers in social science, economy and computer science to gather data about different human aspects, perceptions, behaviors and attitudes [1]. However, when people are asked online questions considering their income, activities and marital status, in the majority of cases, such sensitive questions cause discomfort [2, 3, 4] leading to either untruthful or lack of response. Moreover, in some cases, the participants change the previously provided information in order to either prevent an unwanted and shameful exposure, or to avoid any repercussions. This behavior is triggered by the lack of awareness considering access control management and utilization of the provided information. Another point that should be taken into consideration is the fear about potential information leakage of sensitive information that has been already provided online. In this context, multiple concerns regarding the accuracy of the collected information, as well as the results of conducted surveys, are likely to be raised.

The wide development of secure hardware devices changes the management of sensitive data. Secure Portable Tokens (SPT) [5, 6, 7] are personal servers that can combine hardware security and large quantities of NAND Flash memory storage in a portable form factor. Such devices, with adequate software, allow their owners to manage and control their sensitive data. SPTs constitute a secure repository where the stored data can be accessed upon owner's authentication and following user's access control rules. Their role is decisive since they can serve as an answer to privacy deficiencies in different sectors of everyday life such as education, transportation or healthcare systems.

Regarding the case of sensitive questionnaire surveys, the answers given could remain in the SPT. Similarly, privacy invasive computations could be done inside the secure hardware. For instance, let us consider questionnaire survey with weighted answers, e.g. a questionnaire aiming to suggest careers to students, based on intrusive questions on their likes and dislikes, attitudes at work, etc. A particular value representing the suggested career could be calculated in the SPT, using answers and weights. This score will be available to the researchers of the experiment to perform a quantified analysis, while the *precise* answers will remain private and will be only accessible by the participant.

According to the above considerations, the arising question here is whether individuals would rather disclose their sensitive information by answering a questionnaire survey in a system that provides more tangible security than simply using an online survey web site. On the one hand, it is noticeable that conducting surveys online [8, 9] is both a fast and cheap method to gather data, compared to other methods such as paper and face-to-face questionnaires surveys [10] or questionnaires that need any special equipment to be conducted [11]. However, it includes the risk of receiving no answer from participants concerning sensitive questions. On the other hand, compelling challenges emerge from the domain of secure hardware devices. Our objective is to figure out whether people can trust a hardware device (decentralized-storage module) that supports secure storage and management of personal data (SPT) as well as their willingness to deliver more information supposing that sensitive information is stored on their private device.

The motivation for the system described in this paper emerges from a consideration of the instrumental role of privacy in individuals' lives. We designed an experimentation that involves students who would like to discover their future ideal job. This process includes a questionnaire survey which is submitted to the

students. The collected personal data (including sensitive data) will be used to perform skills' evaluation and statistical analysis. With this in mind, we developed a system that can contribute to perform this experimentation. More specifically, both a secure and a vulnerable version of the same system are described in order to point out the importance of sensitive information protection. In the former case, the questionnaire surveys and participants' answers will be stored in the SPTs. Information that could be disclosed includes some scores that are calculated based on answers' weights. In the latter case, a central server will be used to keep participants' answers. In order to avoid any influence, the same user interface is used by both versions of the system.

II. PLUGDB TOKEN FOR SENSITIVE QUESTIONNAIRES

A Secure Portable Token (SPT) (Fig. 1) is a low-cost tamper-resistant hardware device that combines the following: a microcontroller that is equipped with a 32 bit RISC CPU clocked at about 120 MHz running the main code; a SIM card running the cryptographic code and keeping the secret keys and a micro-SD card storing the encrypted on-board database. The communication of SPTs with the outside world can be achieved through USB or Bluetooth communication protocols. Furthermore, SPTs are equipped with a fingerprint reader. This module allows the owners of the SPTs to access their own personal data by using their fingerprint as credentials, thus providing strong security promises for authentication

A database kernel has been developed on this platform in order to provide data/metadata storage and indexing, SQL-like query execution, users' and application's authentication, as well as access control rules' enforcement, data encryption and decryption. A JDBC bridge facilitates the process of sending SQL commands to this DBMS kernel. Thus, SPTs can be characterized as a full-fledged data server, running on any device that is equipped with USB port or Bluetooth, such as personal computers, tablets and smartphones. The architecture is called PlugDB (Fig. 1).

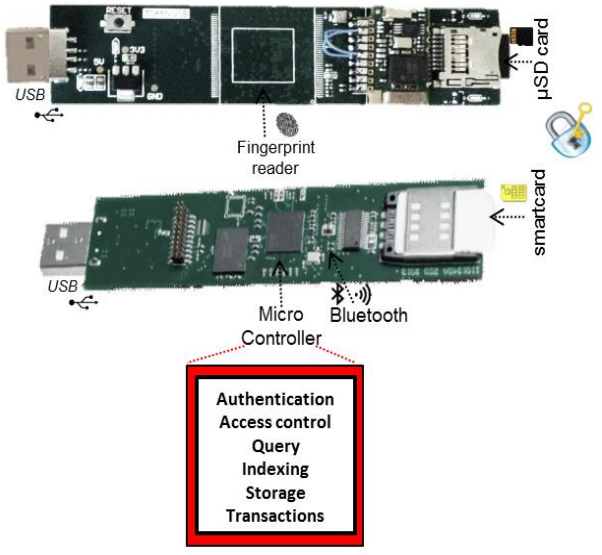

Figure 1: Secure Portable Token (PlugDB Engine)

PlugDB server (Fig. 1) is personal, self-administered, and pluggable on demand. It provides security guarantees and can be used without network connection. Concerning the questionnaire surveys, the sensitive answers remain inside the personal server, as well as the computations based on weights of each answer. These computations will allow us to perform the participants' profile analysis that could be available to the surveys' administrators. This analysis does not reveal any sensitive information beyond the suggested jobs.

PlugDB server is trustworthy compared to a central server, since the cost/benefit ratio of attacks is very high. Indeed, the attack cost is high, given the device tamper-resistance, while the benefit of the attack is reduced since it discloses data of only a single individual. Hence, PlugDB server could be the answer to sensitive questionnaire surveys.

III. EXPERIMENTAL PLATFORM AND METHODOLOGY

A. *Experimental Platform*

In this section, we describe our experimental platform. We developed both a secure and a vulnerable version of a system dedicated to questionnaire surveys. More specifically, the secure version introduces SPTs in the experimentation process while the vulnerable one involves a central server. MySQL Server was chosen as this central server in our system because it is widely used for questionnaires surveys [8, 9]. Two groups of students are asked to answer a questionnaire survey related to job-seeking, using our system.

The questionnaire evokes situations related to competence identification. These situations include, for instance, objectives related to: project launching, financial management, team spirit, confidence levels, risk perception, aggressiveness levels, creativity and self-discipline. It is a representation of the questions' categories; containing a number of questions. In order to compute the privacy and profile score and thus, to show a profile description to participants, each of the answers will be scored <*privacy score, profile score*>. Scores can be fixed at will, depending on how intrusive are the questions (privacy score) and how the answer impacts the profile (highly dependent of the questionnaire and of the chosen questionnaire methodology). Fig.2 represents a part of the questionnaire survey used in the demonstration.

-Self-discipline
Q1.In a project, I consider all the consequences of my actions.
A1.not learned skill;9;0
A2.early acquisition skills;8;1
A3.competence acquisition in progress;5;2
A4.acquired skill;3;3
-Confidence levels
Q1.I am not afraid to face the unknown situations.
A1.not learned skill;8;0
A2.early acquisition skills;7;1
A3.competence acquisition in progress;4;2
A4.acquired skill;2;3

Figure 2: Sample Questionnaire Survey

The first group is given SPTs (1 per participant) containing the survey that they answer through this secure device (secure edition) (see Fig. 3a). All the answers will remain in the SPT while scores, based on the weights of the given answers, will be disclosed. We assume that these scores do not reveal any sensitive information. The second group will respond to the survey by connecting to a central server (vulnerable edition) (see Fig. 3b). All the answers will be stored in the central server. The same graphical interface is used (see Fig. 7) for both groups and they will be reassured that their sensitive information is securely stored. The survey administrators will be responsible for the initialization of the system by providing the appropriate survey as input (see Fig. 2 and Fig. 6).

Hardware Used. In this experiment, 100 SPTs (1 SPT per student), connected to several terminal computers are used (see Fig. 3a). Moreover, several stand-alone terminal computers, with MySQL 56 server installed, are used (see Fig. 3b).

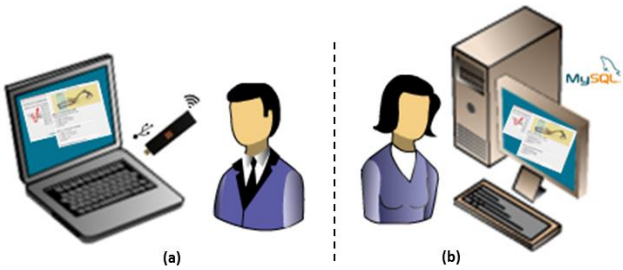

Figure 3: System Architecture

Database schema. For the questionnaire survey application, a common schema was developed in personal and central server. The schema contains the following tables: *Participant, Category, Question, Labels, and Information* (see Fig. 4).

All tables include an id, *Table_id*, which is the primary key. *Category_name* represents the name of questions' category while *Category_next* links a category to the next one. *Question_status* identifies if a question is answered or not (-1: answered, 0: not-answered). *Answer_value* keeps the answer's position, if the tuple is a question (otherwise the value is -1 by default), the *Value_type* identifies if it is a question or an answer (0: question, 1 to 4: answer) while the *Privacy_score* and the *Profile_score* keep the privacy and profile scores, accordingly, associated to an answer.

Creating a survey. From the survey administrators' point of view, the system allows managing the whole experimental procedure. The procedure includes the building of the questionnaires, the initialization of the system and result monitoring. The results include a profile description that represents participants' skills and strengths. When the participants complete the process, the final score can be computed based on profile scores of each question (the formulae being dependent of the questionnaire itself). In the demonstration we simply computed the sum of the profile scores and provided a characterization of the user's profile based on that score (see Fig. 5).

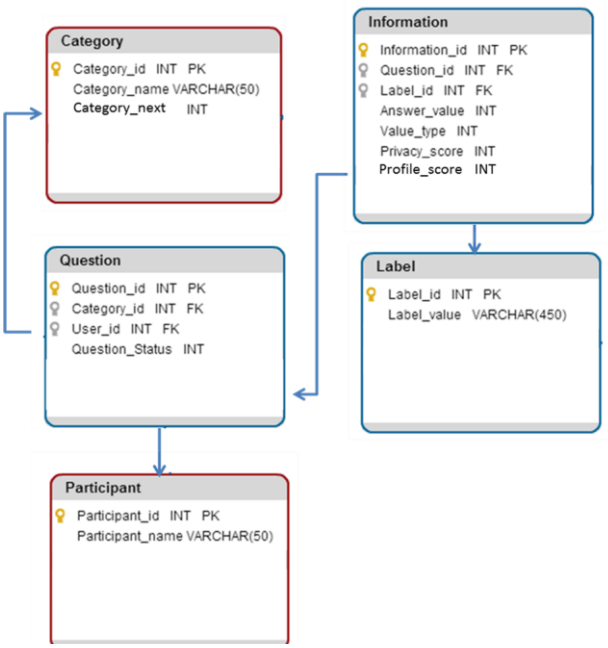

Figure 4: Database Schema

In order to simplify the process of building the questionnaire's forms, the system provides a GUI that allows them to create their questionnaire survey by using a simple text editor.

More specifically, the administrators can create the survey using a simple text editor. Fig. 2 shows a part of the questionnaire survey. The symbol '-' represents a new questions' category, the letter: 'Q' represents a new question and the letter: 'A', a new answer. In order to perform the analysis, privacy and profile scores will be calculated. For this reason, after the desired answer we can add a number followed by ';' and a second number. The first number represents the privacy score and the second one the profile score. For instance, in Fig. 2, "A1.not learned skill;9;0"; the letter: 'A' along with its sequence number signify that this is the first answer, "not learned skill" is the answer label, while the numbers 9 and 0 represent the values for privacy and profile score, accordingly. The higher the privacy score is, the most sensitive the question is. Then, the system allows administrators to transform this file into an *.csv file (see Fig. 6), which is the proper form of input file for system's initialization.

-Adventurer
Enthusiasm is everywhere no matter the situation or experienced live! You want to be totally controlling your future.
-Realist
Your project is carefully studied (cost estimates, constraints, timing and a very precise schedule)….

Figure 5: Profiles' Description

When the experiment finishes, the administrators can monitor the results (see Fig. 5 and Fig. 6). A list of participants, along with their results is available to administrators for any processing.

Answering the survey. Fig. 7 exhibits the graphical interface of the system from participants' side. More specifically, the system allows participants to create their accounts in order to login to the system and answer the questionnaire survey. The questionnaire survey is divided into categories that the participants are being asked to respond to. Having the participants answered all the available questions; they will be able to see their results, obtained after several calculations depending on the weights of the given answers.

B. Methodology

This experiment will be conducted in September 2015 by researchers in experimental economy of the University of Paris-Sud and will involve students interested in characterizing their future professional project. They will be called to participate at some training sessions, without knowing beforehand that they will participate in an experiment.

During the experimentation process, the participants are asked to answer an initial set of questions. These questions include demographic data, pieces of information related to the use of Information and Communications Technology and online services such as digital social networks. Once the participants have answered the first set of questions, a second set of questions, including much more sensitive issues, is available to them. The latter set will allow us to define and propose to them an adaptive job-search strategy. Two different groups of individuals will be created; the first one will be equipped with an SPT while the second will use a central server. The experimentation will be separated into two phases.

During the first phase, the first group is asked to answer the first set of questions using the SPT. They will register their personal data to the given SPT, maintaining their anonymity. In parallel, the second group is being asked to answer the same set of questions, but their information is going to be recorded in a central server. The most important difference between the two approaches is that in the former one the subjects will have the control over their information, while in the latter one the subjects will not be sure about where exactly their answers are going to be stored.

The first phase will lead us to build a self-exposure index, measuring the propensity of the individuals to disclose sensitive information concerning themselves and to test whether there is any statistically significant difference between the indices of the two groups.

The second phase of the experimentation process will be enhanced with several informational shocks, which is a classical protocol in experimental economy. Examples of these shocks include error screens that will appear suddenly (i.e. virus effects, list with instructions for a limited time), actors impersonating survey organizers pretending something is wrong, etc. However, not all of them will be experiencing the informational shocks. We divide each of the groups into two new groups. Table I describes all the possible cases.

		Data Storage	
		Secure Portable Token	***MySQL Server***
Informational shock	Yes	Case1	Case2
	No	Case3	Case4

Table 1: Experimentation protocol

In this phase, a new index is calculated based on sensitive data that has been collected from all the participants (those who faced an informational shock and those who did not). The objective of this phase is to compare the index differences between the cases of participants having the SPT (cases 1 & 3) and those who are using the central server (cases 2 & 4), and observe whether owning a SPT influences negatively the index difference.

At the end of the process, we will be able to assess whether the participants that have been given an SPT and had faced the shock experience remain more confident because their data is stored on a physical object that they hold in their hands.

IV. DEMONSTRATION

In the first part of the demonstration, we will create the Questionnaire Survey and show the system initialization by the surveys' administrator while the second part exhibits the participation in the Questionnaire Survey.

Figure 6: GUI for administrators

Fig. 2 presents a sample of the questionnaire survey, created by the survey's administrators. The administrators will use the graphical interface that is shown in Fig. 6 to transform this text file into the appropriate form. The database initialization for both sides, PlugDB and MySQL server, will be performed in the same manner. In the former case, the SPTs will be plugged in a computer terminal one by one, and the SPTs' database will be initialized (system module for SPT use is chosen) (see Fig 7). In the latter case, the same initialization process will be followed, but it will be made only once (system module for MySQL server use is chosen).

The second part of demonstration will focus on the answers to the questionnaire and on results exploitation. We will show the two modes, will give some examples of informational shocks and examples of final results.

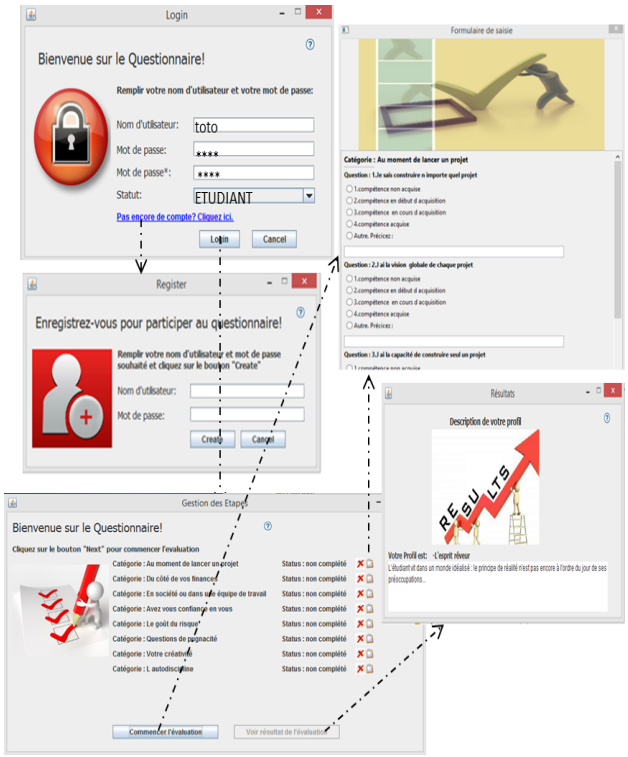

Figure 7: GUI for participants

V. Conclusion and Future Work

In this paper we proposed both secure and vulnerable approaches, of a system dedicated to Sensitive Questionnaire Surveys. We have designed an experimentation process with students in the context of proposing job-searching strategies, showing that the secure approach could potentially be fitted better to individuals. We plan to conduct the experiment in collaboration with experimental economists, on groups of students, by September 2015.

In future work, considering the Sensitive Questionnaire Surveys along with the Secure Portable Token (SPT) context, we plan to conduct the experimentation by enabling the fingerprint module of the SPT, and observing whether the individuals will react positively to this effort. We believe that the area of Surveys, containing Sensitive Questions could benefit from SPTs, while potentially leading to unique challenges coming from various application domains.

References

[1] Andrews, D., Nonnecke, B., Preece, J. Electronic survey methodology: A case study in reaching hard to involve Internet Users. International Journal of Human-Computer Interaction. 16, 2, 185-210, 2003.

[2] Roger Tourangeau, Ting Yan. Sensitive Questions in Surveys. Psychological Bulletin Vol. 133, No. 5, 859 – 883, 2007.

[3] Anthony Ong, David Weiss. The Impact of Anonymity on Responses to Sensitive Questions, Journal of Applied Social Psychology, Volume 30, Issue 8, pages 1691–1708, August 2000.

[4] Hisako Matsuo, Kevin P McIntyre, Terry Tomazic, Barry Katz. The Online Survey: Its Contributions and Potential Problems, JSM 2004, Toronto, 2004

[5] Nicolas Anciaux, Luc Bouganim, Yanli Guo, Philippe Pucheral, Jean-Jacques Vandewalle, and Shaoyi Yin. Pluggable Personal Data Servers, SIGMOD'10, June 6–10, 2010, Indianapolis, Indiana, USA.

[6] Nicolas Anciaux, Luc Bouganim, Benjamin Nguyen, Iulian S.U Popa. Trusted Cells: A Sea Change for Personal Data Services, 6th Biennial Conference on Innovative Data Systems Research (CIDR '13) January 6-9, 2013, Asilomar, California, USA.

[7] Tristan Allard, Nicolas Anciaux, Luc Bouganim, Yanli Guo, Lionel Le Folgoc, Benjamin Nguyen, Philippe Pucheral, Indrajit Ray, Indrakshi Ray, and Shaoyi Yin. Secure personal data servers: a vision paper. PVLDB, 3(1):25-35, 2010

[8] Ronnie Schaniel. Design and Implementation of an Online Questionnaire Tool, ETH Master Th., 2014.

[9] Zurina Saaya, Anusuriya Devaraju, Nuridawati Mustafa, Chew Choon Leong. The implementation of Questionnaires Design Principles via online questionnaire builder, 2007.

[10] Doyle, J. K. Face-to-face surveys. In B.S. Everitt and D. Howell, eds., The Encyclopedia of Statistics in Behavioral Science. New York: Wiley, 2005.

[11] Yehuda Dayan. Responding to sensitive questions in surveys: A comparison of results from online panels, face to face, and self-completion interviews, World Association for Public Opinion Research, Berlin, September, 2007.