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 Fname.Lname@insa-cvl.fr
[4]CozyCloud France paul@cozycloud.cc

I. MOTIVATION

Nowadays, surveys and questionnaires are frequently used by scientists and researchers such as social scientists, economists, computer scientists 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 the cases, such sensitive questions cause discomfort [2, 3, 4] leading to either untruthful or lack of respond. 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 a 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, 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 in education, transportation and healthcare systems.

Regarding the case of sensitive questionnaire surveys, the given answers could remain in the SPT, as well as specific computations could be done inside the SPT. For instance, let us consider questionnaire survey with weighted answer. A particular score will be calculated in the SPT, according to the given answers, based on these weights. This score will be available to the researchers of the experiment to perform a quantified analysis, while the given 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 online. On the one hand, it is noticeable that conducting surveys online [8, 9] is both 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 no answer from participants to sensitive questions. On the other hand, compelling challenges emerge from the domain of secure hardware devices. Our objective is to answer three questions: (a) whether a hardware device that supports secure storage and management of personal data (SPT) can have positive influence on individuals' behavior, (b) whether they trust a decentralized-storage module (i.e. personal server), when they choose to deliver their sensitive personal information, and (c) 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 that the students are being asked to answer. 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, we designed the same user interface for both sides.

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, 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; providing strong security promises for queries'

authentication. This module allows the owners of the SPTs to access their own personal data by using their fingerprint as credentials.

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 and 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, PDAs and smartphones. The architecture is called PlugDB (Fig. 1).

PlugDB server (Fig. 1) is personal, self-administered, pluggable on demand, additionally it does not need any network connection, while providing security. Concerning the sensitive questionnaire surveys, the sensitive answers will remain inside the personal server, as well as the computations based on weights of each answer will be done in it. 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.

PlugDB server (Fig. 1) 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 a single individual. Hence, PlugDB server could be the answer to Sensitive Questionnaire Surveys.

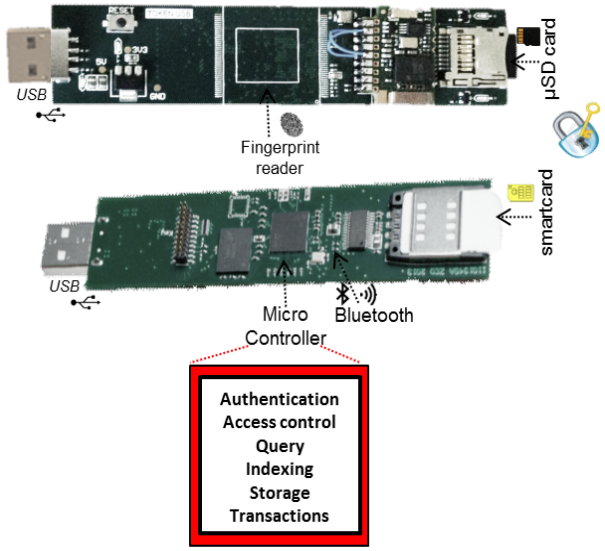

Figure 1 Secure Portable Token (PlugDB Engine)

III. EXPERIMENTAL PLATFORM AND METHODOLOGY

A. *Experimental Platform*

In this section, we are going to describe our experimental platform. We developed a secure and a vulnerable edition of a system dedicated to questionnaire surveys. More specifically, the secure edition introduces SPTs in the experimentation process while the vulnerable one involves a central server. MySQL Server was chosen as a central server in our system because it is preferred for questionnaires surveys [10, 11, 12]. Two groups of students are being asked to answer a questionnaire survey related to job-seeking, using our system. The first group will be given SPTs (1 per participant) containing the survey that they will answer through this secure device (secure edition) (see Fig. 3a). All the answers will remain in the SPT. However, authorized scores calculated based on the weights of the given answers, will be disclosed. These scores will 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. However, both groups will be reassured that their sensitive information is securely stored, as well as they will use the same graphical interface (see Fig. 6). The survey administrators will be responsible for the initialization of the system; providing the appropriate survey as input (see Fig. 2 and Fig. 5).

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).

Database schema. For the questionnaire survey purposes, 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).

```
- At the moment you start a project
Q1.I know how to build any project
A1.not learned skill;9;2
A2.early acquisition skills;8;4
A3.competence acquisition in progress;5;6
A4.acquired skill;3;8
A5.other;1;9
Q2.I have the overall vision of each project
A1.not learned skill;6;4
A2.early acquisition skills;4;6
A3.competence acquisition in progress;2;8
A4.acquired skill;1;6
A5.other;1;4
```

Figure 2 Sample Questionnaire Survey

All tables include an id: *Table*_id, which is the primary key. *Category table* includes the *Category_name* that is the name of questions' category and the *Category_next* that links a category to the next one. *Question table* includes the *Category_id* (foreign key), *Participant_id* (foreign key) and the *Question_status* that identifies if a question is answered or not (-1: answered, 0: not-answered). *Label table* includes the *Label_value* that is the label of the question or answer. *Information table* includes the *Question_id* (foreign key), the *Label_id* (foreign key), the *Answer_value* that keeps the answer's position, if the tuple is a question (otherwise the value is -1 by default), the *Value_type* that identifies if it is a question or an answer (0: question, 1 to 4: answer) and the *Privacy_score* that keeps the privacy score associated to an answer.

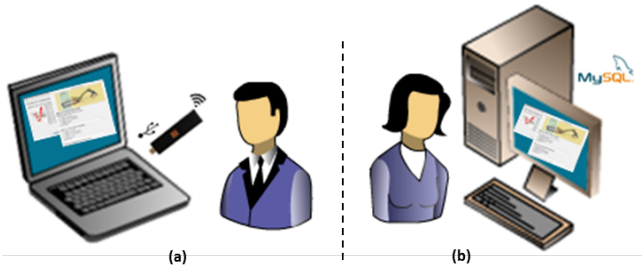

Figure 3 System Architecture

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 results monitoring. Thus, administrators are not able to access the participants' given answers, but only their results. 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.

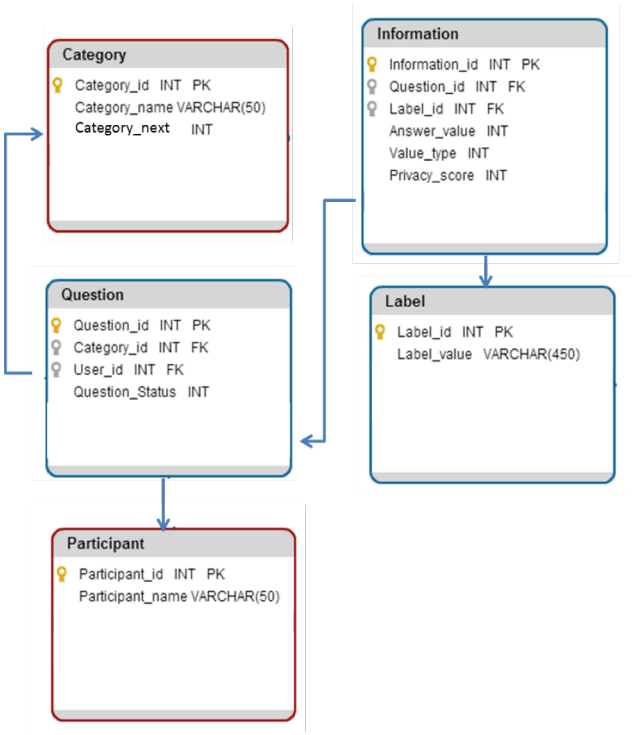

Figure 4 Database Schema

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 represents the profile score. For instance, in Fig. 2, "A1.not learned skill;9;2"; the letter: 'A' along with its sequence number signify that this is the first answer, "not learned skill" is the exact answer, while the numbers 9 and 2 represent the values for privacy and profile score, accordingly. Then, the system allows administrators to transform this file into an *.csv file (see Fig. 5), which is the proper form of input file for system's initialization.

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

Answering the survey. From the participants' point of view, the system provides information related to their potential future work, by answering a questionnaire survey. Fig. 6 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. The results are obtained after several calculations, depending on the weights of the given answers.

B. Methodology

This experiment has not taken place, yet. We plan to conduct it by September 2015. The basic principle used for this study involves individuals who fall into the case of job searching. Individuals are called to participate at some training sessions, without knowing beforehand that this process could be an experiment.

During the experimentation process the participants are being asked to answer an initial set of questions. These questions include demographic data, pieces of information related to ICT (Information and Communications Technology) use and use of online services such as digital social networks. Having the participants 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 a 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 being 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 without using the SPT but instead 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.

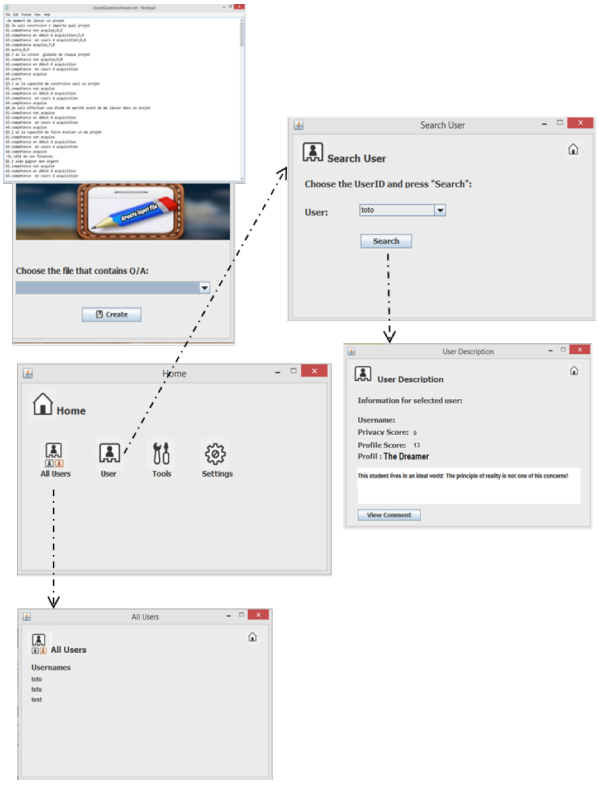

Figure 5: GUI for Administrators

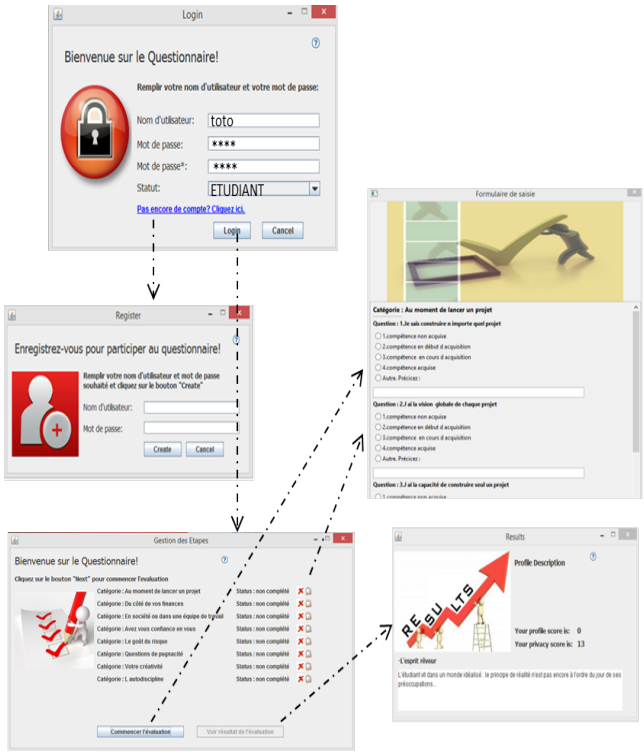

Figure 6: GUI for Participants

The second phase of the experimentation process will be enhanced with several informational shocks. An interesting example of these shocks include error screens that will be appeared suddenly (i.e. virus effects, list with instructions for a limited time), frightening the participants. However, not all the participants will be experiencing the informational shocks. We will divide each of the groups into two new groups. Table. I describes all the possible cases that have arisen.

		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 the possession of a SPT influences the index difference, negatively.

At the end of the process, we will be able to assess whether the participants that have been given a SPT and had faced the shock experience, remain more confident since they feel that 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.

Fig. 2 presents a sample of the questionnaire survey; created by the survey's administrators. This text file will be transformed into the appropriate form of our system's input file (see Fig. 5). 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 terminal computer, one by one and the SPTs' database will be initialized (system module for SPT use is chosen) (see Fig 6). 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). Having the participants answered the survey; the profile and privacy scores along with the profiles' description can only be visible by the survey's administrators (Fig. 5), while their answers remain private.

The second part of demonstration has educational placement. On the one hand, we will identify whether the SPTs' holders will be more confident when they provide their answers to the system, than the users connecting to a central server. On the other hand, the participants that are given SPTs to answer the survey; having the control of their data could be able to realize that the risk of data intrusion and its harmful use is minimized.

V. Conclusion and Future Work

In this paper we proposed two approaches (secure and vulnerable) 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, using 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, 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.

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

[5] Nicolas Anciaux, Luc Bouganim, Yanli Guo, Philippe Pucheral, Jean-Jacques Vandewalle, and Shaoyi Yin. Pluggable Personal Data Servers.

[6] Nicolas Anciaux, Luc Bouganim, Benjamin Nguyen, Iulian S.U Popa. Trusted Cells: A Sea Change for Personal Data Services.

[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, 2014.

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

[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.