UTILIZING HIPPOCRATIC DATABASE FOR PERSONAL INFORMATION PRIVACY PROTECTION

Zailani Mohamed Sidek[1], Norjihan bt Abdul Ghani[2]

[1]Centre for Advanced Software Engineering (CASE)
Faculty of Computer Science and Information Systems
Universiti Teknologi Malaysia
City Campus,Jalan Semarak,
54100 Kuala Lumpur, Malaysia

[2]Department of Information Science
Faculty of Computer Science and Information Technology
University of Malaya
50603 Kuala Lumpur, Malaysia

Email: [1]zailani@citycampus.utm.my, [2]norjihan@um.edu.my

Abstract: In today's digital world, privacy protection over personal information has become a major element in web based application. Both parties involved in a web based application transaction, either consumer or application provider should be ensured with this privacy. Protecting privacy are always related with personal information. Personal information is an information type that usually needs to keep as a private. Because of the important of privacy concerns today, we need to design a database system that suits with privacy. Agrawal et. al. has introduced Hippocratic Database. This paper will explain how HDB can be a future trend for web-based application to enhance their privacy level of trustworthiness among internet users.

Keywords: Hippocratic Database, privacy, privacy-aware, security, metadata.

1. INTRODUCTION

Web-based applications owes its dominance over traditional paper-based systems to the ability of information manipulation – to store endlessly, to sort efficiently, to locate effortlessly, and to make decision effectively [1]. More and more data have been exchanged in order to complete a task in a web-based application. By using a web-based application, people can completed any transaction types via online; no traffic jam, more convenient and of course by only fingertips. Because of that, web-based application has increased the quality of services provided by organizations.

Unfortunately, behind all this advantages of web-based application, the risks of privacy violation are increasing. Not only to keep the databases secure, but at the same time we need to consider the database privacy. It's because the more data disclosure, the more protection should be applied. When dealing with a web based application more personal information are being disclosed in order to complete the transaction. Personal information has become an important element to make sure the completion of this transaction.

Not many people realize that easy access to personal information will cause the misuse of data, no control over the information and others. Because of this, it's important to protect the information not only from external threats but also from insider threats. It is important in order to ensure the data security requirements because data disclosure when performing a task in web-based application should be ensured by data security mechanisms. A complete solution to data security must meet three requirements [2] : 1) *secrecy* or *confidentiality* refers to protection of data against unauthorized disclosure, 2) *integrity* refers to the prevention of unauthorized and improper modifications, and 3) *availability* refers to the prevention and recovery from hardware and software errors from malicious denials that make the database system unavailable. In [3], confidentiality involves sharing of information while secrecy is a type of blocking that makes the information unavailable. Ensuring the secrecy means protection of data involved in highly protected environment such as military environment. Confidentiality are always refers to type of sharing of private information to third parties.

Because of this reason, personal information privacy protection is a growing challenge for database security and privacy experts. Privacy protection is a process of finding appropriate balances between privacy and multiple competing interests [8]. We suggest that in today's era, database system should be secure and private in order to get consumer's trust towards web-based application. The new concept of database system with privacy has been introduced in [7]. Hippocratic database (HDB) has been introduced in responding to significant privacy threats that always caused by inference and multilevel security problems. Hippocratic database should be architected to regulate use and disclosure of private information in strict accordance with privacy & security laws, enterprise policies & individual choices. This means that it'll find an agreement between consumer and the organization itself towards privacy. Hippocratic database will protect individual privacy without impeding legitimate and beneficial uses of information. This paper will explain what privacy is and the privacy problems in the current database system and the introduction of HD as a privacy-aware database system.

The structure of the rest of this paper is as follows. Section 2 will give an overview of privacy followed by Section 3 will discuss on personal information and why it's important to protect

them. We define what is Hippocratic Database in Section 4 and how Hippocratic database can be viewed as privacy-aware database system in section 5. Section 6 will conclude the paper discussion.

2. OVERVIEW OF PRIVACY

Privacy, generally is a central to our dignity and our basic human rights. Privacy is the right of individuals to determine for themselves when, how, and to what extent information about themselves is communicated to others [7]. It's the ability to control collection, retention and distribution of themselves [9].

It's the right of for the data owner to determine for what purposes their personal information is stored and used. Privacy is concerned with confidentiality of the information. When user thinks that there's a need to keep any information from anybody, means that they want to keep the information as their privacy. Now, privacy becomes a major concerns when individual interacting with corporation including in a web based application. Organizations and business needs to protect an individual's privacy. In today's era, there is a must or compulsory for organizations to protect the personal information. The main reason is, to seek the user's trust towards services offer by them. They want users know that they protect their private information in a trustworthy ways. And from the user's perspective, they are realizing that it's is become a compulsory for them to make sure their private information keep as private.

A number of common privacy dimensions have been defined that have gained wide acceptance [8]. They are as follows :

1. *privacy of the person*, sometimes referred to as 'bodily privacy'. This is concerned with the integrity of the individual's body. Issues include compulsory immunisation, blood transfusion without consent, compulsory provision of samples of body fluids and body tissue, and compulsory sterilisation;

2. *privacy of personal behaviour*. This relates to all aspects of behaviour, but especially to sensitive matters, such as sexual preferences and habits, political activities and religious practices, both in private and in public places. It includes what is sometimes referred to as 'media privacy';

3. *privacy of personal communications*. Individuals claim an interest in being able to communicate among them, using various media, without routine monitoring of their communications by other persons or organizations. This includes what is sometimes referred to as 'interception privacy'; and

4. *privacy of personal data.* Individuals claim that data about themselves should not be automatically available to other individuals and organizations, and that, even where data is possessed by another party, the individual must be able to exercise a substantial degree of control over that data and its use. This is sometimes referred to as 'data privacy' and 'information privacy'

In this paper, we will concentrate the fourth dimension of privacy; privacy of personal data. As discussed earlier, confidentiality is about controlling the access to information and its release according to certain agreement, normally from owner, organizations that own the information and third parties that get the accesses. In the next section, we will briefly explain about personal information.

Privacy ensures and protects our information from unneeded disclosure. Most people only use web-based application if they feel that their information is secured and private enough. The main challenge in privacy is to share the information while complying the data owner privacy preferences. Web-based applications require techniques in privacy-preserving data management. There are three techniques in privacy-preserving data management [2] [8]. Our research will focus on the third one; database tailored to support privacy policy, such as the policies that can be expressed by using the well-known P3P. we will explain more on how HDB is well suited to enforce the personal information privacy protection in Section Four.

3. PERSONAL INFORMATION

Data is the most important element in any transaction; either off-line transaction or online transaction. Without data, a transaction can't be completed or can't be done at all. But, in the previous section, noted that privacy protection over personal information is important in today's digital world. Unfortunately, not many people realize and understand how important and valuable their information. Most of them aren't bother information that have been disclosed. They just think to complete or finish the transaction without any failure. Actually, the first and the most important thing that we should consider are; types of information available. Some personal information can be classified as sensitive and need to keep as a private information. Some of personal information are sensitive but, no need to keep it private. Besides that, certain people think in a different perspective of privacy with others. So, here information play a fundamental role in privacy domain as they shall be collected, manipulated, stored, and disclosed according to their needs.

Before we explain further more on personal information, let's examine four types of personal information involved in processing [8]:

i) *Personal information* : any data that can be used to identify a person such as name, address,

telephone number.

ii) *Sensitive information* : any data that disclose information about racial or ethnic origin, religious, philosophical or other belief, political opinion, membership of parties, as well as personal data disclosing health such as health history, race.

iii) *Identification information*: personal data that permit the direct identification of the data subject such as DNA, identity card number

iv) *Anonymous information*: any data that cannot be associated to any identified or identifiable data subject such as gender, type of disease

From the above classification, the first three types of data can be considered as sensitive information. Sensitive information is information that requires protection due to risks that could result from its disclosure, alteration, or destruction. This sensitive information should be protection to ensure the privacy. The fourth type of information is not really sensitive data, but, it can be used by query manipulation in order to retrieve sensitive information.

Personal information is defined as any information that is related to the individual person [14]. Meaning that personal information contains information that is associated to a person. It can be used to identify an individual. Personal information can be private information or non-private information. Private personal information is personal information that refers to uniquely identifiable individual of a possession of a person, meanwhile non-private personal information is personal information that doesn't refer to uniquely identifiable person.

In web-based environment, personal information is created, collected, processed before it will be disclosed by the data owner or organizations. The organization will collect, store, manipulate information to fulfill their organizations needs. Previous personal information flow model introduced by Al-Fedaghi [15] noted that there are four main phases of personal information; creating, collecting, processing and disclosing. But, as we discussed earlier in the previous section, we introduced another phase; controlling. Controlling phase is important in order to control the personal information before disclosed it. Figure 1 shows how personal information is collected, stored, used, controlled and disclosed.
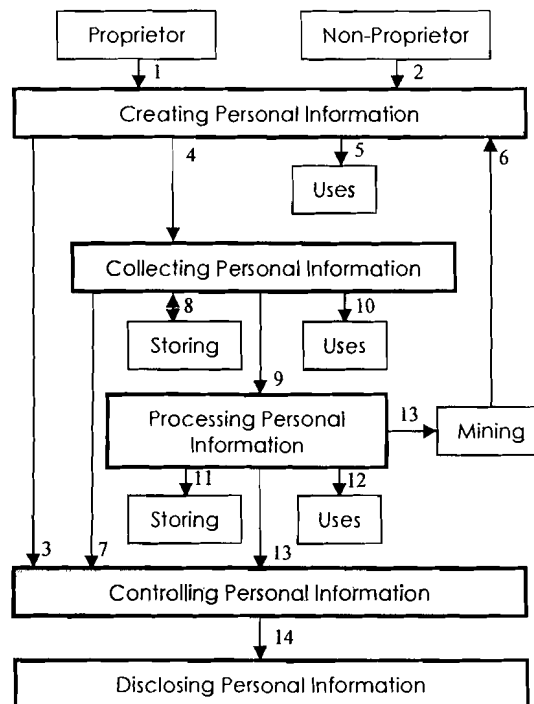
Figure 1. Personal Information Flow Model

4. HIPPOCRATIC DATABASE

Hippocratic Database is a technique dealing with database management systems (DBMS) specifically tailored to support privacy policies, like the policies that can be expressed by using the well known P3P standard [7]. The concept of Hippocratic Database (HDB) had been introduced by Agrawal et al. [7]. Hippocratic Database is the new evolution of database which incorporates privacy protection in relational database systems. The main feature of HDB is accessing the data based on purposes. In a web based application, personal information privacy protection can be achieved by applying the concept of HDB. In this section we'll elaborate on HDB and how it has been referred as a privacy-aware database.

4.1   Ten Hippocratic Database Principles

Ten guiding principles of Hippocratic Databases and initial designs to provide limited disclosure and compliance audition were introduced in [7] :

1. **Purpose Specification** For personal information stored in the database, the purposes for which the information has been collected shall be associated with that information.

2. **Consent** The purposes associated with personal information shall have consent of the donor of the personal information.

3. **Limited Collection** The personal information collected shall be limited to the minimum necessary for accomplishing the specified purposes.

4. **Limited Use** The database shall run only those queries that are consistent with the purposes for which the information has been collected.

5. **Limited Disclosure** The personal information stored in the database shall not be communicated outside the database for purposes other than those for which there is consent from the donor of the information.

6. **Limited Retention** Personal information shall be retained only as long as necessary for the fulfillment of the purposes for which it has been collected.

7. **Accuracy** Personal information stored in the database shall be accurate and up-to-date.

8. **Safety** Personal information shall be protected by security safeguards against theft and other misappropriations.

9. **Openness** A donor shall be able to access all information about the donor stored in the database.

10. **Compliance** A donor shall be able to verify compliance with the above principles. Similarly, the database shall be able to address a challenge concerning compliance.

4.2   Architecture

Main consideration of HDB is to enforce the privacy policy in database system. It has been introduced by Agrawal et al. to incorporate privacy protection in relational database system. An important feature of HDB is the use of metadata, consisting of *privacy policies* and *privacy authorizations* stored in privacy-policies tables and privacy-authorizations table respectively.

In this architecture, purpose is used as the central concepts. Purpose is a major role in access control [11]. The HDB performs privacy checking during the query processing. Every query submitted to the database with the intended purpose. Then, the system will checks either the user who issued the query is authorized to access the information or not by checking either he/she present in the list of authorized user for that purposes in the privacy-authorizations table. Then, if yes, the system will ensures that the query accessed only the attributes that are explicitly listed for query purposes in the privacy-authorizations tables. If yes, the system will ensure that only records whose purposes attribute includes the query purpose are visible to the query. The conclusion is, only attributes for certain purposes are allowed to be accessed by authorized users. Fig. 2 shows the Strawman architecture of a HDB system.
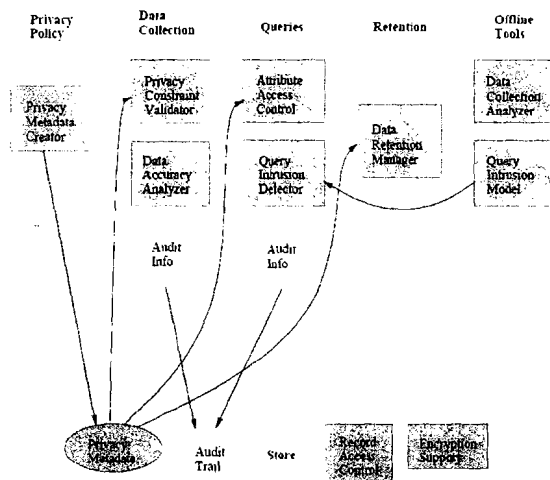
Figure 2. An Architecture of Hippocratic Database [7]

4.3 Privacy Metadata

The *privacy metadata* tables define for each purpose, and for each piece of information (attribute) collected for that purpose [7] :

- the *external-recipients*: whom the information can be given out to,
- the *retention-period*: how long the information is stored, and
- the *authorized-users*: the set of users (applications) who can access this information.

In [7], Agrawal et. al, proposed to split the above information into two separate tables, whose schemas are as shown in Table 1. From the above schema, the external-recipients and retention attributes are in the *privacy-policies* table, while the authorized-users attribute is in the *privacy-authorizations* table. Table 1(a-d) shows an example of the schema of the two tables, *user* and *driving_license*, that store the personal information.

Table 1a. Privacy Metadata Schema

table	Attributes
privacy-policies	purpose, table, attribute, {external-recipients}, retention
privacy-authorizations	purpose, table, attribute, {authorized-users}

Table 1b. Database Schema

table	attributes
user	purpose, user_id, name, address, tel_num, email, ct_card_info, license_num
driving_license	purpose, license_num, expiration date, class

Table 1c. Privacy-Policies Table

purpose	table	attribute	external-recipients	retention
license renewal	user	user_id	*all*	1 month
license renewal	user	name	{insurance-company, bank}	1 month
license renewal	user	address	{insurance-company, bank}	1 month
license renewal	user	tel_num	{bank}	1 month
license renewal	user	email	{insurance-company, bank}	1 month
license renewal	user	ct_card_info	{bank}	1 month
license renewal	user	licence_num	{insurance-company}	1 month
notification	driving_licence	license_num	{insurance-company}	1 year
notification	driving_licence	expiration_date	*empty*	1 year
notification	driving_licence	class	*empty*	10 years

Table 1d. Privacy-Authorizations Table

purpose	table	attribute	authorized-users
license renewal	user	user_id	All
license renewal	user	name	{insurance, payment, RTD officer}
license renewal	user	address	{insurance, payment, RTD officer}
license renewal	user	tel_num	{payment, RTD officer}
license renewal	user	email	{insurance, payment, RTD officer}
license renewal	user	ct_card_info	{payment}
license renewal	user	licence_num	{RTD officer
notification	driving_licence	license_num	{insurance}
notification	driving_licence	expiration_date	{RTD officer}
notification	driving_licence	class	{RTD officer}

The main concept of HDB is purpose. Users are allowed to access the information that is available for them as authorized users. During the HDB design, we'll collect for what purposes the information was collected and to whom the information can be disclosure.

Hippocratic Database as discussed in previous section is also known as privacy-aware database. This purpose concept has been introduced to implement the privacy-aware access in database system. It has been designed to maximize the privacy protection factor in database system. Whenever the system will limit the access only for intended purposes means that users who not have the authorization access to that purposes can't access the information in a database. The access to database in only permitted based on purposes. This will consider that purposes can limit the access and at the same time, privacy can be obtained. In HDB, in order to preserve the privacy of information providers, every information access must comply with the privacy policies on which information providers have agreed. A typical privacy policy for a HDB includes purpose(s), retention and authorized users. It states that the particular information can be accessed only for the specific purpose(s) on the specific condition. The retention indicates how long the information can be retained.

6. CONCLUSION AND FUTURE WORK

As the world are moving forward to a new era of web-based application, it's important to highlight the privacy. Hippocratic Database exists when the world really needs something for privacy. Hippocratic Database, based on purpose concept, identify who can access our information, which information and for what purposes. On the way, it's also depends on owner preferences. This paper discuss on the important of personal information privacy protection and how HDB can be applied as a solution for designing a privacy aware database which is suitable for a web based application. It also discussing on how accessing the information through Hippocratic Database can be considered as private information.

REFERENCES

[1]   J. H. Moor. "Towards a Theory of Privacy for the Information Age". *Computers and Society*, 27(3):27-32, 1997.

[2]   Elisa Bertino, and Ravi Sandhu, "Database Security—Concepts, Approaches, and Challenges", *IEEE Transactions on Dependable and Secure Computing*, Vol. 2, No. 1, January-March 2005, pp 2 – 19.

[3]   Al-Fedaghi, S,. "Privacy as a base for Confidentiality". Presented in the Fourth Workshop on the Economics of Information Security, Harvard University, Cambridge, MA, 2005.

[4]   Csilla Farkas, Sushil Jajodia. "The Inference Problem : A Survey". *SIGKDD Explorations*, Volume 4, issues 2, pp 6-11.

[5]   N. R. Adam & J. C. Wortman. "Security-control methods for Statistical Databases". *ACM Computing Surveys*, 21(4):515 – 556, Dec 1989.

[6]   A. Shoshani. "Statistical Databases : Characterictics, Problems and Some Solutions". In Proc. of the Eighth International Conference on Very large Databases, pages 208 -213, Mexico, September 1982.

[7]   R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu. Hippocratic databases. In *The 28th International Conference on Very Large Databases (VLDB)*, 2002.

[8]   Clarke, R. 1999. Introduction to Dataveillance and Information Privacy, and Definitions and   Terms.[Online]   Available   : http://www.anu.edu.au/people/Roger.Clarke/DV/Intro.html#Priv

[9]   Goldberg, I., Wagner, D., Brewer, E. "Privacy-Enhancing Technologies for the Internet". Proceedings of IEEE COMPCON '97, 1997, 103 – 109.

[10] Marx, G. T., 2001. "Identity and Anonimity: Some Conceptual Distinctions and Issues for Research", *In J. Caplan and J. Torpey, Documenting Individual Identity* (Princeton University Press, 2001)

[11] Ji-Won Byun, Ninghui, "Purpose Based Access Control for Privacy protection in relational Database Systems". *The VLDB Journal,* 2006.

[12] (Book Chapter) Sabrina De Capitani di Vimercati, Sarah Foresti, Pierangela Samarati, "Authorization and Access Control". *Privacy & Trust in Modern Data Management.*

[13] Silcana Castano, Mariagrazia Fugini, Giancarlo Martella, Peirangela Samaranti, "Database Security", Addison Wesley1994.

[14] Heikinen, K., Juha E., Pekka J., and Jari, P. Personalized View of personal information. *WSEAS Transactions on Information Science and Applications,* vol. 2, No. 4, 2004.

[15] Al-Fedaghi, S. Aspects of Personal Information Theory, *Proceedings of the 2006 IEEE Workshop on Information Assurance,* United States Military Academy, West Point, NY. (2006b).