Journal of Basic and Applied Engineering Research
Print ISSN: 2350-0077; Online ISSN: 2350-0255; Volume 1, Number 7; October, 2014 pp. 93-96

http://www.krishisanskriti.org/jbaer.html

A Literature Review on Database Privacy in Social Networks Using Hippocratic Databases

Sonali Ganguly[1], S.P. Singh[2]

[1]*Student, BIT, Mesra (Noida Center)*
[2]*BIT, Mesra (Noida Center)*

Abstract: **The 21st century is the age of netizens. Internet has become a source of knowledge and platform for social media. Social networking sites have gained huge momentum among all age groups and the popularity has attracted users to devote hours on social networking and contributing to a huge repository of information at every minute. Privacy of this information becomes a critical function of the database. Since personal information of the user is subject to cyber crimes therefore, database level privacy is a fundamental requirement to protect the data. Principles of Hippocratic databases are used to manage privacy of data that rests in database. This paper provides a literature review of 12 papers published in various journals and conferences and highlights the work done in database privacy with Hippocratic Databases. This paper will provide insights on various domains where principles of Hippocratic Databases have been implemented.**

1. INTRODUCTION

An online social network is a network of geographically dispersed users i.e. individuals, groups or organizations who connect and interact through messages, chats, comments, images and share relationships like friends, family, business partners over online media. A social networking website is an online community that provides a platform for online social networking to users. According to The Guardian article published online [10], Facebook has claimed to have 1.23 billion active users as in December, 2013. With the popularity of social networking websites, the numbers of social network users have increased significantly.

A social network user can make friends, search for friends, share thoughts, images and videos, use diverse applications, publish contents, and send messages and chats to each other. Thus, social networking sites collect huge amount of information including user preferences like sport teams, movies, television series, hobbies, games, and personal and sensitive information like name, address, email address, contact details, gender, date of birth, educational background, work details, profile pictures, languages known, religious views, political views etc. creating a massive repository of user data. All this information can reveal a lot about the user. Therefore, a major concern while maintaining this huge repository is privacy of user's personal information. Personal data is easily accessible without user's knowledge leading to privacy breach. Social Networking sites have enforced privacy and security features which are activated based on the settings applied by the user but lack of knowledge about privacy policies and willingness of the user to share information leads to user profiles being attacked by corporate and businesses to gather data and use it. Users may not be aware that their information is being collected or what information is being collected exactly.

Privacy is the right of individuals to determine for themselves when, how and to what extent information about them is communicated to others [1]. In the age of social networking, privacy violations have subsequently increased. Users make friends easily over online community without much acquaintance and end up disclosing their private data.

2. PRINCIPLES OF HIPPOCRATIC DATABASE

The objective is to apply the principles of Hippocratic Database on Social Networks to maintain database privacy. The principles of Hippocratic Databases originated from 'Hippocratic Oath' of medical or law profession. Every principle of Hippocratic Database has been laid down by Agrawal et al. [2002] [13]. The sample Hippocratic Oath is given below to get a better understanding of the base concept of Hippocratic principles.

"What I may see or hear in the course of the treatment or even outside of the treatment in regard to the life of men, which on no account one must spread abroad, I will keep to myself, holding such things to be unutterable"

The ten principles of Hippocratic Databases are:

1. Purpose Specification

For every information stored in the database, the purpose for which the information was collected should be associated with the information. For example, when a table is created, the purpose of creating the table and reason of collecting the table

information (i.e. attributes) must be stored. This will also ensure that the relevance of the table is known and any modification or updates are implemented keeping the purpose in mind.

2. Consent

The purpose associated with personal information shall have consent of the donor of the personal information. Consent is needed to determine what information should be disclosed to which group. For example: Donor may share personal photographs with only family and keep them invisible for friends.

3. Limited Collection

The information collected should be limited to minimum necessary for accomplishing the specified purposes. Social networking sites collect "as much data as possible". There should be difference between collection and use of data.

4. Limited Use

The database should run only those queries that are consistent with the purposes for which the information has been collected. For example: the profile view may be different for a stranger and a friend depending upon the privacy settings which will affect the query to retrieve customized information.

5. Limited Disclosure

The personal information stored in the database should not be communicated outside the database for purposes other than those for which there is consent from the donor of the information. Donor's privacy preferences must be enforced to ensure limited disclosure.

6. Limited Retention

Personal information should be retained only as long as necessary for the fulfilment of the purposes for which it has been collected. The period of retention should be reasonable. It is possible that the information is retained for an extended period of time due to statistical computation. In such cases, information for personal identification can be removed.

7. Accuracy

The personal information stored in the database should be accurate and up-to-date. Accuracy is needed to eliminate data inconsistency.

8. Safety

Personal information should be protected by security safeguards against theft and other misappropriations.

9. Openness

A donor should be able to access all information about the donor stored in the database.

10. Compliance

A donor should be able to verify compliance with the above principles. Similarly, the database should be able to address a challenge concerning compliance.

3. LITERATURE REVIEW

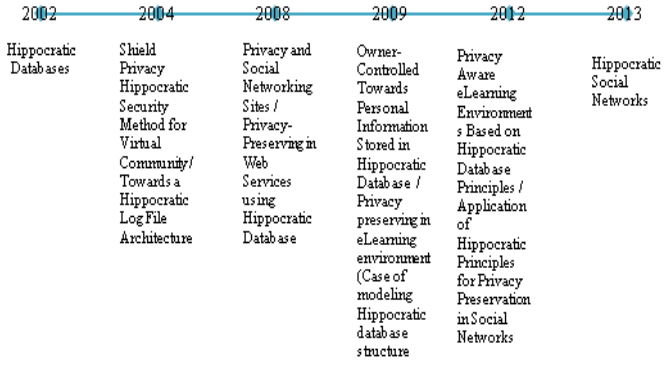

Fig. 1: Papers published during the years

Hippocratic Database Concepts: Agrawal et al [13] has laid down ten founding principles of Hippocratic databases based on the Hippocratic Oath of physicians that defines the responsibility of the database system to effectively manage and control the private information of users stored in the database. This paper recognises that all the data in the world does not live in database alone and Hippocratic databases with the central focus of privacy will be an additional inducement for privacy sensitive data to move to such storage structures. The paper has covered the fundamentals of database systems highlighting statistical and secure databases, US Privacy Regulations and Guidelines. The ten principles have been explained in the previous section. 'Purpose' is the central concept around which database level privacy is built on. The paper has presented strawman architecture of Hippocratic database emphasizing on privacy metadata tables. The principles have been explained using an example of a customer placing order for any item and how each principle plays a role in keeping the information safe or tracking of information. The paper lastly describes the challenges faced like language for P3P, efficiency issues in multilevel secure databases, limited collection, retention, disclosure, openness and safety.

Shield Privacy Hippocratic Security Method for Virtual Community[12]: This paper mainly incorporates four aims i.e. a) A better privacy-preserving method not only for data mining but also in protecting of information resources; b) Protect data at rest such as in database structures and other forms of information resources; c) Apply the Hippocratic database principle to enforceable Hippocratic security policies and procedures; and d) Implement the solution and validate it in real world situation. To incorporate these goals, the

foundation pillars are Privacy, Information Security and Authentication & Access Control. The paper proposed a conceptual framework for shield privacy in a virtual community based on four steps: a) Separation of Data where personal/private information of user and operational information are stored in separate databases; b) Recognization of user through anonymous / pseudo anonymous/ non-anonymous identity and Memorization of the identity; c) Protection of information stored in database and Restriction on the usage of this information in database though access controls and secret sharing scheme; d) Mediation and Coordination between user and the community in terms of privacy of data, accessibility. It also presented architecture and implementation strategies for a virtual community, and testing methods with a real industry strength collaborative environment for knowledge sharing and discovery.

Towards a Hippocratic Log File Architecture: The need for information logging for security and computer forensics is undeniable raising the issues of privacy concerns for the owner of the information. A Rutherford et al [11] has proposed architecture for maintaining privacy in log files through the implementation of the ten principles of Hippocratic database. A layered view of Hippocratic log file architecture has been introduced where log files are 'surrounded' by metadata. Access to log files access through metadata layer ensuring information is accessed based on the purpose defined and user consent giving higher control over private information. The paper covers the operation of query processor and explains the process of aggregation and sanitization.

Privacy and Social Networking Sites [14]: The publication discusses about the growing popularity of social networking sites among college students to connect with people and understanding how personal information may be accessed and used against individuals. The author has cited various examples showcasing that any information posted online is no longer private. Insights of ethical implications through real world examples have been cited. Basically the author highlights how social networking sites have impacted the lives of youngsters who do not understand the criticality of privacy.

Privacy Aware eLearning Environments Based on Hippocratic Database Principles [5]: The author has created a prototype model of e-learning environment that implements the principles of the HDB database. In order to prove the usability and viability of the model, a comparison of the performance of the production eLearning system with prototype model has been done.

Hippocratic Social Network [2]: The authors performed privacy literacy survey to study the privacy awareness among Facebook users. The statistics revealed that 50% users of Facebook never read Facebook privacy policy and though most of the Facebook users don't want to share their profile information, most of them never read the permission message before granting the permission. They also proposed a framework using Hippocratic principles to enhance personal level privacy in social network consisting of following four components i.e. watermarking module, OSN Interaction module, Data Perturbation Module and HSN Repository. The authors have explained the principles of Hippocratic Databases like retention, compliance, consent, sharing using table schemas.

Application of Hippocratic Principles for Privacy Preservation in Social Networks [4]: The paper highlights various types of privacy breaches in social networks like identity, attribute, link and candidate degree disclosure. Authors have categorized information as must-needed (mandatory information to join social network), least-needed (rarely needed information) and moderately-needed information (additional information) based on which two scenarios have been mentioned classifying the need for this segregation. Based on these scenarios, the authors have proposed a privacy preserving model based on Hippocratic principles specifically for Purpose, Limited Disclosure, Consent and Compliance. Authors have quoted that the mechanism proposed can be applied to legacy applications and database query construct upon customization.

Owner-Controlled Towards Personal Information Stored in Hippocratic Database [7]: The paper highlights the importance of preserving privacy and introduces strawman architecture of owner-controlled Hippocratic database that focuses on the owner of information the right to decide what information of him can be stored in the database.

Privacy-Preserving in Web Services using Hippocratic Database [9]: Web services collect customer's personal information. The authors emphasized on data anonymization whenever data is released to some third party, privacy preservation in context of data mining and use of Hippocratic database to incorporate privacy protection in relational database systems. The authors have produced quantitative results and performed analysis. Authors have implemented the principles of Hippocratic database on a section of e-learning environment that keeps student information and have proposed an efficient model for the same. The result of preserving privacy in eLearning model is a normalized relational model. The paper has also highlighted the issues in database systems i.e. privacy, security and access control.

Extending Relational Database Systems to Automatically Enforce Privacy Policies [15]: The paper has highlighted that the security offered by commercial databases is not sufficient to implement privacy compliance. The main contributions of the paper are: a) Row, Column and cell level restrictions in terms of language construct; b) semantics of combining multiple restrictions; c) Transition algorithm from P3P to the proposed construct.

SQL Privacy Model for Social Networks [6]: Authors have identified the key elements of privacy as purpose, visibility, generalization, and retention. The contribution of the paper is twofold: one is to extend SQL to preserve privacy in mandatory access control models and another is to extend the SQL security model to preserve privacy in discretionary access control models.

An approach for detecting profile cloning in online social networks [3]: The paper has emphasized on the issue of profile cloning in social networks where an attacker forges the user profile information and friend circle to deceive other social users. The authors have identified two kind of profile cloning in OSNs: profile cloning and cross site profile cloning. Authors have presented a detection process and profile similarity.

4. CONCLUSION

Though the concept of Hippocratic Database emerged in 2002 to protect the privacy of data that rests in database, the principles of Hippocratic Database have been implemented in various different platforms and data sets. Since social networking sites have a growing demand in today's society to connect with people, the principles can lock the personal information of users at database level. Therefore, social networks that accumulate huge amount of private information need a secure database design for privacy. *Hippocratic Databases* [13], *Shield Privacy Hippocratic Security Method for Virtual Community* [12], *Towards a Hippocratic Log File Architecture* [11], *and Hippocratic Social Network* [2] are the papers that allow to extent research on implementing the principles of Hippocratic databases in Social Networking environment.

REFERENCES

[1] Oberholzer Hendrik JG, Ojo Sunday O and Olugbara Oludayo O, "*A PET evaluation framework for relational databases*" SocialCom, IEEE 2013

[2] Rajneesh Kaur Bedi, V.M. Wadhai and Nitinkumar Rajendra Gove, "*Hippocratic Social Network*", Fifth International Conference on Computational Aspects of Social Networks, 2013

[3] Mohammad Reza Khayyambashi, Fatemeh Salehi Rizi, "*An approach for detecting profile cloning in online social networks*", 7th International Conference, Kish Island, Iran, IEEE, 2013

[4] Rajneesh Kaur Bedi, V.M. Wadhai and Nitinkumar Rajendra Gove, "*Application of Hippocratic Principles for Privacy Preservation in Social Networks*", World Congress on Information and Communication Technologies, IEEE, 2012

[5] Jasmin Azemović, "*Privacy Aware eLearning Environments Based on Hippocratic Database Principles*", BCI, Nova Sad, Serbia, 2012

[6] Maryam Majedi, Kambiz Ghazinour, Amir H. Chinaei and Ken Barker, "*SQL Privacy Model for Social Networks*", Advances in Social Network Analysis and Mining, IEEE 2009

[7] Norjihan Abdul Ghani and Zailani Mohamed Sidek, "*Owner-Controlled Towards Personal Information Stored in Hippocratic Database*", International Conference on Computer Technology and Development, 2009

[8] Vanja Bevanda, Jasmin Azemović and Denis Mušić, "*Privacy preserving in eLearning environment (Case of modeling Hippocratic database structure)*", Fourth Balkan Conference in Informatics, 2009

[9] Norjihan Abdul Ghani, Zailani Mohamed Sidek, "*Privacy-Preserving in Web Services using Hippocratic Database*" IEEE, 2008

[10] http://www.theguardian.com/news/datablog/2014/feb/04/facebook-in-numbers-statistics

[11] Andrew Rutherford, Reinhardt Botha and Martin Olivier, "*Towards a Hippocratic Log File Architecture*", Proceedings of SAICSIT, 2004

[12] G Skinner, E Chang, M McMahon, J Aisbett and M Miller, "*Shield Privacy Hippocratic Security Method for Virtual Community*", The 30th Annual Conference of the IEEE Industrial Electronics Society, Busan, Korea, 2004

[13] Rakesh Agrawal, Jerry Kiernan, Ramakrishnan Srikant and Yirong Xu, "*Hippocratic Databases*", Proceedings of 28th VLDB conference, Hong Kong, China, 2002

[14] Dianne M. Timm, Carolyn J. Duven, *"Privacy and Social Networking Sites*", Published online in Wiley InterScience, 2008

[15] Rakesh Agrawal, Paul Bird, Tyrone Grandison, Jerry Kiernan, Scott Logan, Walid Rjaibi, "*Extending Relational Database Systems to Automatically Enforce Privacy Policies*", Proceedings of the 21st International Conference on Data Engineering (ICDE 2005) IEEE

Ms. Sonali Ganguly is working as a Project Engineer (IT) – I in CDAC, Noida. She received her MCA degree from Guru Gobind Singh Indraprastha University and presently pursuing MTech (CS) part time from Birla Institute of Technology, Mesra (Noida campus). She has three years of work experience in Java, Databases and Quality Assurance. She has published 2 research papers in national/international conference.

Dr. S. P. Singh is working as an Associate Professor in Birla Institute of Technology, Mesra (Noida Centre). He has received his MSc, MTech and Doctorate degree and has a total 13 years of experience. His subject specialization includes DBMS, Parallel & Distributed Computing, Management Information System and System Analysis and Design. He has contributed in more than 15 papers published in various national/ international conferences and journals.