Searchable Encryption for Sensitive Data in Social Network

A Confirmation Report submitted in fulfilment of the requirements for the candidature of

Doctor of Philosophy

By

Shangqi Lai

Supervisor: Dr. Joseph Liu

Co-supervisor: Dr. Ron Steinfeld, Dr. Dongxi Liu

Faculty of Information Technology

Monash University

20/11/2017

Contents

Chapter 1

Research Problem

This report outlines the doctoral research on keeping sensitive data privacy in Online Social Network (OSN) by using 'searchable encryption' technique.

1.1 Background of Problem

OSN is highly popular for many years. Services such as Facebook, Twitter and WeChat allow individuals to create their online profiles, make cyberspace connection with friends via these platforms.

A result of the huge growth of OSN is more and more sensitive personal data is revealed while OSN users enjoy using it. In 2005, two researchers from Carnegie Mellon University conducted a study in the online behaviour of more than 4,000 students in the university who had signed up in Facebook [1]. In their study, they found that the majority of Facebook users at CMU provided astonishing amount of sensitive information in their Facebook profiles: it may contain their real names (over 90%), date of birth (87%), personal images (80%) and so on; actually, privacy preference settings are provided for OSN user to control the searchability, visibility and access privilege of their profiles, but it is sparingly used, which make them searchable and identifiable to the vast majority of anybody else in the network.

Due to the richness and variety of sensitive information disclosed in OSN, the privacy of these data is of critical importance to OSN users, as the permissive and unconcerned uses of sensitive information will put themselves at risk for many attacks from the cyberspace or even the real world: the precise personal information helps potential adversaries to construct

more deceitful fraud messages (e.g. Context-Aware Spam [2]); it also can be used to re-identify some anonymous datasets like hospital medical information by matching the common attributes [3]; additionally, OSN users may be easily stalked as they revealed their location or timetable when they shared their activities to their friends.

Although the main purposes of OSN are to help its members to communicate with others and maintain social relations in a more convenient and affordable way, OSN service providers also design diverse Social Network Services(SNS) for its user interaction virtually: it includes updating activities and location information, sharing multimedia (e.g. photo, video) and events, getting updates and comments on activities by friends, etc.. All of these services retain a huge amount of data from its users: until 2013, there were 1.11 billion monthly active users and they put 4.75 billion shared items in Facebook, these contents received over 4.5 billion 'Likes' from their friends [4].

Making these data searchable to further enhance to capabilities of OSN had become an interesting topic in recent years. In 2013, Facebook introduced their social network search engine – Graph Search [5]. This search engine aims to return the information based on the content from user's social network of friends and connections: for example, you may search "The restaurants visited by friends live in Melbourne" to get more personally relevant results. The concept of Graph Search is also introduced by other OSN service provider: LinkedIn has Job Search Engine which is based on user's location, skill as well as the information from their colleague and alumni; WeChat also can search user-specific content from Moments and Official Accounts.

In a nutshell, Graph Search-like search engines is a powerful tool from the aspect of searching: people with similar or the same attributes are likely to become friend [6] it extracts results from the crowd(i.e. friends) with

similar or the same attributes(e.g. geographical location, hobbies, etc.)

1.2 Research Questions

1.3 Significance of Research

Chapter 2

Literature Review

2.1 Privacy Preserving Scheme for OSN

There have been several schemes proposed to preserve the data privacy for data sharing in OSN: Guha et al. proposed NOYB [7] which can break personal profile to multiple atoms and mix them into the crowd, it then provides confused profile to OSN. In this system, authorised users possess a part of secret can recover the real personal information associated with specific users while unauthorised users only get a mixture of personal data from the crowd. NOYB is able to protect the privacy of user profile but it doesn't work for user relations and interactions. In comparison, Persona [8] solves above issue by applying cryptographic primitive (i.e. Attribute-based Encryption (ABE)). ABE can help to apply access control over a group of users in OSN. The ability to associate the attributes of user with a key provides a convenient way for group manager to grant different access privileges to different groups. Users within those groups use the key to access the group manager's sensitive data, so each group's access to the sensitive data is properly restricted, which guarantees the privacy of group manager. As NOYB and Persona put encrypted content into OSN, OSN is hard to provide services to the users of these external tools. Lockr [9] provides Social Attestation mechanism and Social Access Control Lists to control the access of sensitive data, hence, it can achieve fine-grained access control without data encryption. Furthermore, Social Attestation mechanism makes it easier for key revocation as it has a explicit expire date. In contrast, Persona needs to re-issue a new key to all remaining users in the group. Another benefit for using Lockr applies proof of knowledge protocol. to ensures the

non-transferability of sensitive data – third parties sites cannot abuse these sensitive data because they will not receive the actual identifier of users under the protocol.

There are lots of research works with various primitives aiming to preserve the sensitive data privacy in OSN, such as homomorphic encryption [10], peer-to-peer (P2P) overlays [11].

2.2 Searchable Encryption

Searchable Encryption (SE) aims to provide abundant search functionalities on the server side, without decrypting the data itself.

- Searchable Symmetric Encryption (SSE): Song et al. [5] proposed an efficient scheme for storing and retrieving encrypted data from remote database. The encryption and search algorithms only need O(n) (n is the size of query) to perform its work, and the scheme is indistinguishable against chosen plaintext attacks [6];

- Publickeyencryptionwithkeywordsearch(PEKS):TheschemeproposedbyBonehet al. [7] allows others to read the message (e.g. e-mail services). They further revised it [8] to hide access pattern from service providers. However, this approach?s overhead in search time is O(n2), far less efficient than SSE.

Curtmola et al. [9] extended definition of SSE because they found that the security of index and trapdoor has inherently link. To guarantee the keyword will not leak from trap- door, [9] introduced two new model, a non-adaptive and an adaptive one. The adaptive one allow query as a function of previously obtained trapdoors and search outcomes, and is considered a strong security scheme. Recently, Cash et al. [10] proposed dynamic and multi-keyword search in large database in cloud system. Cash?s work [10] encrypts a clear-text database to get encrypted database (EDB). By randomly storing data in database, the database can hide the relationship of

clear- text database. Their evaluation shows the new search method is very effective for very large dataset (10s in MySQL database with terabytes-scale and billions of record-keyword pairs). Nonetheless, it also exists leakage in [10], as it discloses many information to server, and the server may be able to restore sensitive data of user.

5. Summary the link between sensitive data privacy and Searchable Encryption

Chapter 3

Progress

3.1 Courseworks

To meet the coursework requirements, I've finished two compulsory modules as well as the compulsory events and workshops in FIT 5144 in the first year. Table 3.1 shows the detailed information about my coursework activities.

Course Code	Status
FIT5143	Exempted
FIT6021	Finished

Table 3.1: Coursework Activties and Claimed Hours

3.2 Research

References

[1] Ralph Gross and Alessandro Acquisti. Information Revelation and Privacy in Online Social Networks. In *Proceedings of the 2005 ACM workshop on Privacy in the electronic society*, pages 71–80. ACM, 2005.

[2] Garrett Brown, Travis Howe, Micheal Ihbe, Atul Prakash, and Kevin Borders. Social Networks and Context-Aware Spam. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, pages 403–412. ACM, 2008.

[3] Latanya Sweeney. k-Anonymity: A Model for Protecting Privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05):557–570, 2002.

[4] Facebook. Facebook's Growth In The Past Year. [online] `https://www.facebook.com/media/set/?set=a.10151908376636729.1073741825.20531316728&type=3theater`, 2013.

[5] Facebook. Facebook Graph Search. [online] `https://www.facebook.com/graphsearcher/`, 2013.

[6] Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of A Feather: Homophily in Social Networks. *Annual review of sociology*, 27(1):415–444, 2001.

[7] Saikat Guha, Kevin Tang, and Paul Francis. NOYB: Privacy in Online Social Networks. In *Proceedings of the first workshop on Online social networks*, pages 49–54. ACM, 2008.

[8] Randy Baden, Adam Bender, Neil Spring, Bobby Bhattacharjee, and Daniel Starin. Persona: An Online Social Network with User-Defined Privacy. In *ACM SIGCOMM Computer Communication Review*, pages 135–146. ACM, 2009.

[9] Amin Tootoonchian, Stefan Saroiu, Yashar Ganjali, and Alec Wolman. Lockr: Better Privacy for Social Networks. In *Proceedings of the 5th international conference on Emerging networking experiments and technologies*, pages 169–180. ACM, 2009.

[10] Josep Domingo-Ferrer, Alexandre Viejo, Francesc Sebé, and Úrsula González-Nicolás. Privacy homomorphisms for social networks with private relationships. *Computer Networks*, 52(15):3007–3016, 2008.

[11] Leucio Antonio Cutillo, Refik Molva, and Thorsten Strufe. Safebook: a Privacy Preserving Online Social Network Leveraging on Real-Life Trust. *IEEE Communications Magazine*, 47(12), 2009.