A Content Privacy-Preserving Protocol for Energy-Efficient Access to Commercial Online Social Networks

Yi-Hui Lin[1], Chih-Yu Wang[2], Wen-Tsuen Chen[3]
Institute of Information Science
Academia Sinica
Taipei, Taiwan 115
Email:yihui1223@gmail.com[1], tomkywang@gmail.com[2], chenwt@iis.sinica.edu.tw[3]

***Abstract*—The privacy issue of online social networks (OSNs) has been getting attention from the public, especially when data privacy has caused the disagreement between users and OSN providers. While the providers utilize users' data as a commercial usage to make profit; on the other hand, users feel their privacy has been violated by such behavior. In this paper, we propose a privacy preserving protocol for users' data sharing in OSNs, where the OSN provider cannot retrieve the users' social content while the users can efficiently add or remove a social contact and flexibly perform the data access control. Moreover, we prove that the users would allow the OSN provider to perform keyword search over the encrypted content for advertising profit, so that the OSN provider can commercialize its products without the knowledge of content.**

I. Introduction

The privacy problems of online social networks (OSNs) have received much public attention in recent years because the exposure of personal information through online social networks has affected people's life [15]. For example, a prospective employer may try to evaluate job applicants by searching their information through OSNs. Moreover, OSN providers are not trust-worthy to users because they make profit from collecting large amount of social information regardless of revealing users' personal information. Therefore, a content privacy-preserving OSN protocol requires the posted data unaccessible not only to the unauthorized users but also to the OSN providers.

The most computational efficient solution is to protect the content with symmetric encryption, such as AES because the computation time of symmetric encryption is thousand times faster than public key encryption. However, in addition to the cost of key distribution and management for content owner, communication cost is the major concern in this solution. If a content owner wants to share a photo of size 500KB with 100 friends, the owner needs to encrypt the photo respectively with each friend's key. The total transmission size increases from 500KB to 50MB. Thus, only using symmetric encryption to solve the content privacy problem in OSNs is impractical.

Baden et.al. proposed a data sharing mechanism with user-defined privacy over OSNs [2]. The work uses attribute-based encryption (ABE) to hide data from unauthorized users and the OSN providers and, in the meanwhile, to share the data to the users who satisfy the attribute defined in the encrypted data. In other words, an attribute is a token of a social group for a user, such as family or colleague, and the user assigns an attribute key to each group and then distributes the key to the group members through traditional public-key cryptography (e.g.,RSA encryption). Therefore, the users can share their data with a specified group by encrypting the data with the attribute key. However, revoking a member from a group would require re-computation and re-distribution of the attribute key of the group and re-encryption of previous messages shared with the revoked member, resulting in much energy consumption for user devices.

Jahid et. al. [12] solved the above revoking problem by adopting the approach of [13] in attribute-based encryption scheme. The concept of the solution involves the help of a proxy. With a proxy key given by the data owner, the proxy provides a part of decryption information for the social contacts. The data owner changes the proxy key and re-distributes the proxy information if a contact is revoked, so that the revoked one cannot retrieve a valid information from the proxy. Although the computation and communication of re-keying and re-encryption of previous shared data are spared, it still costs the communication of proxy information distribution and at least one extra paring computation for each decryption in the user side.

To achieve efficient revocation, Sun et. al. and Raji et.al. proposed privacy preserving schemes for OSNs with broadcast encryption schemes [17][14]. Basically, the data owner performs the encryption with the public keys owned by a certain group of social contacts. Anyone who owns one of the private key to those public keys can decrypt the message. In this way, there is no need to renew any key when revocation occurs because there is no group key in this mechanism. Still, the re-encryption of previous shared data for the data owner is still necessary.

The technique of proxy re-encryption (PRE) is suitable to construct a privacy-preserving protocol for OSNs with user efficiency. One of the reasons is that the users do not need to spend extra cryptographic computation or incur communication overhead when revoking a social contact. Tran et. al. used PRE in the proposed social networks to protect data privacy [18]. However, rather than using a generic model of PRE, the work limits to ElGamal-based proxy re-encryption as the fundamental cryptographic component. Since the first concept of proxy re-encryption proposed in 1998 [4], many researchers have developed the idea by improving its security and efficiency [1][7][8][10][11][16]. Therefore, the work of Tran et. al. cannot benefit from improvement of the PRE technique.

Our work also adopt the technique of searchable encryption [5][9] for users to perform online keyword search over encrypted data. Users only need to give service provider trapdoors of keywords to perform search, so the technique does not reveal keywords of content to service provider. There are two types of searchable encryption scheme, searchable public-key encryption scheme (SPKE) [5] and searchable symmetric encryption scheme (SSE) [9]. Two works of privacy-preserving OSNs [18] [17] has adopted SPKE for online data search. However, to achieve user efficiency, SSE is more suitable in mobile environment because SSE requires less computation cost than SPKE. Therefore, we prefer to adopt SSE in order to fulfill the properties of user efficiency, data privacy, and controlled keyword search capability in our proposed protocol.

OSN providers may be reluctant to offer privacy-preserving OSNs because losing the ability of accessing users' data could ruin their business. It would motivate the OSN providers if users allow them to acquire some information with commercial interest. For example, the OSN providers can have the advertising service if the users permit them to search the encrypted users' data for the keywords related to the advertisements. However, service provider cannot assure that if users honestly give the correct trapdoors of keywords because the trapdoors does not reveal any information of keywords. It is possible that users may cheat service provider by giving the wrong trapdoors to protect their data privacy. For drawing users to give the trapdoors of those keywords, we suggest that service provider can share the advertising interest with users so that users can benefit from doing it. Our work also theoretically prove that, under such an advertising profit sharing model, users would honestly give trapdoors of advertising related keywords.

To the best of our knowledge, the paper is the first work that not only devise a user-efficient content privacy-preserving protocol for OSNs but also propose a solution that OSN provider can still make advertising profit without knowing users' social content. To achieve user efficiency, the users do not incur any extra computation cost or communication overhead when revoking a social contact or changing the access policy of the data. Moreover, with very minimal computation cost for users' devices, the users can search their encrypted data on OSN server and honestly offer trapdoors of commercially interested keywords if the advertising profit is properly shared between service provider and users.

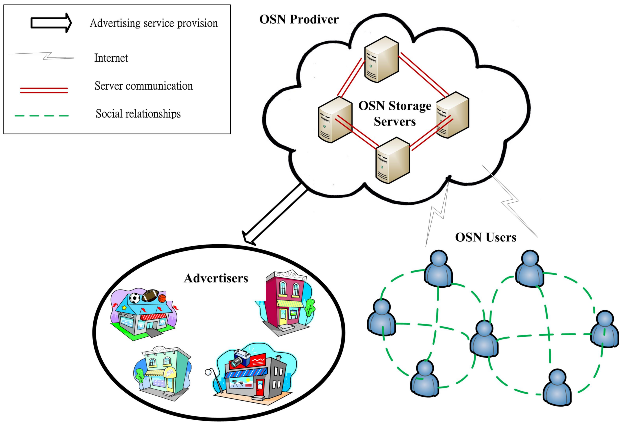


Fig. 1. The Model of Online Social Networks

The rest of our works is divided into the following sections. First, we introduce the OSN environment and the background knowledge of cryptographic components in Section II. In Section III, the proposed protocol is described. Then, user content privacy is defined and proved in Section IV. In Section V, users are theoretically proved honest in advertising profit sharing model. The security and performance is analyzed in Section VI. Finally, conclusions are given in Section V.

II. Preliminaries

In this section, we first introduce the communication and business model of OSNs. Then, two cryptographic components adopted in the protocol are depicted. One is unidirectional PRE scheme, used to protect users' content from the OSN provider. The other is SSE scheme, used to perform keyword search and enable commercialization of OSNs.

A. The Model of Online Social Networks

There are three types of participants in the OSN model, the OSN provider, the users, and the advertisers. Based on the structure of the modern social network providers, such as Facebook, Twitter, Youtube, etc., we introduce the relationships among each participant as follows.

- The relationship between the OSN provider and the users: For users to share information with contacts in social networks, the OSN provider offers the storage for them to upload, view, and give comments to shared data. Besides this benefit, it also provides the data access control service so that the data owners are able to make access policies by themselves. For example, the OSN provider restricts user Eve from viewing an article owned by Bob if he disallows her to see it.
- The relationship between the OSN provider and the advertisers: The OSN provider has the ownership of users' data, such as uploaded videos and posted messages, which are important market information for advertisers. Once advertisers request advertising service from the

OSN provider, they would search related advertising messages in users' data. By providing commercial messages for advertisers to target users, the providers gain enormous profits from the advertisers. In this paper, we suppose that OSN provider charge advertisers by pay-per-click model. That is, the advertisers pay the service provider only when the ads are clicked.

- The relationship among the users: The users build a new social relation by adding a new friend in their list or cut an old one by removing a friend from the list. By privacy setting, the users can set privacy restrictions of data and change them dynamically. For example, the users may want to allow new friends reading the posts written before adding them in the list so the user needs to change the access policies of those posts.

B. Unidirectional proxy re-encryption (PRE) scheme

An unidirectional PRE scheme is composed of five polynomial-time algorithms. We introduce them as follows.

- $KeyGen(1^k) \rightarrow (pk, sk)$: Given a security parameter 1^k, a public key pk together with an associated private key sk are generated with the key generation algorithm $KeyGen$.
- $Enc(pk, m) \rightarrow C$: Given a public key pk and a message m, a cipher C is produced with the encryption algorithm Enc.
- $Dec(sk, C) \rightarrow m$: Given a private key sk and a cipher C, a message m is obtained with the decryption algorithm DEC.
- $ReKeyGen(sk_1, pk_2) \rightarrow rk_{1\rightarrow 2}$: Given a private key sk_1 of user U_1 and a public key pk_2 of user U_2, an unidirectional re-encryption key $rk_{1\rightarrow 2}$ is generated with the re-encryption key generation algorithm $ReKeyGen$.
- $ReEnc(rk_{1\rightarrow 2}, C_1) \rightarrow C_2$: Given a re-encryption key $rk_{1\rightarrow 2}$ and a cipher C_1, a cipher C_2 is generated with the re-encryption algorithm $ReEnc$.

C. Searchable symmetric encryption (SSE) scheme

A SSE scheme is composed of four polynomial-time algorithms. We introduce them as follows.

- $SearchKeyGen(1^k) \rightarrow K$: Given a security parameter 1^k, a symmetric key K is generated with the searchable symmetric key generation algorithm $SearchKeyGen$.
- $BuildIndex(K, D) \rightarrow I$: Given a set of documents D and a symmetric key K, a secure index I is generated with the secure index building algorithm $BuildIndex$.
- $Trapdoor(K, w) \ :\rightarrow \ T_w$: Given a symmetric key K and a keyword w, a trapdoor T_w is generated with the trapdoor making algorithm $Trapdoor$.
- $Search(I, T_w) \rightarrow D(w)$: Given a secure index I of documents D and a trapdoor T_w of a keyword w, $D(w)$, a set of identifiers of documents containing the keyword w, is found with the search algorithm $Search$.

III. The proposed protocol

We will use the PRE scheme and the symmetric encryption scheme to devise a privacy preserving protocol for commercial OSNs. Moreover, the property of user efficiency is taken into consideration. The protocol composed of 1) registering in OSN server, 2) adding friend, 3) removing friend, 4) uploading and sharing data, 5) downloading data, 6) modifying data sharing, 7) making searching index, 8) keyword searching, and 9) advertising, is introduced as follows.

- **Registering in OSN provider:** User U_i chooses ID_i as the identity and computes a public key pair (pk_i, sk_i) and a symmetric key k_i by performing $KeyGen$ in the PRE scheme and $SearchKeyGen$ in the SSE scheme. After that, U_i will send $(\text{``}Register\text{''}, ID_i, pk_i)$ as the request to the OSN provider for registration. Then the provider sends a list of keywords $W' = \{w'_1, w'_2, ..., w'_n\}$ for providing advertising service. Finally, U_i will return a list of trapdoors $T_W = \{T^i_{w'_1}, T^i_{w'_2}, ..., T^i_{w'_n}\}$ by computing $Trapdoor(k_i, w'_m)$, where $1 \leq m \leq n$, if U_i finds little content privacy loss from the keywords.
- **Adding friend:** If user U_i would like to add user U_j with identity ID_j in the friend list, the following actions are executed. First, U_i computes a re-encryption key $rk_{i\rightarrow j} \leftarrow ReKeyGen(sk_i, pk_j)$, where pk_j, the public key of U_j, is obtained from the OSN provider. Then U_i sends the request $(\text{``}AddFriend\text{''}, ID_j, rk_{i\rightarrow j})$ to the OSN provider.
- **Removing friend:** If the user U_i would like to remove user U_j with identity ID_j out of the friend list, U_i sends the request $(\text{``}Removefriend\text{''}, ID_j)$ to the OSN provider to delete the re-encryption key $rk_{i\rightarrow j}$.
- **Uploading and sharing data:** If user U_i would like to share data M_t with users $U_1, U_2, ..., U_n$ in his/her friend list, the following actions are executed. First, U_i randomly selects a string $mk_t \in \{0,1\}^l$ as a key of a symmetric encryption scheme (e.g., AES). Second, U_i computes the cipher $C_t \leftarrow E(mk_t, M_t)$ and the header $H^t_i \leftarrow Enc(pk_i, mk_t)$, where E is the encryption algorithm of the symmetric encryption scheme and Enc is the encryption algorithm of the PRE scheme. Then U_i sends the request $(\text{``}UploadAndShare\text{''}, \{ID_1, ID_2, ..., ID_n\}, H^t_i, id_t, C_t)$ to the OSN provider, where $\{ID_1, ID_2, ..., ID_n\}$ is the share list of user identifiers and id_t is the identifier of M_t. After receiving the request, the OSN provider computes and stores $H^t_1 \leftarrow ReEnc(rk_{i\rightarrow 1}, H^t_i)$, $H^t_2 \leftarrow ReEnc(rk_{i\rightarrow 2}, H^t_i)$,..., $H^t_n \leftarrow ReEnc(rk_{i\rightarrow n}, H^t_i)$ for sharing data M_t with users U_1, U_2,..., U_n. Therefore, the OSN provider stores $(ID_i, H^t_i, id_t, C_t, \{ID_1, ID_2, ..., ID_n\}, \{H^t_1, H^t_2, ..., H^t_n\})$ for U_i's shared data M_t.
- **Downloading data:** If user U_i would like to view data M_s with identifier id_s, U_i will find and obtain (H^s_i, C_s) from the OSN provider. Therefore, U_i can obtain M_s by computing $Dec(sk_i, H^s_i) \rightarrow mk_s$ and $D(mk_s, C_s) \rightarrow$

M_s, where Dec is the decryption algorithm of the PRE scheme and D is the decryption algorithm of the symmetric encryption scheme.

- **Modifying data sharing:** If U_i would like to add/remove friend U_j to/from the share list of data M_t, U_i sends the request (“$modifyAdd$”, ID_j, id_t)/(“$modifyRemove$”, ID_j, id_t) to the OSN provider, where id_t is the data identifier of M_t. Then, the OSN provider stores $H^t_j \leftarrow ReEnc(rk_{i\rightarrow j}, H^t_i)$/deletes the stored H^t_j.
- **Making searching index** After collecting a certain amount of Data $M = \{M_1, M_2, ..., M_q\}$ with the corresponding identifier $Id = \{id_1, id_2, ..., id_q\}$, user U_i will compute and send $I_i \leftarrow BuildIndex(k_i, M)$ to the OSN provider, where $BuildIndex$ is the secure index building algorithm in the SSE scheme.
- **Keyword searching** If U_i would like to find which data is related to keyword w, U_i compute trapdoor $T^i_w = Trapdoor(k_i, w)$ and then send it to the OSN provider. OSN provider performs $Search(I_i, T^i_w) \rightarrow D^i(w)$ and return the result $D^i(w)$ to U_i. Since the trapdoor does not reveal the in formation of the keyword, the OSN provider does not know what he searches for.
- **Advertising** If the OSN provider would like to find which data is related to keyword w'_m, for each user U_i, it can perform $Search(I_i, T^i_{w'_m}) \rightarrow D^i(w'_m)$, where the output is a list of U_i’s data identifiers containing the keyword w'_m. (Note that U_i has given the provider the pair of $(w'_m, T^i_{w'_m})$ in the registration.) Therefore, the provider can insert the advertisement to U_i’s encrypted data with identifiers $D^i(w'_m)$.

IV. Formal Proof of Content Privacy over OSNs

In this section, we propose and explain the security definition of content privacy over OSNs. Then, based on the security of PRE, we theoretically prove that our protocol satisfies content privacy against the OSN server.

Definition 1: (Content Privacy over OSNs) Assume that an OSN protocol protect the data by an encryption algorithm E with the inputs, encrypted key κ and content m, and output $c \leftarrow E(\kappa, m)$. $\mathbb{E}$, a semi-honest OSN adversary, follows the protocol and only performs eavesdropping at most q_e times. Besides, $\mathbb{E}$ is allowed to queries the encrypted keys of the chosen encrypted messages at most q_{key} times.

After finishing information collection, $\mathbb{E}$ begin to attack the protocol by distinguishing the correct cipher of the chosen message. $\mathbb{E}$ chooses a message $\widehat{m}$ for user U_i to run the protocol and then receives two ciphertexts C_b and C_{1-b}, where $b \in_R \{0, 1\}$, $C_b = E(\widehat{\kappa}, \widehat{m})$, the message encrypted with the real key $\widehat{\kappa}$, and $C_{1-b} = E(r, \widehat{m})$, the message encrypted with a random string r. Then, $\mathbb{E}$ will decide the correct cipher by returning a bit b'. If the advantage $advantage^{\mathbb{E}}(k) = (Pr[b' = b] - 1/2)$, where k is the security parameter, is negligible, the protocol is an OSN protocol with content privacy.

Intuition. Content privacy over OSNs is all about how a user shares group keys, which use to encrypt social content, so the principals who are not in the group, including the OSN server, cannot obtain the content. Therefore, our definition of data privacy concerns the security of the group key κ. We think that the group key sharing should be independent from session to session. That is, revealing some keys of contents does not affect the security of the others. That is why, in the above definition, the adversary is given the encrypted keys expect the one he is going to attack. Moreover, contents over OSN are sometimes predictable. For example, people would usually click ”like” on facebook when giving a comment of a post. Therefore, it is easy to recognize and collect the ciphertexts of ”like”. Instead of saying that the attacker does not know a bit of plaintext from ciphertext, the definition says that the attacker cannot recognize the ciphertext from the known plaintext. The definition also guarantees secure keys. If the keys are not well-protected in the protocol, attacker can easily recognize which ciphertext is the one with the correct encrypted key through the known message and the key.

Definition 2: (CCA Security of PRE) The adversary $\mathbb{A}$ can make at most q_{rk} re-encryption key generation queries with the chosen public key, at most q_{re} re-encryption queries with the chosen ciphertexts and public key, and at most q_d decryption queries with the chosen ciphertexts. After that, $\mathbb{A}$ gives two strings (x_0, x_1) and then receives y_c, where $c \in_R \{0, 1\}$. If $c = 0$, y_c is the ciphertext of x_0. Otherwise, y_c is the ciphertext of x_1. Then, $\mathbb{A}$ can continue making queries until it returns a bit c'. If the advantage $advantage^{\mathbb{A}}(k) = (Pr[c' = c] - 1/2)$, where k is the security parameter, is negligible, The PRE scheme satisfies CCA security.

Theorem 1: Our proposed OSN protocol satisfies the definition of data privacy over OSN if the PRE scheme used in the protocol is CCA secure.

Proof: We prove it by contradiction: There are three participants in the proof, $\mathbb{E}$ is a semi-honest adversary who act the role of OSN server, $\mathbb{T}$ is a simulator who represents the users of the protocol, and $\mathbb{C}$ is a role who let attacker to challenge CCA Security of PRE. If there exists $\mathbb{E}$ who can break user’s content privacy of our protocol, $\mathbb{T}$ can use the ability of $\mathbb{E}$ to break CCA security of PRE given by $\mathbb{C}$.

For giving the same ability as OSN server, $\mathbb{T}$ allows $\mathbb{E}$ to have all users’ friend lists generated as follow: T generates the identities ID_i’s and PRE key pairs (pk_i, sk_i)’s of all users $U_1, U_2, ..., U_m$ expect U_q. Then, $\mathbb{T}$ randomly decide the friend lists and generate the corresponding re-encryption keys $rk_{i\rightarrow j}$ with sk_i, where $i \neq q$ and U_j is in U_i’s friend list. Otherwise, $\mathbb{T}$ generates the corresponding re-encryption keys $rk_{q\rightarrow j}$ for U_q by making re-encryption key generation queries to $\mathbb{C}$. Finally, $\mathbb{T}$ gives $\mathbb{E}$ all users’ friend lists along with the corresponding re-encryption keys. With holding all public keys pk_i’s, $\mathbb{T}$ can compute $H^t_i = Enc(pk_i, mk_t)$’s and $C_t = E(mk_t, M_t)$, respond (ID_i, H^t_i, id_t, C_t), and reveal mk_t when $\mathbb{E}$ performs eavesdropping for some session t of U_i and queries for the encrypted keys of the message M_t’s.

After at most q_e times eavesdropping and q_{key} times encrypted key query, $\mathbb{E}$ gives a chosen message $\widehat{M}$ for U_q. $\mathbb{T}$ simulates the response of U_q as follow: $\mathbb{T}$ gives $\mathbb{C}$ two strings, $\widehat{mk_0}$ and $\widehat{mk_1}$, and then receives $\widehat{H_q}$. Then, $\mathbb{T}$ sends $\mathbb{E}$ $(ID_q, \widehat{H_q}, \{ID_1, ID_2, ..., ID_n\}$, and $(\widehat{id}, \widehat{C_0}, \widehat{C_1})$, where $\widehat{C_0} = E(\widehat{mk_0}, \widehat{M})$ and $\widehat{C_1} = E(\widehat{mk_1}, \widehat{M})$. According to the definition of content privacy, $\mathbb{E}$ will returns a bit b'.

$\mathbb{T}$ will take the advantage of $\mathbb{E}$ by responding $\mathbb{C}$ $c' = b'$. If $\mathbb{E}$ can corrupt content privacy of U_q by distinguishing the correct ciphertext between $\widehat{C_0}$ and $\widehat{C_1}$ with non-negligible advantage ϵ, $\mathbb{T}$ can also break CCA security of PRE with non-negligible advantage ϵ. ■

V. Game-Theoretic Proof of Honest User

In this section, we will prove that a user would honestly give the service provider the right trapdoors of keywords. In the perspective of game theory, players in a game would decide to be honest if behaving honest can gain more utility than behaving dishonest. Based on this concept, the proof adopts the Nash bargaining game to model the share of the advertising profit between the service provider and a user. With the Nash bargaining solution, the result will demonstrate that the user will obtain more profit if the OSN service provider is given the correct trapdoors.

In the the following part, we briefly introduce a game and define the Nash bargaining problem and its solution.

A game consists of the following components[20].

- $\mathcal{N}$: a set of players
- $\mathcal{A}$: a set of actions that the players perform
- $\mathcal{C}$: a set of consequences from the actions
- $g : \mathcal{A} \to \mathcal{C}$: a consequence function that associates consequences with actions.
- $\mathcal{U} : \mathcal{C} \to \mathbb{R}$: a set of utility functions that defines the preferences of players on the consequences. For example, player 1 would prefer x to y if $U_1(x) > U_1(y)$.

Definition 3: The Nash bargaining problem is a tuple (S, d, U_1, U_2), where $S \subseteq \mathbb{R}^2$ represents all possible outcomes of the bargaining, $d \in S$ is a disagreement point, and (U_1, U_2) are utility functions for player 1 and 2, respectively.
The Nash bargaining solution: If $x^* \in X$ satisfies $U_1(x^*)U_2(x^*) \geq U_1(x)U_2(x)$ for any $x \in X$, x^* is the solution of the bargaining problem.

We prove that a honest user would make more profit than a dishonest one by modeling two games G_1 and G_2. G_1 is played by the service provider and a honest user while G_2 is played by the service provider and a dishonest user. Suppose that the service provider asks a user to reveal the trapdoors of the keywords $\{kw_1, kw_2, ..., kw_{100}\}$. The user can decide to be honest or not according to the solution of G_1 and G_2. Before the games start, it is necessary to evaluate the privacy payoff and the advertising payoff. The functions represent the evaluation as follows. All the costs and values below are measured in US dollars.

$P : \{0,1\}^\lambda \times T \to \mathbb{R}^+$: a privacy payoff function that evaluate the privacy value of the revealed keywords during a period of time. Such a privacy payoff function has been discovered in the works of [21][22].

$Z = CTR \times i \times CPC$: the expected advertising payoff, where CTR is the click trough rate, i is the times that the ads are shown, and CPC is the cost per click. A lot of factors can affect CTR value, such as interests of contacts. Because this work only aims at content privacy, we only discuss the keywords of content, one of the crucial factors affecting CTR value. A stronger correlation between keywords and advertisements can make a higher CTR. The work of [23] studies such a correlation in the environment of online social networks.

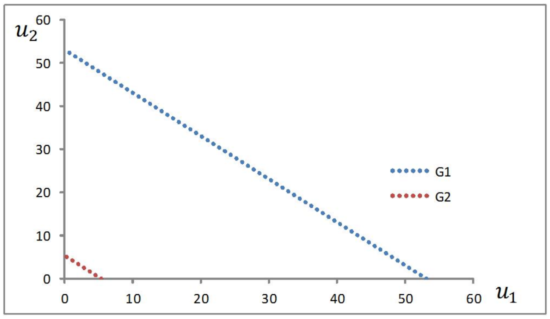


Fig. 2. The solution functions of Game 1 and Game 2

The Solution of G_1. The service provider and the honest user play the bargaining game in G_1. Suppose that the user has 100 contacts [24] and averagely posts 90 pieces of content each month [23], so the user can create approximately 100000 impresion of advertisements in a year. Given $CPC \approx 0.755$ [27] and $CTR \approx 0.08\%$ [26], the expected advertising payoff contributed from the user in a year is $z = 60.4$. Therefore, the service provider and the user shares the advertising profit z. Let's say the user gains x and the service provider gains y, where $z = x + y$. The utility of the user is $u_1 = x - P(kw_1||kw_2||...||kw_{100}, year)$ and the utility of the service provider is $u_2 = y - CPU$, where $P(kw_1||kw_2||...||kw_{100}, year) \approx 5$ is the privacy cost of revealing the mapping of keywords and trapdoors during a year and $CPU \approx 2.31$ [25] is cost per user for service provider. According to the utilities of the user and service provider, we depict the solution function of G_1 in Fig.2. Therefore, to maximize the value of u_1u_2, the Nash bargaining solution (x, y) is $(31.545, 28.855)$.

The Solution of G_2. The dishonest user plays bargaining game with OSN service provider in G_2. The expected advertising value z' is $0.01\% \times 100000 \times 0.755 = 7.55$, where we assume that $CTR = 0.01\%$ [26] if the user gives wrong mapping of keywords and trapdoors. Let's say the user gains x' and the service provider gains y', where $z' = x' + y'$. The utility of the user is $u_1 = x'$ since the user has no content privacy loss and the utility of the service provider is $u_2 = y' - CPU$. The solution function of G_2 is shown in Fig.2. Therefore, to maximize the value of u_1u_2, the Nash bargaining result (x, y) is $(2.695, 5.005)$, where the utility of the user is significantly lower than the one if he plays honest.

We theoretically prove that the profit sharing and bargaining process indeed gives the user the incentive to reveal some keywords with little privacy loss. The bargaining game shows that the user in G_1 gains more profit than G_2, so the user will choose to be honest rather than being dishonest. Therefore, the service provider is able to assure that the user is honest by sharing proper amount of advertising profit.

VI. Security and Performance Analysis

The data privacy of the proposed protocol is based on the security of PRE scheme and SSE scheme. Under the assumption of no collusion between the OSN provider and the social contacts, the content cannot be retrieved unless the content owner gives the permission to the social contacts through the re-encryption algorithm performed by the OSN servers. Based on SSE, the OSN provider retrieves only the data identities when searching the encrypted data for keywords. Therefore, content privacy is achieved in our protocol.

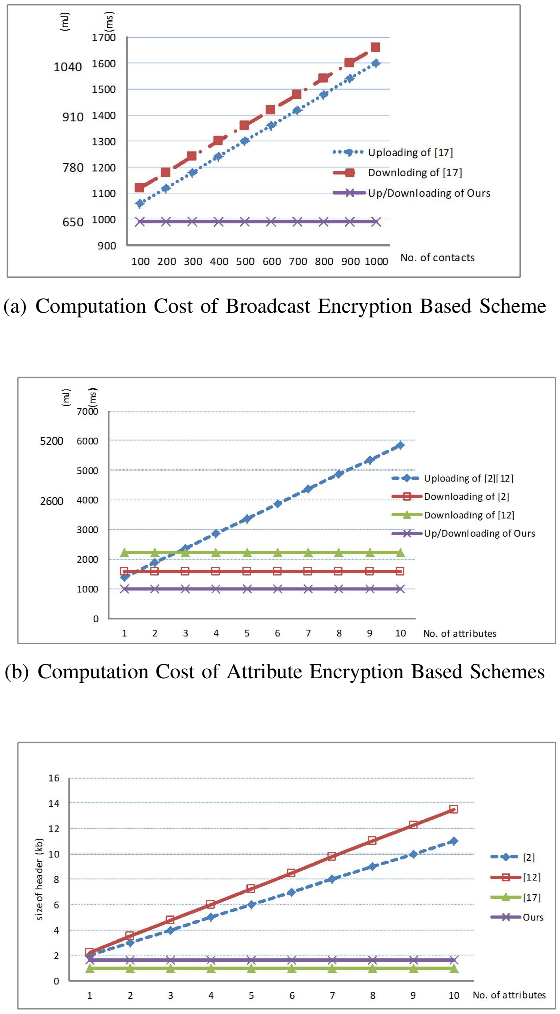
Dynamic group means that the data owners can freely share the data not to a group with fixed members, but to any subset of their social contacts. Furthermore, dynamic group also provides an efficient computation process to users for revoking a social contact without re-computation of common group key. Baden et. al. [2] does not achieve the property of dynamic group because the data owner only builds the groups with fixed members and needs to distribute a group key if formation of a new group or revocation of a member occurs.

Flexible data access control means the data owner can freely modify the access policy of the shared data without encryption again. The BE-based protocols [17][14] need to perform the broadcast encryption again when data owner would like to restrict the previous message from a revoked member or to a new member. When the data owner restricts access of a previous shared data from a revoked member, they do not need to perform attribute-based encryption again in [12]. However, they still need to re-deploy proxy information to the rest of permitted members who are still allowed to access the previous shared data. In our approach, the revoked members cannot access previous shared data online through their private keys sk_i's. However, they may store mk_t's in advance and then intercept C_t's to derive M_t's. Another similar attack case is that an authorized user can share mk_t's to the other unauthorized ones so that they can illegally obtain M_t's. Those attacks can be treated like the inevitable situation that the users directly save M_t's in advance and distribute M_t's without permission.

A trusted third party rarely exists in the real environment, so joining such a party in a protocol is a strong assumption. The work of [18] uses a trusted third party called key manager to divide a secret into two parts and distribute them to the user and to the OSN server, respectively. Using a trusted third party is an easy but impractical way to solve data privacy issue.

Using Generic cryptographic components in a protocol can guarantee that any latest improved components can be adopted in the protocol. The work of [18] looses this advantage because it uses a specific ElGamma based proxy re-encryption scheme.

Only the works of [17][18] adopted searchable encryption to allow OSN servers searching data, but both of them use searchable public-key encryption (SPKE)[5] instead of symmetric one (SSE) [9]. Compared with SSE, SPKE is much more computation consuming; on the other hand, SSE is more efficient because it only performs symmetric algorithms, such as exclusive-or operation and hash function. Therefore, SSE is adopted in our protocol for encrypted data searching.



(a) Computation Cost of Broadcast Encryption Based Scheme

(b) Computation Cost of Attribute Encryption Based Schemes

(c) Communication Cost of Headers

Fig. 3. Computation and Communication Cost of Mobile OSN Users

Figure 3 shows the computation and communication costs of mobile user devices. The experiment is implemented by JPBC library[19] on an HTC One X smart phone with 0.65 watt CPU power consumption. For the broadcast encryption of [6], there is trade-off between key size and ciphertext size. We scale down the ciphertext to constant size in the experiment to make most of communication efficiency. For the attribute-base encryption of [3], the cost of encryption and decryption depends on how a user set the access condition and how the attributes of contacts satisfy the condition. To simplify the differences, our experiment set the access condition is ”or” (e.g., A ”or” B ”or” C for attributes A, B, and C), so the contacts that satisfy only one attribute can perform the decryption. The security parameters of the experiment are set as follows. The group order $|\mathbb{G}|$ in the public-key encryption

TABLE I
PROPERTIES

	Baden et. al. [2]	Jahid et. al. [12]	Sun et.al. [17]	Raji et. al. [14]	Tran et. al. [18]	Ours
Fundamental Component	ABE	ABE	BE+SPKE	BE	PRE+SPKE	PRE+SSE
Dynamic Group		√	√	√	√	√
Flexible Data Access Control		√			√	√
No Trusted Third Party	√	√	√	√		√
Generic Cryptographic Components	√	√	√	√		√
Data Searching			√		√	√

schemes, [6], [3], [8] is 512 bits and the length of symmetric key used to encrypt social content is 128 bits.

Regarding BE based scheme of [17], The experiment result in Figure 3(a) demonstrates that computation cost linearly increases with the number of shared contacts when a user uploads or downloads social contents. Regarding ABE based schemes of [2] and [12], Figure 3(b) and 3(c) shows that computation cost of user uploading and communication cost increase with the number attributes. Our PRE based scheme increases efficiency for mobile users by remaining computation and communication cost constant.

VII. CONCLUSIONS

With proxy re-encryption scheme, we proposed a protocol of content privacy-preserving OSNs, where the users in OSNs can only share the data with the permitted social contacts but also restrict the data from the OSN provider. Moreover, under the structure of proxy re-encryption scheme, the protocol achieves dynamic group and flexible data access control which make users share their data much more efficiently. In order to motivate a commercial-oriented OSN provider to offer a privacy preserving OSN, the profitability of OSN providers is also taken into consideration. With utilizing searchable symmetric encryption in our protocol, users can efficiently allow the OSN provider to search users' encrypted data for a set of keywords, so that a commercial OSN provider can insert advertisement with the help of keyword search. Through game-theoretic proof, OSN can believe that the users honestly give the keywords with advertising value if the advertising profit is properly shared. Therefore, our work guarantees that the OSN provider can offer adverting service to advertisers without knowing the users' personal social content. Also we take into account of power efficiency of users' mobile devices so that our scheme is suitable for OSN content sharing applications in mobile environments.

REFERENCES

[1] G. Ateniese, K. Fu, M. Green, and S. Hohenberger, "Improved proxy re-encryption schemes with applications to secure distributed storage," ACM Trans. Inf. Syst. Secur., 9(1):1-30, 2006.

[2] R. Baden, A. Bender, N. Spring, B. Bhattacharjee, and D. Starin, "Persona: an online social network with user-defined privacy," ACM SIGCOMM, 2009.

[3] J. Bethencourt, A. Sahai, B. Waters, "Ciphertext-policy attribute-based encryption," In IEEE Security and Privacy, 2007.

[4] M. Blaze, G. Bleumer, and M. Strauss, "Divertible protocols and atomic proxy cryptography," In EUROCRYPT. Springer-Verlag, 1998.

[5] D. Boneh, G. D. Crescenzo, R. Ostrovsky, G. Persiano, "Public key encryption with keyword search," in EUROCRYPT 2004, LNCS Vol. 3027, Springer, 2004.

[6] D. Boneh, C. Gentry, B. Waters, "Collusion resistant broadast encryption with short ciphrertexts and private keys", Advance in Cryptology: CRYPTO05, 2005, pp. 258-275, 2005.

[7] R. Canetti and S. Hohenberger, "Chosen-ciphertext secure proxy reencryption," Cryptology ePrint, 2007.

[8] S. Chow, J. Weng, Y. Yang, and R. Deng, "Efficient unidirectional proxy re-encryption," AFRICACRYPT 2010, LNCS Vol. 6055, pp. 316-332, 2010.

[9] R. Curtmola, J. Garay, S. Kamara, R. Ostrovsky, "Searchable symmetric encryption: improved defnitions and efficient constructions" In: Proc. of CCS-2006, pp. 79-88, 2006.

[10] R. H. Deng, J. Weng, S. Liu, and K. Chen, "Chosen-ciphertext secure proxy re-encryption without pairings," Cryptology ePrint, 2008.

[11] A. Ivan and Y. Dodis, "Proxy cryptography revisited," In Network and Distributed System Security Symposium, 2003.

[12] S. Jahid, P. Mittal, and N. Borisov, "Easier: Encryption-based access control in social networks with efficient revocation," in ASIACCS, Hong Kong, 2011.

[13] M. Naor, B. Pinkas, "Efficient trace and revoke schemes," Financial Cryptography, 2001.

[14] F. Raji, A. Miri, M.D. Jazi, B. Malek, "Online social network with flexible and dynamic privacy policies," 15th CSI International Symposium on ComputerScience and Software Engineering (CSSE 2011), 2011.

[15] D. Rosenblum, "What anyone can know: The privacy risks of social networking sites," IEEE Security and Privacy, Vol. 5, No. 3, pp. 40-49, 2007.

[16] J. Shao and Z. Cao, "CCA-secure proxy re-encryption without pairings," Cryptology ePrint, 2009.

[17] J. Sun, X. Zhu, Y. Fang, "A privacy-preserving scheme for online social networks with efficient revocation,"IEEE INFOCOM, 2010.

[18] D. H. Tran, H. L. Nguyen, W. Zha, W. K. Ng, "Towards security in sharing data on cloud-based social networks," Information, Communications and Signal Processing (ICICS), 2011.

[19] Java Pairing Based Cryptography http://gas.dia.unisa.it/projects/jpbc/

[20] M. J. Osborne and A. Rubinstein. "A Course in Game Theory," MIT Press, 1994.

[21] S. Preibush, "Implementing Privacy Negotiation Techniques in ECommerce," Proceeding of the seventh IEEE International Conference on E-Commerce Technology, (CEC05), 2005

[22] A. Yassine and S. Shirmohammadi, "Measuring Users' Privacy Payoff Using Intelligent Agents," IEEE Computational Intelligence for Measurement Systems and Applications, 2009.

[23] Chi Wang, Rajat Raina, David Fong, Ding Zhou, Jiawei Han, Greg Badros, "Learning Relevance from Heterogeneous Social Network and its Application in Online Targeting," Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval, Pages 655-664, 2011.

[24] statista,number of contacts on social networks, http://www.statista.com/statistics/166554/number-of-contacts-on-social-networks-by-country/

[25] technology creek, facebook costs per monthly active user, http://www.technologycreek.com/facebook-costs-per-mau/

[26] Places to Play, facebook CTR and CPC, http://yoheinakajima.com/2013/05/13/stats-about-facebook-ads-click-thru-rates-ctr-cost-per-click-cpc/

[27] statista, facebook CPC, http://www.statista.com/statistics/226059/cost-per-click-on-facebook-in-selected-countries/