S-Match: An Efficient Privacy-preServing Profile Matching Scheme

Xiaojing Liao, Selcuk Uluagac, Raheem A. Beyah
CAP Lab, School of Electrical and Computer Engineering, Georgia Institute of Technology
xliao@gatech.edu, {selcuk,rbeyah}@ece.gatech.edu

***Abstract*—Profile matching is a fundamental and significant step for mobile social services to build social relationships and share interests. Given the privacy and efficiency concerns of mobile platforms, we propose a cost-effective profile matching technique called S-Match for mobile social services in which matching operations are achieved in a privacy-preserving manner utilizing property-preserving encryption (PPE). Specifically, in this poster, we first analyze the challenges of directly using PPE for profile matching. Second, we introduce a solution based on entropy increase. Our initial results, with three real-world datasets, show that S-Match achieves at least an order of magnitude improvement over other relevant schemes.**

I. Introduction

Mobile social services utilize profile matching to help users find friends with similar social attributes (e.g., interests, location, background). Due to privacy concerns, most techniques proposed in the literature use homomorphic encryption to enable servers to perform matching of attributes without knowing the details of the attributes. However, the computational complexity of the homomorphic encryption process is unacceptable for resource-constrained mobile platforms. In addition to the computational complexity of the schemes that rely on homomorphic encryption, confirming that the untrusted server performed the profile matching process correctly is a challenging process. Thus, the resulting matched profiles from the supposedly *untrusted* server are *trusted*.

In this ongoing work, we propose a novel technique for privacy-preServing profile Matching based on PPE [1], S-Match, which significantly reduces the computation overhead compared to that of homomorphic encryption. The contributions of our proposed work are as follows: First, we highlight the challenges of using PPE for profile matching by analyzing the entropy and the landmark nodes of three real-world social profile datasets. Second, we propose an efficient technique to increase the entropy so that PPE can be used in privacy-preserving profile matching. Our experiments, based on three real-world datasets, illustrate that our scheme achieves at least one order of magnitude better computational performance than the techniques that use homomorphic encryption.

II. Challenges of PPE scheme

In PPE, ciphertexts leak the property information associated with the plaintexts, which make them vulnerable when the number of the plaintexts is limited. As shown in Figure 1, in order-preserving encryption (an example of a PPE [2]), an attacker can learn the order of the plaintext from the ciphertext, making a chosen-ciphertext attack significantly easier.

The information leakage problem associated with the PPE

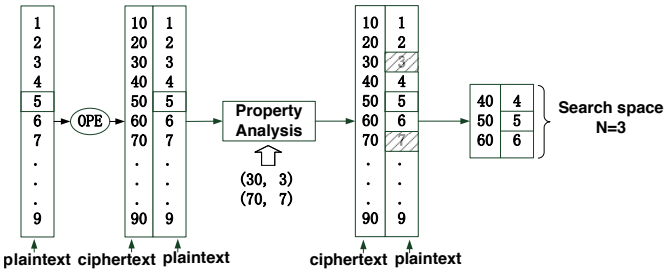


Fig. 1. A simple illustration of information leakage of OPE schemes, where (30,3) and (70,7) are the known ciphertext-plaintext pairs. An untrusted server tries to obtain the ciphertext of 5 (plaintext) analyzing the order of ciphertexts.

TABLE I. The properties of datasets

Dataset	# of Node	# of Attr.	Entropy			Landmark Attribute	
			AVG	MAX	MIN	$\tau = 0.6$	$\tau = 0.8$
Infocom06	78	6	3.10	5.34	0.82	2	1
Sigcomm09	76	6	3.40	5.62	0.86	3	1
Weibo	1 million	17	5.14	9.21	0.54	5	3

is exacerbated when used on social network data. To show this, we analyzed three real-word datasets (Infocom06 [3], Sigcomm09 [4] and Weibo [5]) and presented our observations in Table I. First, as seen in the table, these datasets have small number of attribute values. Moreover, landmark attributes, which is an attribute with value whose probability larger than threshold τ [6], are prevalent in these datasets. The existence of landmark attributes undermines the anonymization of the data [6]. Therefore, the user profile data used in a privacy-preserving profile matching scheme should have high entropy to enlarge the domain of values for attributes and decrease landmark attributes.

III. S-Match: Privacy-preserving Profile Matching based on PPE

With the observation in the previous section, in this section, we propose a technique to increase the entropy of the profile data to enable private profile matching based on PPE. The idea behind our scheme is to construct a one-to-N mapping and make the probability of new attribute value uniform. However, the string size of the attribute values increases further after the mapping procedure. This leads to high communication cost as well as computation cost. To overcome this problem, we propose an attribute chaining technique. In this technique, we chain and randomize the order of the attribute data while guaranteeing the accuracy of the profile matching accurate. The overview of S-match is shown in Figure 2, which describes a four-step profile matching process.

First, we introduce the technique to increase the entropy.

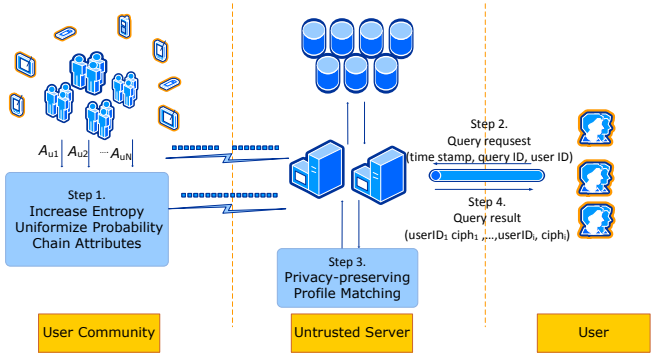


Fig. 2. Overview of S-Match. Step 1: users bootstrap the scheme by increasing the entropy so that the privacy-preserving profile matching is processed on the high-entropy attribute chains encrypted by the PPE scheme. Step 2: the query user requests a profile matching result from the untrusted server. Step 3: the untrusted server conducts the efficient privacy-preserving profile matching process. Step 4: the untrusted server returns the matching result.

A k-bit-binary string is used to represent the attribute value, where $k > 2 \cdot max\{n_i : i \in \{1, 2 \dots d\}\}$ and d is the number of attributes. For a discrete attribute (e.g., gender) $\mathcal{A}_i$ with n_i possible values $a_j^{(i)}$ and the corresponding probability $p_j^{(i)}$, where $j \in \{1, 2 \dots n_i\}$, each attribute value is mapped to $p_j^{(i)}\Delta$ binary strings with equal Hamming weight, where Δ is a constant value and $n_i < \Delta \leq k$ is used to guarantee the one-to-N mapping. And, users with attribute value j choose any one of the $p_j^{(i)}\Delta$ binary strings with equal probability $\frac{1}{p_j^{(i)}\Delta}$ as their mapping attribute value. Hence, after mapping, each attribute value is chosen with equal probability of $\frac{1}{\Delta}$. For continuous attributes (e.g., interest), we apply a two-phase mapping method: probability flatten and entropy increase. For the first phase, we use the distribution flatten technique to make the distribution uniform [2]. Then, we introduce a linear one-to-N mapping to increase the entropy of the new uniform distribution, where the coefficient of the linear mapping is $2^{K-h(X)}$, where K is the new entropy and $h(X)$ is original entropy of the attribute value.

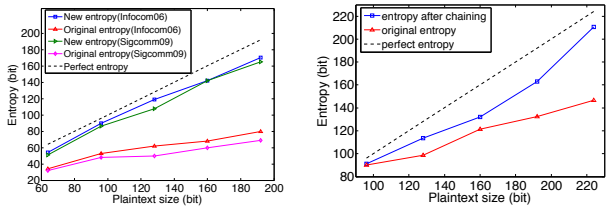
Finally, discrete and continuous attributes are chained (i.e., combined) separately in random order to decrease the overhead, and the the attribute value chains are encrypted using OPE on the mobile device before being sent to the server. Accordingly, the server is able to measure the distance between the stored encrypted profiles and the encrypted profile of the interested user with the following equation:

$$d(u, v) = \frac{O_d(u, v)}{N_d} + \frac{O_c(u, v)}{N_c} \quad (1)$$

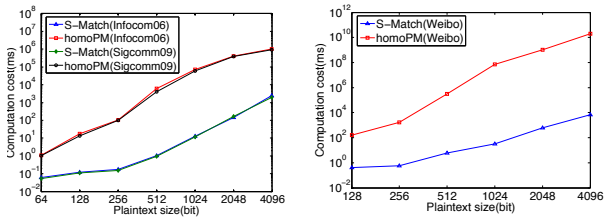
where $O_d(u, v)$ and $O_c(u, v)$ are the order distance of discrete and continuous attributes between user u and user v, N_d and N_c are the number of discrete and continuous attributes.

IV. Initial Evaluation

We evaluated the computation cost of our schemes on a machine with two 3.10 GHz Intel Core i5-2400 processors running the Linux 3.5 kernel. The users and server are evaluated on the same machine. The OPE, homomorphic encryption, AES-256 and SHA-256 algorithms were implemented in C++ based on open source code from CryptDB. Figure 3(a) and 3(b) shows the entropy of Infocom06, Sigcomm09, and Weibo datasets after the entropy increase and chaining procedures compared with that of the original entropy and perfect entropy. Overall, the entropy of the original data increases as the plaintext size k increases. Also, we observe that the rate of the entropy increase becomes higher with the growth of the plaintext size k. Therefore, the attribute data processed after entropy increase is more suitable for PPE schemes. Figure $3(b)$ indicates the entropy of Weibo dataset after the entropy increase and chaining procedures. Similar to the Infocom06 and Sigcomm09 datasets, the entropy increase becomes larger with the growth of the plaintext size k. Figure 3(c) and 3(d) present the computation cost of our schemes (i.e., S-Match) in comparison with that based on homomorphic encryption (i.e., homoPM [7]) under Infocom06, Sigcomm09, and Weibo datasets. Our scheme achieves at least one order of magnitude better computational performance than the techniques that use homomorphic encryption. As for the security analysis in our work, we assume two threat models: *honest-but-curious server* and *honest-but-curious user*. Our scheme is secure against the first threat via the increased entropy and the PPE. S-Match is secure against the second thereat because the users do not communicate with each other directly and the communication between the user and the server is protected by a secure communication tunnel.



(a) Entropy of the Infocom06 and Sigcomm09 dataset (b) Entropy of the Weibo dataset



(c) Computation cost under Sigcomm09 and Infocom06 (d) Computation cost under Weibo dataset

Fig. 3. Entropies of the three datasets after applying our technique vs. original data are indicated in Fig.3(a) and 3(b). The computation cost under three datasets are shown in Fig. 3(c) and 3(d)

References

[1] O. Pandey, Y. Rouselakis. Property preserving symmetric encryption In Advances in Cryptology-EUROCRYPT 2012, 375-391.

[2] R. Agrawal, J. Kiernan, R. Srikant, Y. Xu. Order preserving encryption for numeric data. In Proceedings of the 2004 ACM SIGMOD, 563-574.

[3] Infocom06 Dataset. http://crawdad.cs.dartmouth.edu. [2009/05/29].

[4] Sigcomm09 Dataset. http://crawdad.cs.dartmouth.edu. [2012/07/15].

[5] WEIBO API. http://open.weibo.com/wiki/. [2013/07/10].

[6] M. Srivatsa, M. Hicks. Deanonymizing mobility traces: using social network as a side-channel. In Proceedings of the 2012 ACM CCS, 628-637.

[7] R. Zhang, Y. Zhang, J. Sun, and G. Yan. Fine-grained private matching for proximity-based mobile social networking. In Proceedings of the 2012 IEEE INFOCOM, 1969-1977.