Are You Really My Friend? Exactly Spatiotemporal Matching Scheme in Privacy-Aware Mobile Social Networks

Ben Niu[1](✉), Xiuguang Li[1,2], Xiaoyan Zhu[1], Xiaoqing Li[1], and Hui Li[1]

[1] State Key Laboratory of Integrated Services Networks,
Xidian University, Xi'an, China
xd.niuben@gmail.com, lixiuguang00@126.com,
{xyzhu,xqli,lihui}@mail.xidian.edu.cn

[2] Key Laboratory of Information and Network Security,
Engineering University of Chinese Armed Police Force, Langfang, China

Abstract. We propose an exactly spatiotemporal matching scheme for privacy-aware users in MSNs. Based on the carefully designed spatiotemporal profile, our scheme employs a weight-aware pre-matching module to filter out the users with less similarity and some potential adversaries, thus guarantees that no useful information is revealed before determining the best matches. Further, we propose a privacy-preserving exchanging module against Honest-But-Curious users. Finally, the similarity computing module computes the exact matching result to each candidate to determine the best match. Thorough security analysis and evaluation results indicate the effectiveness and efficiency.

Keywords: Mobile social networks · Private matching · Spatiotemporal

1 Introduction

The success of Mobile Social Networks (MSNs) and location-aware mobile devices has resulted in many popular applications. With these applications, mobile users can either communicate with existing friends or make new social interactions, to share news or funny things. In these applications, people always need to release their personal attributes to others, which conflicts with the increasing privacy concerns of mobile users and may lead to serious privacy disclosure.

Early solutions always rely on the trusted third parity [9] to process the matching work between users. However, they may become the single point of failure in the whole system. Although many follow-up works avoid this problem by employing Private Set Intersection [2,4], Secure Multi-party Computation (SMC) [5] or Paillier Cryptosystem [10], they pay much attention to computing the common attributes/interests privately but ignore the heavy computation cost. Sun *et al.* [8] pointed out the importance of the spatiotemporal information in plenty of social activities, thus proposed the first privacy-preserving spatiotemporal matching scheme to provide more opportunities for mobile users to

J. Tian et al. (Eds.): SecureComm 2014, Part II, LNICST 153, pp. 33–40, 2015.
DOI: 10.1007/978-3-319-23802-9_5

make new social interactions in MSNs. However, neither of them takes a fully consideration on user's priority on each attribute. Niu *et al.* proposed a series of schemes [6,7] to achieve better matching by employing the similarity-based solutions. Unfortunately, all these schemes need to assume the adversary cannot obtain the processing data from the running protocol.

In this paper, we first present some preliminaries in Sect. 2. Then, based on the carefully constructed *spatiotemporal profile*, our scheme in Sect. 3 uses a weight-aware pre-matching module, which combines a random permutation function with commutation encryption function, to compute a coarse-similarity with other users and filter out the users with smaller similarity as well as some potentially malicious adversaries. We design a reorganized profile exchanging module to guarantee that both the *initiator* and the *responder* in our scheme can only obtain the cells in common as well as the corresponding frequencies. Finally, based on the exchanged *reorganized profile*, our similarity computing module finds out the best match by computing the similarity exactly. We show the security analysis and evaluation results in Sects. 4 and 5, respectively. Finally, the conclusions is drawn in Sect. 6.

2 Preliminaries

2.1 Problem Statement

Our scheme aims to achieve privacy-preserving friend discovery based on spatiotemporal profile private matching, which involves several users and relies on no TTP. Mobile users in our scheme periodically record their own locations (i.e., every 5 minutes), each of which is then assigned into a geographic cell to construct the *spatiotemporal profile* within a predefined grid once the spatiotemporal matching is needed. Let's use a users *Alice* for example, she holds a set of messages $\{\langle timeperiod, c_i, freq_i\rangle\}$, where *timeperiod* represents the time period that is considered in the matching scheme, c_i denotes the index of the geographic cell within the predefined grid and $freq_i$ means the total number of times that a user visits c_i. Since users usually stay in different places with different time period, e.g., *Alice* may stay at home and office with more time but less time in a certain shopping mall, then the $freq_i$ may changes with c_i. Our problem is how to find the best match who has more common cells and with similar visiting frequencies, while guaranteeing that no useful personal information is released.

2.2 Adversary Model

In our work, we mainly consider the Honest-But-Curious model (HBC) as some existing work [5,10], which happens within legitimate users, they will infer private information from running protocol but honestly follow the protocol.

2.3 Motivation and Our Basic Idea

Spatiotemporal matching has been one of the most popular technique employed in current social activities [8] such as friend discovering, social collaboration, etc.

Our work is thus motivated by a set of observations. (1): most of existing schemes [5,7,10] use user's attributes information to process matching, but ignore the spatiotemporal information, which is important and has been thoroughly studied in [8]. (2): although Sun *et al.* [8] proposed the privacy-preserving spatiotemporal matching scheme for MSNs, their schemes fail to consider an important social fact that mobile users in our real life always spend more time or higher frequencies on some particular locations such as their homes. As the result, cells with different visiting frequencies will be considered as the same, which does not make sense in reality. (3): schemes in [1,6–8] fail to protect legitimate user's information before identifying the identifier of the other party in the matching phase. Therefore, we argue that, if a potential adversary exists, he may add extra attributes as much as possible to match with legitimate user.

Our main idea is to perform spatiotemporal matching, exactly and privately. We design a weight-aware pre-matching module before computing the exactly spatiotemporal matching results between users. Based on our module, the candidates with less similarity and the potential adversaries can be filtered out effectively. Followed, with the help of our similarity computing module, which considers frequencies assigned on all the cells within the *spatiotemporal profile*, the best match can be selected from the others, exactly.

3 Our Proposed Scheme

3.1 System Architecture

Our scheme is a distributed solution, which allows users to process spatiotemporal matching freely without relying on any third party. Users in our scheme communicate with others in vicinity through some short-range communication techniques such as Bluetooth or WiFi. At the beginning, each user reorganizes the own *spatiotemporal profile* mentioned in Sect. 2.1 into a new *reorganized profile*, which is used for further matching. We set several priority levels for each user, and let users reorganize their *spatiotemporal profile*s based on $freq_i$. For example, we set three priority levels, which are denoted as $Level_1$, $Level_2$ and $Level_3$. $Level_1$ contains the cells that the user visited with high frequencies, $Level_2$ includes the cells with less frequencies, while $Level_3$ means the cells that the user seldom appeared. Based on *Alice*'s willingness, she can assign all the elements in her *spatiotemporal profile* into these three levels. Through this way, all the records can be assigned, we further compute the sum of the frequencies ($F^{A_j} = \sum_{r=1}^{m_j} freq_r^{A_j}$) for each priority level. Finally, *Alice*'s *reorganized profile* RP_A can be constructed. We then present our weight-aware pre-matching, reorganized profile exchanging and similarity computing modules in turn.

3.2 Weight-Aware Pre-matching Module

We suppose that all the entities share a prime q and a hash function $h(\cdot)$ through secure communication channels, which is a wildly used assumption in existing solutions. For any $c_r^{A_j} \in C^{A_j}$, *Alice* computes $(h(c_r^{A_j}))^{k_A}$ offline and sends to *Bob* together with F^{A_j}. *Bob* first transforms the elements in priority level in his *reorganized profile* $(c_t^{B_j})$ into $(\hat{c}_1^{B_j}, \cdots, \hat{c}_t^{B_j}, \cdots, \hat{c}_{n_j}^{B_j})$ offline based on a random permutation function $\prod$, which aims to disorder the elements within the array randomly. Then, he accomplishes some computation work based on the received information as follows. He computes $((h(c_r^{A_j}))^{k_A})^{k_B}$ and performs the random permutation function on them to obtain $\prod(((h(c_1^{A_j}))^{k_A})^{k_B}, \cdots, ((h(c_r^{A_j}))^{k_A})^{k_B}, \cdots, ((h(c_{m_j}^{A_j}))^{k_A})^{k_B})$, namely $(((h(\hat{c}_1^{A_j}))^{k_A})^{k_B}, \cdots, ((h(\hat{c}_r^{A_j}))^{k_A})^{k_B}, \cdots, ((h(\hat{c}_{m_j}^{A_j}))^{k_A})^{k_B})$. Next *Bob* generates a random number r_N, and then computes the hash values to obtain $(h(((h(\hat{c}_1^{A_j}))^{k_A})^{k_B} + r_N), \cdots, h(((h(\hat{c}_r^{A_j}))^{k_A})^{k_B} + r_N), \cdots, h(((h(\hat{c}_{m_j}^{A_j}))^{k_A})^{k_B} + r_N))$. Finally, this message is sent to *Alice* together with $(h(\hat{c}_t^{B_j}))^{k_B}$ and F^{B_j}. *Alice* then computes $((h(\hat{c}_t^{B_j}))^{k_B})^{k_A}$ and sends it back to *Bob*. *Bob* sends the random number r_N to *Alice*. *Alice* computes $(h(((h(\hat{c}_1^{B_j}))^{k_B})^{k_A} + r_N), \cdots, h(((h(\hat{c}_t^{B_j}))^{k_B})^{k_A} + r_N), \cdots, h(((h(\hat{c}_{n_j}^{B_j}))^{k_B})^{k_A} + r_N))$. Next, *Alice* computes the number of common cells in each priority level by

$$\begin{aligned} k_j = |C^{A_j} \cap C^{B_j}| = |\{&h(((h(\hat{c}_1^{A_j}))^{k_A})^{k_B} + r_N), \cdots, h(((h(\hat{c}_{m_j}^{A_j}))^{k_A})^{k_B} + r_N)\} \\ &\cap \{h(((h(\hat{c}_1^{B_j}))^{k_B})^{k_A} + r_N), \cdots, h(((h(\hat{c}_{n_j}^{B_j}))^{k_B})^{k_A} + r_N)\}|, \end{aligned} \tag{1}$$

and *Bob* executes the same procedure by $k_j = |C^{B_j} \bigcap C^{A_j}|$. Then, we define a weight function on each priority level between *Alice* and *Bob*, it is based on a common sense that the more frequency spent on a particular region in common, the higher similarity will be. Our weight function w_j thus can be computed by

$$w_j = \frac{F^{A_j} + F^{B_j}}{F^{A_1} + F^{A_2} + F^{A_3} + F^{B_1} + F^{B_2} + F^{B_3}}. \tag{2}$$

We compute coarse-similarity of each priority level in pre-matching phase by

$$\begin{aligned} P_j(C^{A_j}, C^{B_j}) &= \frac{|(C^{A_j} \cap C^{B_j})|}{|(C^{A_j} \cup C^{B_j})|} = \frac{|(C^{A_j} \cap C^{B_j})|}{|(C^{A_j}| + |C^{B_j})| - |(C^{A_j} \cap C^{B_j})|} \\ &= \frac{k_j}{m_j + n_j - k_j}. \end{aligned} \tag{3}$$

Based on these formulae, both *Alice* and *Bob* can obtain the weight and coarse-similarity information for each priority level. Finally, the total coarse-similarity can be computed as

$$\begin{aligned} P(C^A, C^B) &= \sum_{j=1}^{3} \{w_j \times P_j(C^{A_j}, C^{B_j})\} \\ &= \sum_{j=1}^{3} \{\frac{F^{A_j} + F^{B_j}}{F^{A_1} + F^{A_2} + F^{A_3} + F^{B_1} + F^{B_2} + F^{B_3}} \times \frac{k_j}{m_j + n_j - k_j}\}, \end{aligned} \tag{4}$$

which indicates a probable similarity between user *Alice* and *Bob*.

3.3 Reorganized Profile Exchanging Module

Since adversaries and users with less similarity can be filtered out in the pre-matching module, our goal here is to launch a privacy-preserving exchanging to exchange the *reorganized profile* with users with high similarity value. Specifically, for each element in *Alice*'s *reorganized profile*, she computes $\langle h(c_r^A)^{k_A}, freq_r^A\rangle$ and sends them to *Bob*. *Bob* computes $\langle h(c_t^B)^{k_B}, freq_t^B\rangle$, and $\langle((h(c_r^A))^{k_A})^{k_B}, freq_r^A\rangle$, then sends these messages back to *Alice*. Once *Alice* receives these messages, she computes $\langle((h(c_t^B))^{k_B})^{k_A}, freq_t^B\rangle$ and replies the computation results to *Bob*. This step is accomplished on both sides of *Alice* and *Bob*. Specifically, *Alice* compares each $((h(c_r^A))^{k_A})^{k_B}$ with $((h(c_t^B))^{k_B})^{k_A}$, if they are equal, the value of $min(freq_r^A, freq_t^B)$ is written into $freq_k^C$, and continues if they are not equal. At last, *Alice* outputs the result of $\sum freq_k^C$. While on *Bob* side, he executes the same procedure to obtain the value of $\sum freq_k^C$.

3.4 Similarity Computing Module

Based on the obtained information from the aforementioned modules, user *Alice* and *Bob* can compute the exactly spatiotemporal profile matching results with each other by the following formula.

$$\begin{aligned} P(C^A, C^B) = \frac{|(C^A \cap C^B)|}{|(C^A \cup C^B)|} &= \frac{|(C^A \cap C^B)|}{|(C^A| + |C^B)| - |(C^A \cap C^B)|} \\ &= \frac{\sum freq_k^C}{\sum freq_r^A + \sum freq_t^B - \sum freq_k^C}. \end{aligned} \tag{5}$$

The obtained $P(C^A, C^B)$ is the exactly matching result between them. According to this number, *Alice* and *Bob* could decide whether to be friends.

4 Security Analysis

Since potential adversaries can be filtered out in the weight-aware pre-matching module, then in the reorganized profile exchanging module, we prove that our scheme is secure under the HBC model.

Theorem 1. *Our scheme is secure if the commutative encryption function is secure.*

Proof. The commutative encryption function and keyed hash function provide end users with a secure channel, it means that only the one who has the secret key can decrypt the message. Suppose user *Alice* is a HBC user, she may illegally construct her reorganized profile in two extreme ways, (1) adding all the possible cells into the first priority level to get more information of *Bob*, (2) putting limited number of cells into each priority level.

For case (1), based on the pre-matching result, users who add all the possible cells into the first priority level will cause smaller matching result and will be filtered out in the weight-aware pre-matching module. Therefore, users with less similarity value cannot be executing the reorganized profile exchanging module with legitimate user. For a special case, if the HBC user has enough abilities to modify the reorganized profile, the Equation in 5 can also outputs a lower similarity value to user to make the decision.

For case (2), this kind of threat happens when *Alice* inputs cells as less as possible, for instance, she just inputs one cell with a higher corresponding frequency to perform our matching algorithm, if she fortunately has one intersection with *Bob*, she could infer that *Bob* goes to this cell (it always refers to a area such as a bar) frequently. But if she input two or more cells and the corresponding frequencies, even she has a intersection with *Bob*, she cannot conclude which cell that *Bob* has been to. So we can see that if the HBC user inputs only one cell with the corresponding frequency, he can learn other's secret information. However, this is a very special case and there are many practical ways to tackle this problem, such as setting a rule to limit the minimum number of input, which meets our real life.

Since all the data are transmitted between entities, and some cryptographic tools such as Public Key Infrastructures (PKI) can be easily adopted onto our scheme, the common cells and the corresponding frequencies can only be seen by the legitimate users with proper keys.

5 Performance Evaluations

5.1 Evaluation Setup

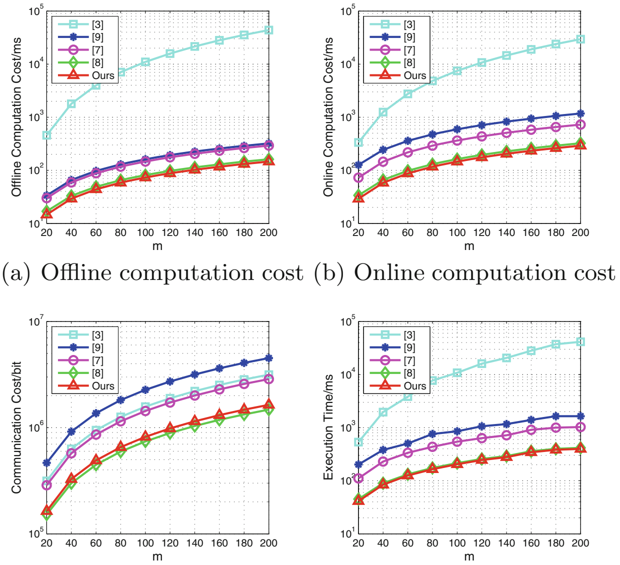
To further study the feasibility of our scheme, we implement aforementioned schemes on a Thinkpad laptop (the cryptography library is Crypto++) with 1.82 GHz CPU, 4 GB RAM to simulate the performance.

5.2 Evaluation Results

In Fig. 1, we first test the offline computation cost when m is changing from 20 to 200. Figure 1 indicates that our scheme has better performance than others [3,7–9]. The computation cost in [3] is high since there are too many exp_1s employed in their schemes.

Figure 1 compares the online computation cost of all the protocols in the log 10 scale for varying m. We can see the efficiency of our scheme over others. For users in our scheme, they need to perform $2m$ exp_1s and m hs on the *initiator* side and same compute cost on the *responder* side. The online cost of the protocols in [3] are much higher since they utilize several exp_1s in their processes.

In Fig. 1, scheme in [8] shows a better performance on the communication cost than our scheme since we just considering two users in the experiment. Now considering a real situation that there are s users in the vicinity around the *initiator*, the scheme in [8] need to transmits $(6m + 4) \times s$ bits in all, on the contrary, our scheme may just need to transmits $4m \times s + 4m$ bits since our scheme filter out the adversaries and users with less similarity.



(a) Offline computation cost (b) Online computation cost

(c) Communication cost (d) The execution time

Fig. 1. Impact of the number of common cells m

Figure 1 provides the total execution time of all the algorithms. Comparing with [3,7,9], our scheme performs better. When we look into our protocol, to get the common cells securely, an *initiator* needs more time to complete the computation. However, it is obvious that our proposed protocol can be finished within about 600 ms in all simulated sceneries.

6 Conclusions

This paper proposed an exactly spatiotemporal matching scheme for privacy-aware users in Mobile Social Networks. Based on the newly constructed spatiotemporal profile, we designed a weight-aware pre-matching module in malicious environment to effectively filter out users with less similarity and the malicious adversaries before determining the best matches. Followed, with executing our reorganized profile exchanging module and the similarity computing module, the best match can be determined exactly against Honest-But-Curious users. Security analysis and evaluation results are also provided.

Acknowledgement. This work was supported by National NSFC under Grant 61003300, Fundamental Research Funds for the Central Universities under Grant K5051201041, and China 111 Project under Grant B08038. The work of Dr. Hui Li was supported by the National Project 2012ZX03002003-002, 863 Project 2012AA013102, IRT1078 and NSFC 61170251.

References

1. Agrawal, R., Evfimievski, A., Srikant, R.: Information sharing across private databases. In: Proceedings of ACM SIGMOD (2003)
2. De Cristofaro, E., Tsudik, G.: Practical private set intersection protocols with linear complexity. In: Sion, R. (ed.) FC 2010. LNCS, vol. 6052, pp. 143–159. Springer, Heidelberg (2010)
3. De Cristofaro, E., Kim, J., Tsudik, G.: Linear-Complexity Private Set Intersection Protocols Secure in Malicious Model. In: Abe, M. (ed.) ASIACRYPT 2010. LNCS, vol. 6477, pp. 213–231. Springer, Heidelberg (2010)
4. Freedman, M.J., Nissim, K., Pinkas, B.: Efficient Private Matching and Set Intersection. In: Cachin, C., Camenisch, J.L. (eds.) EUROCRYPT 2004. LNCS, vol. 3027, pp. 1–19. Springer, Heidelberg (2004)
5. Li, M., Cao, N., Yu, S., Lou, W.: Findu: Privacy-preserving personal profile matching in mobile social networks. In: Proceedings of IEEE INFOCOM (2011)
6. Niu, B., Zhu, X., Liu, J., Li, Z., Li, H.: Weight-aware private matching scheme for proximity-based mobile social networks. In: Proceedings of IEEE GLOBECOM (2013)
7. Niu, B., Zhu, X., Zhang, T., Chi, H., Li, H.: P-match: Priority-aware friend discovery for proximity-based mobile social networks. In: Proceedings of IEEE MASS (2013)
8. Sun, J., Zhang, R., Zhang, Y.: Privacy-preserving spatiotemporal matching. In: Proceedings of IEEE INFOCOM (2013)
9. Wang, Y., ting Zhang, T., zong Li, H., ping He, L., Peng, J.: Efficient privacy preserving matchmaking for mobile social networking against malicious users. In: Proceedings of IEEE TRUSTCOM (2012)
10. Zhang, R., Zhang, Y., Sun, J., Yan, G.: Fine-grained private matching for proximity-based mobile social networking. In: Proceedings of IEEE INFOCOM (2012)