Annet Sahila G *et al*, International Journal of Computer Science and Mobile Applications,
Vol.2 Issue. 2, February- 2014, pg. 106-110 **ISSN: 2321-8363**

Privacy Preserving and Fully Anonymous Protocols for Profile Matching in Mobile Social Networks

Annet Sahila G
M.E. II year, CSE Dept
Government College of Engineering
Tirunelveli, Tamilnadu
annetraj@gmail.com

Dr. P.Latha
HOD, CSE Dept
Government College of Engineering
Tirunelveli, Tamilnadu
platha@gcetly.ac.in

***Abstract*—Social networking makes digital communication technologies sharpening tools for extending the social circle of people. Privacy preservation is a significant research issue in social networking.Here user profile matching with privacy-preservation in mobile social networks (MSNs) is studied and a family of profile matching protocols is introduced. An explicit Comparison-based Profile Matching protocol (eCPM) which runs between two parties, an initiator and a responder is proposed which enables the initiator to obtain the comparison-based matching result about a specified attribute in their profiles, while preventing their attribute values from disclosure.An implicit Comparison-based Profile Matching protocol (iCPM) is then proposed which allows the initiator to directly obtain some messages instead of the comparison result from the responder. The messages unrelated to user profile can be divided into multiple categories by the responder. The initiator implicitly chooses the interested category which is unknown to the responder. Two messages in each category are prepared by the responder, and only one message can be obtained by the initiator according to the comparison result on a single attribute. iCPM is further generalized into an implicit Predicate-based Profile Matching protocol (iPPM) which allows complex comparison criteria spanning multiple attributes.**

Keywords*— *Mobile social network, profile matching, privacy preservation, homomorphic encryption, oblivious transfer

I. INTRODUCTION

Mobile Social networking is where individuals with similar interests connect with each other through their mobile/tablet. They form virtual communities. For example Facebook, Twitter, LinkedIn etc[1]. What makes social network sites unique is not that they allow individuals to meet strangers, but rather that they enable users to articulate and make visible their social networks. On many of the large SNSs, participants are not necessarily "networking" or looking to meet new people; instead, they are primarily communicating with people who are already a part of their extended social network. To emphasize this articulated social network as a critical organizing feature of these sites, we label them "social network sites." some web-based SNSs support limited mobile interactions (e.g., Facebook, MySpace, and Cyworld). Mobile Social Networks is a means of transmitting information (communicating) using a Mixture of voice and data devices over networks including cellular technology and elements of private and public IP infrastructure (such as the Internet). 'Mobile Social Networking' (MSN) refers to all of the enabling elements necessary for the contribution ('posting' and uploading) and consumption (viewing/experiencing) of social media across a mobile network. Key to the definition is the user's implicit or explicit choice of network technologies. If the user accesses a community service platform by way of any device that uses a cellular network, alone or in combination with a commercially-accessible wireless network that has access to cellular network operator-owned resources[2]. Furthermore, mobile community operators and participants are, and can be, influenced by the platforms, trends and members of communities on the Internet.

II. PROFILE MATCHING

Profile matching means two users comparing their personal profiles and is often the first step towards effective PMSN [3]. It, however, conflicts with users' growing privacy concerns about disclosing their personal profiles to complete strangers before deciding to interact with them [4].

III. PRIMITIVES

A. Privacy Preservation

The privacy is "the right to be let alone" and it is the right to keep the disclosure of personal information safe from others" [5]. Privacy implications associated with online social

networking depend on the level of identifiability of the information provided, it's possible recipients, and its possible uses. It is relatively easy for anyone to gain access to it. By joining the network, hacking the site, or impersonating a user by stealing his password. Stalking to identity theft. Personal data are generously provided and limiting privacy preferences are sparingly used [6].

B. Homomorphic Encryption

There are several existing homomorphic encryption schemes that support different operations such as addition and multiplication on ciphertexts. By using these schemes, a user is able to process the encrypted plaintext without knowing the secret keys [7]. Due to this property, homomorphic encryption schemes are widely used in data aggregation and computation specifically for privacy-sensitive content [8]. Here the homomorphic encryption scheme that serves a building block of our proposed profile matching protocols is reviewed.

C. Autoregressive Moving Average (ARMA) Model

Autoregressive model (AR) is a classic tool for understanding and predicting a time series data [9]. It estimates the current term of the series by a linear weighted sum of previous terms (i.e., observations) in the series. The model order is generally much smaller than the length of the series. AR is often combined with Moving-Average model (MA) to obtain complex ARMA model for generally improved accuracy. While AR depends on the previous terms of a time series data, MA describes the current value of the series using a linear weighted sum of white Gaussian noise or random shocks of its prior values [10].

IV. EXPLICIT COMPARISON BASED APPROACH

eCPM protocol allows two users to compare their attribute values on a specified attribute without disclosing the values to each other. But, the protocol reveals the comparison result to the initiator, and therefore offers conditional anonymity. The protocol has a fundamental bootstrapping phase, where the TCA generates all system parameters, user pseudonyms, and keying materials.

A. Bootstrapping

The protocol has a fundamental bootstrapping phase, where the TCA generates all system parameters, user pseudonyms, and keying materials. Specifically, the TCA runs *G* to generate $\langle p, q, R, Rq, Rp, \chi \rangle$ for initiating the homomorphic encryption The TCA generates a pair of public and private keys (*pkTCA*, *skTCA*) for itself. The public key *pkTCA* is open to all users; the private key *skTCA* is a secret which will be used to issue certificates for user pseudonyms and keying materials, as shown below. The TCA generates disjoint sets of pseudonyms (*pidi*) and disjoint sets of homomorphic public keys (*pki*) for users (*ui*). For every *pidi* and *pki* of *ui*, the TCA generates the corresponding secret keys *pski* and *ski*. In correspondence to each pseudonym *pidi*, it assigns a certificate *certpidi* to *ui*, which can be used to confirm the validity of *pidi*. Generally, the TCA uses *skTCA* to generate a signature on *pidi* and *pki*. The TCA outputs *certpidi* as a tuple (*pki*, *SignskTCA*(*pidi*, *pki*)). The homomorphic secret key *ski* is delivered to *ui* together with *pski*; *pki* is tied to *pidi* and varies as the change of pseudonyms.

V. IMPLICIT COMPARISON BASED APPROACH

Here the implicit–based profile matching (iCPM) is proposed by adopting the oblivious transfer cryptographic technique. It is considered that users have distinct values for any given attribute. The iCPM consists of three main steps. In the first step, an interested category by setting element to 1 and other elements to 0 in a length, vector. Then encrypt the vector by using the homomorphic encryption and sends the encrypted vector but still can process on the ciphertext. In the second step, computes the ciphertexts with input of self-defined messages for 1 ≤message ≤length.

A. Protocol Steps

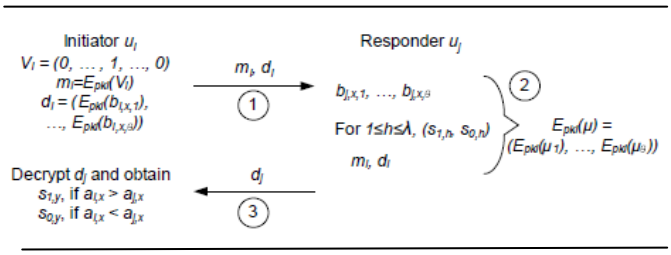


Fig 1. ICPM flow

VI. IMPLICIT PREDICTABLE BASED APPROACH

Both the eCPM and the iCPM perform profile matching on a single attribute. For a matching involving multiple attributes, they have to be executed multiple times, each time on one attribute. In this section, the iCPM is extended to the multi attribute cases, without jeopardizing its anonymity property,

and obtain an implicit Predicate-based Profile Matching protocol, i.e., iPPM. This protocol relies on a predicate which is a logical expression made of multiple comparisons spanning distinct attributes and thus supports sophisticated matching criteria within a single protocol run.

A. Protocol Steps

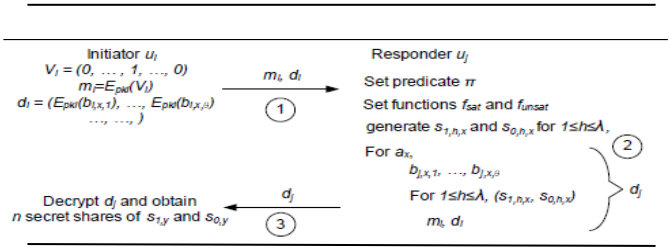


FIG 2. IPPM FLOW

VII. THREE CLASSES OF ANONYMITY

Consider a user has υ possible instances of the profile

Non-Anonymity:

A profile matching protocol provides non anonymity if after executing multiple runs of the protocol with any user, the probability of correctly guessing the profile of the user is equal to 1.

Conditional Anonymity:

A profile matching protocol achieves conditional anonymity if after executing multiple runs of the protocol with some user, the probability of correctly guessing the profile of the user is larger than 1/υ.

Full Anonymity:

A profile matching protocol achieves full anonymity if after executing multiple runs of the protocol with any user, the probability of correctly guessing the profile of the user is always 1/υ.

VIII. HINTS

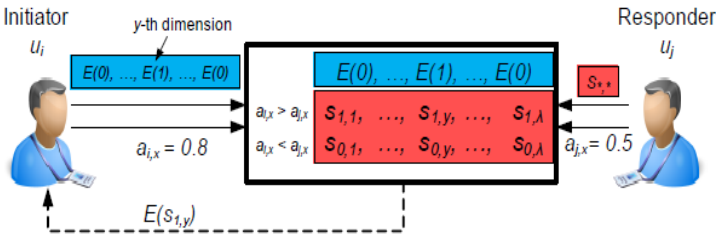
A. Abbreviations and acronyms

MSN	Mobile Social Networks
SNS	Social Networking Sites
OSN	Online Social Network
ARMA	Auto Regressive Moving Average Model
eCPM	explicit Comparison-based Profile Matching
iCPM	implicit Comparison-based Profile Matching
iPPM	implicit predicate-based Profile Matching
TCA	Trusted Central Authority

B. Equations

i. Homomorphic Encryption

A central authority runs a generator G which outputs ⟨p, q,R,Rq,Rp, χ⟩ as system public parameters:
p < q are two primes s.t. q ≡ 1 mod 4 and p ≫ l; Rings R = Z[x]/⟨x2+1⟩, Rq = R/qR = Zq[x]/⟨x2+1⟩; Message space Rp = Zp[x]/⟨x2 + 1⟩; A discrete Gaussian error distribution χ = DZn; with standard deviation σ. user ui has a public/private key pair (pki, ski) such that pki = {ai, bi}, with ai = −(bis +



pe), bi ∈ Rq and s, e ← χ, and ski = s. Let bi;1 and bi;2 be two messages encrypted by ui.

Encryption Epki (bi;1): ci;1 = (c0, c1) = (aiut + pgt +bi;1, biut + pft),

Decryption *Dski (ci;1): If denoting ci;1 = (c0, · · · , c1),bi;1 = (Σ1k=0cksk) mod p.*

ii. ARMA model

ARMA model is notated as ARMA(*p*, *q*) and written as

$$z_k = c + \sum_{i=1}^{P} \phi_i z_{k-i} + \sum_{j=1}^{q} \theta_j \epsilon_{k-j} + \epsilon_k,$$

where *c* is a constant standing for the mean of the series, ϕi the autoregressioncoefficients, θi the moving-average coefficients, and ϵk the zero-mean white Gaussian noise error.

C. Figures and Tables

The concept of profile matching is as follows

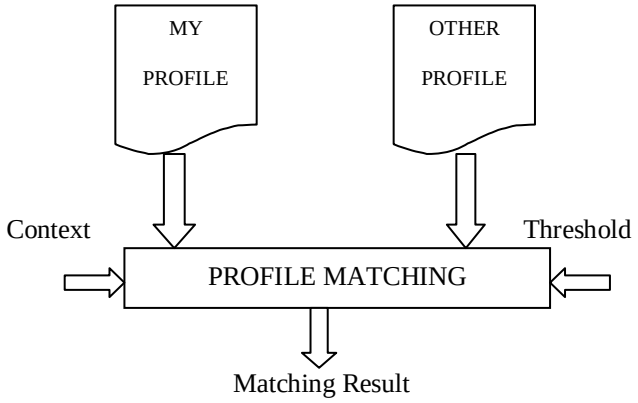


Fig 3. Profile Matching

The working scenario of eCPM is as follows,

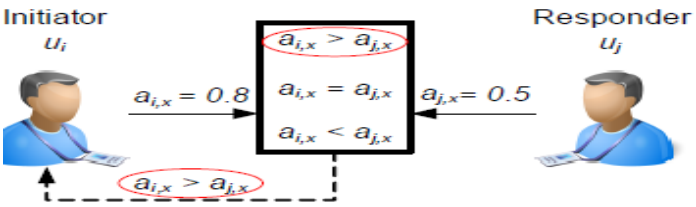


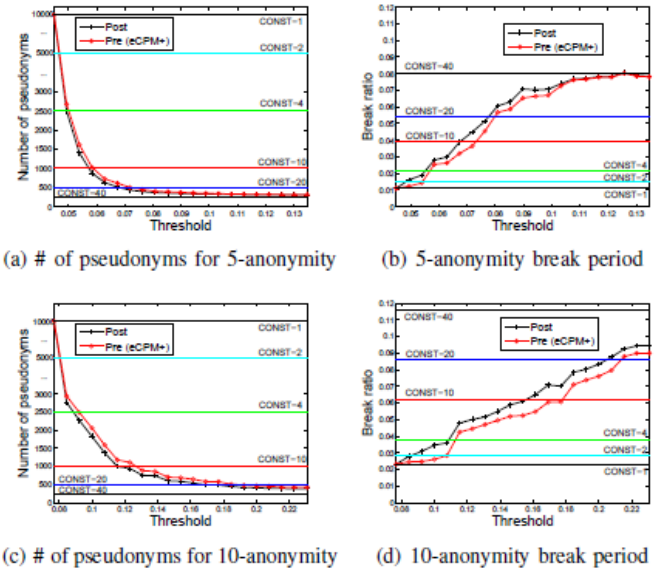
Fig 4. Working scenario of explicit comparison based approach

The working scenario of iCPM is as follows,

Fig 5. Working scenario of implicit comparison based approach

IX. RESULTS

The figure shows the performance of the constant, the postadaptive and the pre-adaptive strategies respectively for 5-anonymity and 10-anonymity, in relation with threshold *th*. The results are obtained with respect to the 32nd user. For the constant strategy, multiple lines are plotted, respectively corresponding to $z = \{1, 2, 4, 10, 20, 40\}$. As z goes up, the user consumes a decreasingly number of pseudonyms and has an increasingly break ratio (the ratio of the number of time slots that the k-anonymity of the 32nd user is broken to 10,000). It can be seen that the number of pseudonyms consumed by the post-adaptive and pre-adaptive strategies are much smaller than those of the constant strategy. For example, in the case of 5-anonymity and *th* = 0.0763, the post-adaptive strategy spends 369 pseudonyms and results in a 514 time slot anonymity break period. The constant strategy consumes 500(> 369) pseudonyms and has a 0.0540(>.0514) break ratio. The post-adaptive strategy outperforms the constant strategy in anonymity protection by using fewer pseudonyms to achieve smaller break ratio. Similar phenomena are observed for other *th* values and 10-anonymity scenario as well. In particular, we find that as expected, the pre-adaptive strategy leads to yet better anonymity performance than the post-adaptive one. It shows that in case of 5-anonymity and *th* = 0.0763, the pre-adaptive strategy consumes 449(> 369) pseudonyms and results in a 0.0445(< 0.0514) break ratio. The pre-adaptive strategy consumes slightly more pseudonyms, but achieves significantly shorter anonymity break period.



(a) # of pseudonyms for 5-anonymity (b) 5-anonymity break period

(c) # of pseudonyms for 10-anonymity (d) 10-anonymity break period

Fig 6. Pseudonyms and break ratio

X. PERFORMANCE ANALYSIS

The performance of three proposed protocols eCPM, iCPM, and iPPM is studied in terms of communication overhead and anonymity strength. When analyzing anonymity, it is considered that users have distinct values for any given attribute. Non-distinct attribute values and comparison operations "≥" and "≤" will be considered in the future work.

A. Communication Overhead

Let |R| be the size of one ring element in *Rq*. In the eCPM, the initiator and the responder both need to send ciphertexts in size of 2|R| and the communication overhead is thus subject only to the system parameter |R|. In order to achieve full anonymity, the iCPM constructs ciphertext in a sequence of operations.It is known that |Enc(b)| = 2|R|. Thus, the communication overhead of the initiator is $2(\theta + \lambda)|R|$ with $\theta = \lceil \log l \rceil$. It can be seen that the initiator's communication overhead increases with system parameters (θ, λ). An addition operation of homomorphic encryption does not increase the ciphertext size, while a multiplication with inputs of two ciphertexts of lengths a|R| and b|R| outputs a (a+b−1)|R|-length ciphertext. Thus, in the iCPM, the communication overhead of the responder increases to 6θ|R|. It is concluded that the communication overhead of the eCPM and the iCPM are constantly dependent on system parameters (θ, λ). The iPPM extends the iCPM by building complex predicates. From the protocol description, we observe that if a predicate includes $n \geq 1$ comparisons, the communication overhead of the iPPM would be approximately *n* times of that in the iCPM.

B. Anonymity

Suppose that user *ui* is currently using pseudonym *pidi* to execute profile matching with others.An adversary aiming to break the *k-anonymity* of *ui is* considered [11]. *k*-anonymity is a classic concept for evaluating anonymity. It implies that a series of comparison results provide *k*-anonymity protection to a user if the user's behavior cannot be distinguished from at least $k - 1$ other users.

XI. CONCLUSION

A unique comparison-based profile matching problem in Mobile Social Networks (MSNs) has been investigated, and novel protocols are proposed to solve it. The explicit Comparison based Profile Matching (eCPM) protocol provides conditional anonymity. It reveals the comparison result to the initiator. Considering the *k*-anonymity as a user requirement; the anonymity risk level in relation to the pseudonym change for consecutive eCPM runs is analyzed. Further an enhanced version of the eCPM, i.e., eCPM+ is introduced, by exploiting the prediction-based strategy and adopting the pre-adaptive pseudonym change. The effectiveness of the eCPM+ is validated through extensive simulations using real-trace data. Two protocols with full anonymity, i.e., implicit Comparison-based Profile Matching (iCPM) and implicit Predicate-based Profile Matching (iPPM) has been devised. The iCPM handles profile matching based on a single comparison of an attribute while the iPPM is implemented with a logical expression made of multiple comparisons spanning multiple attributes. The iCPM and the iPPM both enable users to anonymously request for messages and respond to the requests according to the profile matching result, without disclosing any profile information. In current version of the iCPM and the iPPM, "$>$" and "$<$" operations for profile matching is implemented. One future work is to extend them to support more operations, such as "$\geq$" and "$\leq$". Another future work is to hide the predicate information in the iPPM. Currently, the responder needs to transmit the threshold value of the predicate to the initiator, which may reveal partial information of the responder's interest. Restricting the disclosure of such parameter will be of significance for advancing comparison-based family of profile matching protocols and warrants deep investigation.

REFERENCES

[1]. Comscore,"http://www.comscoredatamine.com/."

[2]. R.Gross, A. Acquisti, and H. J. H. III, "Information revelation and privacy in online social networks," in *WPES*, 2005, pp. 71–80.

[3]. Raad, E. ; LE2I, Bourgogne Univ., Dijon, France ; Chbeir, R. ; Dipanda, A.,"User Profile Matching in Social Networks", 13th International Conference on Network-Based Information Systems (NBiS), 2010

[4]. Rui Zhang, Jinxue Zhang, Yanchao Zhang, Jinyuan Sun, and Guanhua Yan, "Privacy-preserving profile matching for proximitybased mobile social networking," IEEE Journal on Selected Areas in Communications,Special Issue on Emerging Technologies in Communications, 2012

[5]. Wei Dong ; Univ. of Texas at Austin, Austin, TX, USA ; Dave, V. ; Lili Qiu ; Yin Zhang,"Secure friend discovery in mobile social networks",INFOCOM, 2011 Proceedings IEEE.

[6]. Xi Chen Sch., Nanjing Univ., Nanjing, China; Michael, K. Privacy Issues and Solutions in Social Network Sites,IEEE Society on Social Implications of Technology, 2012

[7] P. Paillier, "Public-key cryptosystems based on composite degree Residuosity classes," in *EUROCRYPT*, 1999, pp. 223–238.

[8] M. Naehrig, K. Lauter, and V. Vaikuntanathan, "Can homomorphic encryption be practical?" in *CCSW*, 2011, pp. 113–124.

[9] H. Ltkepohl, *New introduction to multiple time series analysis*. Springer, 2005.

[10] X. Liang, X. Li, Q. Shen, R. Lu, X. Lin, X. Shen, and W. Zhuang, "Exploiting prediction to enable secure and reliable routing in wireless body area networks," in *Proc. IEEE INFOCOM*, 2012, pp. 388–396.

[11] L. Sweeney, "k-anonymity: A model for protecting privacy," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, no. 5, pp. 557–570, 2002.

[12] S. Ioannidis, A. Chaintreau, and L. Massouli´e, "Optimal and scalable distribution of content updates over a mobile social network," in *Proc.IEEE INFOCOM*, 2009, pp. 1422–1430.

[13] R. Lu, X. Lin, and X. Shen, "Spring: A social-based privacy-preserving packet forwarding protocol for vehicular delay tolerant networks," in *Proc. IEEE INFOCOM*, 2010, pp. 632–640.

[14] W. He, Y. Huang, K. Nahrstedt, and B. Wu, "Message propagation in adhoc- based proximity mobile social networks," in *PERCOM workshops*, 2010, pp. 141–146.

[15] D. Niyato, P. Wang, W. Saad, and A. Hjørungnes, "Controlled Coalitional games for cooperative mobile social networks," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 4, pp. 1812–1824, 2011.

