PROFILE MATCHING USING SECURE – PRIVACY DISTRIBUTED METHOD IN MOBILE SOCIAL NETWORKS

Karthik R Rao

Abstract--The Profile matching means two users comparing their personal profiles and is often the first step towards effective PMSN (private mobile social networks). It, however, conflicts with users’ growing privacy concerns about disclosing their personal profiles to complete strangers before deciding to interact with them. Our protocols enable two users to perform profile matching without disclosing any information about their profiles beyond the comparison result. Making new connections according to personal preferences is a crucial service in mobile social networking, where an initiating user can find matching users within physical proximity of him/her. In existing systems for such services, usually all the users directly publish their complete profiles for others to search. However, in many applications, the users’ personal profiles may contain sensitive information that they do not want to make public. In this paper, we propose FindU, a set of privacy-preserving profile matching schemes for proximity-based mobile social networks. In FindU, an initiating user can find from a group of users the one whose profile best matches with his/her; to limit the risk of privacy exposure, only necessary and minimal information about the private attributes of the participating users is exchanged. Two increasing levels of user privacy are defined, with decreasing amounts of revealed profile information. Leveraging secure multi-party computation (SMC) techniques, we propose novel protocols that realize each of the user privacy levels, which can also be personalized by the users. We provide formal security proofs and performance evaluation on our schemes, and show their advantages in both security and efficiency over state-of-the-art schemes. The social proximity between two users as the matching metric, which measures the distance between their social coordinates with each being a vector precomputed by a trusted central server to represent the location of a user in an online social network. By comparison, our work does not rely on the affiliation of PMSN users with a single online social network and addresses a more general private matching problem for PMSN by supports fine-grained personal profiles and a wide spectrum of matching metrics.
Mining high utility Result sets from databases refers to finding the result sets with high profile. Here, the meaning of result set utility is interestingness, importance, or safe ability of an uploader to users.

**Disadvantage:**
- Opens up the possibility for hackers to commit fraud and launch spam and virus attacks.
- Increases the risk of people falling prey to online scams that seem genuine, resulting in data or identity theft.
- May result in negative comments from employees about the company or potential legal consequences if employees use these sites to view objectionable, illicit or offensive material.
- Potentially results in lost productivity, especially if employees are busy updating profiles.
- Existing methods often generate a huge set of PHURs (potential high utility Result sets) and their mining performance is degraded consequently.
- The huge number of PHURs forms a challenging problem to the mining performance since the more PHURs the generates, the higher processing time it consumes.

**IV. PROPOSED SYSTEM**

The Proposed strategies can not only decrease the overestimated utilities of PHURs but greatly reduce the number of candidates. Different types of both real and synthetic data sets are used in a series of experiments to the performance of the proposed algorithm with state-of-the-art utility mining algorithms. Experimental results show that UP-Growth and UP-Growth+ outperform other algorithms substantially in term of execution time, especially when databases contain lots of long transactions or low minimum utility thresholds are set.

In this paper, we overcome the above challenges and make the following main contributions.

1. Two algorithms, named Utility pattern growth (UP Growth) and UP-Growth+, and a compact tree structure, called utility pattern tree (UP-Tree), for discovering high utility Result sets and maintaining important information related to utility patterns within databases are proposed.
2. We formulate the privacy preservation problem of profile matching in MSN. Two levels of privacy are defined along with their threat models, where the higher privacy level leaks less profile information to the adversary than the lower level.
3. We propose two fully distributed privacy-preserving profile matching schemes, one of them being a private set intersection protocol and the other is a private cardinality of set-intersection protocol. However, solutions based on existing PSI schemes are far from efficient. We leverage secure multi-party computation based on polynomial secret sharing, and propose several key enhancements to improve the computation and communication efficiency.

**Advantage:**
- Proximity-based mobile social networking (PMSN) becomes increasingly popular due to the explosive growth of smart phones.
- Two mutually mistrusting parties, each holding a private data set, jointly Compute the intersection or the intersection cardinality of the two sets without leaking any additional information to either party.
- Facilitates open communication, leading to enhanced information discovery and delivery.
- Allows employees to discuss ideas, post news, ask questions and share links.
- Provides an opportunity to widen business contacts.
- Targets a wide audience, making it a useful and effective recruitment tool.
- Improves business reputation and client base with minimal use of advertising.
- Expands market research, implements marketing campaigns.
V. PROJECT ENHANCEMENT

We build UP-Growth outperforms UP-Growth although they have tradeoffs on memory usage. The reason is that UP-Growth+ utilizes minimal node utilities for further decreasing overestimated utilities of result sets. Even though it spends time and memory to check and store minimal node utilities, they are more effective especially when there are many longer transactions in databases. Even though it spends time and memory to check and store minimal node utilities, they are more effective especially when there are many longer transactions in databases. This is because when number of candidates of the two algorithms is similar, UP-Growth+ carries more computations and is thus slower. Finally, high utility Result sets are efficiently identified from the set of PHURs which is much smaller than HTWUIs generated by IHUP. By the reasons mentioned above, the proposed algorithms UP-Growth and UP-Growth+ achieve better performance than IHUP algorithm.

VI. ARCHITECTURE

The Proposed Mining Method

UP-Tree, a basic method for generating PHURs is to mine UP-Tree by FP-Growth. However too many candidates will be generated. Thus, we propose an algorithm UP-Growth by pushing two more strategies into the framework of FP-Growth. By the strategies, overestimated utilities of Result sets can be decreased and thus the number of PHURs can be further reduced.

To address this issue, we propose two novel algorithms as well as a compact data structure for efficiently discovering high utility Result sets from transactional databases. Major contributions of this work are summarized as follows:

1. Two algorithms, named utility pattern growth (UP Growth) and UP-Growth+, and a compact tree structure, called utility pattern tree (UP-Tree), for discovering high utility Result sets and
maintaining important information related to utility patterns within databases are proposed. High-utility Result sets can be generated from UP-Tree efficiently with only two scans of original databases.

2. Several strategies are proposed for facilitating the mining processes of UP-Growth and UP-Growth by maintaining only essential information in UP-Tree. By these strategies, overestimated utilities of candidates can be well reduced by discarding utilities of the Results that cannot be high utility or are not involved in the search space. The proposed strategies can not only decrease the overestimated utilities of PHURs but also greatly reduce the number of candidates.

3. Different types of both real and synthetic data sets are used in a series of experiments to compare the performance of the proposed algorithms with the state-of-the-art utility mining algorithms. Experimental results show that UP-Growth and UP-Growth+ outperform other algorithms substantially in terms of execution time, especially when databases contain lots of long transactions or low minimum utility thresholds are set.

\[
\text{Pu}(\text{<ADC>}; \{\text{B}\}-\text{CPB}) = 10 \times \text{miu(A)} \times \text{<ADC>:count} = 105 \times 1 = 5
\]

and

\[
\text{Pu}(\text{<ADC>}; \{\text{B}\}-\text{CPB}) = 32 \times \text{miu(A)} \times \text{<ADE>:count} = 325 \times 1 = 27;
\]

**Improved Mining Method: UP-Growth+**

UP-Growth achieves better performance than FP-Growth by using DLU and DLN to decrease overestimated utilities of Result sets. However, the overestimated utilities can be closer to their actual utilities by eliminating the estimated utilities that are closer to actual utilities of unpromising Results and descendant nodes. In this section, we propose an improved method, named UP-Growth+, for reducing overestimated utilities more effectively.

In UP-Growth, minimum Result utility table is used to reduce the overestimated utilities. In UP-Growth+, minimal node utilities in each path are used to make the estimated pruning values closer to real utility values of the pruned Results in database.

\[
\text{Nik.nu}_{\text{new}} = \text{Nik.nu}_{\text{old}} + \text{pu}(p, \{\text{im}\} \text{CPB}) \sum_{m} \text{nu}(ij) \times p\text{.count}
\]

J = k+1

VII. CONCLUSION

In this paper, we for the first time formalize the problem of privacy-preserving distributed profile matching in MSNs, and propose two concrete schemes that achieve increasing levels of user privacy preservation. Towards designing lightweight protocols, we utilize Shamir secret sharing as the main secure computation technique, while we propose additional enhancements to lower the proposed schemes’ communication costs. Through extensive security analysis and simulation study, we show that 1) our schemes are proven secure under the HBC model, and can be easily extended to prevent certain active attacks; 2) our schemes are much more efficient than state-of-the-art ones in MSNs where the network size is in the order of tens, and when the number of
query attributes is smaller than the number of profile attributes.

REFERENCES


