A Survey: Tradeoff between Data Sharing and Privacy Preservation Using Trust Parameters

Zainab Yameen

PG Scholar, Department of Computer Science and Engineering, KBN College of Engineering, Kalaburagi

Abstract: **Online social networks have now become the most popular platforms for people to share information with others. Along with this, there is a serious threat to individuals' privacy. One privacy risk comes from the sharing of co-owned data, i.e., when a user shares a data item that involves multiple users, some users' privacy may be compromised, since different users generally have different opinions on who can access the data. How to design a collaborative management mechanism to deal with such a privacy issue has recently attracted much attention. In this paper, we propose a trust-based mechanism to realize collaborative privacy management. Basically, a user decides whether or not to post a data item based on the aggregated opinion of all involved users. The trust values between users are used to weight users' opinions, and the values are updated according to users' privacy loss. Moreover, the user can make a trade-off between data sharing and privacy preserving by tuning the parameter of the proposed mechanism. We formulate the selecting of the parameter as a multi-armed bandit problem and apply the upper confidence bound policy to solve the problem. Simulation results demonstrate that the trust-based mechanism can encourage the user to be considerate of others' privacy, and the proposed bandit approach can bring the user a high payoff.**

Keywords: multi-armed bandit, social trust, voting scheme, collaborative privacy management, online social networks.

1. INTRODUCTION

Online social networks (OSNs), such as Facebook, Google and Twitter have become the most important platforms for people to make social connections with others. Thousands of millions of users post data about their daily lives in terms of text messages, photos, or videos on OSNs. Such data often contain sensitive information of users. The privacy control mechanisms implemented in current OSNs only impose restrictions on users who want to access others' data. While there is no strict restriction on users who post data. A consequence of this one-side restriction is that the user who posts data may unintentionally violate other users' privacy. Consider the following example. Suppose that a user 'A' posts a photo of him/her playing with a friend B, and user ' A' specifies that the photo can be accessed by his/her colleagues. If user B considers this photo to be sensitive and user B is not familiar with user A's colleagues, then user B's privacy will be violated. In the above case, the photo is actually co-owned by the two users. Hence, the privacy policy specified by user A should be compatible with user B's privacy policy, otherwise, user B will suffer a loss in privacy. Data which are co-owned by multiple users are quite common in OSNs. Privacy management of such data requires a collaboration of all involved users. In this system it is assumed that it is the user who wants to post data makes a collective decision based on other users' privacy requirements. Previous studies usually assume that the user who posts the data will tag all the users involved, or the involved users can be identified. In such a case, the mediator is able to notify the involved users about the posting of the data. However, in practice, it is likely that the user posts the data without tagging other users and the involved users are hard to be identified automatically. Considering this, a mechanism is proposed which requires the user to solicit other users' opinions before posting data. And a trust-weighted voting scheme is applied to aggregate different users' opinions. The importance of the vote depends on the trust value between the two users. Only when the aggregation of the votes satisfies a certain condition, the data can be posted. Moreover, the trust values between users are not fixed. A user will lose the trust of others if he/she posts a data item that incurs privacy loss of others. Also, a user can gain more trust from others if he/she adopts others' opinions. The interaction between the trust value and the privacy loss

implies that if the user wants to reduce his/her privacy loss, then when posting a co-owned data item, the user should always consider others' privacy requirements rather than taking a unilateral decision.

2. LITERATURE SURVEY

In [1] authors provided a novel fully distributed and collaborative k-anonymity protocol (LPAF) to protect users' location information and ensure better privacy while forwarding queries/replies to/from untrusted location-based service (LBS) over opportunistic mobile networks (OppMNets). They utilized a lightweight multihop Markov-based stochastic model for location prediction to guide queries toward the LBS's location and to reduce required resources in terms of retransmission overheads and developed a formal analytical model. In [2] authors appeared as taking advantage of real-life social trust between average users (called "trust-based social networks") as well as threshold cryptography. In [3] authors introduced a novel framework of attacks, which they called forest fire attacks. In these attacks, an attacker initially obtains a small number of compromised users, and then the attacker iteratively attacks the rest of users by exploiting trustee-based social authentications. In [4], Y. Tang, H. Wang, and W. Dou gave a brief survey of researches on trust-based incentive in P2P network. By investigating the reputation systems in P2P networks, they outlined some key issues within the design of trust-based incentive in P2P networks. After that they introduced some other approaches addressing the incentive of P2P networks. In [5] authors proposed a trust model for social networks with the aim of building trust communities that inspire members to share their experiences, feelings and opinions in an open and honest way without the fear of being judged. The unique feature of their model was that the trust value is derived from the social capital built in the social networks over a period of time.

3. A STUDY OF UCB ALGORITHM

The basic idea of UCB is to estimate the unknown expected reward of each arm based on previously observed rewards of the arm. During the learning procedure, the policy maintains two quantities, namely n_i and $\bar{r}_i$, for each arm. The first quantity n_i, $tP-1\ \tau=1\ 1\ (I\tau = i)$ denotes how many times the arm has been chosen up to time t. The second quantity $\bar{r}_i$ is the average of the rewards observed for the arm. The average $\bar{r}_i$ is treated as the estimate of the true expected reward with $\bar{r}_i + \alpha$ being the upper confidence bound. The policy always chooses the arm which currently has the maximal upper confidence bound [6].

Algorithm 1 UCB

Require: $\alpha \in \mathbb{R}^+$

for $t = 1$ to K **do**
  Choose arm $I_t = t$
  Observe and record the reward $r_{I_t,t}$
  $r_t \leftarrow r_{I_t,t}$
  $n_{I_t} \leftarrow 1$
end for
for $t = K+1$ to T **do**
  for $i = 1$ to K **do**
    $\bar{r}_i \leftarrow \frac{1}{n_i} \sum_{\tau=1}^{t-1} r_\tau \mathbb{1}(I_\tau = i)$
  end for
  Choose arm $I_t = \underset{i=1,\cdots,K}{\operatorname{argmax}} \left(\bar{r}_i + \alpha\sqrt{\frac{\ln t}{n_i}}\right)$ with ties broken arbitrarily
  Observe and record the reward $r_{I_t,t}$
  $r_t \leftarrow r_{I_t,t}$
  $n_{I_t} \leftarrow n_{I_t} + 1$
end for

4. HOW IS THE ALGORITHM USEFUL IN THIS SCHEME

The rationality of the UCB policy can be explained as follows. The policy estimates the expected reward of each arm and computes the corresponding confidence bound. The width of the confidence bound indicates the uncertainty of the estimated expected reward. The wider the confidence bound is, the less accurate the estimate is. For a given arm, the width of the confidence bound depends on how many times the arm has been chosen. After initialization (i.e. choosing each arm once), the policy always chooses the arm corresponding to the maximal upper confidence bound. If the chosen arm was seldom used in past rounds, which means it has a wide confidence bound, then we can say the policy makes a risky explorative decision. If the chosen arm has been used multiple times, which implies the corresponding confidence bound is relatively tight and the average reward is high, then

we can say the policy makes a conservative exploitative decision. As time goes by, the confidence bound of each arm narrows down. Hence, it is more likely that the arm corresponding to the maximal upper confidence bound is the arm which has the maximal average reward. Meanwhile, the average reward gets closer to the true expected reward. That is to say, after sufficient number of trials, the policy can determine the real best arm. The UCB policy was proposed for the stochastic multi-armed bandit problem. The performance of the learning policy is measured by regret. It has been shown that the UCB policy can achieve a logarithmic regret uniformly over the number of trials. And when the support of every reward distribution is in [0, 1], the upper bound of the expected regret after any number n of trials can be explicitly expressed as a formula of n. Though we apply the UCB policy to the proposed problem, it is difficult to analyze the regret theoretically. Instead, we consider an empirical method to evaluate the performance of the learning policy.

To verify the feasibility of the proposed methods, a series of simulations are conducted.

1. Dataset: simulations are conducted on both synthetic data and real-world data. we generate a scale-free network and a small-world network. The scale-free network contains 1000 nodes and 20021 undirected edges. The small-world network contains 1000 nodes and 20,000 undirected edges.
2. Data Sharing Simulation: we simulate users' data sharing behaviors via the following way. Suppose that time evolves in rounds. At each round $t \in \{1, 2, \cdots, T\}$, a certain number of users are selected as owners, i.e., they want to post data items that involve multiple users. To select the owners, we first pick a number $prob_i$ uniformly at random from [0, 1] for every user v_i . If $prob_i$ is smaller than a pre-specified threshold ρ, then user vi will act as an owner. The threshold ρ actually denotes the ratio of owners to all users. Given an owner o, we determine the corresponding stakeholders via the following two approaches respectively: randomly select several users from the set of o's friends (i.e. users who are directly connected to o); randomly select several users from the whole user set (other than the owner himself/herself). In the latter case, we define that the number of stakeholders is at most 20, considering the average node degree of the network. Each stakeholder's opinion is picked from {0, 1} uniformly at random. given a pair of users vi and vj , where vi is the owner or the stakeholder of a data item, only if the distance between users vi and vj is no greater than a threshold dis_{th}, user v_j is allowed to access the data item.
3. Trust Evaluation: for any 2 users the the trust values are set say a=b, though a=b it does not mean that the trust is reciprocal. Firstly the distance between the 2 users is determined and the values are initializedand are updated at each round of simulations.

CONCLUSION

In this scheme the privacy issue caused by the sharing of co-owned data in OSNs is studied. To help the owner of data collaborate with the stakeholders on the control of data sharing, a trust-based mechanism is proposed. When a user is about to post a data item, the user first solicits the stakeholders' opinions on data sharing, the more the user trusts a stakeholder, the more the user values the stakeholder's opinion. If a user suffers a privacy loss because of the data sharing behavior of another user, then the user's trust in another user decreases. Simultaneously, considering that the user needs to balance between data sharing and privacy preserving, a bandit approach to tune the threshold in the proposed trust-based mechanism is applied, so that the user can get a high long-turn payoff which is defined as the difference between the benefit from posting data and the privacy loss caused by other users.

5. REFERENCES

[1] S. Zakhary, M. Radenkovic, and A. Benslimane, "Efficient location privacy-aware forwarding in opportunistic mobile networks,"

IEEE Transactions on Vehicular Technology, vol. 63, no. 2, pp. 893–906, February 2014.

[2] S. Xu, X. Li, T. P. Parker, and X. Wang, "Exploiting trust-based social networks for distributed protection of sensitive data," IEEE Transactions on Information Forensics and Security, vol. 6, no. 1, pp. 39–52, March 2011.

[3] N. Z. Gong and D. Wang, "On the security of trustee-based social authentications," IEEE Transactions on Information Forensics and Security, vol. 9, no. 8, pp. 1251–1263, Aug 2014.

[4] Y. Tang, H. Wang, and W. Dou, "Trust based incentive in p2p network," in IEEE International Conference on E-Commerce Technology for Dynamic E-Business, September 2004, pp. 302–305.

[5] S. Nepal, W. Sherchan, and C. Paris, "Strust: A trust model for social networks," in 2011 IEEE 10th International Conference on Trust, Security and Privacy in Computing and Communications, November 2011, pp. 841–846.

[6] Lei Xu1, Chunxiao Jiang2, Nengqiang He3, Zhu Han4 and Abderrahim Benslimane, "Trust-based Collaborative Privacy Management in Online Social Networks", IEEE Transactions on Information Forensics and Security, 2018.