SCIENCE CHINA
Information Sciences

• LETTER •

March 2019, Vol. 62 039106:1–039106:3
https://doi.org/10.1007/s11432-018-9483-x

UMBRELLA: user demand privacy preserving framework based on association rules and differential privacy in social networks

Chunliu YAN[1], Ziyi NI[1], Bin CAO[1*], Rongxing LU[2], Shaohua WU[1] & Qinyu ZHANG[1]

[1]*Harbin Institute of Technology (Shenzhen), Shenzhen* 518055*, China;*
[2]*Faculty of Computer Science, University of New Brunswick, Fredericton* E3B 5A3*, Canada*

Received 10 April 2018/Accepted 8 June 2018/Published online 18 October 2018

Citation Yan C L, Ni Z Y, Cao B, et al. UMBRELLA: user demand privacy preserving framework based on association rules and differential privacy in social networks. Sci China Inf Sci, 2019, 62(3): 039106, https://doi.org/10.1007/s11432-018-9483-x

Dear editor,

The past several years have witnessed that people deeply rely on online services, such as shopping, transportation, and medical services. At the same time, privacy leakage has been a hot topic in social networks [1, 2]. Morever, the increasing applications of location based services (LBSs) indeed make our lives more convenient. However, LBSs also lead to significant privacy leakage. On the one hand, location information of a single point is sensitive. On the other hand, location trajectories can excavate more sensitive information despite location. In this regard, location privacy-preserving mechanisms have been extensively studied, such as Dummy, spatial confusion [3, 4], encryption technology [5, 6], differential privacy (DP) [7, 8] and references therein.

It is noted that the user demand based services can also be utilized to obtain sensitive information. In this study, a malicious attacker can infer private information such as living habits, health status, and beliefs from the user demand. For example, after searching for a house, it is likely that solar energy, house insurance and other house related ads are pushed to the user. In more sensitive situations, the privacy implicated by user demand should be protected, i.e., malicious data mining with respect to the user demand should be prohibited. We aim to propose a framework and the key enabling techniques to achieve this goal.

Inspired by the famous marketing strategy of 'beer and diaper' [9], we incorporate the association rule and DP into our system model and mechanism design. Consider a case that, a single mom needs to search for stores selling diapers for her illegitimate child, while she is unwilling to expose that she has a child, i.e., keeping no search history revealing her child, how can she find the stores using her smart phone? i.e., how to protect this privacy led by user demand disclosure? Recall that beer and diaper can be linked in the essence of popular shopping habit. According to this interesting association relationship, it is more likely that the store selling beer also places diapers around. Therefore, it is reasonable to search for beer rather than diaper, and beer itself is less sensitive than diapers. Nevertheless, the searching results may not satisfy her demand, i.e., only beer, which we deem as the quality-of-service (QoS) loss. As an aim to address user demand privacy in more generalized application scenarios, the motivation of this work focuses on designing a framework and techniques which can strike a balance between user demand privacy and QoS loss. To this end, we propose UMBRELLA: a user demand privacy preserving scheme based on association rules and DP in so-

* Corresponding author (email: caobin@hit.edu.cn)

cial networks.

Basically, the association rules can be categorized into two types. The first type is commonly known as support, which is the probability that several data associations occur. The other is the confidence representing the conditional probability of different data sets. In the above case, diaper is replaced by beer thanks to the high confidence between diapers and beer. In addition to the association rule, we also leverage Laplace noise to further obfuscate the real demand and the replacement, which is known as DP. To the best of our knowledge, this is the first work that studies the user demand privacy and QoS in social networks.

System framework. Our model can be mainly divided into three layers, namely entity layer, strategy layer, and evaluation layer, respectively.

• Entity layer mainly involves users, service providers and attackers. A user sends a service request to a service provider, and the service provider returns the result to the user. Attackers directly attack the service providers to obtain the query content. It is possible that the service provider is untrustworthy. In addition, the attacker may also intercept the communication between users and service providers.

• Strategy layer mainly establishes mathematical model by combining association rules and DP technology. The quantitative indicators of association rules are the support degree and confidence level. The DP is performed via Laplace mechanism, i.e., adding noise to the confidence.

• Evaluation layer analyzes and optimizes the system performance in terms of mathematical formulation and optimization, e.g., game theory and convex optimization, to strike a balance between privacy and QoS.

The main focus of our work is on the strategy layer and the evaluation layer.

Association rules and DP. The form of the association rule is: $X \to Y$, where both X and Y are item sets, $X \cap Y = \phi$, and ϕ is an empty set, showing X can infer to Y. X is named the condition of the rule, and Y is called the conclusion. Let S be the support of the rule $X \to Y$ representing the ratio of X and Y to all transactions in the database, i.e.,

$$S(X,Y) = P_r(XY) = \frac{\text{number}(XY)}{\text{number}(\text{AllSamples})}. \quad (1)$$

Define C as the confidence of the rule $X \to Y$ which represents the ratio of X and Y to transactions Y, i.e.,

$$C(X \to Y) = P_r(X|Y) = \frac{P_r(XY)}{P_r(Y)}. \quad (2)$$

In our system model, we assume that strong association rules are available from advanced date mining techniques, and focus on the confidence issues.

In our model, Laplace mechanism is used to add noise to the confidence. The pdf of Laplace is given as follows:

$$f(x|\mu, b) = \frac{1}{2b} \mathrm{e}^{-\frac{|x-\mu|}{b}}, \quad (3)$$

where μ and $b = \frac{\Delta f}{\epsilon}$ represent the position parameter and the scale parameter, respectively, and Δf is the sensitivity of f.

UMBRELLA. A real demand is denoted by s, and S is a set of all possible items consisting of s. The replacement is $o \in O$, and $O = S$. Let prior knowledge ω be the probability distribution of s, which is written as

$$\omega(s) = P_r\{S = s\}. \quad (4)$$

A probabilistic protection mechanism is to replace real demand with other replacements, i.e.,

$$p(o|s) = P_r\{O = o|S = s\}. \quad (5)$$

Since the real demand is replaced by other items, despite the strong association rules will also lead to QoS loss. Let $c(o, s)$ be the confidence between s and o[1)], and denote $\log_2(1/c(o, s))$ as the confidence distance to reflect QoS loss. We add the Laplace noise to the $c(o, s)$, and the confidence with noise is expressed as $c^*(o, s)$. Basically, the greater the confidence, the less the QoS loss, and vice versa. QoS loss is written as

$$\sum_s \omega(s) \sum_o p(o|s) \cdot \log_2(1/c^*(o, s)). \quad (6)$$

Particularly, $c(o, s)$ equaling to 1 shows the real demand is replaced by itself, hence no QoS loss. Our goal is to minimize QoS loss.

Attacker aims to obtain a possible demand $\hat{s}$ through o by the following inference mechanism:

$$q(\hat{s}|o) = P_r\{S = \hat{s}|O = o\}. \quad (7)$$

We define the inference distance function as $\log_2(1/c(\hat{s}, s))$. Given an inference function q, user's distortion privacy is given as

$$\sum_o p(o|s) \sum_{\hat{s}} q(\hat{s}|o) \cdot \log_2(1/c(\hat{s}, s)) \quad (8)$$

reflecting attacker's inference error.

Statistically, the expected distortion privacy is

$$\sum_s \omega(s) \sum_o p(o|s) \sum_{\hat{s}} q(\hat{s}|o) \cdot \log_2(1/c(\hat{s}, s)). \quad (9)$$

1) Both the uppercase C and the lowercase c represent confidence.

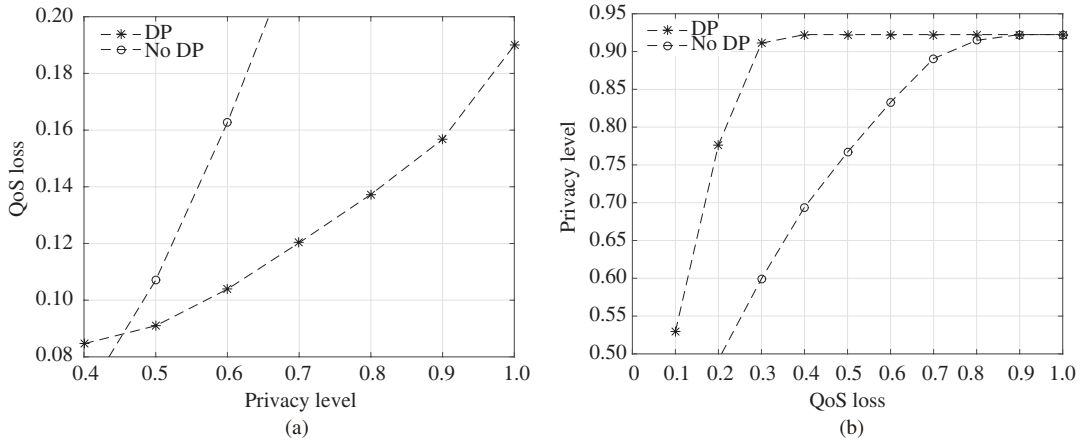

Figure 1 (a) QoS loss vs. privacy, with and without DP in minimizing QoS loss model; (b) privacy vs. QoS loss, with and without DP in Stackelberg game.

Game theory analysis. To perform effective and robust solution to the above problem, game theory is utilized to achieve the optimal strategies. Note that the attacker adapts his/her strategy according to that of the user, it is reasonable to formulate the problem as a two-tied game, namely, Stackelberg game. In this regard, the user is a leader, and the attacker is a follower, respectively, in the game. As the user and the attacker conflict with each other in terms of privacy, it can be further formulated as a zero-sum game.

Experimental evaluation. We conduct experiments to evaluate the performance of our UMBRELLA. Firstly, we aim to minimize QoS loss under distortion privacy. In Figure 1(a), the result proves that QoS loss increases with privacy level. Then we adopt the zero-sum Stackelberg game method to perform effective and robust solution. Similarly, privacy increases with QoS loss at the initial stage. Further, privacy remains unchanged due to that game reached equilibrium as shown in Figure1(b). The performance of UMBRELLA with DP is better than that without DP as a whole as shown in both Figures 1(a) and (b).

Conclusion. A novel demand privacy preserving scheme based on association rules and DP (UMBRELLA) is proposed. We leverage Stackelberg game to formulate the trade-off between privacy and QoS loss, and examine the system performance in terms of privacy and QoS levels. Through experimental results and detailed analyses, the proposed UMBRELLA not only achieves demand privacy preservation but also strikes a balance between privacy and QoS.

Acknowledgements This work was supported by National Natural Sciences Foundation of China (Grant No. 61501211), Basic Research Project of Shenzhen (Grant Nos. JCYJ20160531192013063, JCYJ20170307151148585), Natural Sciences Foundation of Guangdong (Grant No. 2017A030313372), Natural Scientific Research Innovation Foundation in Harbin Institute of Technology, Natural Sciences Foundation of Jiangxi (Grant Nos. 20151BAB217001, 20151BAB217018), and S&T Foundation of Jingdezhen.

References

1 Li H X, Zhu H J, Ma D. Demographic information inference through meta-data analysis of wi-fi traffic. IEEE Trans Mobile Comput, 2018, 17: 1033–1047
2 Li H X, Chen Q R, Zhu H J, et al. Privacy leakage via de-anonymization and aggregation in heterogeneous social networks. IEEE Trans Depen Secur Comput, 2017. doi: 10.1109/TDSC.2017.2754249
3 Peng T, Liu Q, Wang G J. Enhanced location privacy preserving scheme in location-based services. IEEE Syst J, 2017, 11: 219–230
4 Shahid A R, Jeukeng L, Zeng W, et al. PPVC: privacy preserving voronoi cell for location-based services. In: Proceedings of International Conference on Computing, Networking and Communications, Santa Clara, 2017. 351–355
5 Ma X D, Li H, Ma J F, et al. APPLET: a privacy-preserving framework for location-aware recommender system. Sci China Inf Sci, 2017, 60: 092101
6 Zhu H, Lu R X, Huang C, et al. An efficient privacy-preserving location-based services query scheme in outsourced cloud. IEEE Trans Veh Technol, 2016, 65: 7729–7739
7 Andrés M E, Bordenabe N E, Chatzikokolakis K, et al. Geo-indistinguishability: differential privacy for location-based systems. In: Proceedings of the 2013 ACM SIGSAC Conference on Computer and Communications Security, Berlin, 2013. 901–914
8 Shokri R. Privacy games: optimal user-centric data obfuscation. In: Proceedings of the 15th Privacy Enhancing Technologies, Philadelphia, 2015. 299–315
9 Feng L, Dillon T, Liu J. Inter-transactional association rules for multi-dimensional contexts for prediction and their application to studying meteorological data. Data Knowl Eng, 2001, 37: 85–115