Journal of Network and Computer Applications 191 (2021) 103115

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

journal homepage: www.elsevier.com/locate/jnca

Privacy protection among three antithetic-parties for context-aware services

Yan Huang [a,*], Wei Li [c], Jinbao Wang [b], Zhipeng Cai [c], Anu G. Bourgeois [c]

[a] College of Computing and Software Engineering, Kennesaw State University, 1100 South Marietta Pkwy, Marietta, GA, 30060, USA
[b] Academy of Fundamental and Interdisciplinary Science, Harbin Institute of Technology, Harbin, China
[c] Department of Computer Science, Georgia State University, 25 Park Place, Atlanta, GA, 30303, USA

ARTICLE INFO

Keywords:
Privacy protection
Game theory
Context-aware services

ABSTRACT

The popularity of context-aware services is improving the quality of life, while raising serious privacy issues. In order for users to receive quality service, they are at risk of leaking private information by adversaries that are possibly eavesdropping on the data and/or by the untrusted service platform selling off its data to adversaries. Game theory has been utilized as a powerful tool to achieve privacy preservation by strategically balancing the trade-off between profit (service) and cost (data leakage) for the user. However, most of the existing schemes cannot fully exploit the power of game theory, as they fail to depict the mutual relationship between any two (of the three) parties involved: user, platform, and adversary. Existing schemes are also not always able to provide specific guidance for a user to reduce the impact of the joint threats from the platform and adversary. In this paper, we design a privacy-preserving game to quantify the three parties' concerns and capture interactions between any two of them. We also identify the best strategy for each party at a fine-grained level, i.e. specific settings, not simply binary choices. We validate the performance of our proposed game model through both a theoretical analysis and experiments.

1. Introduction

Thanks to the rapid development and popularity of context-aware services, such as recommendation, navigation, and social association, individuals' lives have become more comfortable and convenient than ever before (Zheng and Cai, 2020; Cheng et al., 2021). We can use Yelp to find a popular restaurant, use Facebook to keep up with our friends, and use Google Maps to find the way to a destination. When enjoying such personalized services, we need to provide these service/application platforms with our personal data, e.g., location, weight, age, and income. Unfortunately, service platforms cannot always be trusted by users raising serious privacy issues, which lies in two aspects. On one hand, more personal data is needed to acquire higher quality of service, resulting in that even more private/sensitive information could be inferred from our submitted data. On the other hand, users' personal data may be shared or resold by service platforms to adversaries, as is common practice (Cai et al., 2018; Armstrong, 2016; Cai and He, 2019). Besides having a risk of leaking personal data via the platform, the data may be captured via malicious attacks, such as eavesdropping by an adversary. According to the statistics from (Wang et al., 2016), 55% of iOS applications and 59.7% of Android applications surreptitiously leak user's personal data.

Based on this, *users are suffering joint threats of privacy leakage from untrusted platforms as well as adversaries*, which we depict in Fig. 1. There is no doubt that, in the era of information, the collection and the use of personal data are major privacy concerns for individuals (Kokolakis, 2017), and such concerns will only grow over time (TRUSTe/NCSA, 2016). *Thus, the ongoing progress of context-aware services, the increasingly serious privacy leakage, and the growing privacy concern together make data privacy preservation imperative for users.*

In the past years, privacy-preserving mechanisms have received a lot of attention from researchers. Besides cryptography, game theory has been widely applied as a strategic methodology to search for optimal strategies balancing the trade-off between the benefit of sharing data and cost of privacy disclosure (Sfar et al., 2017; Wu et al., 2017; Hussain et al., 2018). Note that most of the existing research only focuses on the interaction between two opposite parties (Shokri et al., 2017), i.e., using a defender-attacker game model. In (Li et al., 2018; Vakilinia et al.,

[*] Corresponding author.
E-mail addresses: yhuang24@kennesaw.edu (Y. Huang), wli28@gsu.edu (W. Li), wangjinbao@hit.edu.cn (J. Wang), zcai@gsu.edu (Z. Cai), abourgeois@gsu.edu (A.G. Bourgeois).

https://doi.org/10.1016/j.jnca.2021.103115
Received 26 June 2020; Received in revised form 13 April 2021; Accepted 19 May 2021
Available online 29 June 2021

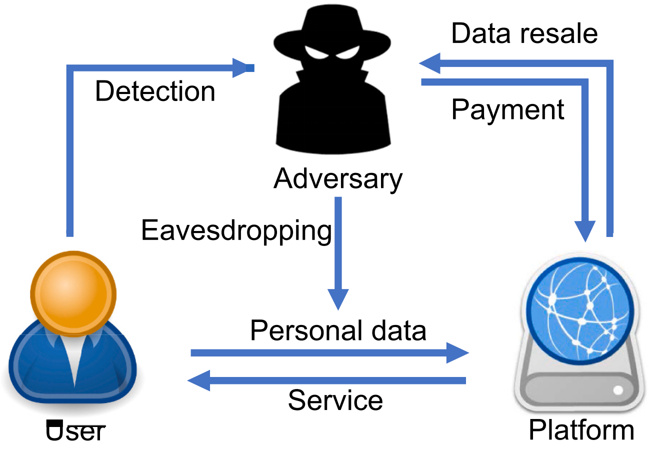

Fig. 1. Structure of thee-party game.

2017), various three-party game models are proposed. But, the game models of (Wang et al., 2017a, 2017b) are not "real" three-party models, because they fail to depict the interaction between any two of the three parties, i.e. they considered either data resale by the platform or attacks by the adversary. Additionally, the schemes in (Wang et al., 2017a, 2017b; KarimiAdl et al., 2012) only provide a binary solution. Specifically, the schemes only determine whether the user should submit their data to receive services (and risk loss of privacy) or not submit any data (and risk poor, to no quality of service).

Further exploring the mutual relationships among user, platform, and adversary would be more helpful for the user to defend against both the untrusted platform and the adversary. Moreover, it would be beneficial to produce a more fine-grained solution, so that a user could have the option to provide obscured data and still receive adequate service. For this purpose, this paper aims to *design a three-party game model among the three antithetic parties for users to simultaneously protect their privacy from untrusted service platforms and adversaries*. Such a realistic and complicated game model challenges us in the following aspects: (i) *Complicated game structure*. As shown in Fig. 1, the interaction occurs between any two of the three parties, increasing the difficulty in addressing the three parties' individual concerns – how does the user assess the potential risk of privacy loss and determine the granularity when submitting personal data; how does the platform determine data resale with consideration of the risk of reputation loss; and how does the adversary make a choice between purchase and eavesdropping? (ii) *Joint threats*. In such a complicated game, the user has to defend the joint threats from both the platform and the adversary, which may be hard to accomplish. (iii) *Multiple data attributes*. For many services, it is common that users need to submit multiple data attributes that could be correlated together. Any obscurity applied to one attribute would need to be correlated accordingly. (iv) *Theoretical analysis & solution*. Designing, analyzing, and solving the proposed three-party game are destined to be difficult due to the complexity of the game structure and correlated data attributes.

Our research endeavor to overcome the above challenges is briefly introduced as follows. Firstly, in our game model, we link the three parties by elaborately quantifying their concerns and mutual interactions, such that they are inseparable. Secondly, based on our game model, we perform a theoretical analysis to rigorously prove the optimal strategies of the three parties, including the optimal data release granularity for the user, the optimal data resale strategy for the platform, and the optimal probability to purchase data (or launch an attack) for the adversary. Finally, we conduct simulations with abstracted privacy protection settings from surveys to validate the effectiveness of our proposed game model.

To the best of our knowledge, we are the first to provide a fine-grained analysis on the behaviors and interactions for the user, platform, and adversary with considering resistance to the joint threats. Our major contributions are summarized as below:

- We design a three-party game to capture the complicated interactions among the user, platform, and adversary, with a goal to defend against the joint threats to the user's privacy from both untrusted platform and adversary.
- We present an in-depth theoretical analysis to identify the best strategy of each of the three parties: user, platform and adversary.
- We perform comprehensive simulations with abstracted privacy protection settings from surveys to evaluate the performance of our game model, regarding the optimal strategy, cost, and utility for each of the three parties.

The rest of the paper is organized as follows. Section 2 summarizes the related work. Our game model is introduced in Section 3. The optimal strategy of each party and the performance of our game are analyzed in Section 4 and Section 5, respectively. Finally, Section 6 briefly concludes this paper and discusses our future work.

2. Related work

Game theory is a popular and efficient methodology to capture interaction between defender and adversary. In this section, we mainly summarize the most related literature in the area of game-theoretical privacy preservation, in which according to the type of game model, the existing work can be classified into two major categories, i.e., two-party and three-party game.

Most of the existing work investigates the interaction between two parties: user/data owner and adversary. In (Chorppath and Alpcan, 2013; Shokri et al., 2012, 2017; Rottondi et al., 2017; Sfar et al., 2017), games are based on a two-player model, i.e., one-against-one. When there are multiple users trying to maintain a certain privacy preserving level, the user-adversary game can be modeled as an n-player game (Wu et al., 2017; Liu et al., 2013; Freudiger et al., 2013; Ma et al., 2017; Ying and Nayak, 2017; Xu et al., 2017), but with the drawback that all users must have the same settings. Another drawback is that the two-party game cannot depict the interactions among three antithetic parties.

Recently, three-party game models have been proposed to study complicated privacy issues among user/data owner, service provider/data requester, and adversary. In (Li et al., 2018), Li et al. designed a hierarchical game, incorporating a user-service provider game and a user-attacker game, to maximize the service provider's utility while assisting the user in defending the attacker. Adl et al. (KarimiAdl et al., 2012) proposed a three-party sequential game to analyze the interactions among a data provider, a data collector, and a data user (i.e., the adversary), which can guide the data provider and the data collector to find the optimal strategies deciding whether to cooperate with the data user. In (Wang et al., 2017a, 2017b), Wang et al. studied the interactions among a user, an application, and an adversary to answer two questions: whether the user should submit data and whether the application should resell the user's data? To resolve the trade-off between sharing advantages and privacy exposure of cybersecurity information exchange system, Vakilinia et al. (Vakilinia et al., 2017) designed a three-party game for privacy-preserving cybersecurity information exchange framework consisting of an attacker, an organization, and a cybersecurity information exchange system. However, the three-party games in (Li et al., 2018; KarimiAdl et al., 2012; Wang et al., 2017a, 2017b; Vakilinia et al., 2017) fail to build the mutual interaction between any two of the three parties, and the strategy of each party in (KarimiAdl et al., 2012; Wang et al., 2017a, 2017b) is coarse-grained, or binary, by indicating "whether to cooperate with opponents or not".

Contrasting from the existing work, we establish a three-party game to capture the mutual interaction between any two of the three antithetic parties (including user/data owner, service provider/data requester, and adversary) and aim to identify their strategies on *"whether and how to defend (or cooperate with) others"*, which can offer *fine-grained guidance to the three parties.*

3. Three-party game model

In this section, the interaction among user, platform, and adversary is modeled as a three-party game, in which their strategies, benefits, and costs are mathematically formulated.

3.1. User model

We consider the following scenario: a user submits personal dataset, denoted by $D = \{d_1, d_2, \ldots, d_n\}$, to a platform to acquire data-based service, where the dataset could contain one or more attributes and d_i ($1 \le i \le n$) is the data of attribute i. Due to privacy concerns, the user may report data attributes with different data release granularity. Formally, the data release granularity of attribute i is defined as $g_i \in [0, 1]$, and the corresponding data granularity set is $G = \{g_1, g_2, \ldots, g_i, \ldots, g_n\}$. Specifically, with a larger g_i, the data of attribute i is less obscured, revealing more personal/sensitive information; for examples, $g_i = 0$ if d_i does not contain any personal data, and $g_i = 1$ if d_i is fully accurate. For example, when the user provides age information to the platform. It can use the age range instead of the exact range. If the user (assume age 24) sets the granularity to 0, the user will provide its age range 1–100. The provided user information actually provides no personal age information. If the user sets the granularity to 0.9, the user will provide an age range of 10, like 20–29 years old. If the user sets the granularity to 1, the user will provide the accurate age 24 to the platform. In this paper, we use data granularity as a measurement of data quality/obscurity.

As the data release granularity increases, the quality of user's requested service is increased with diminishing marginal benefit (Xu et al., 2015). Suppose that the quality of attribute i-based service can achieve a maximum value q_i when $g_i = 1$. Then, the relationship between the quality of attribute i-based service and data release granularity g_i can be formulated to be $2q_ig_i - q_i(g_i)^2$. In addition, any two data attributes may correlate with each other, and such correlation can be exploited to infer more sensitive information (Zhu et al., 2015; Zhang et al., 2016). Let e_{ij} represent the correlation between attribute i and attribute j. Due to correlations among data attributes, the data of attribute i not only contributes to the quality of attribute i-based service, but also contributes to the quality of attribute j-based service. Thus, given the user's dataset D, data release granularity set G, and data correlation $\{e_{ij}\}$, the overall service quality can be estimated as follows.

$$\sum_{i=1}^{n}\left(1 + \sum_{j=1, j\neq i}^{n} e_{ij}g_j\right)\left(2q_ig_i - q_i(g_i)^2\right) \tag{1}$$

While enjoying the service provided by the platform, privacy leakage incurred by data submission brings privacy loss to the user. One possible method for this privacy loss could be due to a malicious attack by an adversary that eavesdrops on the data submitted by the user. In real-world scenarios, the working efficiency of information retrieval is restricted by many factors, such as equipment performance and retrieval technique. Different adversaries may have different work efficiency and different type of work efficiency. For example, when an adversary cannot always succeed when they launch an attack, the work efficiency could be the success rate. In another scenario, an adversary cannot acquire all the data in an attack. The work efficiency can be the expected percentage in an attack. The working efficiency of eavesdropping at the adversary side is denoted by $\varphi \in [0, 1]$, so the granularity of eavesdropped data is φg_i. Assume the adversary purchases data from the platform with probability b and the probability of eavesdropping is $1 - b$. Then, the expected cost due to eavesdropping of dataset D is defined as

$$(1-b)\sum_{i=1}^{n} c_i\varphi g_i,$$

where c_i is the unit privacy cost when $g_i = 1$.

Another possible method for privacy loss could be that the user's submitted data is resold by the platform to a third-party (e.g., adversary) for more profit. To avoid too much sensitive information been sold, the platform will add noise or use privacy project techniques such as θ-differential privacy or k-anonymity before they sell the data. In this scenario, the selling strategy of the platform is the privacy protection level θ of differential privacy or $\frac{1}{k}$ of k-anonymity. We define the set of platform's resale strategy as $S = \{s_1, s_2, \ldots, s_n\}$, where $s_i \in [0, 1]$ and s_ig_i is the resold data granularity of attribute i. The platform does not resell d_i if $s_i = 0$ but resell all collected d_i if $s_i = 1$. The expected privacy cost due to data resale at the platform side can be computed by

$$b\sum_{i=1}^{n} c_is_ig_i.$$

By combining the received service quality and the experienced privacy cost, the user's utility can be calculated in Eq. (2).

$$\begin{aligned} U_u = \ & \lambda\sum_{i=1}^{n}\left(1 + \sum_{j=1, j\neq i}^{n} e_{ij}g_j\right)\left(2q_ig_i - q_i(g_i)^2\right) \\ & -(1-b)\sum_{i=1}^{n} c_i\varphi g_i - b\sum_{i=1}^{n} c_is_ig_i, \end{aligned} \tag{2}$$

where λ is the convention rate between service quality and privacy cost, i.e., one unit of privacy cost is equivalent to λ units of service quality loss. Moreover, λ is also used to measure the user's privacy preference; that is, the user would care more about privacy cost than service quality if λ is large, and the service quality outweighs the privacy cost if λ is small.

In our proposed three-party game, the user aims to maximize its utility by balancing the trade-off between service quality and privacy cost by strategically setting the granularity set G. Accordingly, the optimization problem at the user side is

$$\begin{aligned} & \max_{G} U_u, \\ \text{s.t. } & g_i \in [0, 1], i \in [1, n]. \end{aligned}$$

3.2. Platform model

The platform provides users with requested services based on their submitted data. For instance, Google provides navigation service to users based on their input location.

While providing service to the user, the platform has its private valuation, defined to be V_p, for the collected data from the user. With user's data, the platform can obtain profit from data-based production, such as data statistic analysis and new product development. From the viewpoint that data is a type of potential productivity, the value of data can be computed according to the standard form of Cobb-Douglas production function (Meeusen and van Den Broeck, 1977) as

$$\theta_p\left(\sum_{i=1}^{n} g_i\right)^{\zeta_p},$$

where θ_p is the total value productivity of the platform, and $\zeta_p \in (0, 1)$ is the platform's value output elasticities of G.

To get extra benefits, the platform may resell the collected data to a third party (i.e., the adversary). Assume that p_i is the unit data price of attributes i with $g_i = 1$, so the expected payment received from the adversary is

$$b\sum_{i=1}^{n} p_i s_i g_i, \tag{3}$$

in which b is the adversary's purchase probability and s_i represents the platform's resale strategy.

However, reselling the user's data may cause the risk of reputation loss at the user side and/or in public. According to the instantaneous risk function (Fershtman and Kamien, 1987; Hu et al., 2015), we can define the risk of reputation loss due to data resale of attribute i as

$$l_1 s_i g_i + l_2 (s_i g_i)^2,$$

where l_1 and l_2 are constant parameters of the risk estimation function. Since there may exist a correlation between two data attributes, the adversary can infer more personal/sensitive information from one data attribute to another, leading to an increase in the reputation loss at the platform side. Accordingly, the risk of reputation loss can be estimated as

$$\sum_{i=1}^{n}\left(1+\sum_{j=1,j\neq i}^{n} e_{ij} s_j g_j\right)\left(l_1 s_i g_i + l_2 (s_i g_i)^2\right)$$

In addition, there exists a data processing cost c_p at the platform side. Since the data processing cost may be determined by the processing technology, which is out of the scope of this paper, we assume c_p is a system parameter for simplicity. Therefore, the platform's utility, denoted by U_p, can be defined to be

$$\begin{aligned} U_p = \quad & b\sum_{i=1}^{n} p_i s_i g_i + \theta_p\left(\sum_{i=1}^{n} g_i\right)^{\zeta_p} - c_p \\ & -\sum_{i=1}^{n}\left(1+\sum_{j=1,j\neq i}^{n} e_{ij} s_j g_j\right)\left(l_1 s_i g_i + l_2 (s_i g_i)^2\right) \end{aligned} \tag{4}$$

One can see that the platform faces a struggle between benefit and reputation cost from data resale. More specifically, reselling more accurate data can enhance the profit while damaging reputation, but reselling less accurate data can reduce reputation loss while losing attractiveness of data resale. Thus, to improve utility via balancing the trade-off between benefit and cost, the platform needs to choose a proper resale strategy S. Formally, the optimization problem of the platform is formulated as

$$\begin{aligned} & \max_{S} U_p, \\ \text{s.t. } & s_i \in [0,1], i \in [1,n]. \end{aligned}$$

3.3. Adversary model

To retrieve the user's private information, the adversary could purchase data from the platform with probability b or eavesdrop on the communication between the user and the platform with probability $1 - b$. With respect to each data attribute i, the granularity of purchased data is $s_i g_i$, and that of the eavesdropped data is φg_i.

The adversary also has private valuation for the obtained data. With the analysis similar to that in Section 3.2, we can utilize Cobb-Douglas production function (Meeusen and van Den Broeck, 1977) to compute adversary's private valuation as

$$b\theta_a\sum_{i=1}^{n}(s_i g_i)^{\zeta_a} + (1-b)\theta_a\sum_{i=1}^{n}(\varphi g_i)^{\zeta_a},$$

where θ_a is the data productivity of the adversary and $\zeta_a \in (0, 1)$ is the adversary's value output elasticities of data.

We suppose that the adversary can obtain all the data in D through eavesdropping at a cost (e.g., equipment and time) that can be quantified by a quadratic cost function(Osborne, 2009; Mohsenian-Rad et al., 2010), i.e.,

$$\sigma_1(1-b)^2 + \sigma_2(1-b) + \sigma_3,$$

where $\sigma_1 > 0$, $\sigma_2 \geq 0$, and $\sigma_3 \geq 0$ are constant parameters of the quadratic cost function. Note that when the adversary does not eavesdrop, there still is a cost because it needs to purchase equipment and resources for eavesdropping. If the adversary chooses to purchase data from the platform, the expected payment paid to the platform is formulated in Eq. (3).

To sum up, the utility of the adversary, denoted by U_a, can be computed to be

$$\begin{aligned} U_a = \quad & b\theta_a\left(\sum_{i=1}^{n} s_i g_i\right)^{\zeta_a} + (1-b)\theta_a\left(\sum_{i=1}^{n}\varphi g_i\right)^{\zeta_a} - b\sum_{i=1}^{n} p_i s_i g_i \\ & -\left(\sigma_1(1-b)^2 + \sigma_2(1-b) + \sigma_3\right) \end{aligned} \tag{5}$$

In the three-party game, the adversary faces the trade-off between data purchase and data eavesdropping, i.e., the probability to purchase/eavesdrop data. Therefore, to improve utility, the adversary has to choose a proper purchase probability b to maximize its utility, which can be formulated as the following optimization problem.

$$\begin{aligned} & \max_{b} \ U_a, \\ \text{s.t. } & b \in [0,1]. \end{aligned}$$

4. Nash Equilibrium analysis

In this section, we conduct in-depth theoretical analysis of the three parties' strategies and the relationships among their strategies.

4.1. Nash Equilibrium

In game theory, a Nash equilibrium is a strategy profile E^* with the property that no party can unilaterally do better by choosing an action different from E^*, given that other parties adhere to E^* (Osborne, 2009). Accordingly, the Nash equilibrium of our proposed three-party game can be defined in Definition 1.

Definition 1. *A strategy profile* $\mathrm{E}^* = (\mathrm{G}^*, \mathrm{S}^*, \mathrm{b}^*)$ *is called Nash Equilibrium for the proposed three-party game if the following properties simultaneously hold:*

$$\begin{aligned} & U_u(G^*, S^*, b^*) \geq U_u(G, S^*, b^*); \\ & U_p(G^*, S^*, b^*) \geq U_p(G^*, S, b^*); \\ & U_a(G^*, S^*, b^*) \geq U_a(G^*, S^*, b). \end{aligned}$$

4.2. Strategy analysis of user

To solve the optimization problem of the user, we analyze the concavity of its utility function. The first-order partial derivative and the second-order partial derivatives of Eq. (2) are obtained, respectively.

$$\begin{aligned} \frac{\partial}{\partial g_i} U_u = \quad & -(1-b)c_i\varphi - bc_i s_i + \lambda\sum_{j=1,j\neq i}^{n} e_{ij}\left(-q_j (g_j)^2 + 2q_j g_j\right) \\ & +\lambda\left(1+\sum_{j=1,j\neq i}^{n} e_{ij} g_j\right)\left(-2q_i g_i + 2q_i\right) \end{aligned}$$

$$\frac{\partial^2}{\partial g_i^2} U_u = \quad -2q_i\lambda\left(1+\sum_{j=1,j\neq i}^{n} e_{ij} g_j\right)$$

$$\frac{\partial^2}{\partial g_i g_j} U_u = \lambda e_{ij}\left(-2q_j g_j + 2q_j\right) + \lambda e_{ij}\left(-2q_i g_i + 2q_i\right)$$

To find the maximum value, we need to solve the following system of equations.

$$\begin{cases} \dfrac{\partial}{\partial g_1} U_u = 0; \\ \dfrac{\partial}{\partial g_2} U_u = 0; \\ \quad \dots \\ \dfrac{\partial}{\partial g_n} U_u = 0. \end{cases} \tag{6}$$

All the solutions of the system of equations are the extreme points of user's utility. To find the global maximum value, we create the corresponding Hessian matrix:

$$H_u = \begin{vmatrix} \dfrac{\partial^2}{\partial {g_1}^2} U_u & \dfrac{\partial^2}{\partial g_1 \partial g_2} U_u & \dots & \dfrac{\partial^2}{\partial g_1 \partial g_n} U_u \\ \dfrac{\partial^2}{\partial g_2 \partial g_1} U_u & \dfrac{\partial^2}{\partial {g_2}^2} U_u & \dots & \dfrac{\partial^2}{\partial g_2 \partial g_n} U_u \\ \vdots & \vdots & \dots & \vdots; \\ \dfrac{\partial^2}{\partial g_n \partial g_1} U_u & \dfrac{\partial^2}{\partial g_n \partial g_2} U_u & \dots & \dfrac{\partial^2}{\partial {g_n}^2} U_u \end{vmatrix}$$

The user has a maximum utility only if the matrix is a negative definite matrix. When either of the following two conditions holds, a matrix is negative definite (Horn and Johnson, 2012): (1) all its eigenvalues are less than 0; and (2) the even order principal minors are larger than 0 and odd order principal minors are less than 0. In other words, when the Hessian matrix of the user's utility function can meet anyone of the above two conditions, the user's optimal strategy can be found by solving Eq. (6).

We take the scenario where $e_{ij} = 0$ for $i, j \in [1, n]$ as an illustrative example. In this scenario, the first-order partial derivative and the second-order partial derivatives of the utility function are as follows.

$$\frac{\partial}{\partial g_i} U_u = \lambda(-2q_i g_i + 2q_i) - (1-b)c_i\varphi - bc_i s_i.$$

$$\frac{\partial^2}{\partial g_i^2} U_u = -2q_i\lambda.$$

$$\frac{\partial^2}{\partial g_i g_j} U_u = 0.$$

Then we derive the corresponding Hessian matrix, i.e.,

$$H_u = \begin{vmatrix} -2q_1\lambda & 0 & \dots & 0 \\ 0 & -2q_2\lambda & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & -2q_n\lambda \end{vmatrix}$$

Because the even order principal minors are larger than 0 and the odd order principal minors are less than 0, the matrix H_u is a negative definite matrix. Therefore, the utility function has a maximum value and the maximum points can be calculated by solving Eq. (6), i.e.,

$$\begin{cases} g_1 = \dfrac{-c_1\varphi + (\varphi - s_1)bc_1 + 2q_1\lambda}{2q_1\lambda}; \\ g_2 = \dfrac{-c_2\varphi + (\varphi - s_2)bc_2 + 2q_2\lambda}{2q_2\lambda}; \\ \quad \dots \\ g_n = \dfrac{-c_n\varphi + (\varphi - s_n)bc_n + 2q_n\lambda}{2q_n\lambda}. \end{cases} \tag{7}$$

Since $g_i \in [0, 1]$, the best data release granularity for attribute i is $g_i^* = \max\{\min\{g_i, 1\}, 0\}$.

From the results, to preserve data privacy, the user should decrease the data granularity g_i of attribute i if the platform increases data resale strategy s_i.

4.3. Strategy analysis of platform

We can compute the Hessian matrix to analyze the concavity of the platform's utility function as follows.

$$H_p = \begin{vmatrix} \dfrac{\partial^2}{\partial {s_1}^2} U_p & \dfrac{\partial^2}{\partial s_1 \partial s_2} U_p & \dots & \dfrac{\partial^2}{\partial s_1 \partial s_n} U_p \\ \dfrac{\partial^2}{\partial s_2 \partial s_1} U_p & \dfrac{\partial^2}{\partial {s_2}^2} U_p & \dots & \dfrac{\partial^2}{\partial s_2 \partial s_n} U_p \\ \vdots & \vdots & \dots & \vdots \\ \dfrac{\partial^2}{\partial s_n \partial s_1} U_p & \dfrac{\partial^2}{\partial s_n \partial s_2} U_p & \dots & \dfrac{\partial^2}{\partial {s_n}^2} U_p \end{vmatrix}$$

where

$$\frac{\partial^2}{\partial {s_i}^2} U_p = -2l_2 g_i^2 \left(1 + \sum_{j=1, j\neq i}^{n} e_{ij} s_j g_j\right),$$

and

$$\frac{\partial^2}{\partial s_i \partial g_j} U_p = -e_{ij} g_j \left(l_1 g_i + 2l_2 g_i^2 s_i\right) - e_{ij} g_i \left(l_1 g_j + 2l_2 g_j^2 s_j\right)$$

The platform has a maximum utility only if the Hessian matrix is a negative definite matrix that can satisfy either of the following two conditions (Horn and Johnson, 2012):

(1) all eigenvalues of H_p are less than 0; and (2) the even order principal minors of H_p are larger than 0 and odd order principal minors of H_p are less than 0.

If the maximum value exists, we can find the best strategy of the platform by solving the system of equations as shown below.

$$\begin{cases} \dfrac{\partial}{\partial s_1} U_p = 0; \\ \dfrac{\partial}{\partial s_2} U_p = 0; \\ \quad \dots \\ \dfrac{\partial}{\partial s_n} U_p = 0. \end{cases} \tag{8}$$

In Eq. (8), we have

$$\begin{aligned} \frac{\partial}{\partial s_i} U_p = {} & bp_i g_i - \left(1 + \sum_{j=1, j\neq i}^{n} e_{ij} s_j g_j\right)\left(l_1 g_i + 2l_2 g_i^2 s_i\right) \\ & - \sum_{j=1, j\neq i}^{n} e_{ij} g_i \left(l_1 s_j g_j + l_2 \left(s_j g_j\right)^2\right) \end{aligned}$$

We use the scenario when $e_{ij} = 0$ ($i, j \in [0, 1]$) as an example for demonstration. In this scenario, the first-order partial derivative and the second-order partial derivatives of the utility function are obtained in the following.

$$\frac{\partial}{\partial s_i} U_p = bp_i g_i - l_1 g_i - 2l_2 g_i^2 s_i.$$

$$\frac{\partial^2}{\partial {s_i}^2} U_p = -2l_2 g_i^2.$$

$$\frac{\partial^2}{\partial s_i \partial g_j} U_p = 0.$$

Then we derive the Hessian matrix as:

$$H_u = \begin{vmatrix} -2l_2 g_1^2 & 0 & \dots & 0 \\ 0 & -2l_2 g_2^2 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & -2l_2 g_n^2 \end{vmatrix}$$

Because the even order principal minors are larger than 0 and the odd order principal minors are less than 0, the matrix H_p is a negative definite matrix. Thus, the platform's utility function has a maximum value and the maximum points can be calculated by solving Eq. (8), i.e.,

$$\begin{cases} s_1 = \dfrac{bp_1 - l_1}{2l_2 g_1}; \\ s_2 = \dfrac{bp_2 - l_1}{2l_2 g_2}; \\ \dots \\ s_n = \dfrac{bp_n - l_1}{2l_2 g_n}. \end{cases} \tag{9}$$

As $s_i \in [0, 1]$, the best resale strategy for attribute i is $s_i^* = \max\{\min\{ s_i, 1\}, 0\}$.

According to the above analysis, to avoid too much reputation loss, the platform should decrease the value of s_i if the user increases g_i. Nevertheless, the granularity of resold data, $s_i g_i$, may be increased, bringing a profit increase to the platform. On the other hand, if the adversary prefers to purchase data rather than eavesdropping (i.e., enhance purchase probability b to a sufficiently large value), the platform can increase the value of s_i to earn more profit, in which the reputation loss maybe compensated by the payment from the adversary.

4.4. Strategy analysis of adversary

To maximize the utility, the adversary has to find the best strategy b^*. The first-order partial derivative and the second-order partial derivative of U with respect to b are respectively calculated by

$$\frac{dU_a}{db} = \theta_a \left(\sum_{i=1}^{n} s_i g_i \right)^{\zeta_a} - \theta_a \left(\sum_{i=1}^{n} \varphi g_i \right)^{\zeta_a} - \sum_{i=1}^{n} p_i s_i g_i - \sigma_1 (2b - 2) + \sigma_2$$

and

$$\frac{d^2 U_a}{db^2} = -2\sigma_1.$$

Since $\frac{d^2 U_a}{db^2} = -2\sigma_1 < 0$, the utility function of the adversary is a concave function, which means the maximum value is achievable when $\frac{dU_a}{db} = 0$. Thus, by setting $\frac{dU_a}{db} = 0$, the solution is

$$b = \frac{\theta_a \left(\sum_{i=1}^{n} s_i g_i \right)^{\zeta_a} - \theta_a \left(\sum_{i=1}^{n} \varphi g_i \right)^{\zeta_a} - \sum_{i=1}^{n} p_i s_i g_i + 2\sigma_1 + \sigma_2}{2\sigma_1}.$$

Because $b \in [0, 1]$, the best purchase probability is $b^* = \max\{ \min\{b, 1\}, 0\}$.

According to the above result, one can find that the purchase probability b is reduced when the working efficiency of eavesdropping φ and/or data price p_i increases. This is because the adversary prefers to eavesdrop rather than buying data for cost reduction if the granularity of eavesdropped data is higher than that of the purchased data. However, the relationship between the adversary's strategy and the platform's strategy and the relationship between the adversary's strategy and the user's strategy are not straightforward, because the purchase probability is also affected by the working efficiency φ, the price p_i for attribute i, the data productivity of the adversary θ_a, and the data value output elasticities of adversary ζ_a. These complicated relationships will be investigated in our simulations.

4.5. Deployment in realistic scenario

The theoretical NE analysis provides general guidance for all realistic context-aware services. To deploy the proposed framework in a realistic scenario, we suggest following these steps. Step 1 - Data collection. When any role wants to use this framework, it should collect data of all the other roles as background knowledge to set up the parameters. Step 2 - Parameter setting. The collected data will be used to calculate the actual parameters by using linear regression. Step 3 - Check the NE conditions. We analyze the conditions that make sure NE exists. The parameters will be used to check the conditions to make sure NE exists in the specific scenario. Step 4 - Calculate the optimal strategy. If the NE exists in the specific scenario, the optimal strategy can be calculated by following Eqs. (4.2) and (4.3).

5. Simulation

In this section, we study the interactions among the user, the platform, and the adversary via intensive simulations. In this paper, we assume that the user has multiple attributes in its dataset. However, in some cases, the user's dataset has only one attribute. To provide a detailed simulation result, we study the interactions among three parties in two scenarios: i) the user has one attribute in its dataset; ii) the user has more than one attribute in its dataset.

5.1. Interactions among three parties with one attribute

We first discuss the interaction among the three parties when the user's dataset has only one attribute, $D = \{d_1\}$. The default settings of the parameters are as follows. These value are selected because these parameters can provide better analysis that visualizing the theoretical NE analysis.

The granularity of d_1 is $g_1 = 0.6$. The unit privacy cost due to leakage of d_1 is $c_1 = 3$. Based on d_1, the user can get maximum service quality $q_1 = 50$. The convention rate λ of the user is 0.1. The platform resells d_1 by using reselling strategy $s_1 = 0.6$ with the price $p_1 = 20$. The other parameters in the platform's utility function are: $\theta_p = 15$, $\zeta_p = 0.6$, $l_1 = 5$, $l_2 = 10$, $c_p = 1$. The adversary has a purchase probability $b = 0.6$ and working efficiency $\varphi = 0.2$. The other parameters in the adversary's utility function are: $\sigma_1 = 1.5$, $\sigma_2 = 1$, $\sigma_3 = 1$, $\theta_a = 15$, $\zeta_a = 0.6$.

The simulations that follow depict different strategies by varying certain parameters from the perspective of each of the three parties.

5.1.1. Simulation result of User's utility and optimal strategy

The utility and optimal response of the user are investigated in this subsection. Figs. 2–4 reveal the simulation result of the user.

The results of the user's utility are presented in Figs. 2 and 3, from which we observe that the utility increases at first and then decreases as the granularity increases. The reason lies in two aspects: (i) when the granularity g_1 increases from 0 to a certain value (e.g., 3.46 in line $s_1 = 0.8$ of Fig. 2 and 3.55 in line $b_1 = 0.8$ of Fig. 3), the increase rate of privacy cost is smaller than that of received service quality, therefore the utility increases; and (ii) when g_1 continues increasing from such a certain value, the increase rate of privacy cost is larger than that of received service quality, leading to a decrease in the utility. In fact, such a certain value corresponds to the optimal granularity.

Besides, as shown in Fig. 2, the user's utility U_u decreases as the platform increases the value of its reselling strategy s_1. This is because when the platform increases the value of reselling strategy s_1, the

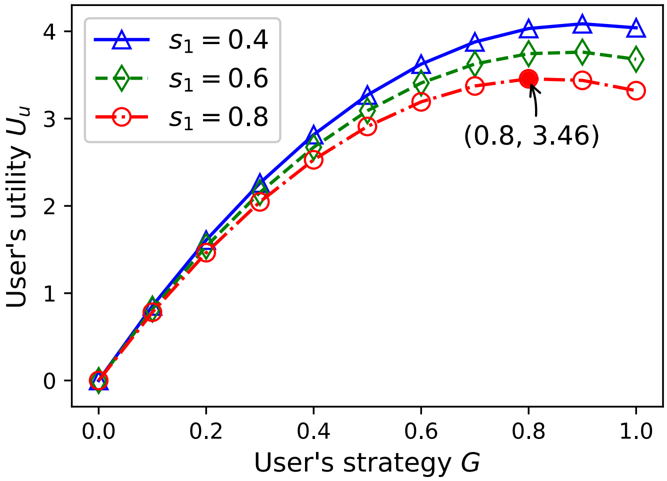

Fig. 2. Utility of user under various G and S.

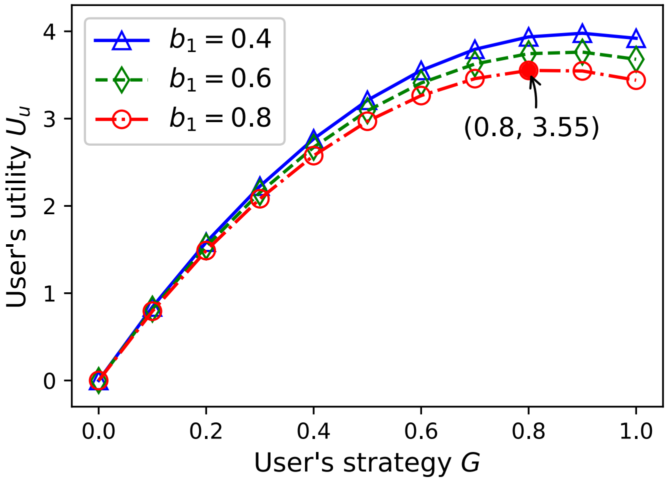

Fig. 3. Utility of user under various G and b_1.

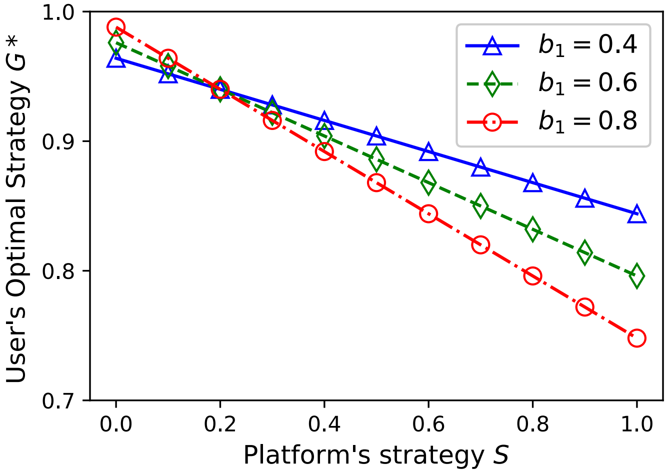

Fig. 4. Optimal strategy of user under various S and b_1.

granularity of the reselling data increases. That increases user's privacy cost, resulting in decreasing of user's utility.

Furthermore, the user's utility U_u also decreases as adversary increases its purchase probability as shown in Fig. 3. Because the purchased data of the adversary has a higher granularity than eavesdropped data, when the adversary increases the probability of purchase and decreases the probability of eavesdropping, the user has more privacy cost, leading to a decrease in the user's utility.

Fig. 4 states the optimal strategy of the user. We can see that the user decreases data granularity if the platform increases the value of reselling strategy s_1. When the platform increases the value of reselling strategy s_1, the granularity of the reselling data increases, thus increasing user's privacy cost. To reduce privacy cost, the user should decrease data granularity as shown in Fig. 4.

Fig. 4 also reveals how the user adjusts its strategy when the adversary uses different strategies. In Fig. 4, we can see that the three lines ($b_1 = 0.4$, $b_1 = 0.6$, and $b_1 = 0.8$) intersect at the point where $s_1 = \varphi = 0.2$. When $s_1 = \varphi$, the adversary's purchased data has the same data granularity with eavesdropped data, the user has the same privacy cost no matter what the adversary prefers, purchase or eavesdropping. Thus, the user does not need to change its data granularity as the adversary change its strategy when $s_1 = \varphi$.

Fig. 4 shows that the user adjusts its strategy according to the adversary's strategy as well as the platform's strategy: (i) the user decreases the granularity of the data as the adversary decreases the probability of data purchase and increases the probability of eavesdropping when the platform resells data with strategy $s_1 < \varphi$. (ii) the user decreases the granularity of the data as the adversary increases the probability of data purchase and decreases the probability of eavesdropping when the platform resells data with strategy $s_1 > \varphi$. The reason lies in two aspects: (i) when $s_1 < \varphi$, the adversary's eavesdropped data has a higher granularity than purchased data. Thus, the eavesdropping causes more privacy cost than data reselling to the user. To decrease privacy cost, the user should decrease the data granularity if the adversary increases the probability of eavesdropping and decreases the probability of data purchase.

(ii) On the contrary, when $s_1 > \varphi$, the adversary's purchased data has a higher granularity than eavesdropped data. Thus, the data reselling causes more privacy cost than eavesdropping to the user. To decrease privacy cost, the user should decrease the data granularity if the adversary decreases the probability of eavesdropping and increases the probability of data purchase.

5.1.2. Simulation result of Platform's utility and optimal strategy

We then study the utility and best response of the platform. The result is shown in Figs. 5 and 6.

From Fig. 5, we can tell that the platform's utility increases at first and then decreases over the increase of reselling strategy $S = \{s_1\}$. When s_1 increases from 0 to a certain value (e.g., 13.34 in line $g_1 = 0.8$), the increase rate of the cost is smaller than that of the profit, resulting in an improvement of utility; however, when s_1 continues increasing from

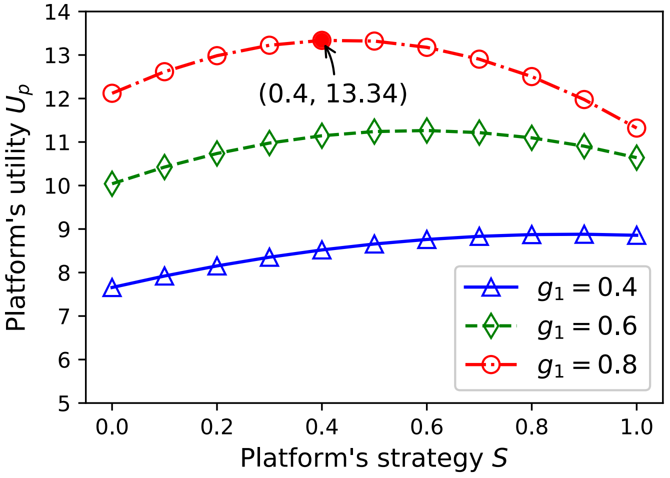

Fig. 5. Utility of platform under various S and G.

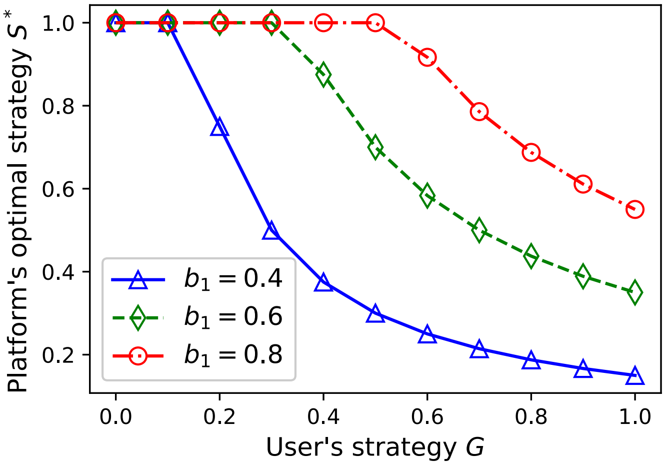

Fig. 6. Optimal strategy of platform under various G and b_1.

such a certain value, the increase rate of the cost is larger than that of the profit, further reducing the utility. In other words, there is an optimal value of s_1 for the platform to balance the profit of data resale and cost of reputation loss.

In Fig. 5, when the data granularity g_1 increases, the granularity of reselling data increases and brings more profit to the platform, leading to the increase of utility of the platform.

Fig. 6 reveals how the platform adjusts its optimal strategy when the user and adversary choose different strategies. When the user's granularity increases, the optimal strategy of the platform decreases. Both the profit and the reputation loss increase if the user increases granularity. However, the higher profit cannot make up the increased reputation loss. Thus, the platform should decrease the value of s_1 to reduce reputation loss.

Moreover, Fig. 6 reveals that the optimal strategy of the platform increases if the adversary increases its purchase probability b and decreases its eavesdropping probability $1 - b$. When the adversary increases the purchase probability b, the expected payment to the platform increases. Thus, to earn more profit, the platform increases the value of s_1 as the adversary increases the purchase probability b, as shown in Fig. 6.

5.1.3. Simulation result of Adversary's optimal strategy

The study of the adversary's optimal strategy is shown in Fig. 7. From this figure, we can see that the optimal strategy of the adversary decreases as the user increases the granularity g_1 or the platform increases the value of its reselling strategy s_1. When the user increases the granularity g_1 or the platform increases the value of its reselling strategy s_1, the granularity of reselling data increases correspondingly, resulting in the increasing of data's price. To decrease the payment, the adversary decreases the probability of data purchase (which also increases the probability of eavesdropping) when the user increases the granularity g_1, or the platform increases the value of its reselling strategy s_1 as shown in Fig. 7.

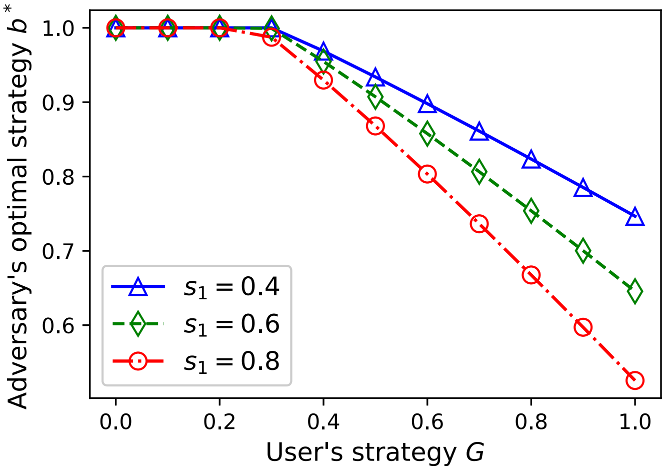

Fig. 7. Optimal strategy of adversary under various G and b_1.

5.2. Interactions among three parties with multiple attributes

According the aforementioned analysis, the theoretical optimal strategies of the user and the platform may not always exist. For computation feasibility, we utilize a parallel machining learning algorithm termed Particle Swarm Optimization (PSO) (Eberhart and Kennedy, 1995), to find the quazi-optimal strategies for the user and the platform, which is performed in the following manner: (i) we run the simulation 100 times; (ii) in each time, each initial strategy and the update vector in each iteration are randomly generated. and (iii) we use the strategy which has the largest utility as the final output. The results derived from PSO are consistent with that in single attribute scenario, thus validating the simulation result. For more details about the implementation of PSO, please refer to (Github).

We use abstracted privacy protection settings from surveys as the inputs of the user and platform. More specifically, based on the privacy survey published by IBM (IBM, 1999) and Data Protection Survey published by SANA (Filkins, 2017), we extract four user's strategies and four platform's strategies. As shown in Table 1, G_r, G_h, G_g, and G_f, are the user's strategies used for Retail applications, Healthcare applications, Government applications, and Financial applications, respectively. Correspondingly, in Table 1, S_r, S_h, S_g, and S_f are the strategies of Retail platforms, Healthcare platforms, Government platforms, and Financial platforms, respectively.

Each extracted strategy contains three data attributes, including income, age, and race. We set the correlation coefficient between income and age as 0.1, the correlation coefficient between income and race as 0.01, and the correlation coefficient between age and race as 0. For the three data attributes, the values of maximum service quality are $q_1 = 60$, $q_2 = 50$, and $q_3 = 40$, the unit privacy costs are $c_1 = 15$, $c_2 = 10$, and $c_3 = 5$, and the unit data prices are $p_1 = 20$, $p_2 = 15$, and $p_3 = 10$. The convention rate λ in Eq. (2) is 0.1. The other parameters in the platform's utility function are: $\theta_p = 15$, $\zeta_p = 0.6$, $l_1 = 5$, $l_2 = 10$, and $c_p = 1$. The adversary has a purchase probability $b = 0.6$ and working efficiency $\varphi = 0.2$. The other parameters in the adversary's utility function are: $\sigma_1 = 1.5$, $\sigma_2 = 1$, $\sigma_3 = 1$, $\theta_a = 15$, and $\zeta_a = 0.6$.

The simulations that follow depict different strategies by varying certain parameters from the perspective of each of the three parties.

5.2.1. Results of User's utility and optimal strategy

The utility and optimal response of the user are investigated through Figs. 8–10 in this subsection.

We analyze user's utility by using real platform's strategy S_r, S_h, S_g, and S_f and increasing the granularity of each attribute from G_0 to G_8 as shown in Table 2. The results of the user's utility are presented in Fig. 8, from which we observe that the utility increases at first and then decreases as the granularity increases. The reason lies in two aspects: (i) when the granularity of each attribute in user's granularity set G

Table 1
Extracted strategies.

Application	Strategy setting of {Income, Age, Race}
Retail	$G_r = \{0.2, 0.3, 0.4\}$, $S_r = \{0.5, 0.7, 0.8\}$
Healthcare	$G_h = \{0.3, 0.4, 0.5\}$, $S_h = \{0.4, 0.6, 0.7\}$
Government	$G_g = \{0.4, 0.5, 0.7\}$, $S_g = \{0.3, 0.5, 0.6\}$
Financial	$G_f = \{0.6, 0.7, 0.8\}$, $S_f = \{0.2, 0.4, 0.5\}$

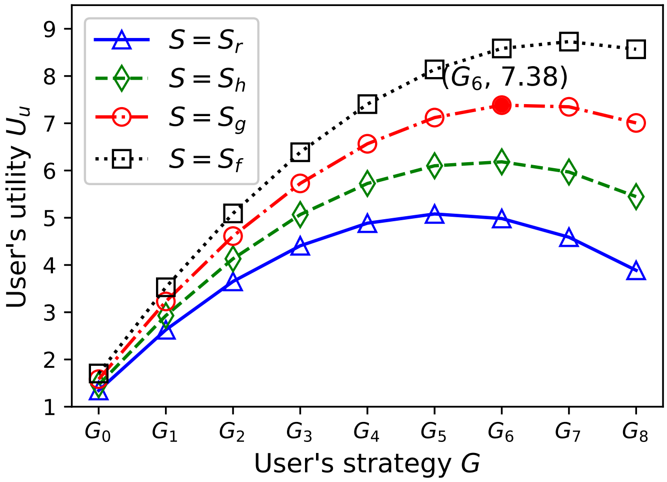

Fig. 8. Utility of user under various G and S.

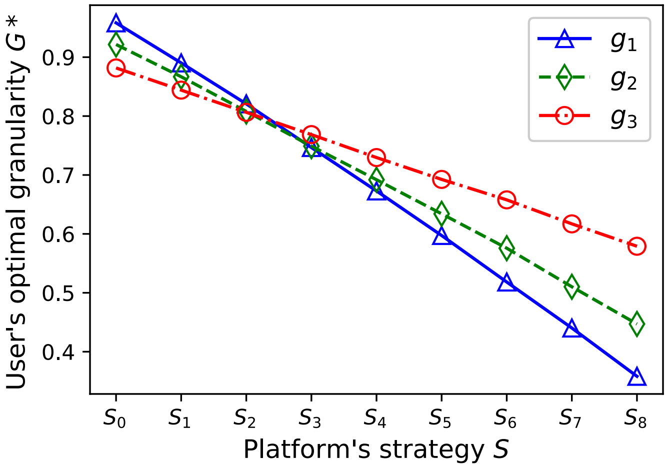

Fig. 9. Optimal strategy of user under various S.

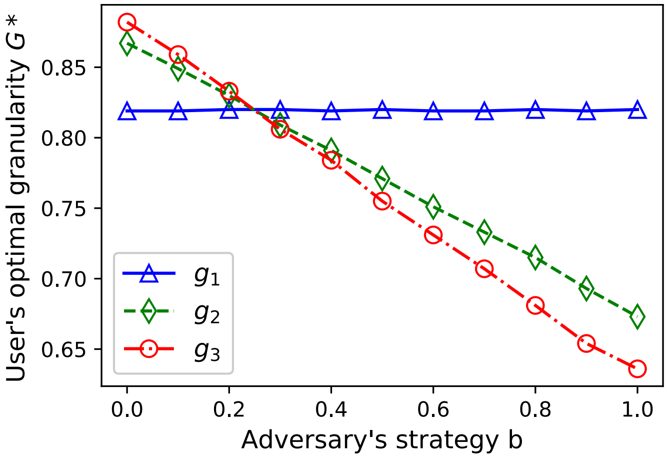

Fig. 10. Optimal strategy of user under various b.

Table 2
Strategy simulation setting.

Notation	Strategy Setting
G_0, S_0	{0.0, 0.1, 0.2}
G_1, S_1	{0.1, 0.2, 0.3}
G_2, S_2	{0.2, 0.3, 0.4}
G_3, S_3	{0.3, 0.4, 0.5}
G_4, S_4	{0.4, 0.5, 0.6}
G_5, S_5	{0.5, 0.6, 0.7}
G_6, S_6	{0.6, 0.7, 0.8}
G_7, S_7	{0.7, 0.8, 0.9}
G_7, S_7	{0.8, 0.9, 1.0}

increases from G_0 to a certain granularity set (e.g., G_6 in line $S = S_g$), the increase rate of privacy cost is smaller than that of received service quality, therefore the user's utility increases; and (ii) when the granularity of each attribute in user's granularity set G continues increasing from such a certain strategy set, the increase rate of privacy cost is larger than that of received service quality, leading to a decrease in the utility. In fact, such a certain granularity set corresponds to the optimal granularity set among granularity sets from G_0 to G_8.

Also, as shown in Fig. 8, the user's utility U_u decreases as the platform increases the value of resale strategy of each attribute from S_r to S_f. This is because when the platform increases the value of resale strategy, the granularity of the resale data increases, enhancing user's privacy cost and reducing user's utility. In particular, the user can gain a larger utility when using Financial Application than other applications because Financial Platform resells user's data with the lowest granularity.

The user's best strategies defending against the platform's different strategies and adversary's different strategies are respectively shown in Figs. 9 and 10, where each line indicates the user's optimal data release granularity for one attribute. The results of Fig. 9 are obtained when the platform uses the strategies $s_1 = 0.2$, $s_2 = 0.4$, and $s_3 = 0.6$. The results of Fig. 10 are obtained when the adversary adopts the data resale strategies $b = 0.2, 0.4, 0.6$.

Fig. 9 shows that the user decreases data release granularity to protect data privacy as the platform increases the data resale granularity from S_0 to S_8. On the other hand, when the value of resale strategy is small (e.g., S_0, S_1, S_2), the user's optimal release granularity of data attribute 1 (corresponding to g_1) is larger than those of attributes 2 and 3 (corresponding to g_2 and g_3, respectively). Since data attribute 1 has the largest service quality value q_1, the user can get more profit from submitting data attribute 1, which can compensate the cost of privacy loss. On the contrary, when the value of resale strategy becomes large (e.g., S_3, S_4, S_5, S_6, S_7, S_8), releasing attribute 1 causes more privacy cost as data attribute 1 has the largest unit privacy cost c_1. As a result, to reduce privacy cost, the user releases less information regarding data attribute 1, i.e., the optimal release granularity of attribute 1 is less than those of other two attributes.

From Fig. 10, we can see that the optimal release granularity g_1 does not change as the adversary changes its strategy, and the optimal release granularities g_2 and g_3 decreases when the adversary increases the data purchase probability and decreases the eavesdropping probability.

When the platform resells attribute 1 (i.e., the line corresponds to g_1), the platform uses resale strategy $s_1 = 0.2$ that is the same as the working efficiency of eavesdropping φ. This means the granularity of purchased data and that of eavesdropped data are equal at the adversary side. Thus, the change of the adversary's strategy will not affect the user's utility. This is the reason why the best release granularity of data attribute 1 does not change when the adversary changes its strategy b under the scenario $s_1 = \varphi$ as we discussed for Fig. 4.

On the other hand, the adversary can get higher data granularities of attribute 2 and 3 (corresponding to g_2 and g_3, respectively) via purchasing rather than eavesdropping because $s_2 > \varphi$ and $s_3 > \varphi$. Thus, the data resale from the platform causes more privacy cost than eavesdropping to the user. As a result, the best data release granularities g_2 and g_3 decreases as the adversary increases the data purchase probability and decreases the eavesdropping probability.

The changes of user's optimal release strategies in Figs. 9 and 10 confirms that data release strategy is also affected by data diversity (e.g., different data attributes have different privacy costs and resale prices).

5.2.2. Results of Platform's utility and optimal strategy

Fig. 11 reports the results of the platform's utility, in which the platform's strategies are set to be S_0 to S_8 according to Table 2 with user's strategy being G_r, G_h, G_g, and G_f as listed in Table 1. From Fig. 11, we can tell that the platform's utility increases at first and then decreases over the increase of the value of user's data release granularity, as we discussed for Fig. 5. When the data resale granularity of each attribute increases from the value in S_0 to the value in a certain resale strategy set (e.g., S_2 in line $G = G_g$), the increase rate of the cost is smaller than that of the profit, resulting in an improvement of utility to the platform; however, when the data resale granularity of each attribute continues increasing from the value in such a certain strategy set, the increase rate of the cost is larger than that of the profit, further reducing the platform's utility. In other words, there is an optimal resale strategy set for the platform to balance the profit of data resale and cost of reputation loss.

Besides, the platform's utility increases as the user increases the release granularity of each attribute from G_r to G_f. This is because the platform can get more accurate data and more resale profit if the user increases the data release granularity. Particularly, the Financial platform can the highest utility because the user submits data with higher granularity to the Financial platform than other three platforms.

The optimal strategy for reselling each data attribute at the platform side are drawn in Fig. 12, from which one can observe that the optimal resale granularity of each attribute decreases when the corresponding data release granularity rises. Notice that as the data release granularity of each attribute increases, the growth rate of reputation cost from data resale becomes larger than the growth rate of the profit from data resale. Therefore, to reduce reputation cost, the platform decreases the resale granularity of each attribute.

Fig. 13 presents the changes of platform's strategy when the adversary enhances the purchase probability. We can see that the optimal resale strategy for each attribute increases as the purchase probability is increased (i.e., the eavesdropping probability is reduced). When the adversary increases the probability of purchase (decreases the probability of eavesdropping), the increase rate of data resale profit is larger than the cost of reputation loss. Thus, to get more profit, the platform increases the data resale granularity.

In Figs. 12 and 13, the value of optimal resale strategy of attribute 1 is higher than that of other attributes, for which the reasons lie in two aspects: (i) more information regarding data attribute 1 is released from the user (see Fig. 10), indicating that less obscured data is available for resale; and (ii) the unit price of attribute 1 is larger than those of other two attributes, implying that more payment can be received by reselling data of attribute 1. These results reflect the fact that a platform may adopt different resale strategies for different data attributes due to data diversity.

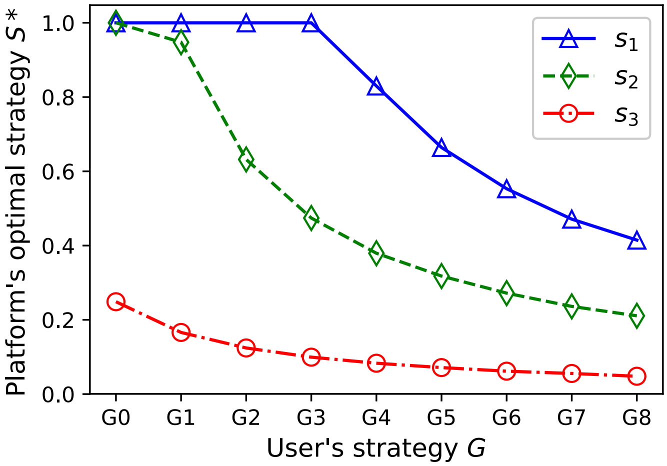

Fig. 12. Optimal strategy of platform under various G.

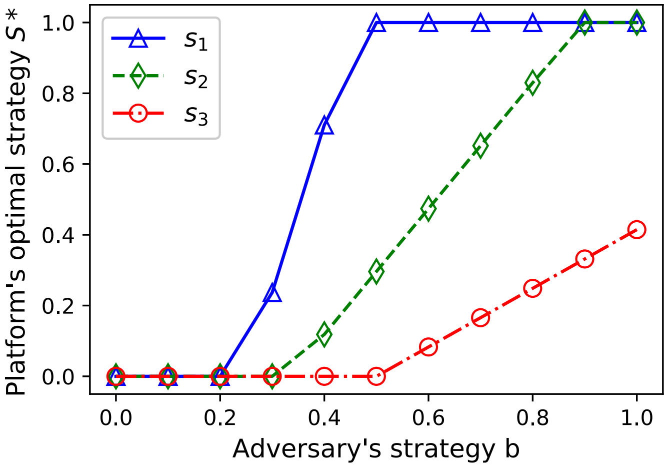

Fig. 13. Optimal strategy of platform under various b.

5.2.3. Simulation result of Adversary's optimal strategy

Fig. 14 shows the adversary's optimal strategy when the user and the platform use the strategies from Table 1. As shown in Fig. 14, the adversary decreases the data purchase probability (i.e., increases the

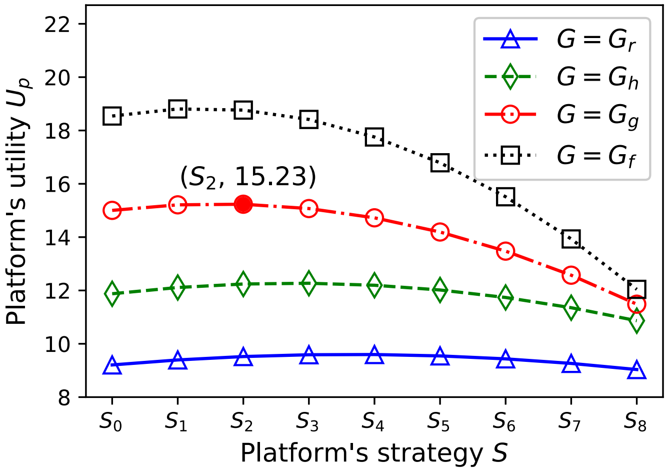

Fig. 11. Utility of platform under various S and G.

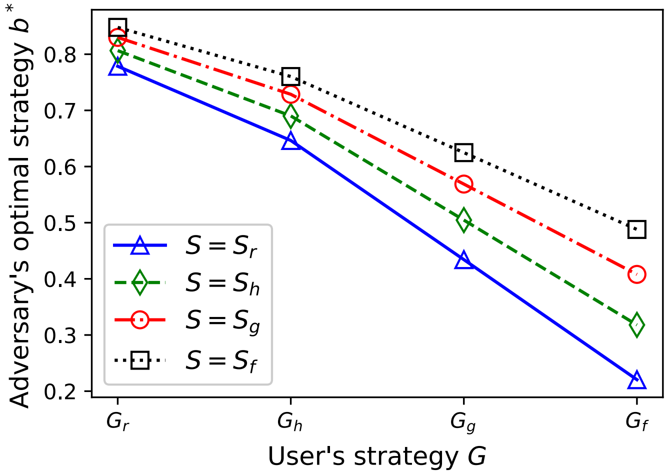

Fig. 14. Optimal strategy of adversary under various G and b.

eavesdropping probability) when the user increases the data release granularity of each attribute in the dataset from G_r to G_f, or the platform increases the data resale strategy of each attribute in the dataset from S_f to S_r. The increase of data release/resale granularity will raise the data price, so the adversary need to decreases the data purchase probability to save cost, which is the same as shown in Fig. 7.

5.2.4. Comparison with two-party game

In this subsection, a comparison between our proposed three-party game and the existing two-party game is performed. According to current research (Chorppath and Alpcan, 2013; Shokri et al., 2012, 2017; Rottondi et al., 2017; Sfar et al., 2017), there are two types of platforms in two-party game: (i) trusted platform that never resells user's data, and (ii) untrusted platform that resells all collected data.

Figs. 15 and 16 show the user's utilities and costs, respectively. On one hand, the user has the highest utility and the lowest cost when the platform is trusted because the potential privacy risk of data resale is ignored and the privacy cost is underestimated. On the other hand, the user has the lowest utility and highest cost when the platform is untrusted because the untrusted platform resells all its collected data and brings more privacy loss to the user. However, the trusted platform is an ideal model, and the untrusted platform, which sells off all data is rare in real life. A more realistic model is a platform that chooses the optimal strategy by balancing the tradeoff between profit and cost. Our three-party game model, captures the actions of this more realistic model.

From the comparison of platform's utility in Fig. 17, it can be seen that a platform can get the highest utility by adaptively reselling user's data; specifically, by adaptively reselling user's data, a platform can get more profit than the trusted platform and suffers less reputation cost than the untrusted platform. This is consistent with the fact that a platform usually owns the ability to adjust its strategy for enhancing profit, further confirming the effectiveness of our proposed game model.

6. Conclusion and future work

This paper studies privacy preservation for context-aware services. To provide realistic optimal strategies for both the user and the platform, we propose a *three-party game model* that captures the interactions between any two of the parties: user, platform and adversary. Interactions include privacy leakage and data resale at the platform side, as well as malicious attacks at the adversary side. Our solution determines an *optimal fine-grained strategy* for the user and platform, so that the user can choose an optimal data granularity to balance service quality and privacy leakage and that the platform can choose the optimal reselling strategy to balance profit and reputation loss. Our model also accounts for the correlations between *multiple data attributes* provided by a user.

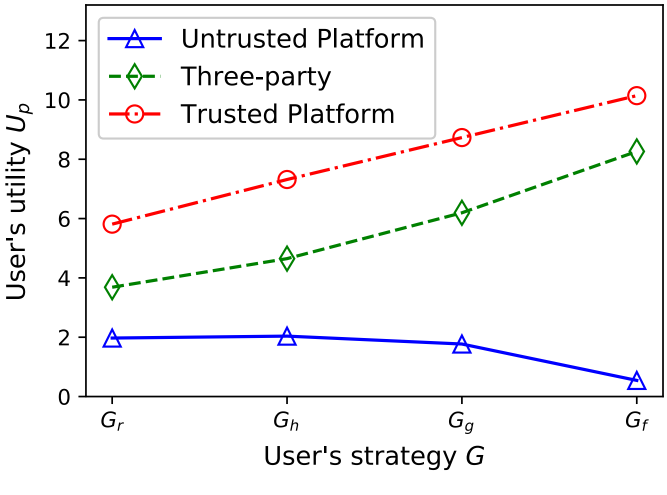

Fig. 15. Utility comparison of user under various platform.

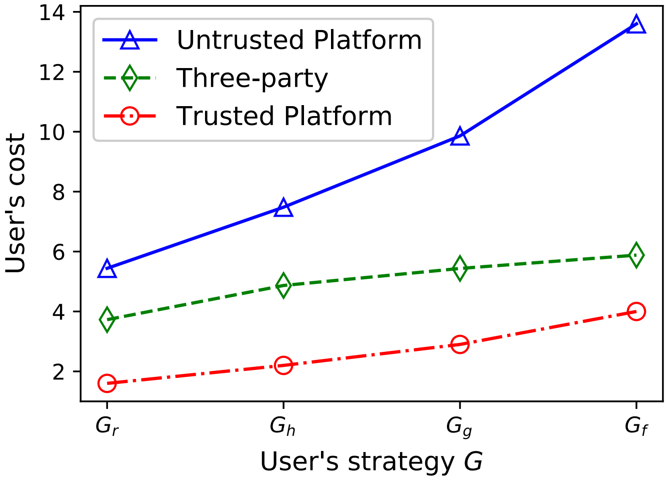

Fig. 16. Cost comparison of user under various platform.

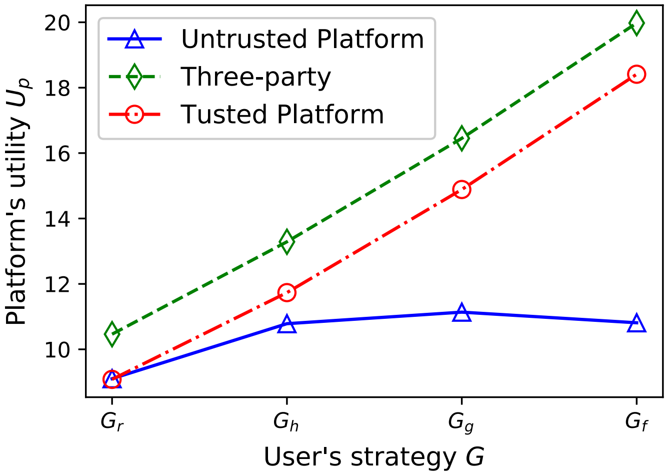

Fig. 17. Utility comparison of various platforms.

To find out the best strategy for each party, we conduct a rigorous theoretical analysis. We also perform simulations using abstracted privacy protection settings from surveys to validate the effectiveness of our proposed game model.

We plan to extend this work to an *m*-user scenario, where the interconnected behaviors of the multiple users, the platform, and the adversary become more complex than the proposed model. This extension will also need to involve another layer of interactions between the users themselves, further complicating the model.

CRediT author statement

Yan Huang: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Writing - Original Draft, Writing - Review & Editing, Visualization. **Wei Li:** Conceptualization, Methodology, Validation, Writing - Review & Editing, Supervision. **Jinbao Wang:** Software. **Zhipeng Cai:** Writing - Review & Editing, Funding acquisition, Supervision. **Anu G. Bourgeois:** Validation, Writing - Review & Editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was partly supported by the National Science Foundation of U.S. (1912753, 1829674, 1704287, 1741277).

References

Armstrong, S., 2016. What happens to data gathered by health and wellness apps? Br. Med. J. 353.

Cai, Z., He, Z., 2019. Trading private range counting over big iot data. In: 2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS), pp. 144–153. https://doi.org/10.1109/ICDCS.2019.00023.

Cai, Z., He, Z., Guan, X., Li, Y., 2018. Collective data-sanitization for preventing sensitive information inference attacks in social networks. IEEE Trans. Dependable Secure Comput. 15 (4), 577–590. https://doi.org/10.1109/TDSC.2016.2613521.

Cheng, B., Wang, M., Lin, X., Chen, J., 2021. Context-aware cognitive qos management for networking video transmission. IEEE/ACM Trans. Netw. 1–13. https://doi.org/10.1109/TNET.2021.3066262.

Chorppath, A.K., Alpcan, T., 2013. Trading privacy with incentives in mobile commerce: a game theoretic approach. Pervasive Mob. Comput. 9 (4), 598–612.

Eberhart, R., Kennedy, J., 1995. A new optimizer using particle swarm theory. In: Micro Machine and Human Science, 1995. MHS '95., Proceedings of the Sixth International Symposium on, pp. 39–43.

Fershtman, C., Kamien, M.I., 1987. Dynamic duopolistic competition with sticky prices. Econometrica 55 (5), 1151–1164.

Filkins, B., Sep. 2017. Sensitive Data at Risk: the Sans 2017 Data Protection Survey.

Freudiger, J., Manshaei, M.H., Hubaux, J.-P., Parkes, D.C., 2013. Non-cooperative location privacy. IEEE Trans. Dependable Secure Comput. 10 (2), 84–98.

https://github.com/chnhuangyan/ThreePartyGameSimulationPSO.

Horn, R.A., Johnson, C.R., 2012. Matrix Analysis, second ed. Cambridge University Press, New York, NY, USA.

Hu, P., Li, H., Fu, H., Cansever, D., Mohapatra, P., 2015. Dynamic defense strategy against advanced persistent threat with insiders. In: Proc. IEEE INFOCOM.

Hussain, R., Kim, D., Son, J., Lee, J., Kerrache, C.A., Benslimane, A., Oh, H., 2018. Secure and privacy-aware incentives-based witness service in social internet of vehicles clouds. IEEE Internet of Things Journal 5 (4), 2441–2448. https://doi.org/10.1109/JIOT.2018.2847249.

IBM, 1999. Ibm Multi-National Consumer Privacy Survey.

Karimi Adl, R., Askari, M., Barker, K., Safavi-Naini, R., 2012. Privacy Consensus in Anonymization Systems via Game Theory. Springer Berlin Heidelberg, pp. 74–89.

Kokolakis, S., 2017. Privacy attitudes and privacy behaviour: a review of current research on the privacy paradox phenomenon. Comput. Secur. 64, 122–134.

Li, W., Hu, C., Song, T., Yu, J., Xing, X., Cai, Z., 2018. Preserving data privacy in context-aware applications through hierarchical game. In: Accepted by IEEE Symposium on Privacy-Aware Computing. Washington DC, USA.

Liu, X., Liu, K., Guo, L., Li, X., Fang, Y., 2013. A game-theoretic approach for achieving k-anonymity in location based services. In: Proc. IEEE INFOCOM.

Ma, R., Xiong, J., Lin, M., Yao, Z., Lin, H., Ye, A., 2017. Privacy protection-oriented mobile crowdsensing analysis based on game theory. In: 2017 IEEE Trustcom/BigDataSE/ICESS, pp. 990–995.

Meeusen, W., van Den Broeck, J., 1977. Efficiency estimation from cobb-douglas production functions with composed error. Int. Econ. Rev. 18 (2), 435–444.

Mohsenian-Rad, A.H., Wong, V.W.S., Jatskevich, J., Schober, R., Leon-Garcia, A., 2010. Autonomous demand-side management based on game-theoretic energy consumption scheduling for the future smart grid. IEEE Transactions on Smart Grid 1 (3), 320–331.

Osborne, M., 2009. Introduction to Game Theory, International Edition. Oxford University Press.

Rottondi, C., Barbato, A., Chen, L., Verticale, G., 2017. Enabling privacy in a distributed game-theoretical scheduling system for domestic appliances. IEEE Transactions on Smart Grid 8 (3), 1220–1230.

Sfar, A., Natalizio, E., Challal, Y., Chtourou, Z., 2017. A game-theoretic approach: a markov game privacy preserving model in retail applications. In: Proc. IEEE MowNet.

Shokri, R., Theodorakopoulos, G., Troncoso, C., Hubaux, J.-P., Le Boudec, J.-Y., 2012. Protecting location privacy: optimal strategy against localization attacks. In: Proc. ACM CCS.

Shokri, R., Theodorakopoulos, G., Troncoso, C., 2017. Privacy games along location traces: a game-theoretic framework for optimizing location privacy. ACM Transactions on Privacy and Security 19 (4), 11–31.

TRUSTe/NCSA Consumer Privacy Infographic – US Edition, 2016 ([online]).

Vakilinia, I., Tosh, D.K., Sengupta, S., 2017. 3-way game model for privacy-preserving cybersecurity information exchange framework. In: MILCOM 2017, pp. 829–834.

Wang, X., Yang, Y., Tang, C., Zeng, Y., He, J., 2016. Droidcontext: identifying malicious mobile privacy leak using context. In: Proc. IEEE Trustcom/BigDataSE/ISPA.

Wang, S., Li, L., Sun, W., Guo, J., Bie, R., Lin, K., 2017a. Context sensing system analysis for privacy preservation based on game theory. Sensors 17 (2), 339.

Wang, S., Huang, J., Li, L., Ma, L., Cheng, X., 2017b. Quantum game analysis of privacy-leakage for application ecosystems. In: Proc. ACM MobiHoc.

Wu, X., Dou, W., Ni, Q., 2017. Game theory based privacy preserving analysis in correlated data publication. In: Proc. ACM ACSW.

Xu, L., Jiang, C., Chen, Y., Ren, Y., Liu, K.J.R., 2015. Privacy or utility in data collection? a contract theoretic approach. IEEE Journal of Selected Topics in Signal Processing 9 (7), 1256–1269.

Xu, L., Jiang, C., Qian, Y., Li, J., Zhao, Y., Ren, Y., 2017. Privacy-accuracy trade-off in differentially-private distributed classification: a game theoretical approach. IEEE Transactions on Big Data, 1–1.

Ying, B., Nayak, A., 2017. Location privacy-protection based on p-destination in mobile social networks: a game theory analysis. In: 2017 IEEE Conference on Dependable and Secure Computing, pp. 243–250.

Zhang, L., Cai, Z., Wang, X., 2016. Fakemask: a novel privacy preserving approach for smartphones. IEEE Transactions on Network and Service Management 13 (2), 335–348.

Zheng, X., Cai, Z., 2020. Privacy-preserved data sharing towards multiple parties in industrial iots. IEEE J. Sel. Area. Commun. 38 (5), 968–979. https://doi.org/10.1109/JSAC.2020.2980802.

Zhu, T., Xiong, P., Li, G., Zhou, W., 2015. Correlated differential privacy: hiding information in non-iid data set. IEEE Trans. Inf. Forensics Secur. 10 (2), 229–242.

Dr. Yan Huang is currently an Assistant Professor in the Department of Software Engineering & Game Development at Kennesaw State University (KSU). Dr. Huang received his Ph.D. degree in the Department of Computing Science at Georgia State University. Before joining GSU, he was the lead of Android Software Development team in the S.F. e-commerce company (Beijing, China). Dr. Huang's research focus on Cyber-Security & Privacy, Federated Learning, and IoT.

Wei Li received the M.S. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 2011, and the Ph.D. degree from George Washington University, Washington, DC, USA, in 2016, both in computer science. She is currently an Assistant Professor with the Department of Computer Science, Georgia State University, Atlanta, GA, USA. Her current research interests include the areas of security and privacy for the Internet of Things and cyber-physical systems, secure and privacy-aware computing, big data, blockchain, game theory, and algorithm design and analysis. Dr. Li was the recipient of the Best Paper Awards in wireless algorithms, systems, and applications (WASA) 2011 and ACM MobiCom cognitive radio architectures for broadband (CRAB) 2013. She is a member of Association for Computing Machinery.

Jinbao Wang received the B.S., M.S., and Ph.D. degrees in database from the School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China, 2006, 2008, 2013, respectively., He is currently an Associate Professor in Database with the School of Computer Science and Technology, Harbin Institute of Technology. His research interests include big data analytic and data privacy.

Zhipeng Cai received his PhD and M.S. degrees in the Department of Computing Science at University of Alberta, and B.S. degree from Beijing Institute of Technology. Dr. Cai is currently an Associate Professor in the Department of Computer Science at Georgia State University. Dr. Cai's research areas focus on Networking, Privacy and Big data. He has published more than 50 journals papers, including more than 20 IEEE/ACM Transactions papers, such as IEEE Transactions on Knowledge and Data Engineering, IEEE Transactions on Dependable and Secure Computing, IEEE/ACM Transactions on Networking, IEEE Transactions on Mobile Computing, etc. Dr. Cai is the recipient of an NSF CAREER Award. He is an editor/guest editor for Algorithmica, Theoretical Computer Science, Journal of Combinatorial Optimization, and IEEE/ACM Transactions on Computational Biology and Bioinformatics. He is a senior member of the IEEE.

Anu G. Bourgeois is an Associate Professor in the Department of Computer Science at Georgia State University. She received her Masters and Ph.D. in Electrical and Computer Engineering from Louisiana State University in 1997 and 2000, respectively. Her research interests include parallel and distributed computing, wireless networks, security and privacy, fault tolerant computing and STEM education. She is a senior member of the IEEE.