Computer Networks 150 (2019) 90–101

Contents lists available at ScienceDirect

Computer Networks

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

Incorporating social interaction into three-party game towards privacy protection in IoT

Kaiyang Li[a], Ling Tian[a,*], Wei Li[b], Guangchun Luo[a], Zhipeng Cai[b]

[a] School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan, China
[b] Department of Computer Science, Georgia State University, Atlanta, GA, USA

ARTICLE INFO

Article history:
Received 26 June 2018
Revised 28 October 2018
Accepted 29 November 2018
Available online 18 December 2018

Keywords:
Internet of Things
Online social network
Data privacy
Social interaction
Three-party game
Nash bargaining

ABSTRACT

By exploiting rich personal information, Internet of Things can provide users with various customized experience and services, improving entertainment, convenience and quality for users' life. However, unavoidably, these users suffer from serious risk of privacy leakage in the presence of untrusted service provider and malicious adversary. Game theory is treated as one of the most promising methodologies to investigate participants' incentive, response, and behaviors and has been widely applied to design privacy preserving schemes. Nevertheless, the complex interactions among users, service provider, and adversary are not fully investigated in the existing work. What's more, users' social connection and interaction are ignored. In this paper, such complex interactions are modeled as a three-party game for the problem of private data trading in IoT with considering user's social interaction in online social network. Particularly, data trading between service provider and adversary is formulated to be a Nash bargaining game, for which Nash bargaining solutions are analyzed via both theoretical analysis and numerical experiments. Our analysis can clearly illustrate data trading strategies between service provider and adversary and offer guidance for designing privacy protection scheme in IoT.

1. Introduction

Internet of Things (IoT) has achieved rapid development and promoted many emerging applications in recent years. With ubiquitous IoT devices, about 2.5 quintillion bytes of data are produced in every day [1]. According to estimation of Ahmed et al. [2], there will be 50 billion IoT devices in 2020. When people use IoT services, a lot of sensitive and personal information is uploaded to the service providers, such as personal profiles, sensors data of mobile terminal, and photos [3]. However, usually, users are not aware of how the service provider use their data and who could access their data clearly. Untrusted service providers may sell user's data to third-party adversaries for extra profit without user's permission. In 2018, Facebook admitted that an App related to Cambridge Analytica to harvest personal data of up to 87 million Facebook users without their consent [4]. And some OSNs, including twitter, Myspace and LiveJournal, also share user's personally identifiable information to third parties [5]. Moreover, the study of Enck et al. [6] found that 15 out of 30 popular network services sent user's information to remote advertisement or analytics servers. *In the presence of untrusted service providers and adversaries, IoT users are suffering from severe risk of privacy leakage.*

As a new kind of IoT, social IoT (SIoT) has become popular [7,8], where the owners of IoT devices are also connected via online social network (OSN). That is, SIoT can be treated as the integration of IoT and OSN. In 2017, there are 2.46 billion social network users around the world, and it is estimated that the number of users will increase to 3.02 billion in 2021 [9]. Since OSNs contain not only various private data but also user's social connection, even a small amount of private data could be used to infer some sensitive information; for example, Facebook Likes can be used to automatically and accurately predict highly sensitive personal attributes, including personality traits, intelligence, happiness, use of addictive substances, parental separation, age, and gender [10]. Moreover, by combining the information in OSN and other sources, adversary could acquire crucial personal data. Prior work [11] found that individual's social security numbers could be predicted by the data from OSN and other open sources. In other words, *privacy leakage in IoT can expand its security hazards through online social networks.*

Meanwhile, at the age of information, users are increasingly concerned about privacy leakage and likely to react to privacy leakage once they realize it. *Better understanding users' response to privacy leakage, their strategies against privacy leakage, and data selling strategies between service provider and adversary can provide help-*

* Corresponding author.
E-mail addresses: kaiyang.li@outlook.com (K. Li), lingtian@uestc.edu.cn (L. Tian), wli28@gsu.edu (W. Li), gcluo@uestc.edu.cn (G. Luo), zcai@gsu.edu (Z. Cai).

https://doi.org/10.1016/j.comnet.2018.11.036
1389-1286/

ful guidelines to design privacy protection mechanism in IoT. A number of schemes have been proposed for private data trading based on game theory. In most of the existing work, only two-party game models are studied [12–16], which can not simultaneously model the complex interactions among users, service provider, and adversary. In [17,18], three-party models are formulated to study more complicated application scenarios. But, these proposed three-party models do not consider user's social connection and interaction, their impacts on user's decision making. Since users are strong connected rather than isolated, users' response and behaviors would be affected by their social connection and interaction, which should be taken into account for improving user's strategy selection. In addition, the potential threat from adversary's attack is ignored, and the data selling strategies between service provider and adversary are simply assumed.

In this paper, we propose a three-party game framework with incorporating user's social connection and interaction, to tackle the following challenges:

- Model complex interactions among users, service provider, and adversary in online social network for the problem of data trading with consideration of privacy leakage.
- Analyze users' response and behaviors when sensing privacy leakage and their impacts on service provider.
- Investigate data trading strategies between service provider and adversary and their received profits.

Our study starts from the proposed game model with taking into account the three parties' different concerns: (1) Users decide whether participate in the OSN or not based on their utility changes due to privacy leakage; (2) service provider operates online social network, collects personal data from users while thinking of whether and how to sell user's data for extra profit; and (3) adversary selects a proper strategy to obtain user's private data, i.e., purchasing data or attacking OSN server. Next, we utilize agent-based model (ABM) to study the network evolution when privacy leakage occurs. Based on our game model, data trading between service provider and adversary is formulated as Nash bargaining problem. Finally, to investigate Nash bargaining solution, rigorous theoretical analysis and comprehensive numerical experiments are well conducted. The major contributions of this paper are addressed below.

- To the best of our knowledge, this is the first work to study the problem of private data trading by incorporating online social network into a three-party game.
- Network evolution is simulated via agent-based model; especially, ABM is used for studying data privacy at the first time.
- Sufficient theoretical proofs and experiment results are presented to analyze Nash bargaining solution for data trading.

The remainder of this paper is organized as follows: Section 2 reviews the related literature. Section 3 introduces preliminaries of our work. The proposed game model and formulation are presented in Section 4. Section 5 describes a simulation for users' response and behaviors based on ABM. In Section 6, the method of calculating Nash bargaining solution and the analysis of Nash bargaining results are proposed. Finally, Section 7 concludes this paper.

2. Related work

2.1. Game theory for private data trading

Game theory has been successfully applied to protect data privacy from multiple perspectives. Most of the existing work focuses on two-party game [12,13,15,19,20], and only a few studies three-party game [17,18]. In this subsection, we mainly summarize the most related work in the field of three-party game for data privacy protection.

Adl et al. [17] modeled data trading among data providers, a data collector, and a data user within a sequential game framework and used the method of backward induction to explore the game equilibrium, in which the underlying assumption is that there is only single-round interaction between the data user and the data collector. Wang et al. [18] formulated the decision-making process of the three parties, including mobile phone users, context-aware application, and malicious adversary. They respectively used extensive form game and repeated game to analyze two typical scenarios, i.e., single-round interaction and multi-round interactions among three parties.

However, in the above game models, social connection among participants (e.g., data providers and mobile phone users) and its impacts are not considered. As a matter of fact, individuals' decision making is also affected by their social connection with others, such their friends. In this paper, *user's social connection in the OSN is taken into account for a more practical game model.* Besides, the above models do not consider adversary could perform attack for private data. Indeed, individuals' privacy may be leaked via data resale from the service provider and/or attack launched by the adversary. That is, *individuals face joint threat from the service provider and the adversary, which is studied in this paper.*

2.2. Online social network for individual interaction

Individuals could share their information through online social network (OSN), which in return, affects each other's decision and behavior. To better understand individuals' interaction and its impacts, agent-based model has been applied to investigate online social network from different aspects. In [21,22], calculation models were proposed to study knowledge sharing behaviors in e-commerce based ABM. Madey et al. [23] used agent-based model to understand the topology and evolution of SourceForge which is a collaborative social network composed of open source software developers. Walter et al. [24] studied how individuals use their trust relationships to filter the information in online social network. The major superiority of ABM over other modeling techniques can be summarized as follows: (i) Many analytic models usually need strong assumptions to be adapted to mathematical tools, but ABM does not need such assumptions. (ii) ABM simulate real-world systems in a natural way so that individuals' behaviors could be more clearly defined.

In this paper, *we apply ABM to analyze individuals' OSN adoption strategies, individuals' response to privacy leakage in OSN and its influence on data trading.*

3. Preliminary

3.1. System model

As illustrated in Fig. 1, the data trade system consists of users, a service provider, and an adversary. The users interact with each other via an OSN. With respect to any user, more friends and/or more involved users in OSN indicate that it becomes more easily for this user to connect with friends, expand her connections, and find enjoyment through interactions within OSN, which brings a higher utility to this user. To receive the requested services, the users connected by the OSN send their IoT data to the service provider continually. However, the service provider is not always trustworthy and may sell the user's data to the adversary without the users' permission for extra profits, which has been confirmed by prior research [4–6]. More specifically, when selling the user's data, the service provider has the following considerations. On one hand, if she sells too much private data, the users may realize their

Table 1
Notation summary.

Symbol	Definiton
N_{bef}	Number of all users in OSN before privacy leakage
N_{aft}	Number of all users in OSN after privacy leakage
ε	Accuracy degree of traded data
ω	Value of one user's data record to adversary
b	Payment of traded data
c	Cost of attacking OSN server
d_{att}	Adversary's attack decision variable
v	Value of one user's data to service provider
N_i	Number of i-th user's friends
N	Number of users in OSN
U_{add}	Additional utility of using OSN
N_{start}	Number of users realizing privacy leakage by themselves
r_i	Privacy preference of i-th user
k	User's sensitivity to privacy leakage
d_i	Variable of i-th user to indicate realize privacy leakage or not
$P_i(t)$	Probability of i-th user to realize privacy leakage at time t
$l_i(t)$	Number of i-th user's friends who knows privacy leakage at time $t-1$

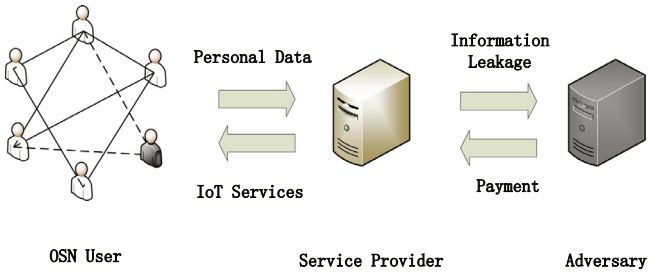


Fig. 1. An example of data trade system.

privacy is leaked and then are likely to leave the OSN, which is indicated by the black user in Fig. 1. A user's departure brings utility reduction to her friends, implying that more users may leave the OSN one after another and that the service provider eventually suffers benefit loss because of such user churn. On the other hand, if the service provider sells little private data, the payment received from the adversary would be very limited. So the service provider needs to strategically determine the amount of private data sold to the adversary for profit maximization. Meanwhile, to balance the tradeoff between the profit from the purchased private data and the payment paid to the service provider, the adversary bargains with the server provider for the accuracy and the price of data. If such data trading benefits both sides, they make a deal; otherwise, according to the attack cost and the data value, the adversary decides to attack the server or not.

For a better presentation, a summary of the notations is shown in Table 1. In this paper, $\varepsilon \in [0, 1]$ is used to denote the accuracy of data sold from the service provider. Particularly, (i) if $\varepsilon = 0$, no user's private data is sold; (ii) if $\varepsilon = 1$, all users' private data is sold; and (iii) a larger ε means more user's privacy is leaked from the service provider. Notice that the methods of determining data accuracy and the adoption of privacy protection techniques at the service provider side are out of the scope of this paper.

3.2. Nash bargaining solution

A *bargain problem* is the situation where (1) participators have conflicts of interest to come to an agreement, (2) they possibly conclude an agreement which benefits both sides, and (3) no participator can be compelled to reach an agreement [25]. In the one-by-one bargaining problem, we assume that each player i has his own preference, represented by a utility function u_i over $X \cup D$, in which X is the set of all the possible outcomes of bargaining and D is the set of disagreement outcomes. Then the possible pairs of utility function are as follows:

$$U = \{(v_1, v_2) \mid v_1 = u_1(x), v_2 = u_2(x) \text{ and } x \in X\} \text{ and } d = (u_1(D), u_2(D)).$$

A pair (U, d) is called *a bargaining solution*, where $U \subset R^2$ and $d \in U$. According to Nash Jr [26], if a bargaining outcome simultaneously satisfies four properties, including Pareto efficiency, symmetry, invariance to affine transformations, and independence of irrelevant alternatives, such outcome is called *Nash bargaining solution* that is unique and formally defined in Definition 1.

Definition 1. A pair of payoffs (v_1, v_2) is a Nash bargaining solution if it solves the following optimization problem:

$$\begin{aligned} \max_{(v_1,v_2)} \quad & (v_1 - d_1)(v_2 - d_2), \\ \text{s.t.} \quad & (v_1, v_2) \in U, \\ & (v_1, v_2) > (d_1, d_2). \end{aligned} \tag{1}$$

4. Game model and formulation

In this section, the complex interactions among the users, the service provider, and the adversary are formulated as a three-party game.

4.1. User model

It is shown that an individual intends to use an OSN once it contains a significant number of participators [27]. If such a number is too small or too large, the impact of one user's participation and departure on the other remaining participators' utilities is very limited and even can be ignored. Accordingly, each user's benefit coming from the number of users in the OSN can be estimated via the following Sigmoid function.

$$f_1(N) = (1 + e^{-\alpha_1(N-\beta_1)})^{-1}, \tag{2}$$

where α_1 determines the steepness of the function and β_1 is the x-coordinate of the symmetric point of the function.

The number of a user's friends in an OSN is another important factor affecting people's decision to adopt the OSN, because the main purpose of most people to join an OSN is communicating with friends easily [28]. For each user, the benefit brought by the number of her friends has a marginal decreasing effect; that is, as the number of her friends in the OSN increases, the benefit of a new friend's participation decreases. Formally, each user i's benefit from her friends in the OSN can be computed as follows:

$$f_2(N_i) = \frac{2}{1 + e^{-\alpha_2 N_i}} - 1, \tag{3}$$

in which α_2 represents the steepness of the function.

In addition, people's desire of using an OSN is also influenced by additional services from the OSN [29]. Many OSNs provide additional services that are not related to the social function, such as querying stock price service and news feeds. However, OSNs also consume user's resources; for instance, APPs cost electric energy and storage space of mobile phones. Suppose that in OSNs, each user can obtain a utility, denoted by U_{add}, consisting of the benefit generated from additional services and the cost of using these additional services.

With respect to privacy leakage from the OSN, the users can be classified into two types: (1) The users who know their privacy is leaked; and (2) the users who don't know that. Formally, a binary variable d_i is used to imply the type of a user. More specifically, if i-th user realizes her privacy leakage, $d_i = 1$; otherwise, $d_i = 0$. Since the preference for preserving data privacy varies in person, we denote $r_i \in [0, 1]$ as privacy preference degree of i-th user to indicate the weight of privacy in i-th user's utility function, which will be investigated in the next section. In conclusion, i-th user's utility function is formulated as:

$$U_i(N, N_i, \varepsilon) = \theta_1 f_1(N) + \theta_2 f_2(N_i) + U_{add} - \theta_3 d_i r_i \varepsilon, \tag{4}$$

where θ_1, θ_2 and θ_3 are the weight parameters. If $U_i(N, N_i, \varepsilon) \geq 0$, i-th user continues to stay in the OSN; otherwise, she leaves the OSN.

4.2. Service provider model

A service provider is an entity who provides OSN service to users while collecting users' personal data. In reality, a service provider is not always trusted and may sell users' data to a third party (i.e., the adversary) for extra profits. The service provider's profit includes the benefit from operating the OSN and the income from selling private data. The income of data sales can be increased if (i) the number of users in the OSN is larger; and/or (ii) the accuracy degree of traded data is higher. However, as the accuracy degree of traded data increases, the number of users in the OSN may be reduced, because it becomes more easily for more users to learn privacy leakage. Thus, to maximize the received utility, the service provider should balance the trade-off between the income of data sales and the benefit loss of user churn. As aforementioned, the adversary may either purchase user's data from the service provider or launch attacks to retrieve user's data. Correspondingly, the service provider faces two possible scenarios.

Scenario 1: Adversary purchases data. Let v denote the average value of users to the service provider and $N_{aft}(\varepsilon)$ be the number of users in the OSN after the service provider sells their data with an accuracy degree ε. In this scenario, the service provider's benefit from users is $vN_{aft}(\varepsilon)$ and the income from data trading is b. Accordingly, her utility can be computed as

$$U_{s1}(b, \varepsilon) = vN_{aft}(\varepsilon) + b. \tag{5}$$

Scenario 2: Adversary launches attack. Suppose that the adversary can obtain all the accurate data when launching attack towards the OSN, which indicates the privacy leakage in the worst case. In this scenario, the number of users in the OSN after privacy leakage is $N_{aft}(1)$, and the service provider's utility is calculated as follows.

$$U_{s2}(b, \varepsilon) = vN_{aft}(1). \tag{6}$$

By combining the above two scenarios, the service provider's expected utility is obtained from Eq. (7).

$$U_s(d_{att}, b, \varepsilon) = (1 - d_{att})(b + vN_{aft}(\varepsilon)) + d_{att} vN_{aft}(1), \tag{7}$$

in which $d_{att} \in \{0, 1\}$ represents the adversary's decision variable. If the adversary decides to attack the OSN, $d_{att} = 1$; otherwise, $d_{att} = 0$.

4.3. Adversary model

An adversary is an entity interested in obtaining personal information for his purposes, such as analyzing people's behavior pattern and pushing individual advertisement. The adversary can obtain the OSN user's data by either purchasing it from the service provider or attacking the OSN. When trading data with the service provider, the adversary would like to get accurate data with less money. But, a higher data accuracy degree usually means a larger payment. Therefore, to receive the maximum utility, the adversary tries to bargain with the service provider regarding data accuracy degree as well as the data price. Meanwhile, the adversary possesses capability attack the OSN server for all the accurate data. If the attack cost is smaller than the data value, the adversary would launch attacks once the bargaining breaks down; otherwise, the adversary does not perform any attack. According to the adversary's strategies, there are two scenarios.

Scenario 1: Adversary purchases data. The adversary purchases user' data from the service provider. In this paper, we assume that the service provider just need to consider whether sell all users' data or not, and the service provider's decision on the amount of data for sales will be studied in our future work. Let ω denote the monetary value of each accurate data record to the adversary, and the total value of accurate data is ωN_{bef}. Moreover, with considering the accuracy degree, ε, of data sold from the service provider, the data value is expressed to be $\omega\varepsilon N_{bef}$. Thus, the adversary's received utility is computed as

$$U_{a1}(b, \varepsilon) = \omega\varepsilon N_{bef} - b, \tag{8}$$

where b is the data price paid from the adversary to the service provider.

Scenario 2: Adversary launches attack. The adversary attacks the OSN server for users' data. Because the adversary can get all the fully accurate data, her profit is ωN_{bef} and can obtain the following utility.

$$U_{a2}(b, \varepsilon) = \omega N_{bef} - c, \tag{9}$$

in which c represents the attack cost.

To sum up, the adversary's expected utility is

$$U_a(d_{att}, b, \varepsilon) = d_{att}(\omega N_{bef} - c) + (1 - d_{att})(\omega\varepsilon N_{bef} - b). \tag{10}$$

5. User's response to privacy leakage

Based on the above game model, in this section, we generate an Agent-based Model to simulate the network evolution with considering privacy leakage. Agent-based model could help us to understand complex and dynamics systems [30] as it describes the autonomy and interaction of systems and brings micro-level results to macro-level conclusions. Then, we deeply investigate the impacts of network parameters on the number of users in the OSN.

5.1. Information diffusion model

When the adversary attacks the users using their personal information, the users may be able to aware of privacy leakage based on the attack activities [31]. For example, if a user frequently receives advertisements closely related to her private information in the OSN, she may believe that her personal data is leaked from the OSN. Let N_{start} be the number of users who initially realize privacy leakage by themselves, and the value of N_{start} can be determined as follows.

$$N_{start}(\varepsilon) = \left\lfloor \frac{e^{k\varepsilon} - 1}{e^{\psi}} N_{bef} \right\rfloor, \tag{11}$$

where k is user's sensitivity degree about privacy leakage, i.e. how sensitive a user could be aware of privacy leakage. And e^{ψ} is used

to normalized the value of $\frac{e^{k\varepsilon}-1}{e^{\psi}}$ into the range [0, 1]. Moreover, $\frac{e^{k\varepsilon}-1}{e^{\psi}}$ represents the probability that a user learns privacy leakage by herself with data accuracy degree ε, and this probability grows at an exponential rate with the increase of data accuracy degree.Then, the information about privacy leakage diffuses from these initial N_{start} users along the connections within the OSN based on the weighted cascade model [32]. The information spreads and the network structure correspondingly changes once in every unit time slot.

In the OSN, if i-th user does not know privacy leakage and her friend, j-th user, knows this in round $t-1$, the probability that i-th user begins to learn privacy leakage through j-th user in round t is $1/N_i$ where N_i is the number of i-th user's friends. In the weighted cascade model, the probability that different users persuade a common friend is independent, and a user only attempts to persuade her friends in the next round after knowing privacy leakage. So, the probability that i-th user starts to realize privacy leakage in the round t is

$$P_i(t) = 1 - \left(1 - \frac{1}{N_i}\right)^{l_i(t)}. \tag{12}$$

5.2. Simulation setting

As social network is a type of scale-free network [33], Barabsi–Albert (BA) model can be adopted to generate a scale-free network for our simulations. When using the BA model to generate a scale-free network containing $N = 3000$ users, we first randomly connect $M_0 = 10$ initial nodes and then add the remaining $N - M_0$ nodes to network one by one; especially, the probability of each new node connecting to the nodes existing in the network is proportional to the degree of the existing nodes.

In this paper, we investigate three typical kinds of distribution of r_i. (1) The first one is Beta distribution with shape parameters (0.149, 0.109), which is considered the closest to the actual survey results with respect to user's privacy preference [34]. (2) The second one is Gaussian distribution with mean value being 0.5 and standard deviation being 0.1938. Because r_i is in [0, 1], the generated r_i will be discard if $r_i \notin [0, 1]$. (3) The last one is uniform distribution over the interval [0, 1].

In the simulation experiments, the network parameters are set as follows. According to the behavioral research [35], we set $\theta_1 = 0.122$ and $\theta_2 = 0.5036$. The remaining setting is: $\theta_3 = 1$, $\alpha_1 = 0.003$, $\beta_1 = 1500$, $\alpha_2 = 0.6$, $\lambda = 0.2$, $\psi = 8$, and $U_{add} = -0.12$

At the beginning of the simulation, all the users are in the OSN. Then, as time goes by, each user has an opportunity to make decision (i.e., stay or leave) in each time slot. Thus, each user's utility changes correspondingly in each time slot and will leave the OSN when her received utility is less than 0.

5.3. Simulation results

In each scenario, to mitigate the impact of random parameters, N_{aft} is the average of 100 experiment results.

Impact of accuracy degree of traded data. To examine the impact of data accuracy on N_{aft}, ε is increased from 0 to 1 with an interval of 0.05, and the remaining parameters are fixed. From Fig. 2, as ε increases from 0 to 1, N_{aft} is gradually reduced from N_{bef} to a constant value. In particular, when ε is small enough, almost no user realizes privacy leakage, and thus all the users stay in the OSN; but, if ε is big enough, almost all the users are able to aware of privacy leakage and then leave the OSN, which is indicated via Fig. 2(b) and (c).

Impact of user's sensitivity. The mean values of N_{aft} under different scenarios are presented in Fig. 2. The results of Fig. 2 indicate that more users decide to leave the OSN when k becomes bigger. This is because with a higher sensitivity degree, it is much easier for the users to sense privacy leakage and then leave the OSN.

Impact of distribution of r_i. Fig. 2 also shows the impact of different distributions of r_i. In Fig. 2(a) where r_i follows Beta distribution, N_{aft} is finally decreased to about 500. When r_i is drawn from Beta distribution with shape parameters (0.149, 0.109), the probability of r_i close to 0 or 1 is big and that of r_i around 0.5 is small. This means a lot of users do not care about their data privacy, so these user still stay in the OSN even if they realize privacy leakage. While, in Fig. 2(b) and (c) where r_i is respectively generated from Gaussian distribution and uniform distribution, the mean values of N_{aft} are reduced to 0, i.e., all the users leave the OSN.

In this paper, we assume N_{start} users who realize privacy leakage by themselves are the information dissemination sources and set the values of N_{start} randomly in each round. Under Beta distribution with shape parameters (0.149, 0.109), most of user's sensitivity degrees are around 0 or 1. Regardless of the selection of N_{start}, the users whose r_i is close to 0 stay in the OSN with a high probability, but the users whose r_i is close to 1 are highly likely to leave the OSN. While under Gaussian distribution and uniform distribution, most of user's sensitivity degrees, r_i, are around 0.5, so the user's departure is severely affected by the selection of N_{start}. In other words, Gaussian distribution and uniform distribution make user's departure decision more random than Beta distribution, resulting in a large standard deviation in N_{aft} as shown in Fig. 3.

Impact of weight of privacy leakage cost. In this scenario, to check the influence of θ_3, the value of k is fixed at 6, and the results are presented in Fig. 4. With a larger θ_3, the cost of privacy leakage is higher, the corresponding user utility is lower, and thus the users are more likely to leave the OSN.

6. Game analysis based on Nash bargaining

In this section, we investigate the bargaining interaction between the service provider and the adversary and their bargaining outcomes.

6.1. Theoretical analysis of Nash bargaining problem

Recall that the Nash bargaining result can be obtained by solving the optimization problem Eq. (1). Specifically, the adversary would take action according to the following two scenarios. (1) **Scenario A: High attack cost.** The adversary's attack cost is more than the value of all the accurate data, i.e., $c \geq \omega N_{bef}$. In this scenario, the adversary will not attack the OSN server if the bargaining fails. (2) **Scenario B: Low attack cost.** The attack cost is less than the value of all the accurate data; that is, $c < \omega N_{bef}$. Therefore, the adversary will attack the OSN server if the bargaining fails. In this subsection, the Nash bargaining problem of data trading between the service provider and the adversary are solved according to the above two scenarios.

Scenario A: The adversary will not attack the OSN server after the bargaining breaks down. Consequently, the Nash bargaining problem can be formulated via Eq. (13a).

$$\max_{(b,\varepsilon)} \quad (U_a(0,b,\varepsilon) - U_a(0,0,0))(U_s(0,b,\varepsilon) - U_s(0,0,0)), \tag{13a}$$

$$s.t \quad U_a(0,b,\varepsilon) \geq U_a(0,0,0), \tag{13b}$$

$$U_s(0,b,\varepsilon) \geq U_s(0,0,0), \tag{13c}$$

$$0 \leq b \leq c, \tag{13d}$$

$$0 \leq \varepsilon \leq 1. \tag{13e}$$

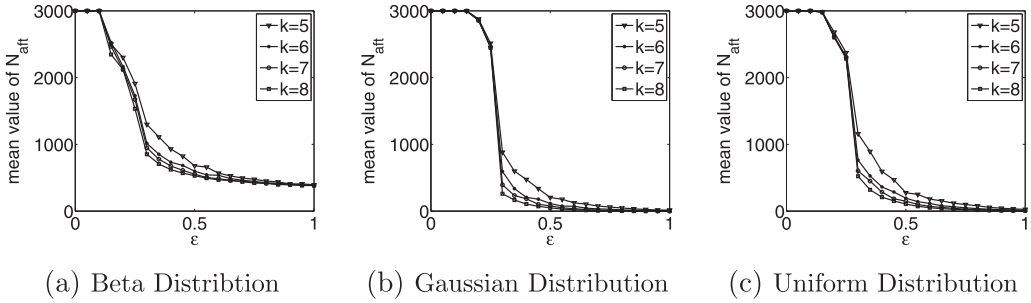


Fig. 2. Relationship between ε and mean value of N_{aft} with different k.

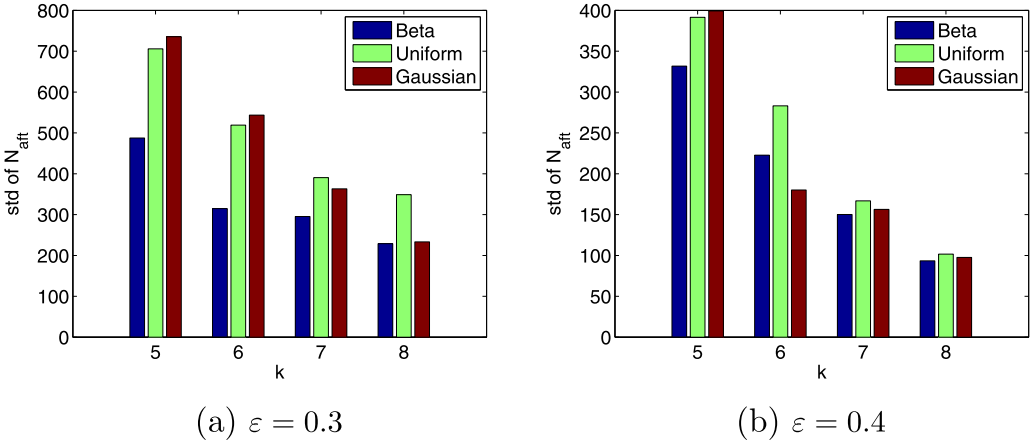


Fig. 3. Standard deviation of N_{aft} with different k.

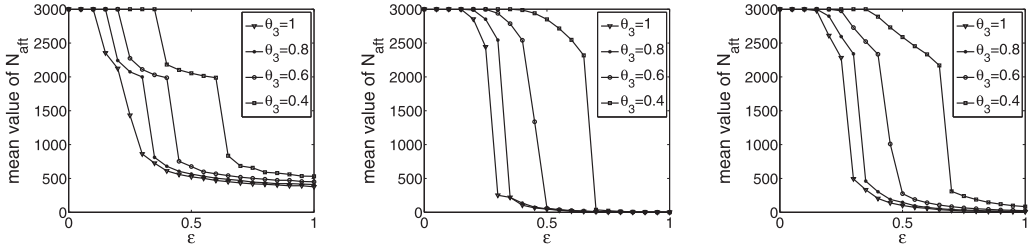


Fig. 4. Relationship between ε and mean value of N_{aft} with different θ_3.

By respectively substituting Eqs. (10) and (7) for U_a and U_s, Eq. (13a) can be equivalently rewritten as

$$\max_{(b,\varepsilon)} \quad (\omega\varepsilon N_{bef} - b)(b + v(N_{aft}(\varepsilon) - N_{bef})), \tag{14a}$$

$$s.t \quad \omega\varepsilon N_{bef} - b \geq 0, \tag{14b}$$

$$b + v(N_{aft}(\varepsilon) - N_{bef}) \geq 0, \tag{14c}$$

$$0 \leq b \leq c, \tag{14d}$$

$$0 \leq \varepsilon \leq 1. \tag{14e}$$

In this scenario, because $c \geq \omega N_{bef}$ and $0 \leq \varepsilon \leq 1$, we have $c \geq \omega\varepsilon N_{bef}$. Therefore, $b \leq c$ if b satisfies Eq. (14b). In addition, $N_{bef} \geq N_{aft}(\varepsilon)$, thus $v(N_{bef} - N_{aft}(\varepsilon)) \geq 0$; that is, $b \geq 0$ if b satisfies Eq. (14c). As a result, we can further simplified the optimization problem Eq. (14) as Eq. (15b) by removing constraint Eq. (14d).

$$\max_{(b,\varepsilon)} \quad f(b,\varepsilon) = (\omega\varepsilon N_{bef} - b)(b + v(N_{aft}(\varepsilon) - N_{bef})), \tag{15a}$$

$$s.t. \quad \omega\varepsilon N_{bef} - b \geq 0, \tag{15b}$$

$$b + v(N_{aft}(\varepsilon) - N_{bef}) \geq 0, \tag{15c}$$

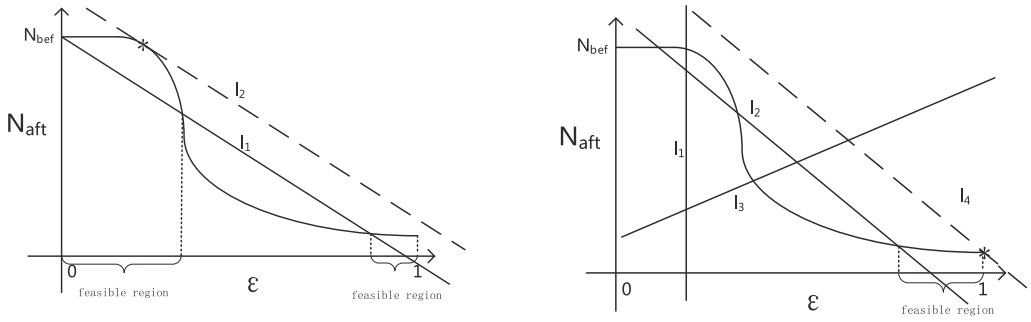
$$0 \leq \varepsilon \leq 1. \tag{15d}$$

Theorem 1. *If $v > 0$, $\omega > 0$, $c \geq \omega N_{bef}$, $N_{aft}(\varepsilon)$ is a non-increasing function and its maximum value is N_{bef} in [0, 1], we can obtain ε^* by solving optimization problemEq. (16b) and compute $b^* = \frac{1}{2}(\omega\varepsilon^* N_{bef} + v(N_{bef} - N_{aft}(\varepsilon^*)))$ forEq. (15b), where (b^*, ε^*) represents an optimal solution toEq. (15b).*

$$\max_{\varepsilon} \quad \frac{1}{4}(\omega\varepsilon N_{bef} - v(N_{bef} - N_{aft}(\varepsilon)))^2, \tag{16a}$$

$$N_{aft}(\varepsilon) \geq N_{bef} - \frac{\omega\varepsilon N_{bef}}{v}, \tag{16b}$$

$$0 \leq \varepsilon \leq 1. \tag{16c}$$



(a) Scenario A: high attack cost

(b) Scenario B: low attack cost

Fig. 5. Feasible region of ε^* in Nash bargaining problem.

Proof. According to Eq. (15c) and (15d), ε in any feasible solution of Eq. (16b) must satisfy $\omega\varepsilon N_{bef} \geq v(N_{bef} - N_{aft}(\varepsilon))$, i.e.,

$$N_{aft}(\varepsilon) \geq N_{bef} - \frac{\omega\varepsilon N_{bef}}{v}. \tag{17}$$

Let us consider a constant ε_0. Then, the quadratic function $f(b, \varepsilon_0)$ in Eq. (15b) can be maximized if Eq. (18) holds.

$$b = \frac{1}{2}(\omega\varepsilon_0 N_{bef} + v(N_{bef} - N_{aft}(\varepsilon_0)). \tag{18}$$

Obviously, b in Eq. (18) satisfies the constraints of Eq. (15b). Via substituting Eq. (18) for b in $f(b, \varepsilon_0)$,we obtain

$$f(\varepsilon_0) = \frac{1}{4}(\omega\varepsilon_0 N_{bef} - v(N_{bef} - N_{aft}(\varepsilon_0)))^2. \tag{19}$$

Since $f(\varepsilon_0) = \max_b f(b, \varepsilon_0)$, for a general case, we have $\max_\varepsilon f(\varepsilon) = \max_\varepsilon \max_b f(b, \varepsilon) = \max_{(b,\varepsilon)} f(b, \varepsilon)$. That is, $\max_\varepsilon f(\varepsilon)$ and $\max_{(b,\varepsilon)} f(b, \varepsilon)$ can achieve the same maximum value with same ε^* if the constraints of the two objective functions are equivalent. Moreover, by combining Eqs. (15b), (17) and (16b) is obtained, and the constraints of Eqs. (15b) and those of (16b) are equivalent.

Therefore, Eqs. (15a) and (16a) have the same maximum value at (b^*, ε^*) with $b^* = \frac{1}{2}(\omega\varepsilon^* N_{bef} + v(N_{bef} - N_{aft}(\varepsilon^*)))$. □

With Theorem 1, we can obtain a Nash bargaining solution (b^*, ε^*) in an efficient manner. Furthermore, from Eq. (16c), we have $\omega\varepsilon N_{bef} - v(N_{bef} - N_{aft}(\varepsilon)) > 0$ and thus get Eq. (20a) for computation simplicity.

$$\max_\varepsilon \quad \omega\varepsilon N_{bef} - v(N_{bef} - N_{aft}(\varepsilon)), \tag{20a}$$

$$N_{aft}(\varepsilon) \geq N_{bef} - \frac{\omega\varepsilon N_{bef}}{v}, \tag{20b}$$

$$0 \leq \varepsilon \leq 1. \tag{20c}$$

The expressions of $N_{aft}(\varepsilon)$ might be various in different scenarios, such as first order piecewise functions and second order piecewise functions. Thus, to compute ε^* effectively, we propose a method to solve Eq. (20a) in the following. By setting $\omega\varepsilon N_{bef} - v(N_{bef} - N_{aft}(\varepsilon)) = A$, we can calculate the explicit expression of $N_{aft}(\varepsilon)$ from Eq. (20a), i.e.,

$$N_{aft}(\varepsilon) = -\frac{\omega\varepsilon N_{bef}}{v} + \frac{A + \omega N_{bef}}{v}. \tag{21}$$

Next, Fig. 5(a) is used to illustrate our method, where the black solid curve represents the function $N_{aft}(\varepsilon)$ and the slope-intercept form of line l_1 is $y_1 = -\frac{\omega N_{bef}}{v}x_1 + N_{bef}$. The solution of Eq. (20a) can be identified based on the intersection point(s) of $N_{aft}(\varepsilon)$ and l_1. According to Eq. (20c) and (20b), for any feasible solution ε, its x-coordinate is within [0, 1] and the corresponding value of $N_{aft}(\varepsilon)$ is above line l_1. As shown in Fig. 5(a), the feasible regions of ε are marked on x-axis. To find an optimal solution ε^*, we parallel move line l_1 up along y-axis until the y-intercept of l_1 is maximum while keeping the set of intersection points of $N_{aft}(\varepsilon)$ and l_1 non-empty. When such movement stops (see line l_2 in Fig. 5(a)), the x-coordinate of each intersection point of $N_{aft}(\varepsilon)$ and l_1 is an optimal solution to Eq. (16b).

Scenario B: The adversary chooses to attack the OSN server if the bargaining breaks down, in which the corresponding Nash bargaining problem is formally expressed in Eq. (22a).

$$\max_{(b,\varepsilon)} \quad (U_a(0, b, \varepsilon) - U_a(1, b, \varepsilon))(U_s(0, b, \varepsilon) - U_s(1, b, \varepsilon)), \tag{22a}$$

$$s.t. \quad U_a(0, b, \varepsilon) - U_a(1, b, \varepsilon) \geq 0, \tag{22b}$$

$$U_s(0, b, \varepsilon) - U_s(1, b, \varepsilon) \geq 0, \tag{22c}$$

$$0 \leq b \leq c, \tag{22d}$$

$$0 \leq \varepsilon \leq 1. \tag{22e}$$

Then, Eq. (22a) can be rewritten to be Eq. (23a) by replacing Eqs. (10) and (7) with U_a and U_s, respectively.

$$\max_{(b,\varepsilon)} \quad ((\varepsilon - 1)\omega N_{bef} - b + c)(b + v(N_{aft}(\varepsilon) - N_{aft}(1))), \tag{23a}$$

$$s.t. \quad (\varepsilon - 1)\omega N_{bef} - b + c \geq 0, \tag{23b}$$

$$b + v(N_{aft}(\varepsilon) - N_{aft}(1)) \geq 0, \tag{23c}$$

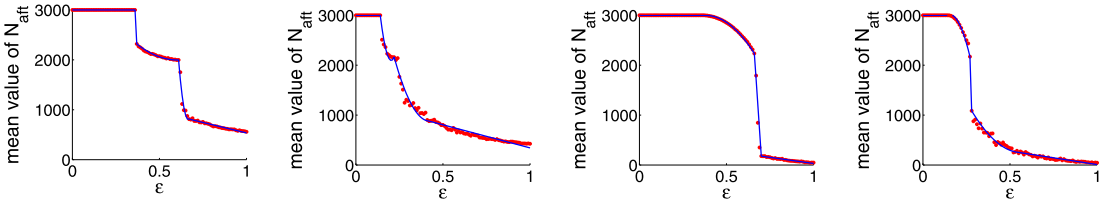
$$0 \leq b \leq c, \tag{23d}$$

$$0 \leq \varepsilon \leq 1. \tag{23e}$$

As $0 \leq \varepsilon \leq 1$, we obtain $c + (\varepsilon - 1)\omega N_{bef} \leq c$. In addition, $N_{aft}(\varepsilon)$ is a non-increasing function and $0 \leq \varepsilon \leq 1$, so we have $v(N_{aft}(1) - N_{aft}(\varepsilon)) \leq 0$, and Eq. (23a) can be equivalently simplified as

$$\max_{(b,\varepsilon)} \quad h(b, \varepsilon) = ((\varepsilon - 1)\omega N_{bef} - b + c)(b + v(N_{aft}(\varepsilon) - N_{aft}(1))), \tag{24a}$$

$$s.t. \quad 0 \leq b \leq c + (\varepsilon - 1)\omega N_{bef}, \tag{24b}$$



(a) $\theta_3 = 0.4$ and r_i follows Beta distribution (b) $\theta_3 = 1$ and r_i follows Beta distribution (c) $\theta_3 = 0.4$ and r_i follows Gaussian distribution (d) $\theta_3 = 1$ and r_i follows Gaussian distribution

Fig. 6. Fitting results of $N_{aft}(\varepsilon)$.

$$0 \le \varepsilon \le 1. \tag{24c}$$

Theorem 2. *If $v > 0$, $\omega > 0$, $0 < c \le \omega N_{bef}$, $N_{aft}(\varepsilon)$ is a non-increasing function and its maximum value is N_{bef} in [0, 1], we can obtain ε^* by solving optimization problemEq. (25a) and calculate $b^* = \frac{1}{2}((\varepsilon^* - 1)N_{bef}\omega + c + v(N_{aft}(1) - N_{aft}(\varepsilon^*))$ forEq. (24a), in which (b^*, ε^*) is an optimal solution toEq. (24a).*

$$\max_{\varepsilon} \quad \frac{1}{4}((\varepsilon - 1)\omega N_{bef} + c - v(N_{aft}(1) - N_{aft}(\varepsilon)))^2, \tag{25a}$$

$$s.t. \quad N_{aft}(\varepsilon) \ge N_{aft}(1) - \frac{c + (\varepsilon - 1)\omega N_{bef}}{v}, \tag{25b}$$

$$N_{aft}(\varepsilon) \le N_{aft}(1) + \frac{c + (\varepsilon - 1)\omega N_{bef}}{v}, \tag{25c}$$

$$1 - \frac{c}{\omega N_{bef}} \le \varepsilon \le 1. \tag{25d}$$

Proof. From Eq. (24b), ε in any feasible solution of Eq. (24a) must meet

$$\varepsilon \ge 1 - \frac{c}{\omega N_{bef}}. \tag{26}$$

Then, by fixing ε at a constant value ε_0, $h(b, \varepsilon_0)$ becomes a quadratic function of b. According to constraint Eq. (24b), we obtain the values of b to maximize $h(b, \varepsilon_0)$, which is computed in Eq. (27).

$$b = \begin{cases} \hat{b}, & \text{if } \hat{b} \in [0, c + (\varepsilon_0 - 1)\omega N_{bef}]; \\ c + (\varepsilon_0 - 1)\omega N_{bef}, & \text{if } \hat{b} > c + (\varepsilon_0 - 1)\omega N_{bef}; \\ 0, & \text{if } \hat{b} < 0; \end{cases} \tag{27}$$

where $\hat{b} = \frac{1}{2}((\varepsilon_0 - 1)\omega N_{bef} + c + v(N_{aft}(1) - N_{aft}(\varepsilon_0)))$.

Notice that when $b = 0$ and $b = c + (\varepsilon_0 - 1)\omega N_{bef}$, the corresponding values of $h(b, \varepsilon_0)$ are non-positive. But, there exist a feasible solution $(\frac{c}{2}, 1)$ such that $h(\frac{c}{2}, 1) = \frac{c^2}{4} > 0$. Thus, for any $\varepsilon_0 \in [0, 1]$, $(0, \varepsilon_0)$ and $(c + (\varepsilon_0 - 1)\omega N_{bef}, \varepsilon_0)$ are not optimal solutions. Therefore, we just need to consider the case when $\hat{b} \in [0, c + (\varepsilon_0 - 1)\omega N_{bef}]$. Accordingly, for ε_0, we have the following requirement.

$$N_{aft}(1) - \frac{c + (\varepsilon_0 - 1)\omega N_{bef}}{v} \le N_{aft}(\varepsilon_0) \le N_{aft}(1) + \frac{c + (\varepsilon_0 - 1)\omega N_{bef}}{v}. \tag{28}$$

By substituting $\hat{b}$ for b in $h(b, \varepsilon_0)$, we get:

$$h(\varepsilon_0) = \frac{1}{4}((\varepsilon_0 - 1)\omega N_{bef} + c - v(N_{aft}(1) - N_{aft}(\varepsilon_0)))^2. \tag{29}$$

Because $h(\varepsilon_0) = \max_b h(b, \varepsilon_0)$, for a general case, we have $\max_{\varepsilon} h(\varepsilon) = \max_{\varepsilon} \max_{b} h(b, \varepsilon) = \max_{(b,\varepsilon)} h(b, \varepsilon)$. That is, $\max_{\varepsilon} h(\varepsilon)$ and $\max_{(b,\varepsilon)} h(b, \varepsilon)$ can achieve the same maximum value with same ε^* if the constraints of the two objective functions are equivalent. From Eqs. (24a), (26), and (28), we obtain Eq. (25a), in which the constraints of Eq. (25a) and those of Eq. (24a) are equivalent.

Thus, Eqs. (25a) and (24a) have the same maximum value at (b^*, ε^*) where $b^* = \frac{1}{2}(\omega\varepsilon^* N_{bef} + v(N_{bef} - N_{aft}(\varepsilon^*)))$. □

Theorem 2 indicates that the Nash bargaining problem Eq. (24a) can be solved via a simplified problem Eq. (25a). The approach for solving Eq. (25a) is similar to that for solving Eq. (16b), which is addressed as follows. Since ε satisfies Eq. (25d), $(\varepsilon - 1)\omega N_{bef} + c - v(N_{aft}(1) - N_{aft}(\varepsilon)) > 0$ and Eq. (30a) is obtained for computation simplicity.

$$\max_{\varepsilon} \quad (\varepsilon - 1)\omega N_{bef} + c - v(N_{aft}(1) - N_{aft}(\varepsilon)), \tag{30a}$$

$$s.t. \quad N_{aft}(\varepsilon) \le N_{aft}(1) + \frac{c + (\varepsilon - 1)\omega N_{bef}}{v}, \tag{30b}$$

$$N_{aft}(\varepsilon) \ge N_{aft}(1) - \frac{c + (\varepsilon - 1)\omega N_{bef}}{v}, \tag{30c}$$

$$1 - \frac{c}{\omega N_{bef}} \le \varepsilon \le 1. \tag{30d}$$

Let $(\varepsilon - 1)\omega N_{bef} + c - v(N_{aft}(1) - N_{aft}(\varepsilon)) = B$. Then. we have

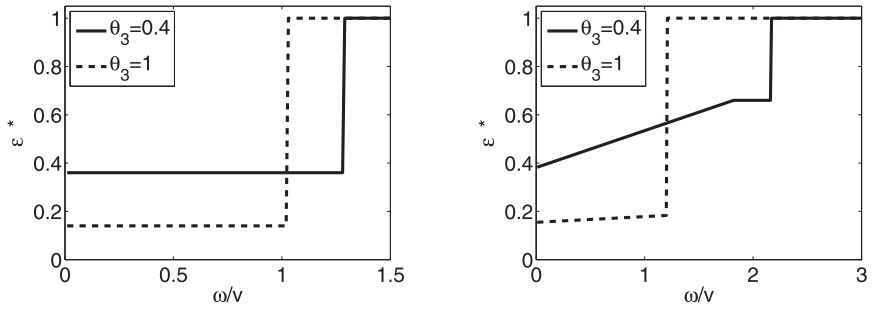
$$N_{aft}(\varepsilon) = -\frac{\omega\varepsilon N_{bef}}{v} + N_{aft}(1) + \frac{B + \omega N_{bef} - c}{v}. \tag{31}$$

An illustrative example is presented in Fig. 5(b), where the black solid curve is the function $N_{aft}(\varepsilon)$. The slope-intercept form of each line is: (1) l_1: $x = 1 - \frac{c}{\omega N_{bef}}$; (2) l_2: $y_2 = -\frac{\omega N_{bef}}{v}x_2 + N_{aft}(1) - \frac{c}{v} + \frac{\omega N_{bef}}{v}$ (i.e., it runs through the point $(1, N_{aft}(1) - \frac{c}{v})$); and (3) l_3: $y_3 = \frac{\omega N_{bef}}{v}x_3 + N_{aft}(1) + \frac{c}{v} - \frac{\omega N_{bef}}{v}$ (i.e., it runs through the point $(1, N_{aft}(1) + \frac{c}{v})$). According to Eq. (25a), for any feasible solution ε, its x-coordinate is within $[1 - \frac{c}{\omega N_{bef}}, 1]$ and the corresponding value of $N_{aft}(\varepsilon)$ is above line l_2 but below line l_3, which is indicated by in Fig. 5(b).

Then, to find an optimal solution ε^*, we parallel move line l_2 up along y-axis until the y-intercept of l_2 is maximum while keeping the set of intersection points of $N_{aft}(\varepsilon)$ and l_2 non-empty. When such movement stops (see line l_4 in Fig. 5(b)), the x-coordinate of each intersection point of $N_{aft}(\varepsilon)$ and l_2 is an optimal solution to Eq. (25a).

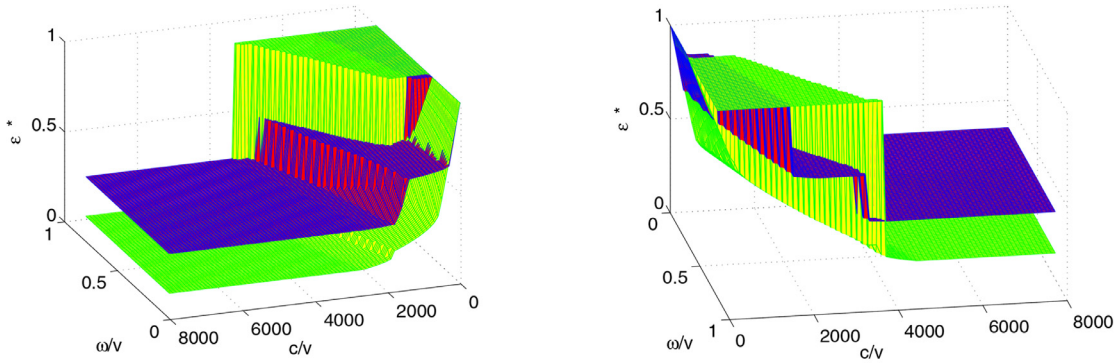
6.2. Numerical analysis of game results

The results of N_{aft} in Section 5.3 can be fitted using a piecewise function that consists of linear and quadratic functions as shown in Fig. 6, where the setting of network parameters is the same as that

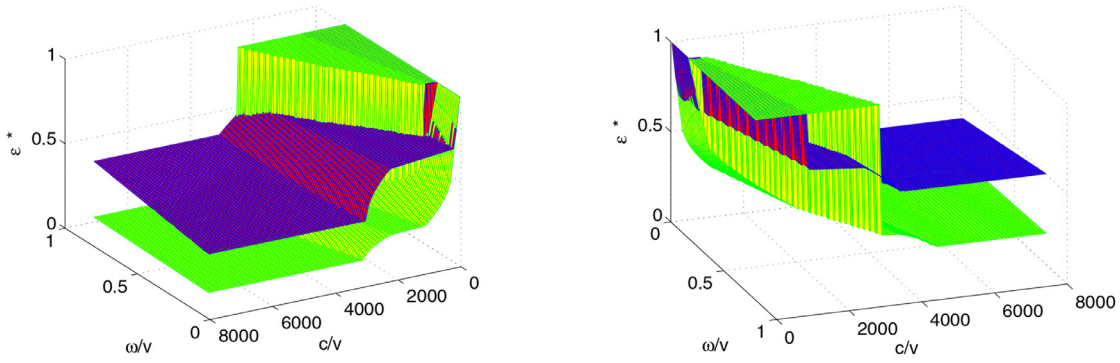


(a) r_i follows Beta distribution (b) r_i follows Gaussian distribution

Fig. 7. Impacts of θ_3 and r_i on Nash bargaining solution in Scenario A.



(a) View 1: r_i follows Beta distribution (b) View 2: r_i follows Beta distribution



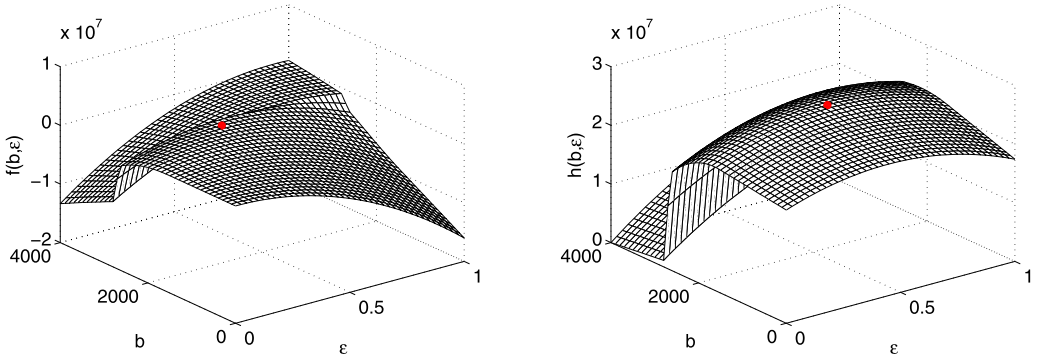
(c) View 1 : r_i follows Gaussian distribution (d) View 2: r_i follows Gaussian distribution

Fig. 8. Impact of θ_3 and r_i on Nash bargaining solution in Scenario B, where blue curved surface is the case when $\theta_3 = 0.4$ and the yellow one is the case when $\theta_3 = 1$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

in Section 5.3 and the fitted results are presented in Appendix A. In this subsection, we calculate ε^* by substituting the fitted equations for $N_{aft}(\varepsilon)$ in Eqs. (16b) and (25a).

First, we study the relationship between ω/v and ε^* under Scenario A, report the results in Fig. 7, and detail our observations as follows.

1. When ω/v changes from 0 to 1, ε^* is always positive. This indicates that user's data is successfully sold from the service provider to the adversary, i.e., such malicious data sales can benefit both the service provider and the adversary.
2. As ω/v increases from 0 to 1, ε^* is increased. A larger ω/v means that user's data has a higher value and brings more profits to the adversary, leading to a higher data price b. Thus, more accurate data is traded to benefit both the service provider and the adversary.
3. There exists a certain value, e.g., 1.02 in Fig. 7(a) and (b). On one hand, when ω/v is smaller than such certain value, the accuracy degree of traded data with $\theta_3 = 0.4$ is higher than that of traded data with $\theta_3 = 1$. Note that θ_3 is the weight of privacy leakage cost at the user side. With a small ω/v, selling user's data to the adversary yields a small profit (i.e., b may be small)



(a) Scenario A (b) Scenario B

Fig. 9. Impacts of b and ε to data trading benefit.

to the service provider but a relatively large cost to the users. Thus, to avoid big decrease in the number of users, the service provider would sell less accurate data if θ_3 is larger, which is consistent with the results of Fig. 6.

On the other hand, when ω/v is bigger than such certain value, the accuracy degree of traded data with $\theta_3 = 0.4$ is lower than that of traded data with $\theta_3 = 1$. With a big ω/v, for the service provider, the payment received from data sales is higher enough to compensate the profit loss brought by user churn. Moreover, when ε and θ_3 increase, Fig. 6 shows that the decrease rate of the number of users is reduced, which means the decrease of the number of users is small even if the service provider sells more accurate data. Thus the growth rate of received payment becomes higher than the growth rate of profit loss. As a result, if θ_3 is bigger, the service provider can earn more by selling more accurate data. This phenomenon indicates that in some situations the users' attention to privacy is not beneficial to their privacy protection, and the panic leaving OSN after privacy leaked may lead to more serious leakage of user's privacy.

Second, we focus on the impacts of ω/v and c/v on ε^* under scenario B and show the experiment results in Fig. 8 with the following analysis.

1. The accuracy degree of traded data, ε^*, reduces when ω/v decreases and c/v increases. The reason is that less accurate data is leaked from the OSN if the value of user's data is smaller to the adversary and/or the attack cost is higher. This indicates that we can protect user's privacy by increasing attack cost at the adversary side and/or enhancing the value of user's data at the service provider side.
2. As shown in Fig. 8, even if c/v is enlarged to 8000 and ω/v approaches 0, ε^* is still more than 0, implying that trading user's private data can benefit the service provider and the adversary.
3. In Fig. 8, the curve surface with $\theta = 1$ is above the curve surface with $\theta = 0.4$ in some regions but such situation changes in some other regions, for which the reasons and inspirations are the same as those for Observation 3 in Fig. 7.

Finally, we look into the impacts of data price b and data accuracy degree ε on the benefits of data trading. From the definition of Nash bargain problem, one can see that the value of the objective function indicates the benefits of the involved players when bargaining successes; that is, in our data trading problem, this implies how data trading benefits the service provider and the adversary.

In Scenario A, the simulation setting is: $\omega = 1.5, v = 1, \theta_3 = 0.4$, and r_i follows Gaussian distribution. Then, to examine the impacts of data price and data accuracy, we calculate $f(b, \varepsilon)$ with b varying in [0,4000] and ε varying in [0, 1] and present the results in Fig. 9(a). As the red point in Fig. 9(a) shows, $f(b, \varepsilon)$ is maximized when $(b, \varepsilon) = (1634.5, 0.61)$, which is consistent with the calculation result of Eq. (20a). In particular, for each fixed b, $f(b, \varepsilon)$ is a convex function of ε, which means there is at least one ε such that $f(b, \varepsilon)$ is maximized.

Similarly, in Scenario B, $\omega = 0.5$, $v = 1$, $c = 1$, $\theta_3 = 0.4$ and r_i follows Gaussian distribution. For all $b \in [0, 4000]$ and $\varepsilon \in [0, 1]$, the values of $h(b, \varepsilon)$ are obtained and shown in Fig. 9(b). When $(b, \varepsilon) = (2139.4, 0.458)$, $h(b, \varepsilon)$ achieves its maximum value indicated by the red point in Fig. 9(b). Particularly, for any fixed b, $h(b, \varepsilon)$ is a convex function of ε, indicating that there exist at least one ε such that $h(b, \varepsilon)$ is maximized.

The relationship between ε and $f(b, \varepsilon)$ (or $h(b, \varepsilon)$) indicates that the service provider needs to deal with the trade-off between the income of data sales and the benefit loss of user churn by selecting an appropriate quality (e.g., accuracy degree) for traded data.

According to the above analyze of experimental results, we can obtain some critical findings for protecting user data privacy, which are detailed as follows:

1. User's personal data is under serious threat and is highly possible sold by service providers, because trading user's private data could always benefit the service provider and the adversary.
2. If the value of one user's data to service provider is larger or the cost of attacking OSN server is higher, user's personal data would be protected better.
3. In some situations, the users' privacy concern may not be beneficial to their privacy protection, and scare leaving OSNs after privacy leakage may lead to more serious leakage of user's privacy. The Observation 3 in both Scenario A and Scenario B shows that the service provider sells a little privacy could incur huge customer churn and thus almost gives up the income from operating the network and mainly earns profit by selling completely accurate data.

7. Conclusion

Privacy leakage is a severe threat to users in IoT and its harm can be expanded through online social networks. This paper formulates the complex interactions among users, service provider and adversary as a three-party game to study data privacy protection with considering user's social connection and interaction. To observe network evolution when privacy is leaked, a simulation is built based on agent-based model. To obtain the Nash bargaining outcomes of data trading problem, thorough theoretical analysis is

performed. Moreover, data trading strategy and its impacts are analyzed through intensive numerical experiments. To the best of our knowledge, this is first time to investigate data privacy via incorporating social connection and interaction into three-party game.

Acknowledgments

This work is supported by the Foundation of Science & Technology Department of Sichuan Province (no. 2017GFW0128, 16FZ0108, 2017JY0027, 2017JY0007, 2017RZ0008, 2017HH0075, 2017JZ0031, 2018HH0075), and by the Sichuan Provincial Economic and Information Commission (no. 2018DS010). And this work is partly supported by the National Science Foundation (NSF) under grant no. 1252292, 1741277, 1704287, 1829674.

Appendix A. Expression of fitting function

(1) Fitting function, $N_{aft}(\varepsilon)$, when $\theta_3 = 0.4$ and r_i follows Beta distribution:

$$N_{aft}(\varepsilon) = \begin{cases} 3000 & 0 \le \varepsilon < 0.36 \\ -68211\varepsilon + 27555.96 & 0.36 \le \varepsilon < 0.37 \\ 5887.57\varepsilon^2 - 7105.28\varepsilon + 4140.83 & 0.37 \le \varepsilon < 0.61 \\ 390370.71\varepsilon^2 - 519417.01\varepsilon + 173584.81 & 0.61 \le \varepsilon < 0.67 \\ -839.66\varepsilon + 1375.4 & 0.67 \le \varepsilon \le 1 \end{cases}$$

(2) Fitting function, $N_{aft}(\varepsilon)$, when $\theta_3 = 1$ and r_i follows Beta distribution:

$$N_{aft}(\varepsilon) = \begin{cases} 3000 & 0 \le \varepsilon < 0.14 \\ 227619.02\varepsilon^2 - 92445.85\varepsilon + 11481.09 & 0.14 \le \varepsilon < 0.22 \\ 31090.16\varepsilon^2 - 26296.55\varepsilon + 6440.24 & 0.22 \le \varepsilon < 0.42 \\ -925.95\varepsilon + 1268.89 & 0.42 \le \varepsilon \le 1 \end{cases}$$

(3) Fitting function, $N_{aft}(\varepsilon)$, when $\theta_3 = 0.4$ and r_i follows Gaussian distribution:

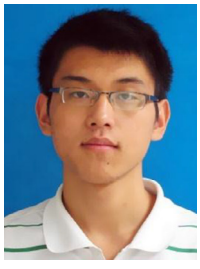
$$N_{aft}(\varepsilon) = \begin{cases} -8.92\varepsilon + 3000 & 0 \le \varepsilon < 0.38 \\ -9800.55\varepsilon^2 - 7474.64\varepsilon + 1571.44 & 0.38 \le \varepsilon < 0.66 \\ -51453\varepsilon + 36194.57 & 0.66 \le \varepsilon < 0.7 \\ -487.35\varepsilon + 518.62 & 0.7 \le \varepsilon \le 1 \end{cases}$$

(4) Fitting function, $N_{aft}(\varepsilon)$, when $\theta_3 = 1$ and r_i follows Gaussian distribution:

$$N_{aft}(\varepsilon) = \begin{cases} 3000 & 0 \le \varepsilon < 0.14 \\ -63315.81\varepsilon^2 + 19575.63\varepsilon 1500.4 & 0.14 \le \varepsilon < 0.27 \\ -108374\varepsilon + 31431.08 & 0.27 \le \varepsilon < 0.28 \\ 10107.07\varepsilon^2 - 11481.41\varepsilon + 3508.76 & 0.28 \le \varepsilon < 0.53 \\ -539.43\varepsilon + 548.59 & 0.53 \le \varepsilon \le 1 \end{cases}$$

References

[1] L. F, Y. W, A. D, et al, A survey on big data market: pricing, trading and protection, IEEE Access 6 (2018) 15132–15154.
[2] E. Ahmed, I. Yaqoob, I.A.T. Hashem, I. Khan, A.I.A. Ahmed, M. Imran, A.V. Vasilakos, The role of big data analytics in internet of things, Comput. Netw. 129 (2017) 459–471.
[3] P. Porambage, M. Ylianttila, C. Schmitt, P. Kumar, A. Gurtov, A.V. Vasilakos, The quest for privacy in the internet of things, IEEE Cloud Comput. (2) (2016) 36–45.
[4] Facebook says cambridge analytica harvested data of up to 87 million users, 2018, https://www.nytimes.com/2018/04/04/technology/mark-zuckerberg-testify-congress.html.
[5] B. Krishnamurthy, C.E. Wills, On the leakage of personally identifiable information via online social networks, in: Proceedings of the 2nd ACM workshop on Online social networks, ACM, 2009, pp. 7–12.
[6] W. Enck, P. Gilbert, B.G. Chun, L.P. Cox, J. Jung, P. Mcdaniel, A.N. Sheth, Taintdroid: an information flow tracking system for real-time privacy monitoring on smartphones, ACM Trans. Comput. Syst. 32 (2) (2010) 1–29.
[7] L. Atzori, A. Iera, G. Morabito, M. Nitti, The social internet of things (siot)–when social networks meet the internet of things: concept, architecture and network characterization, Comput. Netw. 56 (16) (2012) 3594–3608.
[8] A.M. Ortiz, D. Hussein, S. Park, S.N. Han, N. Crespi, The cluster between internet of things and social networks: review and research challenges, IEEE Internet Things J. 1 (3) (2014) 206–215.
[9] Number of social network users worldwide from 2010 to 2021, 2018, https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/.
[10] M. Kosinski, D. Stillwell, T. Graepel, Private traits and attributes are predictable from digital records of human behavior, Proc. Natl. Acad. Sci. 110 (15) (2013) 5802–5805.
[11] A. Acquisti, R. Gross, Predicting social security numbers from public data, Proc. Natl.Acad.Sci. 106 (27) (2009) 10975–10980.
[12] A. Ghosh, K. Ligett, Privacy and coordination: computing on databases with endogenous participation, in: Fourteenth ACM Conference on Electronic Commerce, 2013, pp. 543–560.
[13] A. Ghosh, A. Roth, Selling privacy at auction, in: ACM Conference on Electronic Commerce, 2011, pp. 199–208.
[14] A.K. Chorppath, T. Alpcan, Trading privacy with incentives in mobile commerce: a game theoretic approach, Pervasive Mob. Comput. 9 (4) (2013) 598–612.
[15] Y. Pu, J. Grossklags, An economic model and simulation results of app adoption decisions on networks with interdependent privacy consequences, in: International Conference on Decision and Game Theory for Security, 2014, pp. 246–265.
[16] H. Jin, L. Su, H. Xiao, K. Nahrstedt, Inception: incentivizing privacy-preserving data aggregation for mobile crowd sensing systems, in: Proceedings of the 17th ACM International Symposium on Mobile Ad Hoc Networking and Computing, ACM, 2016, pp. 341–350.
[17] R.K. Adl, M. Askari, K. Barker, R. Safavi-Naini, Privacy consensus in anonymization systems via game theory, in: IFIP Annual Conference on Data and Applications Security and Privacy, Springer, 2012, pp. 74–89.
[18] S. Wang, L. Li, W. Sun, J. Guo, R. Bie, K. Lin, Context sensing system analysis for privacy preservation based on game theory? Sensors 17 (2) (2017) 339.
[19] Y. Wang, Z. Cai, G. Yin, Y. Gao, X. Tong, G. Wu, An incentive mechanism with privacy protection in mobile crowdsourcing systems, Comput. Netw. 102 (2016) 157–171.
[20] Y. Wang, Z. Cai, X. Tong, Y. Gao, G. Yin, Truthful incentive mechanism with location privacy-preserving for mobile crowdsourcing systems, Comput. Netw. 135 (2018) 32–43.
[21] F. Bergenti, E. Franchi, A. Poggi, Agent-based interpretations of classic network models, Comput. Math. Organ Theory 19 (2) (2013) 105–127.
[22] G. Jiang, F. Ma, J. Shang, P.Y.K. Chau, Evolution of knowledge sharing behavior in social commerce: an agent-based computational approach, Inf. Sci. 278 (10) (2014) 250–266.
[23] G.R. Madey, Y. Gao, V. Freeh, R. Tynan, C. Hoffman, Agent-based modeling and simulation of collaborative social networks, in: Americas Conference on Information Systems, Amcis 2003, Tampa, Fl, Usa, August, 2003, p. 237.
[24] F.E. Walter, S. Battiston, F. Schweitzer, A Model of a Trust-based Recommendation System on a Social Network, Kluwer Academic Publishers, 2008.
[25] A. Rubinstein, Perfect equilibrium in a bargaining model., Econometrica 50 (50) (1982) 97–109.
[26] J.F. Nash Jr, The bargaining problem, Econometrica: J. Econom. Soc. (1950) 155–162.
[27] D. Sledgianowski, S. Kulviwat, Using social network sites: the effects of playfulness, critical mass and trust in a hedonic context, Data Processor Better Bus. Edu. 49 (4) (2009) 74–83.
[28] R.K. Baker, K.M. White, Predicting adolescents' use of social networking sites from an extended theory of planned behaviour perspective, Comput. Human Behav. 26 (6) (2010) 1591–1597.
[29] N. Gandal, Hedonic price indexes for spreadsheets and an empirical test for network externalities, RAND J. Econ. (1994) 160–170.
[30] E. Bonabeau, Agent-based modeling: methods and techniques for simulating human systems, Proc. Natl. Acad. Sci. USA 99 (10) (2002) 7280–7287.
[31] W. Wang, Q. Zhang, A stochastic game for privacy preserving context sensing on mobile phone, in: INFOCOM, 2014 Proceedings IEEE, IEEE, 2014, pp. 2328–2336.
[32] W. Chen, Y. Wang, S. Yang, Efficient influence maximization in social networks, in: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 2009, pp. 199–208.
[33] A.-L. Barabási, Scale-free networks: a decade and beyond, Science 325 (5939) (2009) 412–413.
[34] X. Liu, K. Liu, L. Guo, X. Li, Y. Fang, A game-theoretic approach for achieving k-anonymity in location based services, in: INFOCOM, 2013 Proceedings IEEE, IEEE, 2013, pp. 2985–2993.
[35] K.-Y. Lin, H.-P. Lu, Why people use social networking sites: an empirical study integrating network externalities and motivation theory, Comput. Human Behav. 27 (3) (2011) 1152–1161.



Kaiyang Li received the M.S. degree from the School of Energy Science and Engineering University of Electronic Science and Technology of China, Chengdu, China, in 2016. He is currently a Ph.D. student in the School of Computer Science and Engineering, University of Electronic Science and Technology of China. His research interests include game theory and big data.

Ling Tian received the B.S, M.S and Ph.D. degrees from the School of Computer Science and Engineering, University of Electronic Science and Technology of China in 2003, 2006 and 2010, respectively. She is currently a professor in School of Computer Science and Engineering at UESTC. She had been a visiting scholar in Georgia State University during 2013 in United States. Her research interests include big data, signal processing and video coding.

Wei Li currently is an assistant professor in the Department of Computer Science at Georgia State University. Dr. Li received her Ph.D. degree in Computer Science from The George Washington University in 2016, and her M.S. degree in Computer Science from Beijing University of Posts and Telecommunications in 2011. She won the Best Paper Awards in ACM MobiCom Workshop CRAB 2013 and WASA 2011, respectively. Her current research spans the areas of Internet of Things, secure and privacy-aware computing, secure and truthful auctions in dynamic spectrum access, resource management in cognitive radio networks and WiFi-based wireless access networks, game theory, and algorithm design and analysis. She is a member of IEEE and ACM.

Guangchun Luo received the Ph.D. degree in computer science from University of Electronic Science and Technology of China, Chengdu, China, in 2004. He is currently a professor and the Associate Dean of computer science at the UESTC. He has published over sixty journal and conference papers in his fields. His research interests include computer networks and big data.

Zhipeng Cai received his Ph.D. 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.