IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received September 1, 2019, accepted October 19, 2019, date of publication October 24, 2019, date of current version November 7, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2949452

A Mechanism to Improve Effectiveness and Privacy Preservation for Review Publication in LBS

GUANGCAN YANG[1], SHOUSHAN LUO[1], HONGLIANG ZHU[1], YANG XIN[1], KE XIAO[2], YULING CHEN[3], MINGZHEN LI[1], AND YUNFENG WANG[1]

[1]National Engineering Laboratory for Disaster Backup and Recovery, Information Security Center, School of Cyberspace Security, Beijing University of Posts and Telecommunications, Beijing 100876, China
[2]School of Information Technology, North China University of Technology, Beijing 100144, China
[3]State Key Laboratory of Public Big Data, GuiZhou University, Guiyang 550025, China

Corresponding author: Guangcan Yang (yangguangcan@bupt.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFB0802300, in part by the BUPT Excellent Ph.D. Students Foundation under Grant CX2019229, in part by the National statistical scientific research project of China under Grant 2018LY61, in part by the Major Scientific and Technological Special Project of Guizhou Province under Grant 20183001, in part by the Foundation of Guizhou Provincial Key Laboratory of Public Big Data under Grant 2018BDKFJJ008 and Grant 2018BDKFJJ020, and in part by the Applied Sci-Tech Research and Development Special Fund Program of Guangdong Province under Grant 2015B010131007.

ABSTRACT Local business service systems (LBSS), as an essential role of location-based service (LBS), have been gaining tremendous popularity in our daily life. Individuals' reviews in these systems are very important as they not only contribute to building reputations for businesses but also play a guiding role for consumers. However, users' privacy disclosure and the effectiveness of reviews are the urgent problems to be solved for the further development of LBSS. This paper proposes a mechanism to improve effectiveness and privacy preservation for review publication. In users' privacy protection, the mechanism firstly formalizes the model of attackers, then focuses on the identification or inference attack caused by reviews. For improving the effectiveness of reviews, the mechanism introduces users' reputation scores to rank the reviews. We evaluate our mechanism thoroughly by extensive experiments, and the results validate that our mechanism can achieve a better performance.

INDEX TERMS Location-based service (LBS), privacy protection, review publication.

I. INTRODUCTION

Along with the development of latest-generation mobile phones, various types of data (e.g., picture, video, location) can be collected and shared in different applications (Apps) loaded on smartphones [1], [2]. The location-based service (LBS), as an important component of the Mobile Social Networks (MSN), brings remarkable convenience to people's life [3]. For example, people could find businesses or Points of Interests (PoIs) according to the information that has been collected and shared. The crowd-sourced local business service systems (CSLBSSs) such as Yelp, Tripadvisor, Dianping and Facebook, as an important part of LBS, play a vital role in guiding people to choose restaurants and places of entertainment.

The crowd-sourced local business service systems are different from other Apps in LBS, as they are built in a crowd-sourced manner via users' reviews on these businesses or PoIs. People are more likely to choose their entertainment places according to the review lists for businesses or PoIs, especially when they have good reviews or recommendations from most users. From the perspective of businesses, the effectiveness of published reviews provided by users is very important for a business in CSLBSSs, as these reviews are the foundation of the business reputation. From the perspective of users, when a user publishes a review on a business, the visited information such as location is leaked since the location of a business is inherently public information. Some methods, such as pseudo-ID and pseudonym, are usually used to represent users' real identities so that their real identities can not be distinguished, but users' significant privacy may be leaked by

The associate editor coordinating the review of this manuscript and approving it for publication was Malik Najmus Saqib.

identification or inference attack [4], [5]. Therefore, in CSLBSSs, there are two important aspects for review publication: i) How to protect users' privacy caused by published reviews. ii) How to improve the effectiveness of users' published reviews for reputations of businesses.

In CSLBSSs, published reviews may pose some inherent challenges for protecting users' privacy. As we all know, when users' new reviews have been published, the locations of corresponding businesses will be leaked. The location of a business is inherent information, which is usually used to guide interested users. If a user's review on a business is published, it means that he has once visited this location. Moreover, when a user decides to publish his review on a business, he implicitly accepts to reveal the geographical coordinates and the semantic information of the place. For example, when a user's reviews on restaurants that he visited are published, the exact locations and types of these restaurants will be disclosed to other users, including attackers, which might lead to the disclosure of additional private information beyond what he intended to share. Since an adversary could easily read and collect the user's information from the website, the user's information included in his reviews could be analyzed, then the adversary could get more sensitive information such as the user's identity and preference. After that, the adversary may even implement a physical trace attack.

Current researches are incapable of resisting identification or inference attack caused by users' published reviews. For example, Paper [6] proposed a mechanism called LP-Doctor, which was used to prevent users' locations access control according to Apps' functionality. Although this mechanism is useful, it may hinder the publication of users' reviews. To protect users' location information of check-ins, authors in [7] first inferred the motivation for users to share their locations, and then they obfuscated users' location information from the geographical and semantic levels. The paper [8] proposed a mechanism to set the budget of reviews for users in each subregion. The budget was used as a threshold, which was bounded with the size of the subregion to protect users privacy. Nevertheless, the research mechanisms mentioned above do not consider the case where attackers have the ability to collect reviews, and implement an attack to re-identify a target user via the information from the collected reviews. For example, Fig.1 shows three users' location information of reviews in Pittsburgh, PA. We can see that users may have more reviews in a certain region than others, and some reviews even in the same location such as user 1 in region A1 (i.e., dotted box A1). Besides, users may have only one review in a certain region such as user 1 in region A2 (i.e., dotted box A2). The two cases above easily cause users to suffer identification or inference attack which will be explained in detail in the motivation part (i.e., section IV-B).

The effectiveness of users' published reviews is critical to businesses, but how to improve the effectiveness of published reviews is still an urgent problem to be solved. The platforms based on CSLBSSs such as Yelp and Dianping encourage users to review, because good reviews not only help to build the reputations of businesses but provide useful recommendations to more users. However, there are always some fake reviews that can affect the above platforms. The users who leave fake reviews have roughly two purposes: one is to increase the reputation of a business related to his interests by means of positive review fraud; the other is to tarnish the reputation of a business related to his competitors [6]. If fake reviews are accepted by users, it will certainly damage the platforms based on CSLBSSs.

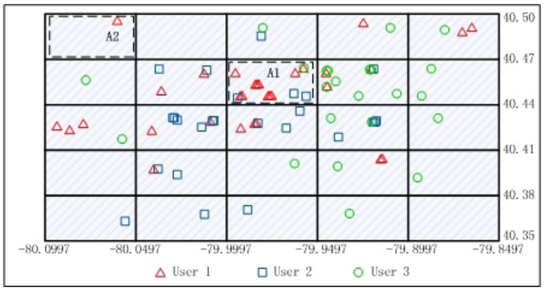


FIGURE 1. Location information of users' reviews in Pittsburgh, PA.

To improve the effectiveness of reviews, recent studies mostly focus on how to identify fake reviews based on characteristics of reviews themselves (e.g., technologies like natural language processing (NLP) and the mechanisms that reviews below 100 words are set to be invalid). For example, authors in [6] proposed a filtering mechanism to discard the reviews that may be fake by analyzing the semantic characteristics of reviews, and the characteristics could be the length or content of reviews. A method based on empirical analysis was also proposed to reduce fake reviews in paper [9]. Instead of analyzing the characteristics of fake reviews to decrease the effect on CSLBSSs caused by them, another important aspect that needs to be considered as a basis for the validity of reviews is the users who leave reviews. In this paper, we improve the effectiveness of reviews by establishing a reputation mechanism for users.

To address the aforementioned problem, this paper studies the threats that may be caused by the reviews to be published. These reviews are considered from the aspect of users to improve their validity. Our objective is to improve the effectiveness of the reviews to be published, while also preventing the identification or inference attack from them. As far as we know, about review publication, there is no research on resisting identification or inference attack and considering the effectiveness of reviews in terms of users' credibility.

Our mechanism, named IEPP, is to Improve Effectiveness and Privacy Preservation for review publication in CSLBSSs. To protect users' privacy and improve the effectiveness of reviews, two important aspects need to be considered before a review is published: i) The possibility of identifying the user who is about to publish his reviews should be assessed. In this paper, we use the concept of similarity probability as a criterion to evaluate the possibility that a user may be identified. The criterion is that other individuals must satisfy a minimum probability of visiting the places that the user has visited. In this way, an adversary can not identify a specific user. Therefore, the possibility that the user suffers identification or inference attack caused by the published reviews can be reduced; ii) The reliability of the user leaving a review should be assessed. We adopt Voting Decision Rule and Beta Reputation Mechanism to build a reputation system for each user. Then we can improve the effectiveness of reviews by ranking reviews according to users' reputation scores.

In this paper, we formalize the model of attackers who are interested in inferring more information about a target user from published reviews. Moreover, the effectiveness of public reviews is also validated by users' reputation system and our algorithm. The main contributions of this paper are summarized as follows:

1. We propose a mechanism (i.e., IEPP) to identify the risk of privacy disclosure and validate the utility of reviews in crowd-sourced local business service systems.

2. To prevent the identification or inference attack caused by reviews to be published, we formalize the model of attackers and formulate the possibility that a user who is about to leave reviews may be identified. We decrease the risk of identification or inference attack from review publication by defining and evaluating the similarity probability.

3. To improve the effectiveness of reviews to be published and decrease the impact of fake reviews, we build a reputation system for each user in CSLBSSs. The reputation system is used to increase availability by ranking reviews.

4. We carry out extensive simulations to evaluate the performance of our mechanism.

The remainder of the paper is organized as follows. After discussing the related work in section II, we present the main theories of our study in section III. Subsequently, the system model, threat model, and framework are introduced in section IV, and our algorithms are presented in section V. We give a discussion about the ways to protect users' privacy and the time complexity in section VI. The experiment results and the limitations are shown in section VII. Lastly, we conclude the paper with giving directions for future work in section VIII.

II. RELATED WORK

In this section, we first present an overview of the current major Location-Privacy Protection Mechanisms (LPPMs). Subsequently, two broad categories of recent research on privacy protection of location-sharing and availability of reviews are discussed.

A. AN OVERVIEW OF LPPMS

Extensive methods have been proposed to design LPPMs, which are targeted for different goals and used to deal with various threats on privacy. The well-known schemes such as Obfuscation and Dummy are usually used to address privacy issues in LBS. Obfuscation is an approach of protecting users' privacy information (e.g., sensitive information including identity and location, etc.) by submitting less accurate information. For example, based on the research of top 750 Apps in Google Play store, authors in [10] proposed a technique called position truncation, which was used to present a user's locations via a grid of fixed points. Obfuscation can not work well in CSLBSSs because a review is always corresponding with a business (e.g., businesses like restaurants), and the location of a business is accurate. Dummy always uses a fake location or a synthesized path to represent a real user's information. Niu *et al.* [11] investigated the privacy level in terms of entropy and proposed an algorithm to generate dummy locations. For the purpose that users' locations can not be distinguished, Chow and Golle [12] designed a scheme to generate more realistic synthesized paths. However, Dummy is not a suitable method in CSLBSSs, because users usually can not stand the false reviews generated by fake locations.

B. PRIVACY PROTECTION OF LOCATION-SHARING

One review potentially shares the location of the corresponding business. There are some studies on privacy protection of location-sharing [13]. In [14], the authors designed a privacy-protection mechanism after investigating the motivation behind users' check-in and divided location information into geographical and semantic levels. The true geographic information was coarsened to its upper level according to varying levels of geographical, and the semantic information was represented as Foursquare's semantic hierarchy. Partial publication and delayed publication are suitable approaches to protecting privacy protection of location-sharing. Thus, the authors formalized the user-adversary model and used the zero-sum Bayesian Stackelberg game to decide whether a true location should be published [15]. Nonetheless, restricting the publication (e.g., partial publication) of users' reviews may adversely affect the availability of CSLBSSs, and influence the enthusiasm of reviewers. The way of delayed publication can release all users' reviews since sensitive information fades over time. However, users usually do not want to see reviews from a long time ago, because the recommended value and reliability of reviews will diminish over time.

C. AVAILABILITY OF REVIEWS

Numerous studies focus on the availability of reviews, especially how to identify fake reviews. Some of them adopted techniques such as semantic filtering, natural language processing (NLP), and text mining [6], [16], [17]. However, these studies mostly focus on the characteristics of fake reviews themselves. For example, to reduce the prevalence of fake reviews, Mayzlin et al. [17] designed a scheme by only allowing verified users to leave a review, and put forward a method to reduce the number of fake reviews by analyzing the fake content. Besides, there are also some researches on developing a series of tools to identify fake reviews [9], [18], [19]. For example, after analyzing Yelp's filtering algorithm which was used to prevent review fraud, the authors in [9] investigated the economic incentives to leave fake reviews on the platform Yelp and designed a method that used two complementary approaches to identifying fake reviews.

III. PRELIMINARIES

In this section, we give the basic concepts and theories used later in this paper.

A. GRID PARTITION

In a crowd-sourced local business service system (CSLBSS), assume that there are m businesses located in a city, and the set of businesses is represented by $\{b_1, b_2, \ldots, b_m\}$. Each business has an exclusive position coordinates $\{x_{bi}, y_{bi}\}$, in which x_{bi} stands for longitude and y_{bi} stands for latitude. In this paper, we use the grid partition to divide a city into a 2D grid with equal-size. In this way, for a city, each region can be represented by each grid, and the set of grids is represented by $\{g_1, g_2, \ldots, g_k\}$. In this paper, note that one region is represented by one grid, and one region (i.e., one grid) may contain multiple businesses. If a user leaves a review on a business which is located in a region, there will be a corresponding record in the grid that corresponds to the region. Fig.1 shows an example that the locations of users' reviews are plotted by graphic symbols. By the way, a city map also can be divided by other strategies such as clustering partition and manual partition [8]. However, clustering partition suffers the drawback that it needs to input the number of clusters [20]. And manual partition has the disadvantage that a designer may have limited knowledge of a city map. The strategy (i.e., dividing a city into a 2D grid) we use in this paper is a better way, because a city map is usually divided by roads and streets, especially in the big cities like Beijing and New York. As we all know, one grid may include many businesses. Therefore, one phenomenon in Fig.1 is that different users' reviews are in the same grid, which elicites the concept of users' similarity probability that will be introduced in the next subsection.

B. SIMILARITY PROBABILITY

Generally speaking, assume that there are n users in a CSLBSS, and the set of users is represented by $\{u_1, u_2, \ldots, u_n\}$. In a period of time, the database of one CSLBSS will record the number of reviews for each user in the grids of a city. Then the distribution probability of each user's reviews in each grid can be calculated. Assume that the distribution probability of u_i's reviews in grid g_k is represented by $Pd_{(u_i,g_k)}$. If u_i never visits grid g_k, the distribution probability of user u_i in grid g_k is 0. Besides, if user u_i leaves reviews in grid g_k, the number of reviews can also be counted, and it is represented by $C_{(u_i,g_k)}$. Thus, if user u_i never leaves a review in grid g_k, the $C_{(u_i,g_k)}$ is 0. We take user u_i in grid g_k as an example. In a period of time, assume that the total number of users' reviews in grid g_k is represented by $C_{(u_{all},g_k)}$ and the number of user u_i' reviews in grid g_k is represented by $C_{(u_i,g_k)}$. Then the probability that u_i' reviews belong to grid g_k can be represented by $Pr_{(u_i,g_k)}$, where $Pr_{(u_i,g_k)} = \frac{C_{(u_i,g_k)}}{C_{(u_{all},g_k)}}$. From the perspective of an adversary, if he is interested in the probability of the reviews belonging to target user u_i and wants to confirm u_i in grid g_k, the probability can be given as follows:

$$P(u_i, g_k) = Pd_{(u_i,g_k)} \cdot Pr_{(u_i,g_k)} \tag{1}$$

For two users u_i and u_j in the same city, the similarity probability can be defined in terms of formula (2), which is similar to the concept of "individuals' closeness" [21]. Therefore, from the perspective of an adversary, the similarity probability of two users u_i and u_j in grid g_k can be given as follows:

$$\epsilon_{(k,min)} \leq \frac{P(u_i, g_k)}{P(u_j, g_k)} \leq \epsilon_{(k,max)} \tag{2}$$

The formula (2) embodies a privacy criterion, which is used to protect the case that a user's reviews can be mapped to his identity. ϵ_k is a threshold that reflects the similarity probability between u_i and other users except u_i (i.e., u_j) in grid g_k. To reduce the possibility of identification or inference attack on user u_i, an appropriate privacy criterion interval $[\epsilon_{(k,min)}, \epsilon_{(k,max)}]$ can be selected according to his $P(u_i, g_k)$. The way to protect users' privacy with similarity probability will be explained in section VI.

C. VOTING DECISION RULE

To improve the effectiveness of reviews, the voting decision rule [22] is used in this paper. In current CSLBSSs, when users leave reviews, there are usually some scoring mechanisms such as 10-score or 5-star level for users to evaluate a certain level of a business. For example, a user may give a 4-star (e.g., in 5-star level mechanism) with his review for a restaurant. In this paper, we take the 5-star level mechanism as an example to represent the users' evaluation on a business, and the evaluation result such as 4-star given by user u_i is represented by Ror_{u_i}. According to users' evaluation results, a binary qualitative decision can be defined. For example, when the evaluation result of user u_i is higher than 3-star, the binary qualitative decision which is represented by d_{u_i} can be defined as approbation. In this paper, we use approbation and disapprobation to represent the binary qualitative decision according to users' evaluation results. Approbation means that a certain level of one business has been achieved by the evaluation result of a user's review. Let h_0 and h_1 be the hypothesis that a certain level of one business is disapproving (i.e., disapprobation via the result of a user's review) and approving (i.e., approbation via the result of a user's review), respectively. Assume that τ is a predefined threshold, such as 3-star, to measure whether a certain level of a business is worthy of approbation. Through the above settings, P_{fp}, P_{fn}, and P_t can be denoted as the probabilities of positive review fraud, negative review fraud, and true review evaluation, respectively, i.e., $P_{fp} = P(Ror_{u_i} > \tau | h_0)$, $P_t = P(Ror_{u_i} > \tau | h_1)$, and $P_{fn} = 1 - P_t$. The binary qualitative decision about approbation or not from each user can be fused by the following fusion rule [23]:

$$Est = \begin{cases} H_1, & \sum_{i=1}^{n} d_{u_i} \geq \lambda, d_{u_i} \in \{0,1\} \\ H_0, & \sum_{i=1}^{n} d_{u_i} < \lambda, d_{u_i} \in \{0,1\} \end{cases} \tag{3}$$

For a certain level of a business, the global decision, which is represented by Est, is worthy of approbation when at least λ out of n users are giving the decision h_1. Besides, $\rho = \frac{\lambda}{n}$ is denoted as the threshold of the global decision.

D. BETA REPUTATION MECHANISM

Beta reputation mechanism [24] is introduced to improve the reliability of voting rule, and it is also used to progressively reduce the impact of malicious users. In general, a user's reputation score can reflect his level of reliability. Therefore, assume that at a time point t, a set of decision vectors $d(t) = [d_{u_1}(t), d_{u_2}(t), \ldots, d_{u_n}(t)]^T$ are obtained from the users who left the reviews about a certain level of one business, where $d_{u_i}(t)$ means the binary qualitative decision about approbation or not from user u_i at the time point t (i.e., $d_{u_i}(t) = 0$ (resp. $d_{u_i}(t) = 1$) means that a certain level of one business is disapproving (resp. approving) via the result of user u_i' review). Then the global decision using the fusion rule at the time point t can be given as follows:

$$Est(t) = f(w(t), d(t)) = \begin{cases} 1, & if \sum_{i=1}^{n} w_{u_i}(t) \cdot d_{u_i}(t) \geq \rho \\ 0, & otherwise \end{cases} \tag{4}$$

The weight vectors $w(t) = [w_{u_1}(t), w_{u_2}(t), \ldots, w_{u_n}(t)]^T$ are based on the reputation score of each user before the time point t. Besides, the positive rating $\zeta_{u_i}(t)$ and negative rating $\eta_{u_i}(t)$ of user u_i are defined as: $\zeta_{u_i}(t) = \zeta_{u_i}(t-1) + v_1(t)$ and $\eta_{u_i}(t) = \eta_{u_i}(t-1) + v_2(t)$. Furthermore, $v_1(t)$ and $v_2(t)$ can be calculated as follows:

$$v_1(t) = \begin{cases} 1, & d_{u_i}(t) = Est(t) \\ 0, & otherwise \end{cases} \tag{5}$$

$$v_2(t) = \begin{cases} 1, & d_{u_i}(t) \neq Est(t) \\ 0, & otherwise \end{cases} \tag{6}$$

Note that $\zeta_{u_i}(t-1)$ represents the number of times that the decisions of u_i are consistent with global decision Est before the time point t, and $\eta_{u_i}(t-1)$ represents the number of times that the decisions of u_i are inconsistent with global decision Est before the time point t. After calculating the global decision $Est(t)$, for user u_i, his reputation score $R_{u_i}(t)$ can be updated by formula (7). The weight vector $w_{u_i}(t)$ used in formula (4) is calculated by formula (8).

$$R_{u_i}(t) = \frac{\zeta_{u_i}(t) + 1}{\zeta_{u_i}(t) + \eta_{u_i}(t) + 2} \tag{7}$$

Note here that the initial reputation score of each user is set as $1/2$ in this paper, because no prior information can be used to judge whether his reviews are always objective or not.

$$w_{u_i}(t) = \frac{R_{u_i}(t-1)}{\sum_{j=1}^{n} R_{u_j}(t-1)} \tag{8}$$

The formula (8) shows that u_i's weight vector $w_{u_i}(t)$ is a relative value. It is variable when different users participate, and it means the proportion of his reputation score among all the users who left a review about a certain level of one business at the time point t. Within the period of time T, u_i's final reputation score can be calculated based on the number of times that the decisions of u_i are consistent with global decision Est. Generally speaking, most users are always honest and usually give objective evaluation by their reviews. To achieve a high reputation score (i.e., R_{u_i}), users need to be honest and provide objective evaluations via their reviews (For example, their reviews are genuine reviews rather than fake ones). Each user has his reputation database stored at the review center which will be described in the next section.

IV. MODELS AND IEPP FRAMEWORK

A. SYSTEM MODEL

The system model of this paper is mainly composed of three parts: Mobile Users (i.e., the set of users $\{u_1, u_2, \ldots, u_n\}$), Review Center (RC), and a specific crowd-sourced local business service system (CSLBSS). The model which describes the process before reviews are published is shown in Figure 2, and the details of each part are shown as follows.

1) MOBILE USERS

Without loss of generality, if user u_i wants to leave a review, he must register himself to a specific CSLBSS such as Dianping. For security reasons, the CSLBSS requires users to provide their specific information such as mobile phone numbers to ensure that they are real users. To protect users' privacy, especially users' identities, the CSLBSS allows each user to identify themselves with a pseudonym by means of some pseudo-IDs technologies [25]. In our paper, one user's pseudonym corresponds to a real user, and users' pseudonyms are not often changed due to the complexity of re-authentication [9]. To ensure the authenticity of users' reviews, we stipulate that the users who leave reviews are the users who have experienced in businesses. This stipulation can be achieved by methods such as check-in and reviews based on consumption code [7]. There are dishonest or malicious users in the CSLBSS. They leave reviews for some aims such as building the reputations for their businesses or putting down the reputations of their competitors.

2) REVIEW CENTER

RC is responsible for protecting users' privacy and improving the effectiveness of reviews. Firstly, it divides a city into a 2D grid with equal-size according to the regions of a city via the method mentioned in section III-A. Secondly, it asks the CSLBSS for users' reviews left to the businesses in a period of time and calculates the similarity probability among users using the way explained in section III-B. Note here that the CSLBSS and mentioned later in the paper is a specific crowd-sourced local business service system such as Dianping because different crowd-sourced local business service systems have different databases. Base on the result of users' similarity probability, RC decides the status of users' reviews (i.e., public or anonymous). In our paper, the main way to protect users' privacy (i.e., preventing users' identities leakage from identification or inference attack) is to set users' reviews to be anonymous. Note here that the anonymity (i.e., setting users' reviews to be anonymous) means users' pseudonyms and avatars related to their reviews are blurred so that users cannot be identified. Thirdly, according to the binary qualitative decisions obtained from users' reviews (i.e., approbation and disapprobation), for a business, RC figures out the global decision via the voting decision rule (i.e., section III-C), then updates and stores users' reputation scores (i.e., section III-D) in its database. Lastly, RC provides the status of users' reviews and the updated users' reputation scores to the CSLBSS. In our paper, RC is a trustworthy entity, which means it faithfully implements the relevant regulations and agreements.

3) CROWD-SOURCED LOCAL BUSINESS SERVICE SYSTEM (CSLBSS)

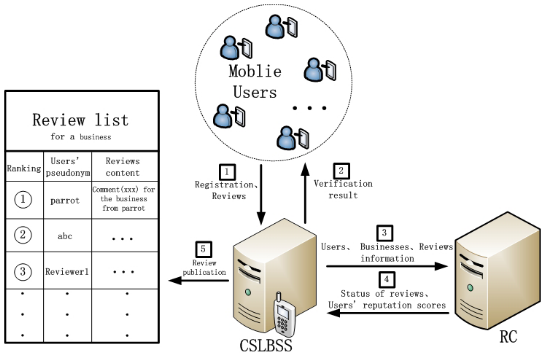
The CSLBSS has four missions in this paper. Firstly, the CSLBSS provides users' registration interface and ensures the authenticity of users registration via some techniques such as binding a user's pseudonym with his phone number. The purpose is to make sure that the pseudonym is the unique identifier of the user's identity. Besides, the CSLBSS should ensure that one user can only have one pseudonym in a system and one review belongs to only one user. Secondly, the CSLBSS offers real information such as users, businesses, and reviews to RC, and regularly maintains and updates the relevant information. Thirdly, according to the status of users' reviews (i.e., public or anonymous) decided by RC, the CSLBSS publishes the corresponding reviews. Fourthly, since most users only read the top-ranked reviews in a review list [8], the CSLBSS sets the top-x public reviews based on the users' reputation scores to improve the effectiveness of reviews. In our paper, the CSLBSS is seen as an entity of trust. On the one hand, it usually agrees on privacy policies with users and helps users to resist attacks from malicious attackers. On the other hand, it also wants to maintain the validity of reviews to enhance users' acceptance.

B. MOTIVATION AND THREAT MODEL

In this subsection, we give our motivation for this paper and the threat model.

1) MOTIVATION

In Fig.1, we can know that user 1 has more than one review on the businesses located in region A1, and user 1 has only one review on the business located in region A2. Our motivation is that if one user's location information from his reviews is distinctly different from other users at a certain moment, the user can be easily identified by an adversary's database. The following two scenarios are used to illustrate the privacy disclosure via reviews publication in a CSLBSS.



Review list
for a business

Ranking	Users' pseudonym	Reviews content
①	parrot	Comment(xxx) for the business from parrot
②	abc	. . .
③	Reviewer1	. . .
.	.	.

FIGURE 2. System model for review publication.

Scenario 1: Over a period of time, for a certain region, if the number of public reviews for one user is significantly more than other users, it is more likely to suffer identification or inference attack. This is because an attacker can easily distinguish this user from other users, and educe more information about the user. This case usually happens in the overall release of reviews over a period of time.

Scenario 2: Over a period of time, if there is only one public review of a user in a certain region, especially there is only one business in a certain region, it is more likely to suffer identification or inference attack. This case usually happens in the instant release of reviews.

2) THREAT MODEL

In our paper, we assume that the adversary is an active one who is interested in inferring more information about the users, especially the users' identities, from all the public reviews in a CSLBSS. The adversary could be any individual who has the ability to read and collect users' public reviews and locations of businesses from a CSLBSS, and he can use the data collection software such as Crawlzilla, Heritrix, Ex-Craw, etc. According to the ability of the adversary, the number of reviews he can collect ranges from some regions to the entire city. The adversary intends to infer the identities of target users based on their published reviews, and even the users' name are anonymous. The adversary also wants to map the users' distribution of visited locations from their reviews to users' identities via the background information or side-information. This is widely referred to as the identification or inference attack [4], [5]. Simply speaking, the adversary aims to match the public reviews from the CSLBSS with one of the individuals (i.e., the target user) in his database according to the characteristics of public reviews.

In our paper, the attack against users' privacy is to find the target users' identities. Therefore, the notion of privacy threat is given as follows.

Privacy Threat. *By collecting the Spatio-temporal information of users' reviews, the adversary can use the information as quasi-identifiers to identify the target users' identities.*

From the system model, we know the RC and CSLBSS are trusted entities. However, users are curious and dishonest. They are curious about other users' information and part of them themselves may be adversaries. They could be dishonest because they want to publish uncertain or fake reviews to achieve their purpose. Hence, there should be some security assumptions in our model as follows:

Security Assumption 1. *One user can only have one identity, and dishonest users can not madly inject fake reviews. The evaluation results of users' reviews (i.e., Ror) can not be maliciously changed by attackers.*

Security Assumption 2 *The collusion among parts does not exist. That is, (i) CSLBSS does not collude with users; (ii) users do not collude with each other; (iii) RC does not collude with users or CSLBSS.*

C. FRAMEWORK

In a CSLBSS, users always wish their reviews to be public. The reason is that they not only want to share their experience via reviews but are willing to provide helpful information for other users. However, on the premise of guaranteeing the guiding role of their reviews, users usually can not easily find a proper way to avoid privacy leaking from public reviews. Hence, our mechanism gives a proper way and focuses on providing anonymization of reviews from the user's perspective. If a user wants privacy protection in a region without losing the utility of his reviews, he can offer a privacy protection request when submitting a review, and the details will be shown in section V and VI. As for the utility of reviews, high ranking reviews have a high impact on users, so setting the reviews of the users who have high reputation scores (i.e., users who usually leave objective and honest reviews) as high ranking reviews in a review list can provide more useful information to other users. Therefore, the starting point of this paper is to use our IEPP mechanism to protect users privacy and improve the effectiveness of reviews. The main method we use to achieve privacy protection is to set reviews to be anonymous, and the way to improve the effectiveness is to rank the reviews of users with high reputation scores as high ranking reviews. As is mentioned in section IV-A-2, the anonymity in our paper is different from the general sense of anonymity (i.e., replacing a user's real name with a pseudonym), which means users' pseudonyms and avatars related to reviews are blurred so that users cannot be identified (i.e., no one can get any information such as pseudonym and picture related to the user except his content of review). Besides, the users with low reputation scores are untrustworthy, because their reviews are more likely to be supported by doubt or dishonest users. Accordingly, setting reviews of these users to low ranking is also a good way to decrease the impact caused by fake reviews. Thus, the framework of our paper is shown in Fig.3.

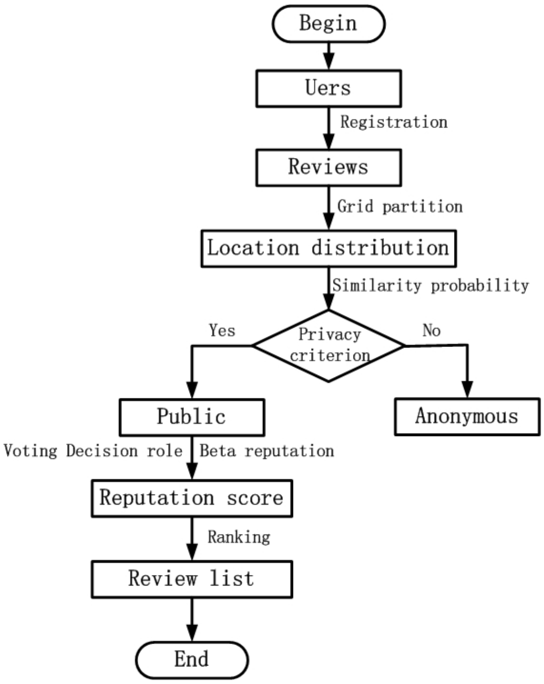


FIGURE 3. **The framework of IEPP mechanism.**

V. ALGORITHM DESIGN

According to the above framework, in this section, we give the main algorithms used in this paper. Based on the flow process of data, our goal is divided into three phases, and the second phase is our core.

A. DATA PREPARATION

Generally speaking, if one user wants to leave reviews on the businesses in a CSLBSS, he needs to register himself with a unique identifier (i.e., the user's ID number and telephone number tied to the user's identity) in the CSLBSS at first. Then he should identify himself with a unique pseudonym. We assume that the pseudonym is the only identifier that reflects the identity of a user in this paper, since some CSLBSSs use different ways such as the combination of pseudonym and avatar to reflect the identity of a user. If one user names himself with a pseudonym that other users have used, the CSLBSS will prompt the user to use another pseudonym that has never been used. This is a common mechanism for verifying the identity of real users in a CSLBSS such as Dianping, etc. After users register and name their pseudonyms, they can leave reviews while giving an x-star level to comment on a business. As we all know, giving an x-star level is a common way to represent users' evaluation results for a business. For example, Fig.4 shows a way of using an x-star level (top is five-star level) while leaving reviews to represent one user's evaluation result for a restaurant. We take u_1 as an example to explain the user who wants privacy protection when he leaves a review on business b_j. The way is that he needs to offer a privacy protection request which is represented by $ppr_{u_1}^{b_j}$ when he submits his review. For

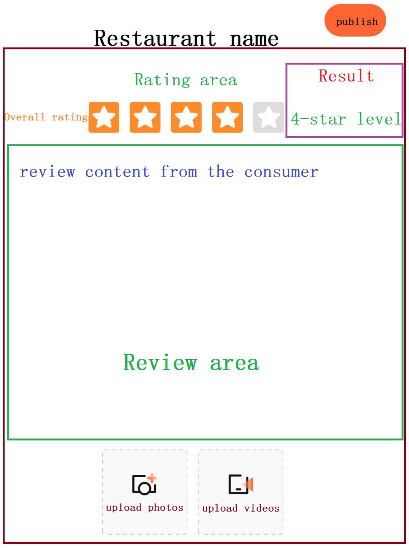


FIGURE 4. The mechanism of x-star level with reviews.

```
Algorithm 1 Data Preparation
Require: Users' pseudonyms and reviews
Ensure: The unique identification of users, Users' evaluation results
  users register in the CSLBSS;
  the CSLBSS guarantees that each user's pseudonym is unique;
  the CSLBSS collects and stores users' data in its database;
```

clarity, assume that the pseudonym of user u_1 is represented by $PSEu_1$, the review of u_1 on business b_j is represented by $REV_{u_1}^{b_j}$, and the evaluation result of $REV_{u_1}^{b_j}$ is represented by $Ror_{u_1}^{b_j}$. In short, the main role of this phase is to confirm users' identities and collect users' data such as the requests for privacy protection and evaluation results of their reviews. The pseudo-code of the data preparation process is described as Algorithm 1. In this phase, the flow process of data is from users to the CSLBSS.

B. DATA PROCESSING

In this phase, the flow process of data is from the CSLBSS to RC, and the data processing is taken by RC. There are two main processes: i) Privacy protection is to calculate users' similarity probability for deciding the status of reviews (i.e., public or anonymous); ii) The calculation of users' reputation scores is to decide the ranking of the public reviews.

1) PRIVACY PROTECTION

Firstly, RC divides a city into a 2D grid with equal-size via the strategy of grid partition because of the advantages which have been analyzed in section III-A. According to the locations of businesses in the CSLBSS, the grid which business b_j locates in can be determined by RC. For simplicity, assume that business b_j is located in grid g_k. Secondly, RC requests the users' data including the number of users and the number of their reviews in each grid within the period of time T from the CSLBSS. Note that T is the length of time that the system updates reviews. According to the users' data, the distribution probability of u_1' reviews in each grid can be calculated. For clarity, assume that the distribution probability of u_1' reviews in grid g_k is $Pd_{(u_1,g_k)}$. Besides, in grid g_k, we also assume that the total number of reviews is $C_{(u_{all},g_k)}$ and the number of u_1' reviews is $C_{(u_1,g_k)}$. Based on the above results (i.e., $Pd_{(u_1,g_k)}$, $C_{(u_{all},g_k)}$, and $C_{(u_1,g_k)}$), the probability $P(u_1, g_k)$ can be figured out by RC. Thirdly, since u_1 offers a privacy protection request $ppr_{u_1}^{b_j}$ and business b_j is located in grid g_k, the similarity probability among u_1 and other users such as u_j in grid g_k should be calculated. For clarity, assume that the set of users in grid g_k is represented by $\{u_1^{g_k}, u_2^{g_k}, \ldots, u_n^{g_k}\}$, so u_1 in grid g_k can be represented by $u_1^{g_k}$, and user u_j who means any user except u_1 in grid g_k can be represented by $u_j^{g_k}$ (i.e., $u_j^{g_k} \in \{u_2^{g_k}, u_3^{g_k}, \ldots, u_n^{g_k}\}$). Besides, we also assume that the privacy criterion interval of u_1 in grid g_k is represented by $[\epsilon_{(k,min)}, \epsilon_{(k,max)}]$. According to $P(u_1, g_k)$ and $[\epsilon_{(k,min)}, \epsilon_{(k,max)}]$, if there is a user in grid g_k (i.e., $u_j^{g_k}$) that meets the privacy criterion interval of $u_1^{g_k}$, all the status of $u_1^{g_k}$' reviews can be set to be public. If not, partial publication which means part of $u_1^{g_1}$' reviews can be set to be public is used to decide the number of public reviews, and the way of partial publication is to reduce the number of $u_1^{g_k}$'s reviews one by one until the given privacy criterion interval (i.e., $[\epsilon_{(k,min)}, \epsilon_{(k,max)})]$ is satisfied. If no user in the set $\{u_2^{g_k}, u_2^{g_k}, \ldots, u_n^{g_k}\}$ satisfies the the given privacy criterion interval (i.e., $[\epsilon_{(k,min)}, \epsilon_{(k,max)})]$, the number of $u_1^{g_k}$'s public reviews is 0. The time complexity and the way of users' privacy protection will be discussed in detail in section VI. Lastly, according to the number of $u_1^{g_k}$'s reviews that can be set to be public, RC selects this number of $u_1^{g_k}$'s reviews and sets their status to be public. Note that if the number of $u_1^{g_k}$'s reviews is more than 1 and not all $u_1^{g_k}$'s reviews can be set to be public, the status of $REV_{u_1}^{b_j}$ is randomly decided by RC (i.e., the status of $REV_{u_1}^{b_j}$ may be anonymous or public). Then RC gives all the status of $u_1^{g_k}$' reviews to the CSLBSS. The pseudo-code of privacy protection is described as Algorithm 2.

2) REPUTATION SCORE CALCULATION

To rank the public reviews, RC should calculate the final reputation scores of the users who have left reviews in the period of time T. We take the calculation procedure of the global decision of review $REV_{u_1}^{b_j}$ as an example to show how to calculate each user's final reputation score. Firstly, RC requests the evaluation results of the users who have left reviews in the period of time T. For clarity, assume that the set of users who have left their reviews on business b_j in the period of time T is represented by $\{u_1^{b_j}, u_2^{b_j}, \ldots, u_n^{b_j}\}$, and the set of their evaluation results is represented by $\{Ror_{u_1}^{b_j}, Ror_{u_2}^{b_j}, \ldots, Ror_{u_n}^{b_j}\}$. Secondly, assume that the threshold, such

Algorithm 2 Privacy Protection

```
Require: Database of the CSLBSS, RC
Ensure: $P(u_1, g_k)$, The status of $u_1$' reviews
the businesses information $\{b_1, b_2, \ldots, b_m\}$, $\{x_{bi}, y_{bi}\}$ $\rightarrow$ RC;
RC forms the set of grids $\{g_1, g_2, \ldots, g_k\}$, and determines that $b_j$ is located in grid $g_k$;
in the period of time $T$, the users' data $\rightarrow$ RC;
$ppr_{u_1}^{b_j} \rightarrow$ RC;
RC calculates $Pr_{(u_1,g_k)}$ and $Pd_{(u_1,g_k)}$;
RC obtains $P(u_1, g_k)$;
while $C_{(u_1,g_k)} > 0$ do
    RC calculates similarity probability between $u_1^{g_k}$ and $u_j^{g_k}$;
    if $[\epsilon_{(k,min)}, \epsilon_{(k,max)}]$ is satisfied then
        the number of $u_1^{g_k}$'s reviews that can be set to be public is $C_{(u_1,g_k)}$;
    else
        $C_{(u_1,g_k)} - 1$;
        if $C_{(u_1,g_k)} = 0$ then
            break, the number of $u_1^{g_k}$'s reviews that can be set to be public is 0;
        end if
    end if
end while
according to the number of $u_1^{g_k}$'s reviews that can be set to be public, RC decides the status of $u_1$' reviews in grid $g_k$;
the status of $u_1^{g_k}$' reviews $\rightarrow$ CSLBSS;
```

as 3-star, for measuring approbation of business b_j is represented by τ_{b_j}, and the predefined threshold of b_j's global decision is represented by ρ_{b_j}. Then the decision vectors of the users who left their reviews on business b_j, which is represented by $\{d_{u_1}^{b_j}, d_{u_2}^{b_j}, \ldots, d_{u_n}^{b_j}\}$, can be figured out based on τ_{b_j}. Thirdly, after getting the set $\{d_{u_1}^{b_j}, d_{u_2}^{b_j}, \ldots, d_{u_n}^{b_j}\}$, RC begins to calculate the global decision for business b_j based on ρ_{b_j}, and this global decision is represented by Est^{b_j}. Based on Est^{b_j}, whether the decision of u_1's review on business b_j is consistent with the global decision Est^{b_j} can be figured out. According to the above calculation process, whether the decisions of u_1's reviews on other businesses are consistent with their corresponding global decisions in the period of time T can be figured out. Assume that the set of users' reputation scores in the previous period of time (i.e., $T - 1$) is represented by $\{R_{u_1}(T-1), R_{u_2}(T-1), \ldots, R_{u_n}(T-1)\}$. Based on the number of times that the decisions of u_1 are consistent with the global decisions in the period of time T and $R_{u_1}(T-1)$, u_1's final reputation score in the period of time T (i.e., $R_{u_1}(T)$) can be figured out. For each user who has left reviews in the period of time T, RC calculates their final reputation scores in the period of time T. For simplicity, assume the users who left reviews in the period of time T is represented by the set $\{u_1^T, u_2^T, \ldots, u_n^T\}$, and u_i^T can be used to represent any user in $\{u_1^T, u_2^T, \ldots, u_n^T\}$. Lastly, RC sends these users' final reputation scores in this period of time T (i.e., $\{R_{u_1}(T), R_{u_2}(T), \ldots, R_{u_n}(T)\}$) to the CSLBSS. The pseudo-code of reputation score calculation is described as Algorithm 3.

Algorithm 3 Reputation Score Calculation

```
Require: Database of the CSLBSS, RC, $\rho$, $\tau$
Ensure: User's final reputation scores
users' data including set of users and set of users' evaluation results in the period of time $T$ $\rightarrow$ RC;
for each user $u_i^T$ do
    for each business of user $u_i^T$ do
        RC calculates the global decision of each business;
        RC counts the number of times that the decisions of $u_i^T$ are consistent with the global decisions;
    end for
end for
RC calculates each user's final reputation score $\{R_{u_1}(T), R_{u_2}(T), \ldots, R_{u_n}(T)\}$;
RC sends $\{R_{u_1}(T), R_{u_2}(T), \ldots, R_{u_n}(T)\}$ $\rightarrow$ the CSLBSS;
```

C. REVIEW PUBLICATION

In this phase, the data processing is taken by the CSLBSS. Firstly, the CSLBSS gets the status of $u_1^{g_k}$' reviews and users' final reputation scores in the period of time T (i.e., $\{R_{u_1}(T), R_{u_2}(T), \ldots, R_{u_n}(T)\}$). Secondly, the CSLBSS will reform a new review list by ranking the reputation scores of the users who have left reviews on business b_j. For clarity, assume that the new review list of business b_j is represented by L^{b_j} and the final reputation scores of the users who have left reviews on business b_j in the period of time T is represented by $\{R_{u_1}^{b_j}(T), R_{u_2}^{b_j}(T), \ldots, R_{u_n}^{b_j}(T)\}$. Note that the meaning of $R_{u_n}^{b_j}(T)$ is different from that of $R_{u_n}(T)$. $R_{u_n}^{b_j}(T)$ represents the reputation score of user $u_n^{b_j}$ who has left reviews on business b_j in the period of T, but $R_{u_n}(T)$ represents the reputation score of user u_n who has left reviews in the period of T, that is, $\{R_{u_1}^{b_j}(T), R_{u_2}^{b_j}(T), \ldots, R_{u_n}^{b_j}(T)\} \subseteq \{R_{u_1}(T), R_{u_2}(T), \ldots, R_{u_n}(T)\}$. If u_1' review on business b_j (i.e., $REV_{u_1}^{b_j}$) is set to be anonymous, the CSLBSS will randomly rank this review behind the public reviews in list L^{b_j}. If u_1' review on business b_j (i.e., $REV_{u_1}^{b_j}$) is set to be public, the CSLBSS will rank this review according to the reputation scores (i.e., $\{R_{u_1}^{b_j}(T), R_{u_2}^{b_j}(T), \ldots, R_{u_n}^{b_j}(T)\}$). Since the formed review list reflects the results in the period of time T, the CSLBSS is responsible for updating the review lists in different period of time. The pseudo-code of review publication is described as Algorithm 4.

Our algorithms reflect the framework shown in Fig.3. Simply speaking, the first thing is to get users' data from the CSLBSS in a period of time. Then RC will decide the status of users' reviews according to users' similarity probability and calculate users' reputation scores. Finally, for the public

Algorithm 4 Reviews Publication

```
Require: The status of $u_1^{g_k}$' reviews, $\{R_{u_1}(T), R_{u_2}(T), \ldots, R_{u_n}(T)\}$
Ensure: Review list $L^{b_j}$
the CSLBSS gets the status of $u_1^{g_k}$' reviews and $\{R_{u_1}^{b_j}(T), R_{u_2}^{b_j}(T), \ldots, R_{u_n}^{b_j}(T)\}$;
if $REV_{u_1}^{b_j}$ is set to be anonymous then
    the CSLBSS ranks $REV_{u_1}^{b_j}$ behind the public reviews in list $L^{b_j}$;
else
    the CSLBSS ranks $REV_{u_1}^{b_j}$ based on $\{R_{u_1}^{b_j}(T), R_{u_2}^{b_j}(T), \ldots, R_{u_n}^{b_j}(T)\}$ in list $L^{b_j}$;
end if
```

reviews, the CSLBSS reforms the review lists by ranking users' reputation scores.

VI. DISCUSSION

The way of using similarity probability to protect users' privacy and time complexity of our mechanism will be discussed in this section.

A. PROTECTING WAY FOR PRIVACY

Before discussing the ways for privacy protection with similarity probability, we give the definitions of the frequent region and general region for users. According to the Yelp dataset [26], we know that the frequency of one user's reviews is usually varying in different regions such as the location information of users' reviews in Pittsburgh, PA. Therefore, for clarity, we give the meaning of concepts and symbols used in this paper.

Frequent Region & General Region. *For a single user, if the number of reviews in a grid is more than 1, this grid is a frequent region for this user. If the number of reviews in a grid is 1, then this grid is a general region for this user.*

In this paper, for a single user, his frequent region is represented by the symbol F, and his general region is represented by the symbol G. Then the concept of F/G can be given as follows.

F/G. For a single user, there are a certain number of reviews in a period of time. The ratio of the number of his reviews in the frequent region to the number of his reviews in the general region is called the ration of one user's reviews in the frequent region and general region. The value of this ratio is represented by the symbol F/G.

From the concept of F/G, the meaning of F/G can be given.

Theorem 1: *For a single user, if the number of his reviews is a fixed value, bigger F/G means more reviews in his frequent region.*

Proof of Theorem 1: Assume that one grid represents one region, and one user has left some reviews in these grids within a period of time. Then during this time period, the user's number of reviews can be seen as a fixed value. If there are more reviews in his frequent region, the number of reviews in his general region is less. Therefore, bigger F/G means more reviews in his frequent region. □

As we all know, different users have different privacy requirements. Therefore, in this paper, two privacy protection levels are provided for users by using similarity probability: regional privacy requirement and global privacy requirement.

1) REGIONAL PRIVACY REQUIREMENT

Generally speaking, if one user wants to reduce the possibility of being identified caused by review publication in some key regions (e.g., the regions where their home and workplace are located), the way that all his reviews in key regions are directly set to be anonymous can protect his privacy well, but this way is like a one-size-fits-all solution which inevitably affects the utility of reviews and his original intention. For example, if grid g_k is the region where u_1's workplace is located, grid g_k usually contains many reviews when u_1 is active. Therefore, the privacy protection requirement in grid g_k is usually eager for u_1. Within the period of time T, if the number of u_1's reviews in grid g_k is significantly different from other users, then the number of u_1's reviews can be used by an adversary as a quasi-identifier to identify his identity. It is not a good choice to set all u_1's reviews in grid g_k to be anonymous for protecting the privacy of u_1 because this will seriously affect the utility of his reviews. Hence, whatever the reason, if u_1 wants to reduce the possibility of being identified caused by review publication without losing the utility of his reviews in grid g_k, he can offer a privacy protection request when he submits his reviews by using our algorithms in section V. Therefore, for a single user, the request of reducing the possibility of being identified caused by review publication in some certain regions is called regional privacy requirement in this paper.

2) GLOBAL PRIVACY REQUIREMENT

For a single user, the request of reducing the possibility of being identified caused by review publication in each of his reviewed region is called global privacy requirement in this paper. For example, if user u_1 requests a global privacy requirement, it means that the privacy criterion (i.e., formula (2)) in all the reviewed grids of u_1 should be satisfied respectively. Therefore, according to the formula (2) which reflects users' similarity probability in a certain grid, the following formula (9) reflects the similarity probability in all the reviewed grids of u_1. For clarity, assume that the set of all u_1's reviewed grids in the period of time T is represented by $\{g_1^{u_1}, g_2^{u_1}, \ldots, g_n^{u_1}\}$. Note that $g_1^{u_1}$ is not the same meaning of g_1, it just represents one of u_1's reviewed grids, that is, $\{g_1^{u_1}, g_2^{u_1}, \ldots, g_n^{u_1}\} \subseteq \{g_1, g_2, \ldots, g_k\}$. The global privacy criterion interval $[\mu_{min}, \mu_{max}]$ can be a privacy criterion which can measure the similarity degree among u_1 and other users in the reviewed grids of u_1. When user u_1 needs the global privacy requirement, RC should calculate the privacy criterion of each u_1's reviewed grid. In short, the global privacy requirement aims to find whether there are users who have almost the same distribution probability with u_1, which is used to guarantee that u_1 is indistinguishable within a theoretical set of users. Furthermore, when the total number of users' reviews in grid $g_n^{u_1}$ (i.e., $C_{(u_{all}, g_n^{u_1})}$) is very large, it means $Pr_{(u_1, g_n^{u_1})}$ is very small. Then we know that the similarity probability of u_1 is basically determined by the distribution probability of u_1's reviews for grid $g_n^{u_i}$ (i.e., $Pd_{(u_1, g_n^{u_1})}$). In formula (9), $\mu_{min} = \prod_{i=1}^{n} \epsilon_{(g_i^{u_1}, min)}$ and $\mu_{max} = \prod_{i=1}^{n} \epsilon_{(g_i^{u_1}, max)}$. n is the number of u_1' reviewed grids, and u_j is any user except user u_1 in grid $g_i^{u_1}$.

$$\mu_{min} \leq \prod_{i=1}^{n} \frac{P(u_1, g_i^{u_1})}{P(u_j, g_i^{u_1})} \leq \mu_{max} \tag{9}$$

B. TIME COMPLEXITY

Authors in [8] proposed (ϵ, δ)-public principle to protect a user's privacy by binding the user's threshold of public reviews, and the threshold is simply related to the number of the user's reviewed grids. Hence, their scheme limits the number of users' public reviews to some extent due to users' thresholds. However, our mechanism theoretically supports any number of users' public reviews as long as users are indistinguishable. The time complexity analysis of our algorithms is given as follows.

1) TIME COMPLEXITY OF PRIVACY PROTECTION

In a city, for a CSLBSS, assume that within the period of time T (i.e., the length of time that the system updates reviews), the number of the users in grid g_k (i.e., the users who have left their reviews on the businesses located in grid g_k) is N_{g_k}. One user, such as u_1, offered a privacy protection request $ppr_{u_1}^{b_j}$ when he left a review on business b_j located in grid g_k (i.e., the assumption in our algorithm 2), and the number of u_1's reviews in grid g_k is $C_{(u_1, g_k)}$. Therefore, this can be seen as an example of regional privacy requirement. To satisfy the privacy criterion interval in grid g_k (i.e., $[\epsilon_{(k,min)}, \epsilon_{(k,max)}]$), RC should calculate similarity probability between $u_1^{g_k}$ and other users except $u_1^{g_k}$ in grid g_k (i.e., user $u_j^{g_k}$ defined in this paper). Therefore, at worst, the algorithm has to run $(C_{(u_1, g_k)})(N_{g_k} - 1))$ rounds to determine how many reviews of $u_1^{g_k}$ can be set to be public, so the time complexity is $O(C_{(u_1, g_k)})(N_{g_k} - 1))$. If user u_1 requests a global privacy requirement, the number of the users who have left reviews in each u_1's reviewed grid should be figured out at first. Assume that the number of u_1's reviewed grids in the period of time T is represented by K_{u_1}, at worst, the algorithm has to calculate similarity probability between u_1 and other users except u_1 in each u_1's reviewed grid, so the time complexity is $O(K_{u_1} C_{(u_1, g_{max})}(N_{g_{max}} - 1))$. $C_{(u_1, g_{max})}$ represents the maximum number of u_1's reviews in u_1's reviewed grids, and $N_{g_{max}}$ represents the maximum number of users in u_1's reviewed grids.

2) TIME COMPLEXITY OF IMPROVING EFFECTIVENESS OF REVIEWS

The method of improving effectiveness of users' reviews is to rank their reviews in review list based on their reputation scores. Therefore, as for the time complexity of ranking reviews for one business such as business b_j assumed in our paper, the time complexity contains two stages. The first stage (i.e., algorithm 3) is to calculate the final reputation scores of the users who have left reviews in the period of time T. Assume that the number of the users who have left reviews in the period of time T is N_{all}, and the maximum number of business among the users who have left reviews in the period of time T is B_{max}, at worst, the time complexity of calculating users' final reputation scores is $O(N_{all} B_{max})$. The second stage (i.e., algorithm 4) is to rank the users' reviews based on users' reputation scores in one business, and we take business b_j as an example. Assume that the number of the users who left reviews on business b_j is N_{b_j}, at worst, the time complexity of ranking users' reviews is $O(N_{b_j}^2)$.

VII. EVALUATION

In this section, the performance metrics for privacy protection and reviews' utility are explained at first. Then the simulation settings are given and the results are analyzed. Lastly, the limitations of this paper are discussed.

A. PERFORMANCE METRICS

The specific attack methods are firstly formalized before introducing our metrics. After a series of analyses, we know that scenario 1 and 2 are more likely to suffer identification or inference attack. Therefore, the two specific attack methods assumed in this paper are as follows.

Attack method one. *Assume that an adversary has side-information or background information about the user who is the only one with the most reviews in scenario 1. In a period of time, if the adversary could find a user who is the only one with the most reviews in a grid, the user's identity can be identified.*

Attack method two. *Assume that an adversary has side-information or background information about the user who is the only one that has left reviews in a certain grid (i.e., similar to scenario 2). In a period of time, if the adversary could find such reviews, the user's identity can be identified.*

Based on the above two specific attack methods, the concept of the vulnerable region is given as follows.

Vulnerable region. *Assume that one grid represents one region, and one user has left some reviews in these grids in a period of time. Then during this time period, i) if the number of one user's reviews in a certain grid is the most, this grid is called the vulnerable region for this user; ii) if one user is the only one who has left reviews in a certain grid, this grid is called the vulnerable region for this user.*

After giving the specific attack methods, three performance metrics used in this paper are given as follows.

First metric: With the same distribution of users' reviews, if an algorithm can resist the above specific attack methods while releasing more public reviews, the algorithm is a better one. To measure the proportion of public reviews in all users' reviews, the concept of the public rate is given.

The public rate. *In a period of time, there are a certain number of users' reviews, and these reviews can be seen as the total number of reviews during this time period. The ratio between the number of reviews that can be set as public reviews and the total number of reviews is called the public rate, which is represented by R.*

From the concept of R, the meaning of R can be given.

Theorem 2: *In a period of time, if the total number of users' reviews is a fixed value, bigger R indicates more reviews that can be set to be public, and also means more contributions to the CSLBSS.*

Proof of Theorem 2: Under the condition that the total number of reviews is a fixed value, the more the number of reviews that can be set to be public, the larger the proportion of public reviews in all reviews. If more reviews can be set as public reviews, more useful information can be provided to users. Therefore, R can be used to measure the degree of reviews that can be set to be public, and bigger R means more public reviews. □

Second metric: In privacy protection, entropy H can be used to measure the degree of uncertainty for a user. When the number of users is k, the maximum value of H is $log_2\ k$ [27]. Therefore, entropy H of vulnerable regions under different algorithms is used to measure the degree of privacy protection.

Third metric: For a business, the first few reviews in its review list usually get more attention because most users only read the top-ranked reviews [8]. However, there are many factors (e.g., membership) can also decide the top-ranked reviews. To measure the effectiveness of reviews via different methods, the relevant criteria are given as follows.

The ranking rule. *In a review list, assume that there is a uniform ranking criterion, and the criterion is that the ranking of reviews is determined by users' reputation scores. The number one review in this review list belongs to the user with the highest reputation score among all users in this list. If the user with the highest reputation score is not unique, the number one review can be randomly selected from the users with the highest reputation score.*

The dubious probability. *In a review list, assume that the reviews are composed of honest users' reviews and dubious users' reviews. Each honest user gives an objective and consistent decision via their reviews, but each dubious user provides a random decision via their reviews. The probability that the number one review belongs to a dubious user is called the dubious probability, which can be used to measure the effectiveness of the review list. The dubious probability is represented by the symbol P.*

From the concept of P, the meaning of P can be given.

Theorem 3: In a review list, smaller P means better effectiveness of the review list.

Proof of Theorem 3: In a review list, the number one review is very important because it usually gets more attention from users. If the number one review belongs to a dubious user, it will inevitably have a negative impact on the credibility of this list. Therefore, if the probability that the number one review belongs to a dubious user can be reduced, it will improve the effectiveness of the review list. In other words, smaller P means better effectiveness of the review list. □

B. SIMULATION SETTINGS

We use Matlab R2018a to conduct our simulations, and run all algorithms on a local machine with an Intel Core-i5 2.5 GHz, 8 GB RAM, and Microsoft Windows 7 OS. To simulate the users' review distribution in a city, a full-mesh network consisting of 5*5 grids is constructed to simulate the regions in city A. Assume that each grid represents one unit region of city A, and each user has 12 reviews. Through the above settings, three scenes are provided to verify the validity of our mechanism. Under the premise that the privacy criterion interval ϵ for each grid is fixed, the first scene (i.e., scene-1) is mainly to verify the effect on the public rate and the entropy under different F/G. While proving the applicability of our algorithm, we also introduce the second scene (i.e., scene-2) to verify the effect on our algorithm under different privacy criterion intervals ϵ for each grid. The third scene (i.e., scene-3) is mainly used to confirm the role of our method in improving the effectiveness of reviews by comparing the way of equal weight combining (EWC).

Scene-1: Each user has only one frequent region, and the F/G of each user is set to 1/3, 1/6, and 1/9, respectively. Each user's frequent and general region are randomly distributed in the 5*5 grids. For each grid, assume the privacy criterion interval ϵ for each grid is [1/2,2].

Scene-2: Each user has only one frequent region. For each user, the number of reviews in the frequent region is set to be a random value from 3 to 9. Each user's frequent and general region are randomly distributed in the 5*5 grids. For each grid, the privacy criterion interval ϵ is set to [1/2, 2], [7/10, 10/7] and [9/10, 10/9], respectively.

The above two scenes are performed one hundred times. The average value of public rate and the average value of entropy in vulnerable regions are used for evaluation.

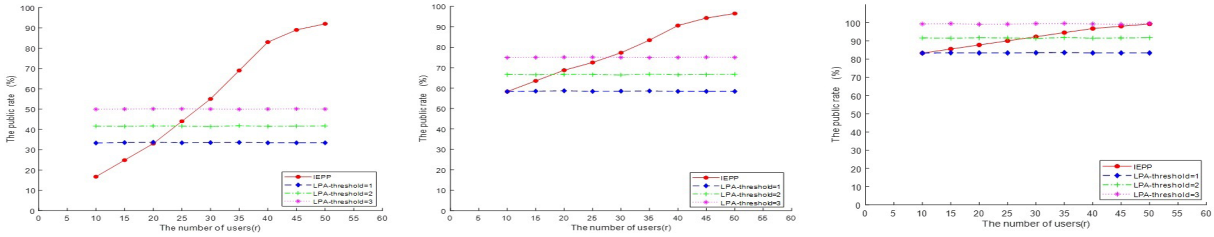
Scene-3: Assume that ten users form a group, and this group will leave reviews on different businesses. In the group, the number of dubious users, which is represented by *NoD*, ranges from one to four. This scene is implemented one hundred times. The average value of dubious probability is treated as a measurement.

C. SIMULATION RESULTS

We compare our algorithm with the function $f(\cdot)$ of LPA proposed in [8] when LPA's threshold for each grid (i.e., the number of public reviews that can be set for each user in each grid) is set to 1, 2, and 3, respectively.

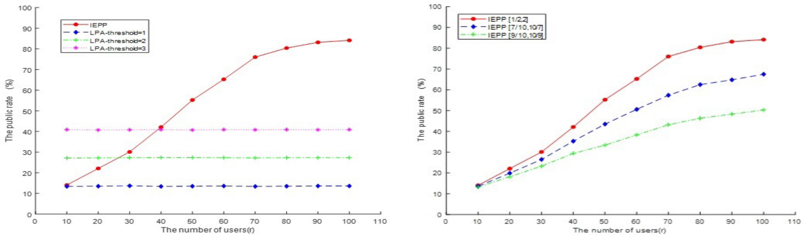
1) PUBLIC RATE

Under the setting of scene-1, Fig.5(a), Fig.5(b) and Fig.5(c) show the results of public rate when F/G is set to 1/3, 1/6



(a) $F/G = 1/3$ (b) $F/G = 1/6$ (c) $F/G = 1/9$

FIGURE 5. The simulation results of public rate for scene-1.



(a) The public rate with random F/G (b) The public rate under different intervals ϵ

FIGURE 6. The simulation results of public rate for scene-2.

and 1/9, respectively. We can draw the following conclusions from observing these figures.

Firstly, we can see that when LPA's threshold for each grid is fixed, the public rate R is almost a fixed value. Moreover, R does not change with the number of users, and bigger LPA's threshold for each grid leads to higher public rate R. This is because the function $f(\cdot)$ of LPA simply sets the number of each user's public reviews according to the number of grids, and this way is effective in defending against identification or inference attack when the number of reviews for all users' in a grid is the same. However, if the scenarios which have been analyzed in the motivation section occur, the method of LPA can not effectively resist identification or inference attack.

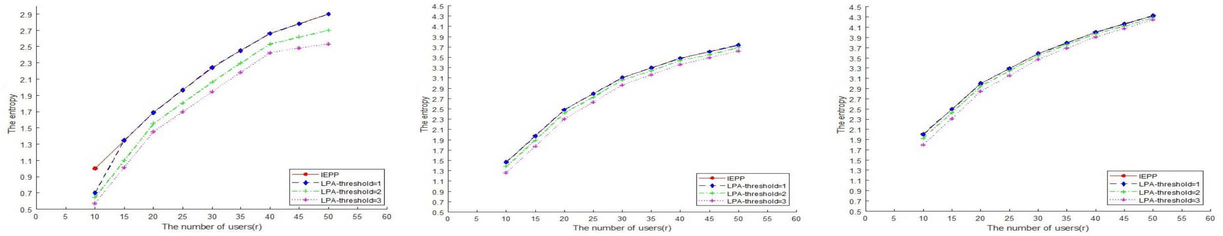
Secondly, as the ratio F/G decreases (i.e., as the number of each user's reviews in the frequent region decreases), the public rate R increases no matter which method is used. This is because there are more users' reviews in the general region, and these reviews are easier to meet both our and LPA's publication strategy. Besides, we can know that the more average the distribution of users' reviews, the more the number of users' reviews that can be set to be public.

Thirdly, as the number of users increases, our algorithm can release more users' reviews. This is because the increasing number of users increases the probability of similar distribution of users' reviews, and the users who meet the condition ϵ can release more public reviews. Though not shown in the figures, most of the unreleased reviews belong to each user's frequent regions which are more likely to be home and work regions. Therefore, our algorithm is more effective than LPA in resisting identification or inference attack.

Fourthly, from Fig.5(a), we can see that when the number of users is less than 20, the public rate obtained by using our algorithm is lower than the public rate obtained by using LPA even when LPA's threshold for each grid is set to 1. This is because the relatively small number of users leads to the emergence of the attack method two defined in this paper, and our algorithm sets the reviews of the users who suffer the attack (i.e., the defined attack method two) to be anonymous for resisting identification or inference attack. However, as the number of users increases, the public rate obtained by using our algorithm can be higher than the public rate obtained by using LPA even when LPA's threshold for each grid is set to 3. From Fig.5(b) and Fig.5(c), we also know that the public rate obtained by using our algorithm is sometimes less higher than the public rate obtained by using LPA when LPA's threshold for each grid is set to 2 or 3. The reason is that when LPA's threshold for each grid is set to 2 or 3, it is at the cost of leaking users' privacy, which means the number of reviews for a user in a certain grid may still be the most and the user with the most reviews in a certain grid still suffers identification or inference attack by using the attack method one defined in this paper.

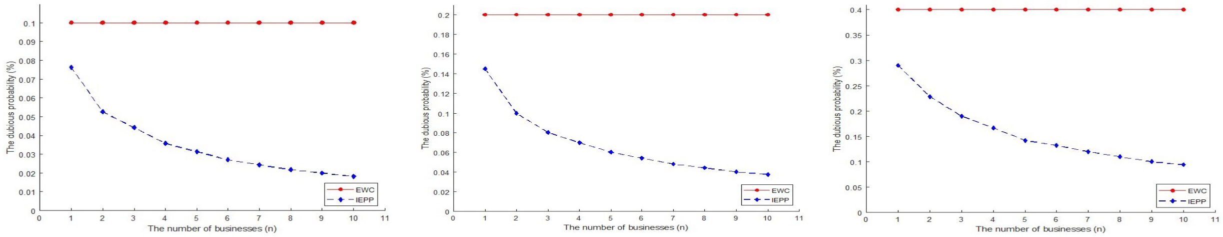
To verify the applicability of our algorithm, the number of reviews in each user's frequent region is set to be random. We study the users with the most reviews in each grid as the objects. Fig.6(a) and Fig.6(b) also show the results of the public rate under the setting of scene-2.

From Fig.6(a), we can see that when LPA's threshold for each grid is fixed, the public rate is almost a fixed value, but the public rate obtained by using our algorithm increases with the number of users. When the number of users is more than 40, the public rate obtained by using our algorithm is



(a) $F/G = 1/3$ (b) $F/G = 1/6$ (c) $F/G = 1/9$

FIGURE 7. The simulation results of entropy for scene-1.



(a) $NoD = 1$ (b) $NoD = 2$ (c) $NoD = 4$

FIGURE 8. The simulation results of dubious probability for scene-3.

bigger than the public rate obtained by using LPA. This is also because the increasing number of users increases the probability of similar distribution of users' reviews. Therefore, more reviews can be set as public reviews.

Fig.6(b) shows the public rate under different intervals ϵ. We can conclude that the smaller the interval ϵ, the lower the public rate. The reason is that the smaller interval requires the target user (i.e., the user with the most reviews in a grid) to find users with a higher similarity probability, which increases the difficulty of finding the users whose review distribution has the approximate probability with the target users. Thus, more reviews have to be set as anonymous reviews.

2) ENTROPY

Under the setting of scene-1, Fig.7(a), Fig.7(b) and Fig.7(c) show the results of entropy when F/G is set to 1/3, 1/6 and 1/9, respectively. We can draw the following conclusions from observing these figures.

Firstly, we can see the value of entropy increases with the number of users no matter which method is used. This phenomenon is due to the increasing number of reviews in each grid. From Fig.7, we can know that when $threshold = 1$, the value of entropy is bigger than the value of entropy when $threshold = 2$ or $threshold = 3$. This is because when $threshold = 1$, users' reviews are more evenly distributed than when $threshold = 2$ or $threshold = 3$, and it is at the cost of reducing the number of public reviews.

Secondly, as the ratio F/G decreases (i.e., as the number of each user's reviews in frequent region decreases), the value of entropy increases. This is because a more even distribution of users' reviews results in more reviews in each grid, which directly leads to bigger entropy.

Thirdly, the value of entropy obtained by our algorithm is almost the same with the value of entropy obtained by LPA when LPA's threshold for each grid is set to 1 (i.e., $threshold = 1$). By the way, $threshold = 1$ means the strictest privacy condition of LPA. The reason is that our algorithm can decrease the occurrence of the attack method one defined in this paper and reduce the probability of being identified.

Fourthly, from Fig.7(a), we can see when the number of users is ten, the value of entropy obtained by using our algorithm is higher than the value of entropy obtained by using LPA. This is because the relatively small number of users leads to the emergence of the attack method two defined in this paper. To resist identification or inference attack, our algorithm sets the review to be anonymous when there is only one review in a certain grid.

3) DUBIOUS PROBABILITY

We compare our beta reputation mechanism with equal weight combining (EWC). EWC assigns the same weight to each user, which means each user's review has the same probability of becoming the number one review in a review list.

Under the setting of scene-3, Fig.8(a), Fig.8(b) and Fig.8(c) show the results of dubious probability when *NoD* is set to 1, 2 and 4, respectively. We can draw the following conclusions from observing these figures.

Firstly, as the number of dubious users increases in the group, the dubious probability increases no matter which method is used. When the number of dubious users is fixed, the dubious probability obtained by using EWC is almost a fixed value regardless of the number of businesses. This is because each user's weight is the same, hence each user has the same probability to be the number one review.

Secondly, as the number of businesses increases, the dubious probability decreases by using our scheme. Furthermore, we can see that our scheme is better than EWC since the dubious probability obtained by our scheme is lower than the dubious probability obtained by EWC in spite of the number of dubious users. The reason is that our scheme of credibility score assigns weighted coefficients to users, which helps eliminate the effect of dubious users. Hence, the objective reviews from honest users can be ranked in the top of review lists, and the effectiveness of reviews can also be enhanced.

D. LIMITATIONS

There are also some limitations to our current work. In our mechanism, we know the binary qualitative decision is a key which can affect the correctness of the global decision, but it is not easy to obtain a qualitative decision when the users' evaluations vary a little for the same object. Besides, different users may have different evaluation standards about the same object, and different conditions such as regional customs as well as the degree of urban development may cause discrepancy of users' evaluations. Lastly, the level of privacy protection is regional in this paper, so the privacy protection of more fine-grained level such as aiming at a single business may not work well by using our mechanism. However, our simulations prove that our mechanism is useful to resist identification or inference attack, and the theoretical analysis is also proved by the results of our experiments.

VIII. CONCLUSION AND FUTURE WORK

The crowd-sourced local business service systems are an important part of Location-based service (LBS). The privacy leaking and effectiveness of reviews are the essential issues as well as key challenges. To resolve the above two problems, this paper proposes a novel mechanism named IEPP. The mechanism prevents users from identification or inference attack by setting the status of reviews and improves the effectiveness of reviews by ranking users' reviews based on their reputation scores. The extensive experiments validate the performance of our mechanism comprehensively.

In future work, in addition to testing our approach in the real-world environment by using some real-world datasets to further increase our contributions, we will also study the privacy protection of more fine-grained for review publication. Moreover, the relationship between genuine and fake reviews will be explored to expand the scope and precision of our mechanism.

REFERENCES

[1] M. Han, L. Li, Y. Xie, J. Wang, Z. Duan, J. Li, and M. Yan, "Cognitive approach for location privacy protection," *IEEE Access*, vol. 6, pp. 13466–13477, 2018.

[2] M. Usman, M. R. Asghar, I. S. Ansari, F. Granelli, and K. A. Qaraqe, "Technologies and solutions for location-based services in smart cities: Past, present, and future," *IEEE Access*, vol. 6, pp. 22240–22248, 2018.

[3] G. Yang, S. Luo, H. Zhu, Y. Xin, M. Li, and Y. Wang, "An efficient approach for LBS privacy preservation in mobile social networks," *Appl. Sci.*, vol. 9, no. 2, p. 316, Jan. 2019. [Online]. Available: http://www.mdpi.com/2076-3417/9/2/316

[4] J. Krumm, "Inference attacks on location tracks," in *Pervasive Computing* (Lecture Notes in Computer Science), A. LaMarca, M. Langheinrich, and K. N. Truong, Eds. Berlin, Germany: Springer, 2007, pp. 127–143.

[5] C. Bettini, X. S. Wang, and S. Jajodia, "Protecting privacy against location-based personal identification," in *Proc. 2nd VDLB Int. Conf. Secure Data Manage. (SDM)*. Berlin, Germany: Springer-Verlag, 2005, pp. 185–199. [Online]. doi: 10.1007/11552338_13.

[6] K. Fawaz, H. Feng, and K. G. Shin, "Anatomization and protection of mobile apps' location privacy threats," in *Proc. 24th USENIX Conf. Secur. Symp. (SEC)*. Berkeley, CA, USA: USENIX Association, 2015, pp. 753–768. [Online]. Available: http://dl.acm.org/citation.cfm?id=2831143.2831191

[7] K. Huguenin, I. Bilogrevic, J. S. Machado, S. Mihaila, R. Shokri, I. Dacosta, and J.-P. Hubaux, "A predictive model for user motivation and utility implications of privacy-protection mechanisms in location check-ins," *IEEE Trans. Mobile Comput.*, vol. 17, no. 4, pp. 760–774, Apr. 2017.

[8] X. Zheng, Z. Cai, J. Li, and H. Gao, "Location-privacy-aware review publication mechanism for local business service systems," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, May 2017, pp. 1–9.

[9] M. Luca and G. Zervas, "Fake it till you make it: Reputation, competition, and yelp review fraud," *SSRN Electron. J.*, vol. 62, no. 2, pp. 3393–3672, Jan. 2013.

[10] V. Gudmundsson, M. Lindvall, L. Aceto, J. Bergthorsson, and D. Ganesan, "Model-based testing of mobile systems—An empirical study on QuizUp Android app," in *Proc. Electron. Theor. Comput. Sci.*, vol. 208, May 2016, pp. 16–30. [Online]. Available: http://arxiv.org/abs/1606.00503

[11] B. Niu, Q. Li, X. Zhu, G. Cao, and H. Li, "Achieving *k*-anonymity in privacy-aware location-based services," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Apr. 2014, pp. 754–762.

[12] R. Chow and P. Golle, "Faking contextual data for fun, profit, and privacy," in *Proc. 8th ACM Workshop Privacy Electron. Soc. (WPES)*. New York, NY, USA: ACM, 2009, pp. 105–108. doi: 10.1145/1655188.1655204.

[13] H. To, G. Ghinita, L. Fan, and C. Shahabi, "Differentially private location protection for worker datasets in spatial crowdsourcing," *IEEE Trans. Mobile Comput.*, vol. 16, no. 4, pp. 934–949, Apr. 2017. [Online]. Available: http://ieeexplore.ieee.org/document/7501846/

[14] A. Zarezade, S. Jafarzadeh, and H. R. Rabiee, "Spatio-temporal modeling of users' check-ins in location-based social networks," Nov. 2016, *arXiv:1611.07710*. [Online]. Available: http://arxiv.org/abs/1611.07710

[15] R. Shokri, G. Theodorakopoulos, C. Troncoso, J.-P. Hubaux, and J.-Y. Le Boudec, "Protecting location privacy: Optimal strategy against localization attacks," in *Proc. ACM Conf. Comput. Commun. Secur. (CCS)*. New York, NY, USA: ACM, 2012, pp. 617–627. [Online]. Available: http://doi.acm.org/10.1145/2382196.2382261

[16] C. Xu and J. Zhang, "Towards collusive fraud detection in online reviews," in *Proc. IEEE Int. Conf. Data Mining*. Atlantic City, NJ, USA, Nov. 2015, pp. 1051–1056. [Online]. Available: http://ieeexplore.ieee.org/document/7373434/

[17] D. Mayzlin, Y. Dover, and J. A. Chevalier, "Promotional reviews: An empirical investigation of online review manipulation," *Amer. Econ. Rev.*, vol. 104, pp. 2421–2455, Aug. 2012.

[18] T. Chang, P. Y. Hsu, M. S. Cheng, C. Y. Chung, and Y. L. Chung, "Detecting fake review with rumor model—Case study in hotel review," in *Intelligence Science and Big Data Engineering. Big Data and Machine Learning Techniques*, vol. 9243, X. He, X. Gao, Y. Zhang, Z.-H. Zhou, Z.-Y. Liu, B. Fu, F. Hu, and Z. Zhang, Eds. Cham, Switzerland: Springer, 2015, pp. 181–192. [Online]. Available: http://link.springer.com/10.1007/978-3-319-23862-3_18

[19] N. Jindal, B. Liu, and E.-P. Lim, "Finding unusual review patterns using unexpected rules," in *Proc. 19th ACM Int. Conf. Inf. Knowl. Manage. (CIKM)*. Toronto, ON, Canada: ACM, 2010, p. 1549. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1871437.1871669

[20] P. Berkhin, "A survey of clustering data mining techniques," in *Grouping Multidimensional Data: Recent Advances in Clustering*, J. Kogan, C. Nicholas, and M. Teboulle, Eds. Berlin, Germany: Springer, 2006, pp. 25–71. [Online]. doi: 10.1007/3-540-28349-8_2.

[21] K. Fawaz and K. G. Shin, "Location privacy protection for smartphone users," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur. (CCS)*. New York, NY, USA: ACM, 2014, pp. 239–250. doi: 10.1145/2660267.2660270.

[22] W. Zhang, R. K. Mallik, and K. B. Letaief, "Cooperative spectrum sensing optimization in cognitive radio networks," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2008, pp. 3411–3415.

[23] M. Grissa, A. Yavuz, and B. Hamdaoui, "An efficient technique for protecting location privacy of cooperative spectrum sensing users," in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Sep. 2016, pp. 915–920.

[24] K. Arshad and K. Moessner, "Robust collaborative spectrum sensing based on beta reputation system," in *Proc. Future Netw. Mobile Summit*, Jun. 2011, pp. 1–8.

[25] R. Lu, X. Lin, Z. Shi, and J. Shao, "PLAM: A privacy-preserving framework for local-area mobile social networks," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Apr. 2014, pp. 763–771.

[26] Yelp Inc. (2015). *Yelp Academic Dataset*. [Online]. Available: https://www.yelp.com/academic

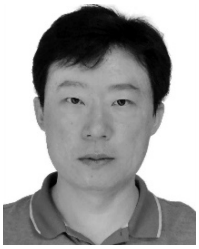
[27] Y. Sun, M. Chen, L. Hu, Y. Qian, and M. M. Hassan, "ASA: Against statistical attacks for privacy-aware users in location based service," *Future Gener. Comput. Syst.*, vol. 70, pp. 48–58, May 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167739X16302023

GUANGCAN YANG was born in 1986. He received the B.Sc. degree in network engineering and the M.Sc. degree in computer application technology from Henan Polytechnic University (HPU), in 2010 and 2013, respectively. He is currently pursuing the Ph.D. degree with the School of Cyberspace Security, Beijing University of Posts and Telecommunications, Beijing, China. His research interests include network and information security with a focus on LBS security.

SHOUSHAN LUO received the B.Sc. degree in mathematics from Beijing Normal University, in 1985, and the M.Sc. degree in applied mathematics and the Ph.D. degree in signal and information processing from the Beijing University of Posts and Telecommunications (BUPT), in 1994 and 2001, respectively. He is currently a Professor with the School of Cyberspace Security, BUPT, Beijing, China. His research interests include cryptography and information security.

HONGLIANG ZHU was born in 1982. He received the Ph.D. degree from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China. He is currently a Vice Director of the Beijing Engineering Lab for Cloud Security Technology and the Information Security Center, BUPT, where he is also a Lecturer and a Master Supervisor. His current research interests include big data security, cloud computing security, and network security.

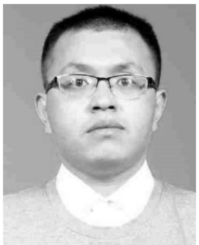
YANG XIN was born in 1977. He received the B.Sc. degree in signal and information system and the M.Sc. degree in circuits and systems from Shandong University, in 1999 and 2002, respectively, and the Ph.D. degree in signal and information processing from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2005. He is currently an Associate Professor with the School of Cyberspace Security, BUPT. His research interests include big data security, cloud computing security, and network security.



KE XIAO received the Ph.D. degree in circuit and system from the Beijing University of Posts and Telecommunications, China, in 2008. He has been a Professor with the North China University of Technology, since 2018. He has long been engaged in research and development and teaching of wireless communications, the Internet of Things, and embedded systems.

YULING CHEN is currently an Associate Professor with the Guizhou Provincial Key Laboratory of Public Big Data, Guizhou University, Guiyang, China. Her recent research interests include cryptography and information security.

MINGZHEN LI was born in 1986. She received the B.Sc. degree in information and computing science from Luoyang Normal University (LYNU), in 2009, and the M.Sc. degree in computer application technology from the Guilin University of Electronic Technology (GUET), in 2012. She is currently pursuing the Ph.D. degree with the School of Cyberspace Security, Beijing University of Posts and Telecommunications, Beijing, China. Her research interests include network and information security with a focus on LBS security.



YUNFENG WANG was born in 1987. He received the B.Sc. and M.Sc. degrees in management science and engineering from Henan Polytechnic University (HPU), in 2012 and 2015, respectively. He is currently pursuing the Ph.D. degree with the School of Cyberspace Security, Beijing University of Posts and Telecommunications, Beijing, China. His research interests include network and information security with a focus on LBS privacy preserving.