Hindawi
Security and Communication Networks
Volume 2021, Article ID 5556155, 22 pages
https://doi.org/10.1155/2021/5556155

Research Article

Cross-Platform Strong Privacy Protection Mechanism for Review Publication

Mingzhen Li,[1,2,3] **Yunfeng Wang**,[1] **Yang Xin**,[1,2] **Hongliang Zhu**,[1] **Qifeng Tang**,[4,5] **Yuling Chen**,[2] **Yixian Yang**,[1,2] **and Guangcan Yang**[1]

[1] *School of Cyberspace Security, Beijing University of Posts and Telecommunications, Beijing 100876, China*
[2] *State Key Laboratory of Public Big Data, College of Computer Science and Technology, Guizhou University, Guiyang 550025, China*
[3] *School of Computer and Information Engineering, Hechi University, Yizhou 546300, China*
[4] *National Engineering Laboratory for Big Data Distribution and Exchange Technologies, Shanghai 200436, China*
[5] *Shanghai Data Exchange Corporation, Shanghai 200436, China*

Correspondence should be addressed to Mingzhen Li; mzli@bupt.edu.cn and Yuling Chen; ylchen3@gzu.edu.cn

Received 23 January 2021; Accepted 21 May 2021; Published 8 June 2021

Academic Editor: Clemente Galdi

As a review system, the Crowd-Sourced Local Businesses Service System (CSLBSS) allows users to publicly publish reviews for businesses that include display name, avatar, and review content. While these reviews can maintain the business reputation and provide valuable references for others, the adversary also can legitimately obtain the user's display name and a large number of historical reviews. For this problem, we show that the adversary can launch connecting user identities attack (CUIA) and statistical inference attack (SIA) to obtain user privacy by exploiting the acquired display names and historical reviews. However, the existing methods based on anonymity and suppressing reviews cannot resist these two attacks. Also, suppressing reviews may result in some reiews with the higher usefulness not being published. To solve these problems, we propose a cross-platform strong privacy protection mechanism (CSPPM) based on the partial publication and the complete anonymity mechanism. In CSPPM, based on the consistency between the user score and the business score, we propose a partial publication mechanism to publish reviews with the higher usefulness of review and filter false or untrue reviews. It ensures that our mechanism does not suppress reviews with the higher usefulness of reviews and improves system utility. We also propose a complete anonymity mechanism to anonymize the display name and avatars of reviews that are publicly published. It ensures that the adversary cannot obtain user privacy through CUIA and SIA. Finally, we evaluate CSPPM from both theoretical and experimental aspects. The results show that it can resist CUIA and SIA and improve system utility.

1. Introduction

With the development of position technology and the widespread use of smartphones, more and more social network applications provide Location-Based Services (LBSs), known as Location-Based Social Networks (LBSNs) [1], such as TripAdvisor, Yelp, Dianping. We can exploit these applications to easily socialize online, plan travel routes, have spatial crowdsourcing [2, 3], and query surrounding Point of Interests (POIs), which greatly facilitates our lives [4]. Among these applications, Crowd-Sourced Local Businesses Service Systems (CSLBSSs), such as Yelp and Dianping, are interactive platforms that provide users with business information, consumer preferences, and consumer reviews in the areas of dining, shopping, etc. CSLBSSs are also special LBSNs that crowdsource review lists of businesses and maintain their reputation [5, 6].

In CSLBSSs, a public review mainly includes attributes such as display name, avatar, and review content (text, image, video, etc.). By browsing the list of reviews, consumers can get a true picture of the quality of the services provided by the business without going to the physical store.

That is, consumers can refer to the business reputation (i.e., business reputation score) and the review list to quickly and easily select POIs, such as restaurants with high scores. Note that the CSLBSSs evaluate the business reputation from both subjective and objective aspects: user rating and business reputation score. The user rating is a score that the system allows the user to make on the business reputation when a user publishes a review and is highly subjective. The business reputation score is a score that the system makes on the business reputation by calculating all user ratings and is highly objective.

However, while consumers enjoy the convenient services brought by CSLBSSs, they also face the risk of privacy leakage. In CSLBSSs, a business corresponds to a unique address. A review for a business implies that a user has visited the business or has related experiences associated with it. Moreover, reviews are public information, which can be obtained by anyone, even adversaries. By collecting and analyzing the reviews published by users [5, 6], some malicious adversaries are even able to infer users' privacy. Furthermore, by using the cross-social network fusion technology [7–9], we can connect user identities across multiple social networks and further infer more privacy, such as occupation, address, e-mail. At the same time, multiple types of platforms, including social networks, will provide users with different services and publish different information about users. The published information always contains the users' real profile and can be easily used to infer users' privacy.

In general, the process by which an adversary obtains a user profile in a CSLBSS includes the following: (1) the adversary uses display name as an initial keyword to search for the user's information, such as using engines to obtain the user's QQ, WeChat, and e-mail from social networks; (2) the adversary uses some information (e.g., e-mail, which is considered less private by users compared to QQ and WeChat), acquired from the first search process, as the keyword to further search for user's real name, educational background, organization. We call the above process that obtains a user's profile connecting user identities attack (CUIA). At present, most researches focus on the protection of user privacy in terms of query content [2, 3, 10–12] and data publication [8, 13, 14]. However, for data publication, few studies are investigating the privacy protection of review publication in CSLBSSs. To the best of our knowledge, only Zheng et al. [5] and Yang et al. [6] have explored the issue. However, both focus on privacy protection within the same platform and do not consider privacy disclosure on the cross-platforms. Therefore, neither of these methods can resist CUIA. Moreover, although the schemes of literature [5, 6] can protect users' privacy to some extent, they do not take into account the behavioral patterns of users. This results in the above approaches being unable to resist statistical inference attack (SIA) due to their inability to prevent adversaries from accessing long-term user behavior data [15, 16].

Currently, schemes of literature [5, 6] mainly protect user's privacy with the combination of suppression publication (partial publication) and anonymous publication. On the one hand, the adversary launches CUIA starting from the display name. While anonymous and partial release can reduce the risk of compromise, adversaries can still gain the user's display names and other private information from public reviews. On the other hand, the CSLBSSs rely on user's reviews to sustain the evolution of the platform; i.e., consumers are more likely to buy goods that have more credible reviews than less credible reviews. That is, consumers' willingness to purchase goods depends on how credible the reviews are. In this paper, we call the ability of the review to influence consumers' willingness usefulness of review. In the CSLBSSs, the systems hope to publish as many reviews that consumers consider as credible as possible. In general, for a system, the more such reviews it publishes means that it has a better ability to sustain development. In this paper, we call the ability to sustain development system utility. However, partial or anonymous reviews will reduce the usefulness of reviews, because partial publication results in a decrease in the number of reviews that the system can publish, or anonymous publication reduces the credibility of reviews. Thus, how to balance privacy protection and system utility becomes an issue that needs to be addressed.

To address the above issues, we need to propose a method to effectively protect a user's cross-platform privacy in the scenario of review publication while maintaining system utility. We first investigated the process of privacy disclosure and the usefulness of review in CSLBSSs. We found that adversaries generally can mine user's profiles to infer user's privacy by launching CUIA and SIA. In this process, the display name is usually the keyword exploited to mine a user's profiles. Based on the information adoption model [17], we found that users' real identity and consensus information, namely, the degree of consistency between user rating and business reputation score, are two key factors that determine the usefulness of the review. When little identity information of user is disclosed, if the user rating is consistent with the business reputation score, the consumer considers the review credible [18]. In other words, even if a user's real identity is not disclosed in the review, the usefulness of the review will not be affected.

Based on the above research, we propose a cross-platform strong privacy protection mechanism (CSPPM) based on the partial publication and complete anonymity mechanism to publish reviews. CSPPM partially publishes public reviews, but all published reviews are anonymous. It is a restricted privacy protection mechanism than [5, 6]. Here, complete anonymity refers to obscuring the display names and avatar or directly replacing them with randomly assigned strings and uniform icons in partial reviews that are allowed to be published. It solves the leakage of display names and minimizes the possibility of users suffering from CUIA and SIA. However, considering the background knowledge of the adversary, the completely anonymous reviews still have the risk of being identified. For example, reviews often contain users' photos and landmarks. In this case, people familiar with the user can easily identify that the review is published by the user. The more information the review discloses, the higher the risk of the user will have. Therefore, we adopt a partial publication mechanism to reduce the disclosure of less useful reviews. Whether a review is published or not depends on the difference between the user rating for a business and the business reputation score (short for score difference). It measures the degree of consistency between the user rating and the business reputation score. For the mechanism, reviews whose score differences fall within the threshold range are published anonymously, while those that exceed the threshold range are not published (these are most likely false reviews and extreme cases of positive and negative reviews). It ensures that the published reviews are those with high consistency with the business reputation score, which best reflect the business reputation. It also ensures the reference value of the review list to users. Besides, the list of reviews is sorted by a combination of score difference and user reputation score; namely, reviews are sorted by their usefulness.

The main contributions of our paper are as follows:

(i) We identify and formalize CUIA and SIA in the scenario of review publication. We also find that the display name is a key factor in user identification (i.e., privacy leakage).

(ii) We propose a stricter privacy protection mechanism based on partial publication and complete anonymity mechanism to protect privacy in the scenario of review publication.

(iii) We propose a method to improve the usefulness of reviews based on consensus information. In cases where the user's real disclosed identity information is too little or where the user is anonymous, the score differences of reviews are used to decide which reviews to publish. In the method, reviews with a small score difference will be published first.

(iv) We conducted experiments to verify the effectiveness of the proposed algorithms in the terms of resisting CUIA and SIA and maintaining system utility.

2. Related Work

In view of the above scenarios, we review the existing technologies from three aspects: how the attacker identifies the user's identity (attack identification), how the existing methods protect location privacy (privacy protection), and how to evaluate the system utility under different schemes (system utility).

2.1. User Identification. User identification [19], also known as linking user identities [20] or connecting user identities [21], refers to connecting user identities across multiple social networks by mining user profiles, relationships, and user-generated content (UGC, i.e., user behavior data, such as social network check-in, blog posts, shared pictures) from different social networks and associates the accounts of the same natural person on different social networks. According to the different types of information that the attacked can obtain, the existing research mainly focuses on user attribute, user relationship, and UGC.

User identification based on user attributes means that an attacker can connect user identity based on user profiles (mainly user names). As an identifier that uniquely identifies a user, because of individuals being accustomed to using the same or similar user names, user names are often used to identify users' accounts in different social networks. Zafarani et al. [21] found that the user homepage URL usually contains the user name, and they are used to adding a prefix or suffix to the user name to form a new user name. This means that these different usernames belong to the same person. Therefore, Liu et al. [20] considered seven characteristics including the length of the user name, special characters, numbers, to determine the user's identity. In response to this problem, the display name is used to replace the user name and become a kind of public information. However, the display name can still be used to identify the user. Li et al. [22] designed a distributed crawler to obtain user profiles containing display names in Foursquare, Facebook, and Twitter and identify users based on the extracted display name feature comparison results. Therefore, the user name or display name has become a key information for the attacker to identify the user's identity.

User identification based on user relationship is a method by which an attacker uses the user's circle of friends to identify multiple different accounts belonging to the same user. The core of this method is to connect users based on overlapping subnets in different social networks and improve user identification accuracy. The higher the degree of subnet overlap is, the higher the identification accuracy is [7, 19]. At present, there are some methods to identify users: related user mining based on prior users, related user mining based on non-priori users [7], and non-priori knowledge user identification algorithm based on friend relationships (FRUI-P) [23]. All these methods show that although the same user has different accounts in different social networks, attackers can still use user relationships to infer that these accounts belong to the same user.

User identification based on UGC mainly uses user behavior data on social networks for cross-social network user identification. Li et al. [24] mined the similarity of space (extracting location), time (extracting timestamp), and content features (counting semantic similarity and the number of identical words in text content) from UGC and then used supervised machine learning method to match the user accounts. Zhang et al. [25] analyzed the user's spoken language, content complexity, content standardization, and the characteristics of user pictures and user time series in multimedia content for the text content, multimedia content, and time series content published by users. And then they proposed text content analysis and identification methods, multimedia content identification methods, and time series content identification methods to identify user organizations/personal identities. As a result, UGC has become the key information for identifying users.

In addition, some scholars have tried user identification methods that combine user attributes, user relationships, and UGC content [9, 26, 27] to identify users' multidimensional identity feature information, in order to solve the drawbacks of using a single attribute to identify user accounts. Although the method of combining user's multiple attributes will result in a high degree of sparse data and a high degree of complexity in extracting features, it can extract more comprehensive user characteristics and increase the probability of recognizing a user's identity.

2.2. Location Privacy Protection. The inability to associate a user's identity with a precise location is a privacy protection method commonly used by current location privacy protection technologies. These techniques include three categories: obfuscation method, dummy method, and pseudonym.

The obfuscation method [28–30] protects the user's location privacy by using imprecise locations or areas instead of the user's real or precise location. It requires users to submit imprecise locations or areas to the server, for example, Gedik et al. [28] submission area, Gedik et al. [29] submission imprecise location, and so on. However, in the review publication scenario, the location of each business is precise. Therefore, it is not suitable for protecting location privacy in this scenario.

The dummy method [16, 31, 32] usually adds false users or false locations to achieve anonymity. For example, Li et al. [16] added fake locations, and Niu et al. [32] added fake users to achieve anonymity. However, in the review publication scenario, if a user has not visited a business, the user is not allowed to publish reviews on the business's services. Therefore, the dummy method is not suitable for protecting location privacy in this scenario.

The pseudonym [15] realizes privacy protection by replacing the user's identity identifier with a pseudonym. The basic assumption of this method is that the identity identifier is the only information that can be used by an attacker to identify a user's identity. For example, to a certain extent, display name can be regarded as a pseudonym that replaces the user's identity identifier. User reviews usually include photos, real-time location, and personal information. According to the aforementioned analysis, an attacker can identify the user's identity through CUIA. Li et al. [14] and Zhang et al. [33] and others have proposed privacy protection methods to balance the needs of users for such information and privacy protection. However, in the review publication scenario, we cannot protect user privacy only by replacing the user ID.

Zheng et al. [5] pointed out that only when the user's location obtained by the attacker exceeds the threshold will the user's privacy be disclosed. Therefore, Zheng et al. [5] and Yang et al. [6] proposed two mechanisms combining partial publication and anonymous publication. These two mechanisms protect user privacy by reducing the number of public reviews published by users. However, they did not anonymize the display name and avatar in the public reviews, making it easy for attackers to use this information to carry out cross-platform CUIA to obtain more private information from users.

2.3. Usefulness of Review. Online consumer reviews (OCRs), known as electronic word-of-mouth (WOM), are the experience, usefulness, performance about the business, brand, product, service, etc., published on the Internet by consumers [34]. Good OCRs can effectively help consumers make choices without being familiar with the business, brand, product, or service. To study the mechanism by which the OCRs influence consumer purchasing behavior, Sussman et al. [17] proposed an information adoption model, as shown in Figure 1. Information usability is an important factor in determining consumers' adoption of information. Therefore, we usually use usefulness to measure the effectiveness of reviews, which is embodied in two aspects: review quality and source credibility.

Generally speaking, evaluation criteria of review quality [35] include the review content, as well as the accuracy, relevance, timeliness, and length of the description of the review content about businesses and products. In general, the more accurate the content of the description is, the greater the relevance is, the stronger the timeliness is, and the longer the length is, the higher the quality of the review is. The existing CSLBSS is used to use ratings (e.g., 5 stars), scores (e.g., 10 points), and thumb-up numbers (thumb-up denotes agreement with the content described in the review) to measure a reviewers' approval of businesses and products [5, 6].

For source credibility, since it is difficult to judge whether the reviewer is credible, the participants judge the credibility of the reviewer's source by obtaining the reviewer's profiles from OCRs. In addition, out of consideration of different interests, the CSLBSS platform, businesses, and consumers intentionally publish some false reviews [36]. This also reduces the credibility of the source to a certain extent. Xu et al. [37] pointed out that profiles, such as photos and reputation [38], can improve the credibility of the reviewer perceived by the recipient. The more profiles disclosed by the reviewer, the higher the credibility is, and the easier it is for consumers to adopt the review [18].

Starting from the privacy risks of the CSLBSS, we studied the strong privacy protection mechanism of partial publication (reviews that meet the publication conditions are anonymous) and how to ensure the system utility under this strong privacy protection mechanism.

3. Preliminary

3.1. Statistical Inference Attack. In typical LBSs, to enjoy the service, the user needs to submit a service request, containing ID, location, POI, etc., to the LBS service provider (LSP). The massive service requests make the user vulnerable to statistical inference attack. For example, LSA attacks and RSA attacks [15] are defined as collecting historical query information of target users, analyzing the geographical distribution probability and time period distribution probability of their query, and inferring the area where their family and company are located, including personal preferences and living habits. In CSLBSSs, each review corresponds to a business, and the business uniquely corresponds to a physical address. The distribution probability of user reviews can also reflect user's sensitive locations, causing them to also easily suffer from statistical inference attack.

In CSLBSSs, reviews reflect where the user has been (a business corresponds to a unique geographic location, and a

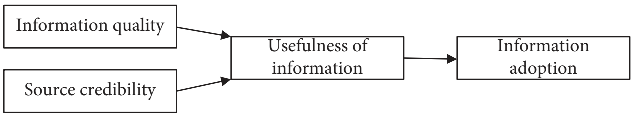


Figure 1: Information adoption model.

user review on a business means that he or she has been there or have related experience). Although CSLBSSs crowdsource reviews, users are still subject to statistical inference attacks, since the distributed probability of reviews can reflect their sensitive locations. For example, a user will review some Chinese restaurant in a mall at noon every day for a long time. If there are companies in the mall's office buildings, it is likely that the target user works in one of the companies and is an employee who loves Chinese food. For statistical inference attacks, Zheng et al. [5] pointed out that adversaries can infer a user's privacy based on the distribution of reviews. Then, Yang et al. [6] proposed IEPP mechanism. In this mechanism, users with similar probabilities of reviews in an area are allowed to publish their reviews. For example, users A and B are allowed to publish reviews if the probability of publishing reviews in area G is in the range $(p - \theta, p + \theta)$. However, neither of the two mechanisms takes into account the behavior patterns of users that characterize user behavior over the long term. Consider the anonymous group containing 3 users: A, B, and C. Assume the probabilities of publishing reviews of them in area G is in the range $(p - \theta, p + \theta)$ during the period T_1. Then, we can formalize the probabilities of publishing reviews of them as $\lim_{\theta \longrightarrow 0} p_A^{T_1} = \lim_{\theta \longrightarrow 0} p_B^{T_1} = \lim_{\theta \longrightarrow 0} p_C^{T_1} = p$.

When these users publish a large number of reviews over the long term, denoted as $T_1 + nT$ (T is the period for the system to update the reputation score of users and the business), we can compute the probabilities of publishing reviews, as shown in the formula:

$$\begin{aligned} \lim_{nT \longrightarrow 0, \theta \longrightarrow 0} p_A^{T_1 + nT} &= p', \\ \lim_{nT \longrightarrow 0, \theta \longrightarrow 0} p_B^{T_1 + nT} &= p'', \\ \lim_{nT \longrightarrow 0, \theta \longrightarrow 0} p_C^{T_1 + nT} &= p'''. \end{aligned} \tag{1}$$

Since individual behavior patterns are not exactly the same, p', p'', and p''' are not exactly the same at the time $T_1 + nT$. For example, if $p' > p'' = p'''$, the adversary will find that user A has a higher probability of publishing reviews in area G and infer that G is the sensitive area of A.

3.2. Connecting User Identities Attack. In our scenario, by analyzing user personal information, relationships, and UGC, the adversaries can launch CUIA to link user identities across social networks and obtain user privacy. CUIA discussed in our paper includes two types: CUIA on the same social network (short for SSN-CUIA) and CUIA across social networks (short for ASN-CUIA). Among them, SSN-CUIA, as a special case of ASN-CUIA, refers to mining and analyzing all the information of a user on the same social network to determine the user's personal information as much as possible. To protect privacy, most users use pseudonyms and fake avatars when publishing reviews. However, privacy will inevitably be leaked when the user publishes reviews that include photos, or organization, etc. Considering that the user profiles on the same social network is limited, if the adversary wants to obtain more dimensional profiles of the user, they need to launch ASN-CUIA to link the user identities. Figure 2 shows the specific process of ASN-CUIA.

When user u_j publishes a review in CSLBSS, the adversary first obtains their attributes $A_{ii} = \{a_j | j = 1, 2, \ldots, m\}$ (a_i is the j-th attribute of u_j) on the CSLBSS pl_i based on the review. Let $\langle u_i, pl_i \rangle$ be the user u_i on the pl_i. In the first stage, the adversary links user identities by obtaining the user's information on other social networks. In this stage, the adversary uses the user's display name as the keyword to search the user's information across different social networks, which is represented as $\langle u_{i+j}, pl_{i+j} \rangle, j = 1, 2, \ldots,$. These searched information is highly similar to the display name and is likely to belong to the same user. Then, for each $\langle u_{i+j}, pl_{i+j} \rangle$, the adversary crawls attributes $A_{i+j,i}$ of u_{i+j} on pl_{i+j} and connects u_{i+j} identities for the first time. Through these processes, the adversary finally obtains the consistent and more dimensional attributes of the user across different social networks. To distinguish from the original attributes, the attributes finally obtained in the first stage is represented as $A_{ii}' = \{a_1, a_2, \ldots, a_m, \ldots, a_{m+n}\}$. In the second stage, the adversary extracts the key attributes of A_{ii}', such as e-mail, phone number, QQ, and WeChat, and conducts the next round of search and user identification. In this stage, the adversary can further obtain and determine attributes (represented as $A_{ii}'' = \{a_1, a_2, \ldots, a_m, \ldots, a_{m+n}, \ldots, a_{m+n+r}\}$) with a higher level of user privacy than the first stage through data mining and analysis.

3.3. The Consistency between the User Rating and the Business Reputation Score. Based on the above statement, consumers still consider a review credible based on the consistency between the user rating and the business reputation score (short for score consistency) when the real identities of most users have not been disclosed. In this section, we define the score consistency as follows.

Definition 1 (Score Consistency): assume the rating of user u_i for business b_j is $r_{ij}(T_1 + nT)$ at time $T_1 + nT$ and the score of business reputation is $R_{b_j}(T_1 + (n-1)T)$. Then the score consistency of u_i and b_j refers to the difference between $r_{ij}(T_1 + nT)$ and $R_{b_j}(T_1 + (n-1)T)$, which is expressed as

$$\psi\Big(r_{ij}(T_1 + nT), R_{b_j}(T_1 + (n-1)T)\Big) = \Big| r_{ij}(T_1 + nT) - R_{b_j}(T_1 + (n-1)T) \Big|. \tag{2}$$

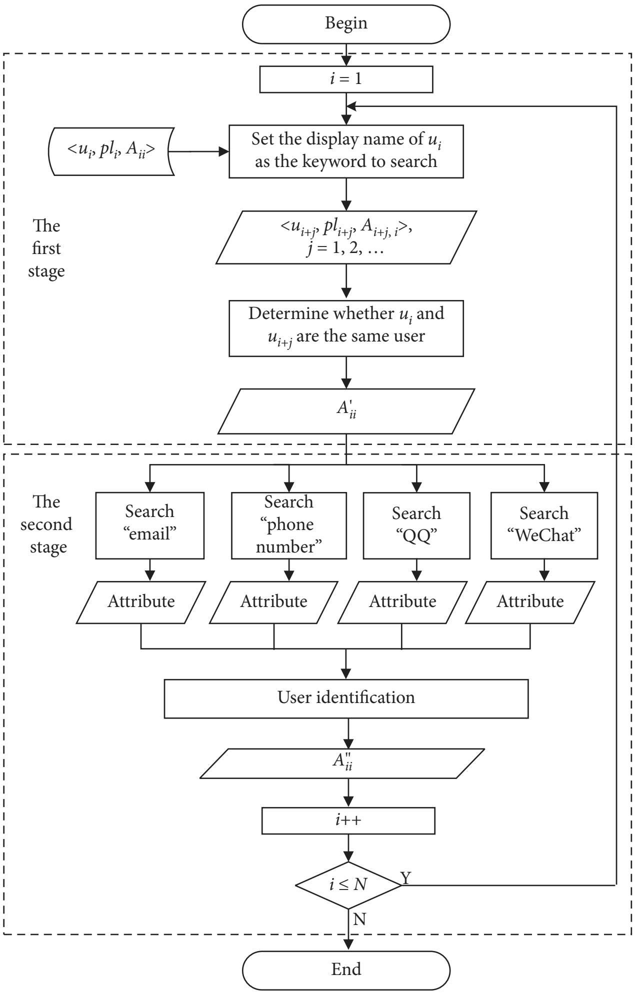


FIGURE 2: ASN-CUIA.

For simplicity, we abbreviate $\psi(r_{ij}(T_1 + nT), R_{b_j}(T_1 + (n-1)T))$ as ψ.

Generally speaking, a smaller ψ means the smaller difference between $r_{ij}(T_1 + nT)$ and $R_{b_j}(T_1 + (n-1)T)$; namely, the user rating and the business reputation score are more consistent. Referencing the paper [18], a smaller ψ also means higher credit of the review. Then, we clarify this conclusion from the perspective of benefit.

Specifically, the reviews for businesses are divided into three categories: false-positive review, false-negative review, and real review. False-positive review means that the business induces or hires users to give ratings that are significantly higher than the business reputation score in order to improve its own reputation score; that is, the ψ is greater. False-negative review means that consumers give a business a lower rating unequally; that is, the ψ is greater. The real review means that the user rating is basically the same as the business reputation score; that is, the ψ is smaller. Therefore, the smaller ψ is, the closer the user rating is to the business reputation score and the more credible consumers consider a review. Note that the business reputation score reflects the majority of consumers' rating for the service of a business service. Therefore, it can reflect the real service quality of the business more objectively than a single user rating. For the extremely low or extremely high user rating given by a very small number of users, although they are not false reviews, they cannot reflect the real service quality of the business due to the large score difference. Therefore, in reality, they will not affect consumer decisions. That is, a review with a higher ψ has low credibility.

3.4. Voting Decision Rule. To improve the usefulness of reviews, Yang et al. [6] used Voting Decision Rule [39] to improve the usefulness of reviews. In the current CSLBSSs, a 10-point scale or 5-star rating is used to rate businesses, and the most commonly used is the 5-star rating. Our paper uses 5-stars rating to rate the business service. Let r_{ij} be the user rating of which user u_i scores business b_j and τ be the threshold at which each user agrees to recommend a business to other users. u_i makes a decision d_{u_i} (use h_1 for approval and h_0 for the opposition) on whether or not to agree to recommend a business to other users, denoted by $d_{u_i} = 1$ (agree) and $d_{u_i} = 0$ (not agree). Then, for users u_i, $r_{ij} \ge \tau$ means they agree to recommend b_j and $d_{u_i} = 1$. In contrast, u_i, $r_{ij} \le \tau$ means they do not agree to recommend b_j and $d_{u_i} = 0$.

We consider the case that u_i and $u_{i'}$ rate business b_j, respectively. For a constant τ, there exists $r_{ij} < \tau$ and $r_{i'j} > \tau$. However, since u_i and $u_{i'}$ have different subjective experiences with b_j's service, $r_{i'j} > \tau$ does not mean that $u_{i'}$ is in favor of recommending b_j. Further, the business reputation score is dynamic process, and a constant τ cannot reflect changing process of the business's service quality. To address the problem, we consider using score consistency as a dynamic indicator to assess the usefulness of review. That is, as long as $\psi \le \psi_d$, i.e., $r_{ij} - R_{b_j} \in [-\psi_d, \psi_d]$, the user rating r_{ij} is considered credible and useful. In other words, it also means that u_i approve of R_{b_j}. Then, the approval or otherwise of the business in literature [6] becomes approval or otherwise of the current business reputation score.

Based on the dynamic τ and evaluation criteria, the overall binary decision of whether users agree or not is expressed as formula (3) [40].

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

Among them, Est represents the overall decision. $\sum_{i=1}^{n} d_{u_i} \ge \lambda$ represents that the choice of at least λ users out of n users is h_1, and Est is the result of approval. This paper uses $\rho = (\lambda/n)$ to indicate the threshold of approval.

3.5. Beta Reputation Mechanism. According to Section 3.4, the user's choice determines the overall decision based on the number of approvals. Intuitively, everyone's user reputation score is different, and the credibility of their reviews is also different. To reduce the influence of false-positive and false-negative reviews and improve the credibility of the overall decision, we compute the user reputation score and business reputation score based on Beta Reputation Mechanism [41]. Suppose that at time $T_1 + nT$, $u_i, i = 1, 2, \ldots, m$ publishes a review for b_j and r_{ij} is the user rating. Then, we get the decision vector, as shown in the following formula:

$$d(T_1 + nT) = \left[d_{u_1}(T_1 + nT), d_{u_2}(T_1 + nT), \ldots, d_{u_m}(T_1 + nT)\right]^T, \tag{4}$$

where $d_{u_i}(T_1 + nT)$ represents the binary decision made by u_i to b_j at time $T_1 + nT$. $d_{u_i}(T_1 + nT) = 1$ and $d_{u_i}(T_1 + nT) = 0$ represent opinions for approval or refusal, respectively. Then, at time $T_1 + nT$, the rule of global decision is shown in the following formula:

$$\text{Est} = f\left(\omega_{u_i}(T_1 + nT), d_{u_i}(T_1 + nT)\right)\begin{cases} 1, & \text{if } \sum\limits_{i=1}^{n} \omega_{u_i}(T_1 + nT) \cdot d_{u_i}(T_1 + nT) \geq \rho, \\ 0, & \text{if } \sum\limits_{i=1}^{n} \omega_{u_i}(T_1 + nT) \cdot d_{u_i}(T_1 + nT) < \rho, \end{cases} \tag{5}$$

Based on (5), we get the weight vector $\omega(T_1 + nT)$, as shown in

$$\omega(T_1 + nT) = \left[\omega_{u_1}(T_1 + nT), \omega_{u_2}(T_1 + nT), \ldots, \omega_{u_m}(T_1 + nT)\right]^T, \tag{6}$$

Here, $\omega_{u_i}(T_1 + nT)$ denotes the weight of the decision made by u_i in the overall decision and it is determined by the user reputation score of u_i before $T_1 + nT$. The formula for calculating the weight is shown in

$$\omega_{u_i}(T_1 + nT) = \frac{R_{u_i}(T_1 + (n-1)T)}{\sum_{j=1}^{n} R_{u_j}(T_1 + (n-1)T)}, \tag{7}$$

where $R_{u_i}(T_1 + (n-1)T)$ represents the user reputation score of u_i at $T_1 + (n-1)T$, and the specific calculation is as in formula (10). According to formula (7), $\sum_{i=1}^{n} \omega_{u_i}(T_1 + nT) = 1$. Here, $\omega_{u_i}(T_1 + nT)$ is a relative value that depends on the user reputation score of m users who published reviews for the business b_j at time $T_1 + nT$ before the time $T_1 + (n-1)T$. Different users who published reviews for b_j have different user reputation score. But for any user among them, the higher the user reputation score at the previous moment is, the greater its weight is, and the greater its impact on the overall decision is.

In addition, we define the positive rating $\varphi_{u_i}(T_1 + nT)$ and the negative rating $\eta_{u_i}(T_1 + nT)$ of u_i, as shown in

$$\varphi_{u_i}(T_1 + nT) = \varphi_{u_i}(T_1 + (n-1)T) + \nu_{u_1}(T_1 + nT), \tag{8}$$

$$\eta_{u_i}(T_1 + nT) = \eta_{u_i}(T_1 + (n-1)T) + \nu_{u_2}(T_1 + nT), \tag{9}$$

where $\varphi_{u_i}(T_1 + (n-1)T)$ denotes the times of which u_i's decision before time $T_1 + nT$ is consistent with the global decision Est. $\nu_1(T_1 + nT)$ denotes whether u_i's decision before time $T_1 + nT$ is consistent with the global decision *Est*, denoted by $\nu_1(T_1 + nT) = 0$ (they are not consistent) and $\nu_1(T_1 + nT) = 1$ (they are consistent). $\eta_{u_i}(T_1 + (n-1)T)$ denotes the times of which u_i's decision before time $T_1 + nT$ is not consistent with the global decision Est. $\nu_2(T_1 + nT)$ denotes whether u_i's decision before time $T_1 + nT$ is not consistent with the global decision Est, denoted by $\nu_2(T_1 + nT) = 0$ (they are consistent) and $\nu_2(T_1 + nT) = 1$ (they are not consistent).

Then, we can calculate the user's score at time $T_1 + nT$, as shown in

$$R_{u_i}(T_1 + nT) = \frac{\varphi_{u_i}(T_1 + nT) + 1}{\varphi_{u_i}(T_1 + nT) + \eta_{u_i}(T_1 + nT) + 2}. \tag{10}$$

In formula (10), because each user publishes no reviews in the initial state, it is impossible to judge the consistency of their decision with the overall decision, so we set the initial value of the user reputation score of each user as 0.5. After calculating the user reputation score $R_{u_i}(T_1 + nT)$, the user rating determines the business reputation score. Considering the difference in the influence (i.e., weight) of users' reviews for business b_j at time $T_1 + nT$ on the overall decision (as it is known in formula (5)), we can get the method to calculate the business reputation score of business b_j at time $T_1 + nT$, as shown in

$$R_{b_j}(T_1 + nT) = \frac{R_{b_j}(T_1 + (n-1)T) + \sum_{i=1}^{m} \omega_{u_i}(T_1 + nT) \cdot r_{ij}(T_1 + nT)}{2}. \tag{11}$$

It can be seen from formula (11) that the business reputation score of a business is jointly determined by the business reputation score at time $T_1 + (n-1)T$ and the weighted sum of the user scores of all users at time $T_1 + nT$ for the business.

4. Motivation and Model

4.1. Motivation and Basic Idea. In CSLBSSs, users submit reviews to rate businesses, and consumers make decisions based on reviews and the business reputation score. Considering that the adversary can legally obtain reviews from the CSLBSSs and user profiles from other social networks, it leads to the inevitably leaking of the user privacy due to CUIA and SIA. Also, the more user profiles the adversary obtains, the more accurate the user privacy can be inferred. Although partial publication and anonymity mechanism can reduce the risk of privacy leakage caused by excessively publishing reviews, users still suffer from CUIA and SIA, especially CUIA. As long as the display name is not anonymized, the adversary can exploit it as the keyword to launch CUIA. However, the partial publication and anonymity mechanism can also reduce the usefulness of review and the utility of the system. Besides, the consumers need more public reviews and more objective business reputation score to enjoy a better service. Hence, how to balance privacy protection and system utility is a problem that needs to be addressed urgently.

To address the above problems, the basic idea of our paper is to partially publish public reviews of which the usefulness of review is high. At the same time, we need to anonymize the display name and avatar of all reviews that are allowed to be publicly published. Based on above two mechanisms, it can achieve a balance between privacy protection and system utility. As we stated in the section Introduction, the system hopes to publish as many reviews that consumers consider credible as possible. Whether consumers consider a review credible depends on the usefulness of the review, which is measured by score consistency. The partial publication mechanism can publish public reviews of which the usefulness of review is high and suppress reviews of which the usefulness of review is low. The complete anonymity mechanism ensures that the adversary cannot infer user privacy by exploiting the display name to launch CUIA. Therefore, the basic idea can protect user privacy by resisting CUIA and SIA while improving system utility.

4.2. Threat Model. The goal of this paper is to protect user privacy while improving system utility. We do this by a partial publication mechanism which anonymously publishes reviews with a high score consistency. In our threat model, the CSLBSSs are entities of "honest but curious". In other words, the CSLBSSs execute the agreement honestly, but they also collect and analyze users' data curiously and infer the POIs that users have visited to provide users with better personalized recommendation services. The CSLBSSs have no subjective maliciousness. However, to further attract users through the social relationship, CSLBSSs tend to establish a personal homepage for each user, which makes it easy for the adversary to get all reviews and personal profiles. By analyzing users' data, the adversary can infer the user's privacy. In addition, we assume that CSLBSSs are credible and cannot be hacked, which is also the basic trust or agreement between users and service providers.

Furthermore, adversaries are subjective, malicious, and highly motivated entities. Their goals are to infer as much privacy as possible about the users, including real identity and organization. The adversary may be any user who can access CSLBSSs. In our attack model, the adversaries can use any tools and methods to collect user reviews on the CSLBSSs, as well as the personal profiles on other social networks including demographic information. They use this information as background knowledge to implement CUIA and SIA such that they can identify as much real personal information of users as possible.

In this paper, the user's privacy we consider mainly includes location privacy, identity privacy, and preference privacy. The user's privacy is considered threatened if the following conditions are met:

(1) The adversary can directly or indirectly infer areas where users frequently visit and can even determine accurate information such as their home address and workplace

(2) The adversary directly or indirectly infers the personal profiles such as real name, photo, and phone number

(3) The adversary directly or indirectly infers the preference information such as consumption habits and behavior habits

4.3. System Model. In this paper, the system model consists of four parts: user/business, CSLBSS, reputation system, and review list. The general business data process includes 5 steps, as shown in Figure 3. Among them, the business publishes services on CSLBSS, and the user obtains services from CSLBSS and reviews the business's services. CSLBSS is the management center of the entire system and mainly contains two functions: (1) provide a service interface for the users/business (including registration, login, and review) and (2) provide the users/business information and review information to the reputation system. The reputation system calculates the user reputation score and the business reputation score according to the user/business information and review information from the CSLBSS to determine the status of the review (published or not, public or anonymous) and give feedback of the calculation results to CSLBSS. The review list shows the calculation results from the reputation system which includes the user reputation score, the business reputation score, and the review content.

(i) User/business: in the CSLBSS, there are M businesses, denoted as $B = \{b_1, b_2, \ldots, b_M\}$. Each business b_i has a unique location denoted by coordinate (x_i, y_i), which is the longitude and latitude. There are N users, denoted as $U = \{u_1, u_2, \ldots, u_N\}$. Each

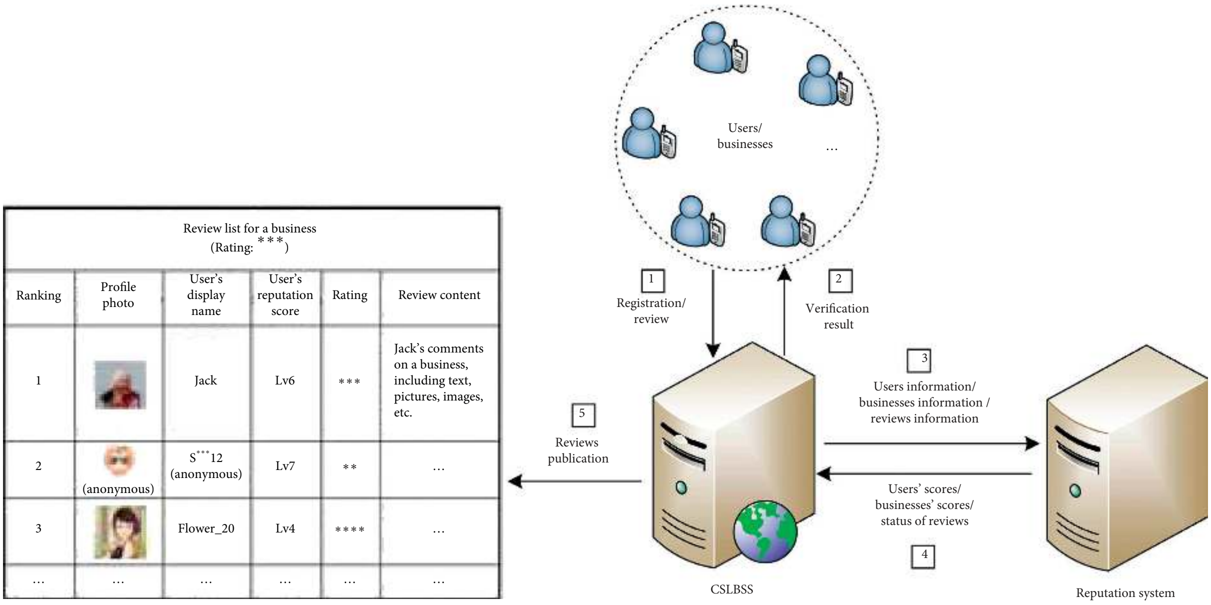


Figure 3: System model.

user u_j publishes a review and gives a user rating (e.g., score or star rating, denoted as r_{ij}) for b_i. Generally speaking, users need to register an account on the system and successfully log in to obtain services and publish reviews. To ensure the authenticity of the user, a mobile phone number or Short Messaging Service (SMS) verification code is required when registering. At the same time, users are allowed to customize their pseudonym (namely the user name displayed in the review list, also called display name), instead of the real name to uniquely identify the user, so as to protect the privacy of the user's identity. In this paper, for reasons such as user naming preference and reauthentication complexity, we assume that each user will not change the pseudonym for a long time. It is agreed that each user can only review on relevant experience to ensure the objectivity and authenticity of the review.

(ii) CSLBSS : the CSLBSS mainly includes 3 functions: (1) CSLBSS provides users with interfaces to register, log in, obtain services and reviews, store and update users' pseudonyms, and bound mobile phone numbers, avatars, reviews, and the user reputation score and other profiles. It also provides a platform for the business to display products, store and update the business reputation score. (2) The CSLBSS provides the user reputation score, the business reputation score, and user review to the reputation system and updates relevant information about users and businesses in real-time. (3) Based on the user reputation score, the business reputation score, the score consistency, and the threshold ψ, the CSLBSS selects reviews that meet the publication criteria and filters out some reviews with low usefulness and credibility. In addition, the published reviews are sorted according to ψ and the user reputation score.

(iii) Reputation system: the reputation system is the core computing part of the entire system. It is responsible for calculating the business reputation score based on user rating and the user reputation score based on the usefulness of the review. Then, it determines the objectivity of the review and whether it will be published based on the score consistency. In a cycle, the reputation system will perform the above process to update the business reputation score, the user reputation score, and the status of reviews.

(iv) Review list: CSLBSS publishes the calculation results of the reputation system and sorts the reviews by the usefulness of review in the form of a web page, namely a review list for users to use. Therefore, the review list is the page display of the reputation system and an important reference for users when they consume. As shown in Figure 3, the review list contains the business reputation score and user reviews that can directly reflect the quality of business services, as well as the user reputation score that indirectly affects users' acceptance of reviews (indicated by user membership levels in the figure).

5. Privacy Protection Framework and Core Algorithms

5.1. Privacy Protection Framework. Zheng et al. [5] and Yang et al. [6] pointed out that users are more willing to publicly publish as many reviews as possible to obtain a higher user reputation score. However, users will suffer CUIA and SIA if any one public review discloses the display name. Therefore, users need to publicly publish as many reviews as possible without disclosing the display name. In this paper, we propose two mechanisms to address the above problem: (1) the partial publication mechanism. This mechanism publishes the reviews of which the usefulness of review is high and suppresses reviews of which the usefulness of review is low according to ψ. Specifically, we give two thresholds ψ_d and ψ_p for the score consistency. ψ_d is related to voting decision rule and the overall decision. For a review, if there exists $\psi \leq \psi_d$, consumers will consider the user rating and the business reputation score highly consistent. That is, consumers consider the business reputation score credible. ψ_p determines which reviews are published publicly. A review can be published anonymously if $\psi \leq \psi_p$. A review, whose user rating is not objective enough and that has low reference value to consumers, cannot be published if $\psi > \psi_p$. On the one hand, reviews whose $\psi > \psi_p$ not being published can reduce disclosure of user reviews and privacy risks. On the other hand, reviews whose $\psi \leq \psi_p$ being retained can improve the usefulness of reviews that are publicly published. (2) Anonymize display names and avatars in publicly published reviews that meet the publication conditions. We find that the display name is the key factor of privacy leakage in review publication, and anonymous treatment of display name can prevent privacy leakage caused by display name in review publication scenario from the root.

Furthermore, we first sort the publicly published reviews according to the score consistency ψ. Then, we sort the reviews with the same value of ψ according to the user reputation score. The purpose of this step is to ensure that useful and reliable reviews are ranked first. The reason why ψ is the main keyword for ranking is that it best reflects the degree of the score consistency. The user reputation score is calculated based on the reviews for different businesses. It reflects the degree of consistency between the user's decision and the overall decision and does not reflect the degree of consistency between a specific user rating and the business reputation score. Moreover, even if the user reputation score is high, it does not mean that a certain user rating can accurately reflect the business service. However, if ψ is the same, the higher the user reputation score is, the higher the usefulness and reliability are. Therefore, in this paper, the score consistency ψ is used as the primary keyword for ranking, and the user reputation score is used as the secondary keyword for ranking. The specific privacy protection process is shown in Figure 4.

As shown in Figure 4, the key part of the privacy protection framework includes calculating the score difference ψ and judging whether the privacy metrics are met. If the privacy metrics are not met, the review will not be published; if the privacy metrics are met, the review can be published after being anonymized. Next step, the system needs to update the user reputation score and the business reputation score and publish review list.

5.2. Core Algorithm

5.2.1. Calculate the Score Difference. In this paper, the score difference ψ is a key parameter to realize privacy protection and improve the usefulness of review and is calculated by the reputation system. At time $T_1 + nT$, the reputation system first obtains the user rating $r_{ij}(T_1 + nT), i = 1, 2, \cdots, m$ from CSLBSS and then obtains the business reputation score $R_{b_j}(T_1 + (n-1)T)$ at time $T_1 + (n-1)T$. Then calculate the score difference $\psi(T_1 + nT)$. The specific process is shown in Algorithm 1.

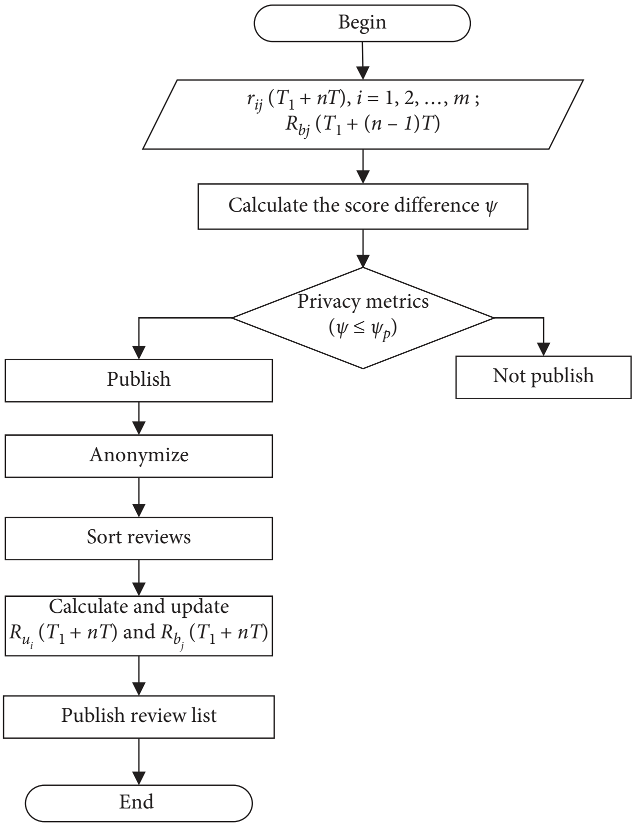


Figure 4: The privacy protection framework.

5.2.2. Determine the Privacy Metric. In this process, determining privacy metric is mainly to determine which reviews can be published according to the threshold ψ_p. In our paper, whether a review can be published depends on the score difference ψ being bound to the threshold ψ_p. In the process, we first obtain ψ_p and observe whether ψ exceeds ψ_p. For a review, if $\psi \leq \psi_p$, we will publish it publicly, namely any user can browse it on the CSLBSS. If $\psi > \psi_p$, we will not publish it publicly; namely, only the user who published it can browse it on his own page. Note that the reviews published in our paper are anonymous, so the threshold ψ_p is used to roughly identify false or extreme reviews (non-objective reviews) and filter those with low usefulness of review to reduce excessive disclosure of user privacy while improving the usefulness of review that can be published. Therefore, ψ_p is an empirical parameter and is relatively larger than ψ_d. In other words, in this case, there will be more reviews that satisfy the condition. Relative to ψ_p, the threshold ψ_d determines whether a user approves of the business reputation score. The smaller the ψ_d is, the more the user approves of the review. If $\psi \leq \psi_b$, the user will approve of the business reputation score. That is, the user considers the current business reputation score objective; otherwise, the user will make an opposing decision. That is, the user considers the current business reputation score inconsistent with the actual service. The specific privacy metric determination algorithm is shown in Algorithm 2.

```
(1) Obtain $r_{ij}(T_1 + nT), i = 1, 2, \ldots, m$;
(2) Obtain$R_{b_j}(T_1 + (n-1)T)$;
(3) $i = 1$;
(4) while $(i \leq m)$ do
    (i) $R(T_1 + nT) \longleftarrow r_{ij}(T_1 + nT)$; //$R(T_1 + nT)$ is the set of user reputation score
        $\psi_i(T_1 + nT) = |r_{ij}(T_1 + nT) - R_{b_j}(T_1 + (n-1)T)|, i = 1, 2, \ldots, m$,
        $\psi(T_1 + nT) \longleftarrow \psi_i(T_1 + nT)$; // the set of the score difference
        i + +;
(5) return $(\psi(T_1 + nT), R(T_1 + nT))$;
```

Algorithm 1: Calculate the score difference.

```
(1) Obtain the set of the score difference $\psi(T_1 + nT)$and the threshold $\psi_p$;
(2) $i = 1, k = 0$;
(3) while $(i \leq m)$ do
        if $(\psi_i(T_1 + nT) < = \psi_p)$
            $\psi'(T_1 + nT) \longleftarrow \psi_i(T_1 + nT)$;//the set of the score difference that satisfies the privacy metric;
            $k = k + 1$;//the number of published reviews
        $i + +$;
(4) Obtain all reviews corresponding to the score difference in$\psi'(T_1 + nT)$ and get the set of reviews $RL'_{ij}(T_1 + nT)$ needed to be
    published;
(5) return $(\psi'(T_1 + nT), RL'_{ij}(T_1 + nT), k)$;
```

Algorithm 2: Determine the privacy metric.

5.2.3. Update the User Reputation Score and the Business Reputation Score. The process of updating the user reputation score and the business reputation score mainly involves calculating the user reputation score and the business reputation score at time $T_1 + nT$. The user reputation score reflects the objectivity of the review for the business's service, which is represented by the ratio of the times that the user's decision is consistent with the overall decision to the total times of decisions made by the user. The business reputation score reflects the quality of the business's service. It is the average value of the business reputation score at time $T_1 + (n-1)T$ and the sum of the user rating' weight at time $T_1 + nT$. The specific process is shown in Algorithm 3.

5.2.4. Publish Review List. Publishing the review list is the final stage of privacy protection. After determining the privacy metric, the system has determined which reviews can be published. Our next step is to anonymize the display name and avatar and sort the review list to be published before publishing. In this paper, we first use uniform characters and pictures to replace the user's original display name and avatar, so as to protect user privacy while reducing computational complexity. Then, according to $\psi'(T_1 + nT)$, we sort publicly published user reviews and get a new review list. For the reviews whose score difference is the same, we then sort them according to the user reputation score $R'_{u_i}(T_1 + nT)$ and get a new review list. Through these two sorts, we can improve the usefulness of review. Finally, we publish the review list publicly; namely, any user can browse it on the CSLBSS. The specific process is shown in Algorithm 4.

6. Scheme Analysis

In this section, we focus on the privacy protection effect and the usefulness of the review of our scheme.

6.1. Privacy Protection Effect. Essentially, a user can be represented by a series of attributes that characterize the user's characteristics, expressed as $\overline{A} = \{a_1, a_2, \ldots, a_m, \ldots, a_{m+n}, \ldots, a_{m+n+r}, \ldots, a_s\}$. Assume that there are $m + n + r$ attributes that can be collected across different social networks and the total number of user attributes is s, and $m + n + r \leq s$. Therefore, the adversary identifying a user is to identify the user attributes. For an individual user, the more publicly published reviews are and the more attributes that are collected, the higher the probability that the individual user will be identified. For the CSLBSS, the more publicly published user reviews are, the more users will be identified and the higher the probability of the recognition rate of the CSLBSS. In this paper, we use public publication rate (PPR), user attribute recognition rate (UARR), and user recognition rate (URR) to describe the privacy protection effect.

Definition 2. (Public Publish Rate) : we define the ratio of the number of publicly published reviews to the total number of reviews as the PPR, expressed as

$$\text{PPR} = \frac{\sum_{i=1}^{N} |u'_i|}{\sum_{i=1}^{N} |u_i|}, \tag{12}$$

```
(1) Obtain the set of the user reputation score $R(T_1 + (n-1)T)$ at time$T_1 + (n-1)T$;
(2) Obtain $\psi_p$ and $\psi_d$;
(3) $i = 1$;
(4) while $(i \le m)$ do
        Calculate the times $\varphi_{u_i}(T_1 + nT)$ of the decision made by user $u_i$ at time $T_1 + nT$
        in the overall decision;
        $i++$;
(5) $i = 1$;
(6) while $(i \le m)$ do
        Calculate the times $\eta_{u_i}(T_1 + nT)$ that the decision made by user $u_i$ is
        inconsistent with the global decision Est before time $T_1 + nT$;
        Calculate the user reputation score $R_{u_i}(T_1 + nT)$ of $u_i$ at time $T_1 + nT$ ; before time $T_1 + nT$;
        $R(T_1 + nT) \longleftarrow R_{u_i}(T_1 + nT)$;
        $i++$;
(7) Calculate the business reputation score $R_{b_j}(T_1 + nT)$ of $b_j$ at time $T_1 + nT$;
(8) Return $(R(T_1 + nT), R_{b_j}(T_1 + nT))$;
```

Algorithm 3: Update the user reputation score and the business reputation score.

```
(1) Input $\psi'(T_1 + nT)R_{u_i}'(T_1 + nT),RL_{ij}'(T_1 + nT)$;
(2) According to the $\psi'(T_1 + nT)$, sort the $RL_{ij}'(T_1 + nT)$and get the new review list$RL_{ij}''(T_1 + nT)$;
(3) According to the $R_{u_i}'(T_1 + nT)$  sort the reviews whose score difference is the same in $RL_{ij}'(T_1 + nT)$ and get the new review
    list$RL_{ij}''(T_1 + nT)$;
(4) Return $(RL_{ij}''(T_1 + nT))$;
```

Algorithm 4: Publish review list.

where $\sum_{i=1}^{N}|u_i'|$ and $\sum_{i=1}^{N}|u_i|$ represent the total number of reviews publicly published and the total number of reviews published by N users, respectively.

Definition 3 (user attribute recognition rate). For a user, each of their attributes contains different privacy information, and we use the privacy level to express this difference. In this paper, we introduce the attribute weight α_i to represent the privacy level. Then, the user attribute recognition rate is defined as

$$\text{UARR} = \frac{\sum_{i=1}^{m+n+r}\alpha_i'\cdot|\alpha_i'|}{\sum_{i=1}^{s}\alpha_i\cdot|a_i|}. \tag{13}$$

Here, $\sum_{i=1}^{m+n+r}\alpha_i'\cdot|\alpha_i'|$ represents the privacy information contained in the attributes that can be collected across different social networks. α_i' represents the attribute weight corresponding to attribute α_i'. Whether the adversary has obtained the user's attribute information α_i' is denoted by $|\alpha_i'| = 1$ (i.e., adversary obtains user's attribute information α_i') and $|\alpha_i'| = 0$ (adversary does not obtain user's attribute information α_i'). $\sum_{i=1}^{s}\alpha_i\cdot|a_i|$ represents the total amount of privacy information contained in all attributes, $\sum_{i=1}^{s}\alpha_i = 1$ and $|\alpha_i'| = 1$. Then, formula (13) can be further simplified as

$$\text{UARR} = \sum_{i=1}^{m+n+r}\alpha_i'\cdot|\alpha_i'|. \tag{14}$$

Definition 4 (user recognition rate). Assume that a user is identified if $\text{UARR} \le \xi$. Let the total number of users be N and the number of users who has been identified be N'. Then, we define the user recognition rate as

$$\text{URR} = \frac{N'}{N}. \tag{15}$$

Conclusion 1. For the same user, the larger the UARR is, the more privacy of the user is leaked and the higher the probability of the user being identified is.

Proof. Based on the aforementioned analysis, the adversary can first obtain the set A_{ii} of attributes of u_i on the pl_i. Let A_{ii} contain m attributes and the corresponding privacy be denote as R_{ii}. Then, the adversary searches the display name of u_i through the Internet and get the set A_{ii} of attributes that can be visited. Let A_{ii} contain $m+n+r$ attributes and the corresponding privacy be denoted as R_{ii}''. We can draw that $R_{ii} \le R_{ii}' \le R_{ii}'' \le RR$ (RR represents the all privacy of the user) due to $A_{ii} \subseteq A_{ii}' \subseteq A_{ii}''$ and then the UARR is greater. Especially, all attributes of u_i will be searched if $R_{ii}'' = RR$. In other words, if all the privacy of u_i is leaked, the adversary can accurately identify them. Therefore, it can be proved that Conclusion 1 is true. Note that, although we use the CUIA as an example to verify the risk of privacy leakage, the obtained privacy is a conservative estimate. That is, the amount of privacy information leaked is at least R_{ii}''. If there exists a better attack tool or method, more private information will be leaked. □

Conclusion 2. In a CSLBSS, the greater the PPR is, the more users will be identified. The greater the *URR* is, the greater is the risk that privacy will be compromised.

Proof. We consider CSLBSS with different PPR, i.e., PPR_1 and PPR_2 $(0 < \text{PPR}_1 \leq \text{PPR}_2)$. The number of users who have been identified and the URR corresponding to PPR_1 is N_1 and URR_1, respectively. The number of users who have been identified and the URR corresponding to PPR_2 is N_2 and URR_2, respectively. For the same privacy protection scheme, we can easily get the conclusion $N_1 \leq N_2$, and then $(N_1/N) \leq (N_2/N)$, namely $\text{URR}_1 \leq \text{URR}_2$. Therefore, it can be proved that Conclusion 2 is true.

According to the derivation process of Conclusions 1 and 2, we know that there will be a potential risk of privacy leakage, as long as the user's display name is disclosed (that is, the review is publicly published). In addition, even if all the display names of the user have not been disclosed, the attacker can obtain some personal privacy information from the review content. Therefore, we adopt a strong privacy protection mechanism of complete anonymity and partial publication. That is, user reviews that meet the threshold range of the score difference are published publicly, and the user's display name and avatar must be anonymized before publication. □

Conclusion 3. The privacy protection framework proposed in this paper can effectively reduce the risk of privacy leakage.

Proof. For the individual user, the user attribute set obtained without using the privacy protection framework proposed in this paper is A''_{ii}, and the corresponding privacy information is R''_{ii}. All reviews are processed anonymously and have no personalized characteristics after using our privacy protection framework. It is difficult to identify the regional characteristics, namely the distribution characteristics of the published reviews, and they are not vulnerable to SIA. In addition, the display name is anonymized. The adversary cannot implement CUIA based on attributes that can uniquely identify users, and the user information across social networks is not easy to obtain. Assume that the adversary can obtain the user attribute set $\overline{A}''_{ii}$ and the corresponding privacy information is $\overline{R}''_{ii}$ when using our privacy protection framework. Then, $\overline{A}''_{ii} \subset A''_{ii}$ and $\overline{R}''_{ii} < R''_{ii}$. For a CSLBSS without using our privacy protection framework, the number of users that have revealed the display name is N'. The number of users who have been identified and the URR corresponding to the CSLBSS is $\overline{N}'$ and URR', respectively. For a CSLBSS using our privacy protection framework, the number of users that has revealed the display name is $N'' = 0$. The number of users who have been identified and the URR corresponding to it is $\overline{N}''$ and URR'', respectively. We can easily conclude that $N' \geq N''$. Then, there will be $\overline{N'} \geq \overline{N}''$; namely $\text{URR}' \geq \text{URR}''$. Therefore, the privacy protection framework proposed in this paper can resist CUIA and SIA and can effectively reduce the risk of privacy leakage. □

6.2. The Usefulness of Review. The factors affecting the usefulness of a review include the accuracy of the review content, relevance, timeliness, length, number of reviews, and reliability of the source. Considering the difference in the users' subjective experience, it is difficult to ensure the accuracy, objectivity, and reliability of individual reviews [18]. Therefore, this paper chooses two indicators, i.e., the score difference ψ and the user reputation score $R_{u_i}(T_1 + nT)$, to judge the usefulness of a review. The business reputation score is the overall score of a business by all users at a certain time and reflects the overall evaluation of the business services by historical users. The score difference ψ reflects the difference between individual user rating and overall evaluation. In other words, it reflects whether a user rating is objective and credible. The lower the ψ is, the more objective the score is and the more credible the review is. Furthermore, the user reputation score $R_{u_i}(T_1 + nT)$ reflects a user's reputation. Using it to measure the usefulness of review can better reflect the objectivity and credibility of a review. In addition, ψ and $R_{u_i}(T_1 + nT)$ are the primary keyword and the secondary keyword in sorting the usefulness of review, respectively. The reason is that $R_{u_i}(T_1 + nT)$ reflects the reputation of the overall credibility of an individual user, but it does not mean that each review of the individual user is objective and credible. However, for the same ψ, the higher the is, the more objective and credible the user's current review is.

7. Evaluation

In this section, we evaluate the utility and privacy of CSPPM on the real review datasets.

7.1. Dataset. We use two datasets, i.e., Dataset1 and Dataset2, to evaluate our scheme. Among them, Dataset1 is collected from Dianping, which contains businesses, user data, and reviews, only containing the text in the reviews but not containing pictures. It has become a dataset used in many scenarios [42]. In this paper, we use it to evaluate our scheme in terms of utility and privacy. The shortcoming of Dataset1 is that the reviews do not contain pictures so that it cannot be used to evaluate the ability of our solution to resist CUIA and SIA. Therefore, as a supplement to Dataset1, we design Dataset2 based on Dataset1. We randomly select 10 businesses and 560 different users from Dataset1. Then, we crawl the 560 users' reviews including pictures on the Dianping and their profiles from other websites to evaluate the ability to resist attacks. All these data make up Dataset 2. The statistical information of Dataset1 and Dataset2 is shown in Tables 1–3.

7.2. Evaluation Metric. In CSLBSS, both users and businesses desire as many highly credible reviews as possible to be published. The goal of the business is that more reviews will attract more consumers. The goal of the user is that more reviews will build more objective reputations for businesses while protecting their privacy. Therefore, we evaluate our scheme with respect to three metrics: system utility, user utility, and privacy. For existing researches, both LPA [5] and IEPP [6] are methods to protect user privacy in the scenario of review publication. Therefore, we select them for comparison with our scheme.

7.2.1. System Utility Metric. In the scenario of review publication, a basic fact is that the more public (non-anonymous) reviews a user has published, the more private information is leaked. To publish as many reviews as possible while protecting user privacy, a feasible idea needs to meet two requirements: (1) limit the number of reviews published publicly by each user; (2) each user publishes the maximum number of reviews allowed for (1). Therefore, all methods (i.e., our and [5, 6]) set different thresholds (the maximum number of reviews that each user is allowed to publish publicly) to implement this idea. Based on this idea, if a threshold is given, the system utility will depend on the number of reviews submitted by each user. For example, suppose the threshold is 3; that is, each user can publish no more than 3 reviews. In this case, the more reviews a user submits, the more reviews will be suppressed, and the lower the system utility will be. In other words, when users submit different numbers of reviews, the greater the difference in the ratio that the reviews are publicly published (Public Publish Rate, PPR), the lower the system utility is. Therefore, we use the PPR to measure the system utility. Considering the different thresholds set by different methods, we analyze the impact of different thresholds on the difference in the PPR under the same method.

Based on the above analysis, we divide the administrative area covered by Dataset1 into a $5 * 5$ grid. For LPA, we set the thresholds of the total number of reviews published publicly as 60, 70, 80, 90, 100, 110, 120, and 130, respectively. For IEPP, we set the threshold interval of user similarity as [1/2,2], [1/2,3], [1/2,4], and [1/2,5], respectively. For CSPPM, we set the thresholds for the score difference as 0.5, 1, 1.5, 2, 2.5, and 3.5, respectively. In addition, to evaluate the impact of the number of reviews submitted by individual users, Dataset1 is divided into Dataset3 (the number of reviews submitted by each user is less than 4) and Dataset4 (the number of reviews submitted by each user is not less than 4).

7.2.2. User Utility Metric. Users hope to publish as many credible reviews as possible. In the scenario of review publication, the usefulness of review reflects whether a review is credible. For a user, the greater the proportion of reviews considered credible, the higher the user utility is. For this, the existing literatures propose some metrics for evaluating the usefulness of review (number of thumbs-up [5] and user rating [6]). However, a metric that is considered credible should meet the two requirements: (1) users have evaluated most of reviews based on the metric; for example, 90% of the reviews are liked by users; (2) for a review, the evaluation result based on the metric should be consistent with the evaluation results of other metrics. Therefore, we measure user utility as the usefulness of a review and compare it with [5, 6] to prove that it is objective and credible.

Based on the above analysis, we evaluate the usefulness of reviews under different values of each metric. For number of thumbs-up, we set the likes interval as 0, [1, 100], [101, 200], [201, 300], [301, +∞]. For the user rating, we set the rating from 0 to 7.

7.2.3. Privacy Metric. The goals of the adversary include (1) identifying individual users with the acquired attributes and (2) identifying as many users in a CSLBSS as possible. The ability of a method to resist attacks reflects how difficult the adversary achieves the two goals. The stronger the ability to resist attacks is, the lower the risk of privacy leakage is. Therefore, we measure the ability to resist attacks as two metrics: UARR and URR. Among them, UARR is used to measure the degree of user attribute leakage. The greater UARR is, the greater the probability of an individual user being identified. URR is defined as the ratio of the number of users identified to the total number of users in a CSLBSS and is used to measure the risk of privacy leakage of a CSLBSS. The larger the URR is, the higher the risk of privacy leakage of a CSLBSS is.

In our scheme, the set of attributes used to evaluate the ability of our system to resist the CUIA and the SIA includes 11 attributes: display name, avatar, real photo, name, location, gender, age, education background, organization, contact information, and home address. We assign attribute weight to each attribute according to its privacy level. The more likely an attribute can uniquely identify a user, the higher the privacy level and the greater its attribute weight. The greater the privacy level of an attribute, the more privacy information it will leak after being acquired by the adversary. Therefore, we use the privacy leakage level to measure the privacy level. The lower the privacy leakage level is, the greater the privacy level is. Specifically, we divide the privacy disclosure level into 5 levels, as shown in Table 4.

7.3. Results

7.3.1. System Utility. Figures 5(a)–5(c) separately show the PPR of LPA, IEPP, and CSPPM on Dataset1, Dataset3, and Dataset4.

LPA sets the total number of reviews published publicly as the threshold. In Figure 5(a), as the threshold increases, the PPR trends on Dataset3 is almost horizontal while both on Dataset1 and Dataset4 are gradually increasing. The corresponding average number of reviews that each user can publish in each grid is about 2 and 3 when the thresholds are 60 and 80, respectively. In Dataset3, the proportion of the total number of reviews published by users whose number of submitted reviews is less than 4 and is 1 or 2 exceeds 90%. Therefore, the PPR is 97.52% when the threshold is 60 and the PPR is even 100% when the threshold is increased to 80. In Dataset4, since the number of reviews submitted by each user is not less than 4 and the total number of reviews published publicly is limited to not exceeding the threshold, some reviews will not be allowed to be published publicly, which makes the PPR in Dataset4 higher than in Dataset3. As the threshold increases, the number of reviews that users

Table 1: The statistical information of Dataset1.

Districts	Baiyun	Conghua	Panyu	Haizhu	Huangpu	Liwan	Tianhe	Yuexiu
Businesses	7	4	17	27	3	14	49	34
Users	13356	3161	16602	37970	7776	47970	81551	83951
Reviews	13929	3222	17562	43065	7843	53594	111805	103784

Table 2: The distribution interval of numbers of user reviews in Dataset1.

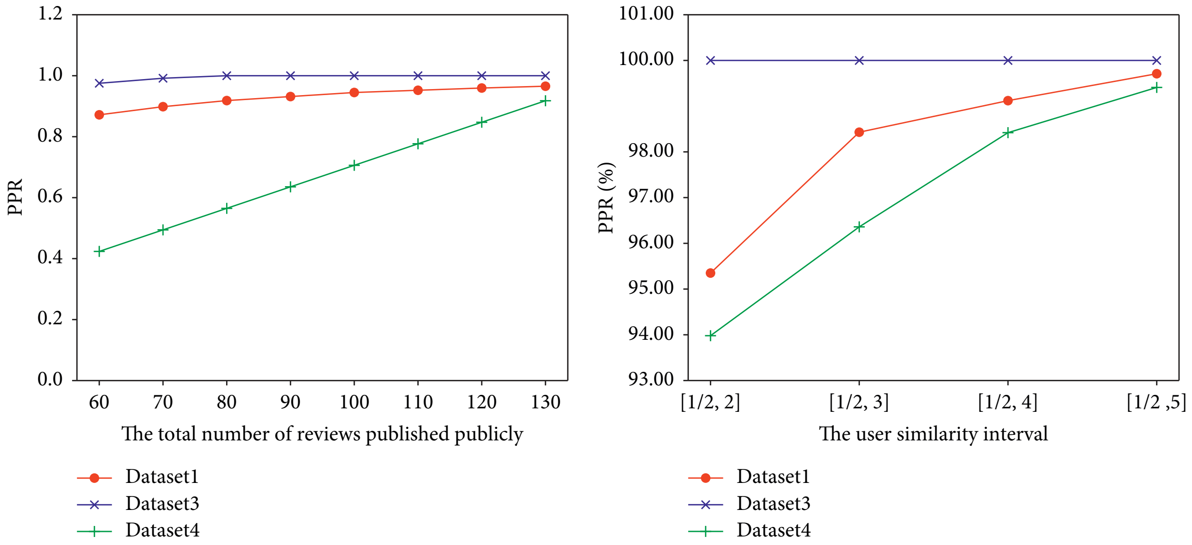
Interval	[1, 5]	[6, 10]	[11, 15]	[16, 20]	[21, +∞)
Users	234862	3297	469	127	41

Table 3: The statistical information of Dataset2.

Dianping	Districts	Businesses	Users	Reviews
	3	10	560	1957
Other websites	**Website**		**Type**	**Number**
	15		Blog	4
			Library	3
			Community	6
			Others	2

Table 4: The privacy leakage level and attributes that the adversary can obtain in different levels.

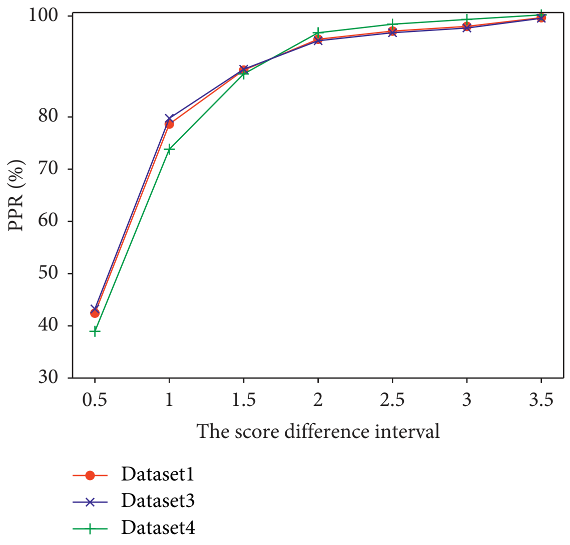
Level	Attributes that the adversary can obtain
0	None
1	Display name, avatar
2	Display name, avatar, real photo, name
3	Attributes involved in level 2 and associate them with the user's attributes across other social networks to determine gender, age, or location
4	Attributes involved in level 3, education background
5	Attributes involved in level 4, organization, contact information, home address



(a)

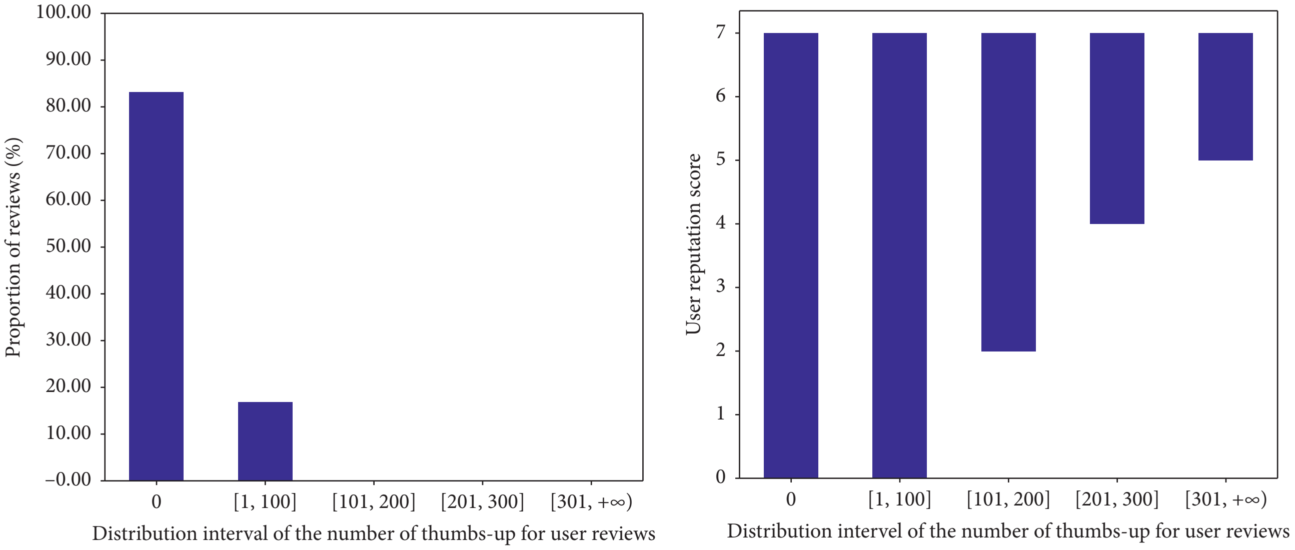
(b)

Figure 5: Continued.



(c)

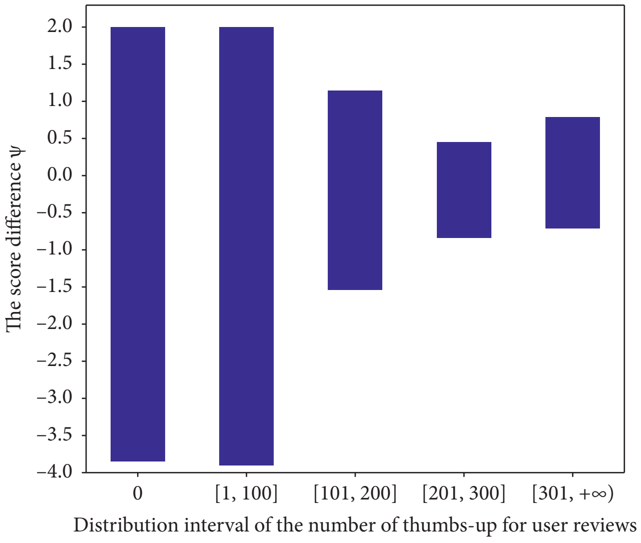
Figure 5: The PPR of LPA, IEPP, and CSPPM in Dataset1, Dataset3, and Dataset4. (a) The PPR of LPA. (b) The PPR of IEPP. (c) The PPR of CSPPM.



(a)

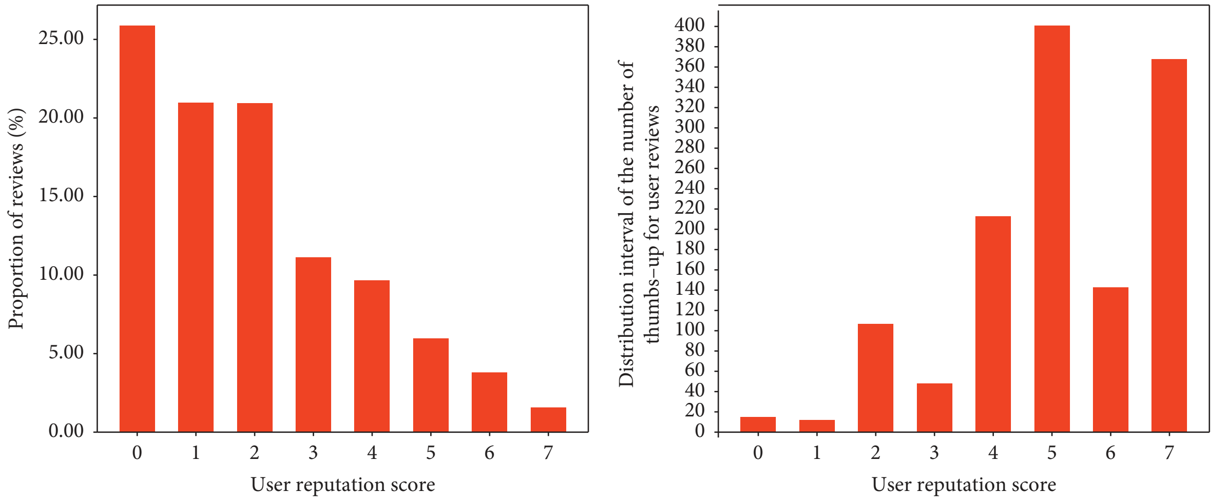
(b)

Figure 6: Continued.



(c)

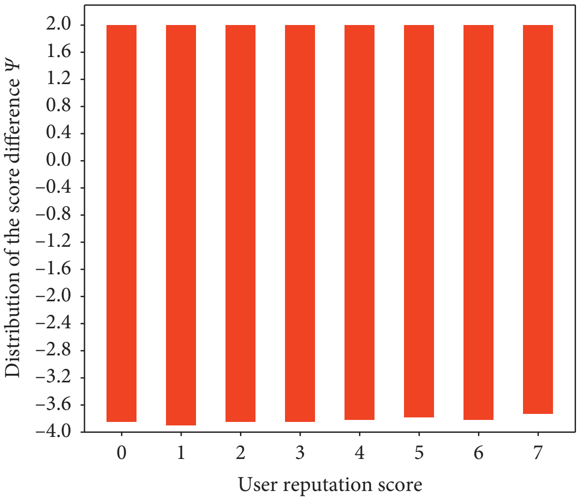
Figure 6: The user rating and the score difference with different distribution intervals of the number of thumbs-up in Dataset1. (a) Proportion of reviews with different distribution intervals of the number of thumbs-up. (b) User rating with different distribution intervals of the number of thumbs-up. (c) Score difference with different distribution intervals of the number of thumbs-up.



(a)

(b)

Figure 7: Continued.



(c)

Figure 7: The distribution intervals of the score difference and the distribution intervals of the number of thumbs-up with different user rating in Dataset 1. (a) Proportion of reviews with different user rating. (b) Distribution intervals of the number of thumbs-up with different user rating. (c) Distribution intervals of the score difference with different user rating.

can publicly publish in each grid increases, and the overall PPR also increases. In Dataset1, the proportion of the number of submitted reviews less than 4 is 81%. Therefore, the PPR in Dataset1 is higher than in Dataset4 and is lower than in Dataset3. The results prove that the number of reviews submitted by individual users can affect the overall review publication. The fewer the number of reviews an individual user submits, the higher the probability that all their reviews will be published, and the higher the overall PPR is. This is also the essence of LPA's privacy protection. That is, privacy protection is achieved by limiting the number of reviews published by individual users.

IEPP sets the user similarity interval as the threshold. In Figure 5(b), as the threshold increases, the PPR trends on Dataset3 is almost horizontal while both on Dataset1 and Dataset4 are gradually increasing. In Dataset3, there are close to 200,000 users and the number of reviews submitted by each user is less than 4. It ensures that there are enough similar users in Dataset3 and the similarities of all users are located in [1/2,2], Therefore, all reviews in Dataset3 can be published publicly; namely, the PPR is 100%. In Dataset4, the number of reviews submitted by each user is not less than 4. It makes the difference in the number of reviews and distribution characteristics between users very obvious and reduces the similarity between users and the PPR. However, the threshold for publicly published reviews increases when the user similarity interval increases. It allows user reviews with greater differences to be published and the PPR also increases. Dataset1 contains all users whose number of reviews submitted is less than 4. Therefore, IEPP can publicly publish all reviews of the users whose number of reviews submitted is less than 4 when the user similarity interval is [1/2.2]. For the users whose number of reviews submitted is not less than 4, only partial reviews of them can be publicly published. Thus, the PPR in Dataset1 is higher than in Dataset4 and is lower than in Dataset3. The results prove that the number of reviews submitted by individual users can affect the PPR of IEPP overall review publication. The fewer the number of reviews an individual user submits, the more the number of users is, the more similar users is, and the higher the overall PPR is.

CSPPM sets the score difference interval as the threshold. In Figure 5(c), as the threshold increases, the PPR curves on Dataset1, Dataset3, and Dataset4 are very close and the PPR trends on three datasets are gradually increasing. On the one hand, the increase in the score difference interval means that more reviews that meet the criteria for review publication can be published, and the overall PPR will increase. On the other hand, the score difference is used to determine whether a review can be publicly published while the score difference is determined by a single review and has no direct relationship with the number of users and the number of reviews submitted by individual users. Therefore, the PPR of CSPPM is relatively close on the three datasets.

Note that LPA and IEPP only allow reviews that meet the conditions to be published publicly, while reviews that do not meet the conditions are published anonymously. Here, the PPR refers to the percentage of the total number of reviews published that are not anonymous. Therefore, for LPA and IEPP, the greater the PPR is, the more display names and avatars of users are disclosed, and the greater the risk of privacy disclosure caused by display names and avatars. Compared with them, CSPPM is a more stringent privacy protection scheme. That is, reviews that meet the publication conditions are published anonymously, and reviews that do not meet the conditions are not published. It can effectively reduce the risk of privacy leakage caused by the display name and avatar. In addition, by comparing the PPR of the three methods on three datasets, it can be seen that our method is basically not affected by the number of reviews submitted by individual users and the total number of users, and the privacy protection effect is more effective, stable, and more universal.

7.3.2. User Utility. LPA, IEPP, and CSPPM use the number of thumbs-up, the user rating, and the score difference to evaluate the usefulness of review, respectively. It can be seen from Figure 6(a) that the proportion of reviews with 0 thumbs-up is more than 80%, indicating that most of the reviews are not liked, and it is not feasible to rank the usefulness of the reviews solely relying on the number of thumbs-up. However, In Figures 6(b) and 6(c), although the proportion of reviews with more than 200 thumbs-up is small, the distribution of score difference ranges from -0.84 to 0.79 and the distribution of user rating ranges from 4 to 7. It proves that some reviews with much more thumbs-up will also have a smaller score difference and a higher user rating.

As shown in Figure 7(a), the proportion of reviews with user rating no more than 4 is nearly 90% and the distribution of score difference corresponding to each rating is very close, but the corresponding to reviews received relatively few thumbs-up. In addition, the proportion of reviews with user rating of more than 4 is about 10%, but these reviews contain most of the reviews with much more thumbs-up and a small part of reviews with lower thumbs-up. It proves that, to a certain extent, user rating can reflect the usefulness of review, but it is not a decisive factor in determining the usefulness of the review. That is, users with high ratings may also make evaluations that are inconsistent with the facts, and users with low ratings may publish reliable reviews. In other words, it is not feasible to rely solely on user rating to determine whether a review is reliable.

CSPPM preferentially publishes reviews with a small score difference, namely the score difference threshold $|\delta| \leq 1$(approximately 78.740%), as shown in Figure 8, which contain the reviews with more than 200 thumbs-up, since the score difference is a reflection of the consistency between user reviews and business services. In other words, reviews with the smaller score difference make consumers feel more objective and credible and thus gives more thumbs-up to these reviews. Therefore, comparing the three metric, the score difference is the most feasible one for evaluating the usefulness of the review. Although the number of thumbs-up reflects the objectivity of the reviews to a certain extent, it is not feasible due to small users evaluating the reviews based on the metric. For the user rating, under the same conditions, it is suitable as a reference metric. Therefore, the score difference and the user rating can be selected to evaluate the usefulness of the review. Specifically, when publishing reviews, CSPPM first sorts the published reviews according to the score difference and then sorts the reviews with the same score difference according to the user rating.

7.3.3. Privacy. Table 5 shows the proportion of users at different privacy leakage levels in Dataset1 and the corresponding UARR.

In Table 5, privacy leakage level 0 represents that users publish reviews anonymously, which can effectively reduce the risk of privacy leakage due to the display name and avatar; the privacy leakage level 1 and 2 represent that reviews on Dianping are not anonymous, while they also cannot link them to the user's profile across other social networks. These levels correspond to the risk of privacy leakage within the same CLBSS; the privacy leakage levels 3, 4, and 5 correspond to the risk of privacy leakage caused by information such as the display name in the publicly published reviews. Especially the privacy leakage levels 4 and 5 can even directly determine the user's name, gender, real photo, occupation, contact information, address, etc. To evaluate the ability of our scheme to resist CUIA and SIA, we evaluate the URR at the different PPR and UARR > 0.2 (0.2 corresponds to the privacy leakage level 3). As shown in Figure 9, the URRof LPA and IEPP both increase with the increase of the PPR. The reason is that more reviews can be published publicly with the increase of the PPR. Thus, the adversary can collect more users' display names and real identities, which can identify more easily users' identities across social networks. However, CSPPM uses a strong anonymity mechanism to anonymize the display name and avatar of reviews that need to be published publicly. Therefore, CSPPM can avoid the privacy leakage caused by the display name. As a result, the URR of CSPPM is close to 0. It proves that, compared with LPA and IEPP, CSPPM has better resistance to CUIA and SIA.

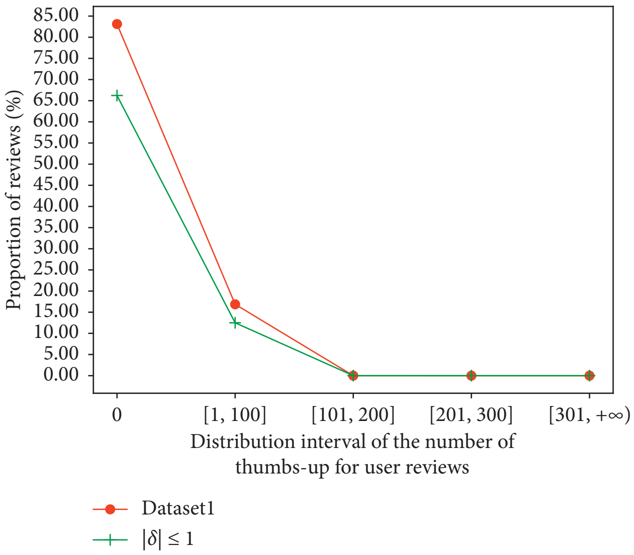


FIGURE 8: The distribution intervals of the number of thumbs-up with $|\delta| \leq 1$.

8. Discussion

We adopt a strong privacy protection mechanism. That is, in order to reduce the privacy risks caused by display name, user reviews meeting the release conditions should be uniformly anonymous before publication. This can achieve better privacy protection effect. But for users, they cannot independently choose whether to publish reviews publicly, and the personalized needs of them cannot be met. This is the limitation of the privacy protection mechanism proposed by us. Aiming at the limitation, in the future research

Table 5: The proportion of users at different privacy leakage levels in Dataset1 and the corresponding UARR.

Privacy leakage level	0	1	2	3	4	5
Proportion of users	5.45%	54.55%	40.00%	9.09%	14.55%	16.36%
UARR	0	0.0606	0.1818	0.1976	0.3083	0.6403

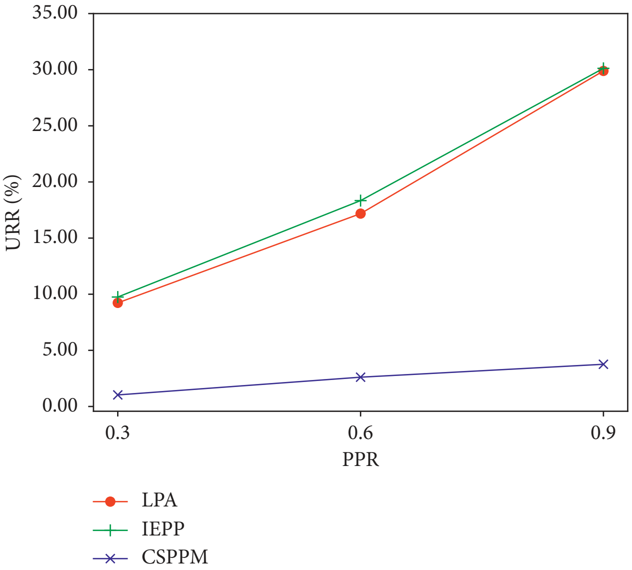


Figure 9: The URR of LPA, IEPP, and CSPPM at the different PPRand UARR > 0.2.

work, we will use ASN-CUIA identification method to implement the personalized privacy protection of users demand, called user privacy risk identification system. This system adopts the technique of artificial intelligence and big data to roughly estimate the privacy risks of user cross-platform and gives feedback of the assessment result to the user as a decision reference. Depending on the privacy risk feedback given by user privacy risk identification system, users can decide whether to publicly publish a review. Besides, it should be noted that we will still consider anonymity for reviews that exceed privacy risk thresholds.

9. Conclusions

In this paper, we proposed a strong cross-platform privacy protection mechanism (CSPPM) based on the partial publication and complete anonymity mechanism to resist connecting user identities attack (CUIA) and statistical inference attack (SIA) on the scenario of review publication. To be specific, on the one hand, we used the consistency between the user score and the business score as a criterion to publicly publish reviews with the higher usefulness of review and filter false or untrue reviews; on the other hand, we anonymized the display name and avatars of reviews that are publicly published. Besides, we evaluate the performance of CSPPM from three aspects: system utility metric (i.e., Public Publish Rate, PPR), user utility metric (i.e., number of thumbs-up, user rating, score difference), and privacy metric (i.e., the privacy leakage level based on user attribute recognition rate (UARR) and user recognition rate (URR)). Based on these metrics, we compared the effectiveness of our scheme with LPA and IEPP by implementing some experiments: (1) we analyze the impact of different thresholds on the difference in the PPR under the same method by considering the different thresholds set by different methods; (2) we evaluate the usefulness of reviews under different number of thumbs-up, user rating, and score difference; (3) we evaluate the URR at the different PRR and UARR > 0.2. Evaluation results show that CSPPM has better system utility and can better avoid the privacy leakage than LPA and IEPP in terms of resistance to CUIA and SIA. The evaluation results also prove that, as a metric to measure the usefulness of review, the consistency between the user score and the business score is more objective and credible than the number of thumbs-up and the user rating.

Data Availability

In this paper, we use restaurant review data on Dianping.com to evaluate CSPPM. The URL of our data is https://doi.org/10.18170/DVN/GCIUN4.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This research was funded by the Major Scientific and Technological Special Project of Guizhou Province (20183001), the Foundation of Guizhou Provincial Key Laboratory of Public Big Data (2017BDKFJJ015, 2018BDKFJJ020, and 2018BDKFJJ021), the National Natural Science Foundation of Guangxi (2019JJA170064), and the Basic Ability Improvement Program for Young and Middle-Aged Teachers in Guangxi (2021KY0615 and 2021KY0620).

References

[1] B. Lee, J. Oh, H. Yu et al., “Protecting Location Privacy Using Location Semantics,” in *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Diego, CA, USA, 2011.

[2] F. J. Wu and T. Luo, “CrowdPrivacy: publish more useful data with less privacy exposure in crowdsourced location-based services,” *ACM Transactions on Privacy and Security*, vol. 23, no. 1, 2020.

[3] J. Y. Dai and K. K. Qiao, “A privacy preserving framework for worker’s location in spatial crowdsourcing based on local differential privacy,” *Future Internet*, vol. 10, no. 6, 2018.

[4] K. Shu, S. H. Wang, J. L. Tang, R. Zafarani, and H. Liu, “User identity linkage across online social networks,” *ACM SIGKDD Explorations Newsletter*, vol. 18, no. 2, pp. 5–17, 2017.

[5] X. Zheng, Z. P. Cai, J. Z. Li et al., “Location-privacy-aware Review Publication Mechanism for Local Business Service Systems,” in *Proceedings of the 36rd IEEE International Conference on Computer Communications*, Atlanta, GA, USA, 2017.

[6] G. C. Yang, S. S. Luo, H. L. Zhu et al., “A mechanism to improve effectiveness and privacy preservation for review publication in LBS,” *IEEE Access*, vol. 7, pp. 156659–156674, 2019.

[7] X. P. Zhou, X. Liang, J. C. Zhao et al., “Correlating user mining methods for social network integration: a survey,” *Journal of Software*, vol. 28, no. 6, pp. 1565–1583, 2017.

[8] H. X. Li, H. J. Zhu, S. G. Du et al., “Privacy leakage of location sharing in mobile social networks: attacks and defense,” *IEEE Transactions on Dependable & Secure Computing*, vol. 15, no. 4, pp. 646–660, 2016.

[9] K. K. Deng, L. S. Xing, L. Zheng, H. Wu, P. Xie, and F. Gao, “A user identification algorithm based on user behavior analysis in social networks,” *IEEE Access*, vol. 7, pp. 47114–47123, 2019.

[10] M. K. Tefera and X. L. Yang, “A game-theoretic framework to preserve location information privacy in location-based service applications,” *Sensors*, vol. 19, no. 7, 2019.

[11] Y. B. Cui, F. Gao, W. M. Li et al., “Cache-based privacy preserving solution for location and content protection in location-based services,” *Sensors*, vol. 20, no. 16, 2020.

[12] M. K. Tefera and X. L. Yang, “Location privacy protection systems in presence of service quality and energy constraints,” *Information*, vol. 10, no. 4, 2019.

[13] C. Xu, X. Xie, L. H. Zhu et al., “A privacy-preserving location-sharing scheme in mobile online social networks,” *Science China Information Sciences*, vol. 63, no. 3, 2020.

[14] C. Li, L. H. Yin, K. Geng et al., “Location privacy preservation approach towards to content sharing on mobile online social network,” *Journal on Communications*, vol. 37, no. 11, pp. 31–41, 2016.

[15] Y. M. Sun, M. Chen, L. Hu, Y. Qian, and M. M. Hassan, “ASA: against statistical attacks for privacy-aware users in Location Based Service,” *Future Generation Computer Systems*, vol. 70, no. 5, pp. 48–58, 2017.

[16] M. Z. Li, Y. F. Wang, G. C. Yang et al., “DGS-HSA: a dummy generation scheme Adopting hierarchical structure of the address,” *Applied Sciences*, vol. 10, no. 2, p. 548, 2020.

[17] S. W. Sussman and W. S. Siegal, “Informational influence in organizations: an integrated approach to knowledge adoption,” *Information Systems Research*, vol. 14, no. 1, pp. 47–65, 2003.

[18] A. Munzel, “Assisting consumers in detecting fake reviews: the role of identity information disclosure and consensus,” *Journal of Retailing and Consumer Services*, vol. 32, no. 9, pp. 96–108, 2016.

[19] X. Li, K. K. Deng, H. H. Wu et al., “A survey of across social networks user identification,” *IEEE Access*, vol. 7, pp. 137472–137488, 2019.

[20] D. Liu, Q. Y. Wu, W. H. Han et al., “User identification across multiple websites based on username features,” *Chinese Journal of Computers*, vol. 38, no. 10, pp. 2028–2040, 2015.

[21] R. Zafarani and H. Liu, “Connecting corresponding identities across communities,” in *Proceedings of the International Conference on Weblogs & Social Media*, DBLP, San Jose, CA, USA, 2009.

[22] Y. J. Li, Y. Peng, W. L. Ji, Z. Zhang, and Q. Xu, “User identification based on display names across online social networks,” *IEEE Access*, vol. 5, pp. 17342–17353, 2017.

[23] X. P. Zhou, X. Liang, X. Y. Du, and J. Zhao, “Structure based user identification across social networks,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 6, pp. 1178–1191, 2018.

[24] Y. J. Li, Z. Zhang, Y. Peng, H. Yin, and Q. Xu, “Matching user accounts based on user generated content across social networks,” *Future Generation Computer Systems*, vol. 83, no. 6, pp. 104–115, 2018.

[25] S. S. Zhang, X. Liang, B. T. Mi et al., “Content-based social network user identification methods,” *Chinese Journal of Computers*, vol. 42, no. 8, pp. 1739–1754, 2019.

[26] K. X. Hu, Y. Liang, H. B. Xu et al., “A method for social network user identity feature recognition,” *Journal of Computer Research and Development*, vol. 53, no. 11, pp. 2630–2644, 2016.

[27] S. Y. Liu, S. H. Wang, and F. D. Zhu, “Structured learning from heterogeneous behavior for social identity linkage,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 7, pp. 2005–2019, 2015.

[28] B. Gedik and L. Liu, “Location privacy in mobile systems: a personalized anonymization model,” in *Proceedings of the 25th IEEE International Conference on Distributed Computing Systems (ICDCS’05)*, pp. 620–629, Columbus, OH, USA, 2005.

[29] B. Gedik and L. Liu, “Protecting location privacy with personalized k-anonymity: architecture and algorithms,” *IEEE Transactions on Mobile Computing*, vol. 7, no. 1, pp. 1–18, 2007.

[30] M. Gruteser and D. Grunwald, “Anonymous usage of location-based services through spatial and temporal cloaking,” in *Proceedings of the 1st International Conference on Mobile Systems, Applications and Services*, pp. 31–42, San Francisco, CA, USA, 2003.

[31] H. Kido, Y. Yanagisawa, and T. Satoh, “An anonymous communication technique using dummies for location-based services,” in *Proceedings of the International Conference on Pervasive Services (ICPS’05)*, pp. 88–97, Santorini, Greece, 2005.

[32] B. Niu, Q. H. Li, X. Y. Zhu et al., “Achieving k-anonymity in privacy-aware location-based services,” in *Proceedings of the 33rd IEEE International Conference on Computer Communications*, pp. 754–762, Toronto, Canada, 2014.

[33] X. Q. Zhang, Q. R. Zhou, C. H. Gu, and L. Han, “The location privacy preserving of social network based on RCCAM access control,” *IETE Technical Review*, vol. 35, no. sup1, pp. 68–75, 2018.

[34] B. Hernández-Ortega, “Don’t believe strangers: online consumer reviews and the role of social psychological distance,” *Information & Management*, vol. 55, no. 1, pp. 31–50, 2018.

[35] C. M. K. Cheung, M. K. O. Lee, and N. Rabjohn, “The impact of electronic word-of-mouth: the adoption of online opinions in online customer communities,” *Internet Research*, vol. 18, no. 3, pp. 229–247, 2008.

[36] Y. Y. Wu, E. W. T. Ngai, P. K. Wu et al., “Fake online reviews: literature review, synthesis, and directions for future research,” *Decision Support Systems*, vol. 132, pp. 1–15, 2020.

[37] Q. Xu, “Should I trust him? The effects of reviewer profile characteristics on ewom credibility,” *Computers in Human Behavior*, vol. 33, no. 4, pp. 136–144, 2014.

[38] H. Park, Z. Xiang, B. Josiam et al., “Personal profile information as cues of credibility in online travel reviews,” *Information and Communication Technologies in Tourism Anatolia: An International Journal of Tourism and Hospitality Research*, no. 1, pp. 13–23, 2013.

[39] W. Zhang, R. K. Mallik, and K. B. Letaief, “Cooperative spectrum sensing optimization in cognitive radio networks,” in *Proceedings of the 2008 IEEE International Conference on Communications (ICC’08)*, pp. 3411–3415, Beijing, China, 2008.

[40] M. Grissa, A. Yavuz, and B. Hamdaoui, “An efficient technique for protecting location privacy of cooperative spectrum sensing

users," in *Proceedings of the IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, pp. 915–920, San Francisco, CA, USA, 2016.

[41] K. Arshad and K. Moessner, "Robust collaborative spectrum sensing based on Beta reputation system," in *Proceedings of the Future Network Mobile Summit*, pp. 1–8, Warsaw, Poland, 2011.

[42] T. Li, *Restaurant Review Data on Dianping*, Peking University, Beijing, China, 2018.