TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214 03/14 pp452–463
DOI: 10.26599/TST.2022.9010003
Volume 28, Number 3, June 2023

A Differential Privacy Protection Protocol Based on Location Entropy

Ping Guo, Baopeng Ye, Yuling Chen*, Tao Li, Yixian Yang, Xiaobin Qian, and Xiaomei Yu

Abstract: A Location-Based Service (LBS) refers to geolocation-based services that bring both convenience and vulnerability. With an increase in the scale and value of data, most existing location privacy protection protocols cannot balance privacy and utility. To solve the revealing problems in LBS, we propose a differential privacy protection protocol based on location entropy. First, we design an algorithm of the best-assisted user selection for constructing anonymity sets. Second, we employ smart contracts to evaluate the credibility of participants, which ensures the honesty of participants. Moreover, we provide a comprehensive experiment; the theoretical analysis and experiments show that the proposed protocol effectively resists background knowledge attacks. Generally, our protocol improves data availability. Particularly, it realizes user-controllable privacy protection, which improves privacy protection and strengthens security.

Key words: Location-Based Services (LBS); smart contract; location entropy; differential privacy; privacy protection

1 Introduction

A Location-Based Service (LBS) refers to services around geolocation data[1], A mobile terminal uses a wireless communication network or satellite positioning system[2] based on a spatial database to obtain a user's geographic position coordinate information and integrate it with other information to provide the user with the required location-related value-added services. In the current era of smart city construction[3, 4], with the development of mobile communication technology and the widespread use of smart terminal devices, the Internet of Things (IoT)[5], which is essential in our daily activities, generates massive location data related to transportation, healthcare, business, and social activities[6]. LBS is widely used in e-commerce, health care, traffic travel, mobile social networking, etc.[7–9] Despite the evident relevance of LBS in our daily lives, Location Service Providers (LSPs) collect location data for deep learning training data or similarity matches[10] to provide customers with more personalized recommendations. However, attacks on deep learning frameworks by malicious internal or external attackers would exert substantial effects on society and life[11], such as revealing home address, identity number, degree, and religion. Hence, Location Privacy Protection (LPP) is a vital part of LBS[12] .

Traditional LPP methods apply only to small datasets, the privacy protection process needs to rely on an attacker's restricted background knowledge and there is no rigorous attack model. Moreover, the amount of randomized noise is related to the data size. Consequently, the security of traditional LPP reduces

• Ping Guo, Yuling Chen, and Tao Li are with the State Key Laboratory of Public Big Data, School of Computer Science and Technology, Guizhou University, Guiyang 550025, China. E-mail: 18768673031@163.com; ylchen3@gzu.edu.cn; litao_2019@qfnu.edu.cn.
• Baopeng Ye is with the Information Technology Innovation Service Center of Guizhou Province, Guiyang 550025, China. E-mail: yebaopeng@yeah.net.
• Yixian Yang is with School of Cyberspace Security, Beijing University of Posts and Telecommnuications, Beijing 100000, China. E-mail: yxyang@bupt.edu.cn.
• Xiaobin Qian is with Guizhou CoVision Science & Technology Co., Ltd., Guiyang 550025, China. E-mail: alanqian@139.com.
• Xiaomei Yu is with School of Information Science and Engineering, Shandong Normal University, Jinan 250000, China. E-mail: yxm0708@126.com.
* To whom correspondence should be addressed.
Manuscript received: 2021-11-09; revised: 2021-12-02; accepted: 2022-01-08

when attackers get all recorded information except for the target of attackers. Differential Privacy (DP)[13], as a privacy protection method that can break away from the limitation of background knowledge, defines a strict attack model. Thus, DP is wildly used in LPP.

Generally, according to the backdrop discussed above, in this study, we propose a user-controllable DP protection protocol based on Location Entropy (LE), improving the security of user location data. In this protocol, we combine DP technology with LE to ensure data security while obtaining location services. Our main contributions are as follows.

(1) Combining LE with DP, we construct an algorithm for the optimal selection of anonymity sets to assist users and propose a DP protection protocol. This protocol resists background knowledge attacks and achieves user-controllable privacy protection, which is a strict privacy protection agreement.

(2) We design a smart contract to evaluate the reputation of users, and it avoids malicious behaviors that assist users. In other words, the location privacy of participating users can be effectively guaranteed.

(3) The proposed protocol leverages Dijkstra's algorithm to protect the privacy of a user's location while guaranteeing data availability. Besides, we optimize the shortest path for the result set returned by the anonymous, and then we return the best result after processing to the requesting user.

The rest of this article is organized as follows. In Section 2, we introduce the related research results in recent years. In Section 3, we describe the concepts of DP, information entropy, and smart contracts. In Section 4, a DP protection protocol based on LE is proposed. The anonymous set construction method, reputation evaluation mechanism, and query-result optimization algorithm are also proposed. In Section 5, the security and correctness of the proposed protocol are proven. Finally, we conclude this work and discuss future work in Section 6.

2 Related Work

Dwork et al.[13–15] proposed DP using Laplace distribution based on indistinguishability, and then proposed the Gaussian mechanism of DP and centralized DP. Austrin[16] proposed a DP index mechanism to achieve nonnumerical DP protection. Geng and Viswanath[17] proposed an optimal ϵ-DP mechanism, based on the balance of privacy and utility in DP. Apple has updated its privacy policy, introduced DP technology, and applied it in the collection of personal information to prevent sensitive information from matching personal real identity information and avoid personal privacy leakage. Li et al.[18] proposed a matrix mechanism to achieve the workload response of linear counting queries; the nature of a given query can adjust the noise distribution. This mechanism adds relevant noise to achieve DP and increases accuracy.

When using LBS, users need to send personally identifiable information to the LSP. If a third-party abuses, resells, or intercepts the information by an attacker, the user's privacy may be leaked. Regarding the privacy problem of LBS, some researchers use traditional data encryption methods, such as K-anonymity, attribute encryption, and homomorphic encryption[19]. Additionally, many scholars have combined the advantages of DP to study it. Shokri et al.[20] proposed an LPP mechanism based on the Stackelberg game and found an optimal LPP mechanism for each user's service quality constraints. The optimal LPP maximizes the level of privacy protection while meeting the user's service quality requirements, but this method requires a lot of running time. Chatzikokolakis et al.[21] introduced DP into location protection based on the scheme of Shokri et al.[20], and constructed a privacy protection mechanism that optimizes service quality. Because DP can resist background knowledge attacks, the constructed mechanism can minimize the loss of service quality under the premise of satisfying location indistinguishability. Shokri[22] further proposed the use of two indicators of DP and distortion privacy to optimize the privacy protection strategy based on the Stackelberg game. DP limits the extent of user privacy leakage, where as distortion privacy measures the error rate of inferring user privacy. Combining these two standards, this privacy protection strategy can resist more types of speculative attacks while ensuring service quality. Ni et al.[23] proposed a user data DP protection clustering algorithm based on DP and the DBSACN algorithm by increasing the Laplace noise disturbance data release, effectively protecting the validity and privacy of user data.

Recently, in the protection of location privacy, generating fake locations instead of real locations is also a research hotspot. Niu et al.[24] proposed a fake location selection algorithm to achieve user privacy protection based on location services and selected virtual locations based on LE measurements. Then, an enhanced DLS algorithm was proposed; DLS generates false positions

according to the query probability of each position to ensure that selected virtual positions are distributed as much as possible. Niu et al.[25] also proposed a cache-based fake location selection algorithm, which measures the impact of cache on privacy protection based on the information and enhances LPP by maximizing query privacy and the contribution of fake location cache levels. To et al.[26] discussed the problem of perturbing the LE of a group of locations based on DP. Because the current solution for calculating LE requires full access to the user's location, it poses a privacy threat to the user. A threshold technology to limit the number of user visits has been employed; this technology can retain the release of the original data but will introduce approximate errors. Wang et al.[27] related utilities with entropy to enhance the utility of the IoT and maintain the security of the system. Han et al.[28] proposed a k-means cluster-based LPP scheme for IoT. To protect the source location, a fake source node is used to simulate the function of the real source. Ni et al.[29] proposed an LPP scheme based on anonymous entropy, which uses location distance and requested content as anonymous entropy to generate false locations to construct anonymous areas to resist adversary background knowledge attacks. Liu et al.[30] proposed a blockchain-based distributed k-anonymity LPP scheme, which regards the construction of an anonymous area as a two-party game between the requesting and collaborative users, and uses the blockchain to record the game between the two parties. A true location is used as evidence, and the users who have location leaks and deceptive behaviors are punished so that they cannot successfully construct an anonymous zone when they are requesters, to restrain their self-interest. When an attacker has strong background knowledge, this scheme cannot resist the attack well. Moreover, blockchain is an emerging technology in many fields, including distributed systems and the IoT; when there is selfish mining in a blockchain, a selfish mining attack uses loopholes in the consensus mechanism to destroy the blockchain system[31–34]. Zhu et al.[35] proposed a privacy protection framework to outsource LBS to the cloud. The framework supports multilocation queries with fine-grained access control, and it performs search using location attributes while providing semantic security. The protocol allows users to control the tradeoff between accuracy and privacy on a dynamic per-query basis. The scenario assumes that the cloud is completely credible. When the outsourcing cloud is breached, the privacy of a user will be completely leaked.

The abovementioned studies have made certain contributions to the research on LPP, but there are still some problems in the solution. In response to these problems, we focus on the issue of user LPP and query-result availability.

3 Preliminary

3.1 Differential privacy

DP[36] is a privacy protection framework supported by a solid mathematical theory. By adding noise to a dataset, an attacker cannot infer the sensitive information on the dataset based on the result. The DP protection model ensures that deleting or inserting a record in dataset will not significantly change the query result, i.e., the privacy disclosure risk caused by a single datum record to a dataset is controlled within a given range to ensure the privacy of individuals in the dataset.

Definition 1 For two datasets D and D', at most one record is different between the two datasets, namely $|DD'| \leqslant 1$, these two datasets are called adjacent datasets.

Definition 2 M means a privacy algorithm, S_M means the set of all possible output results. Algorithm M works on the datasets D and D'. If the output result meets

$$Pr[M(D)] \in S_M \leqslant e^{\epsilon} \times Pr[M(D') \in S_M] \quad (1)$$

then the algorithm M satisfies the ϵ-DP, where ϵ is the privacy protection budget, indicating the degree of privacy protection. The smaller the value of ϵ, the higher the degree of privacy protection, and vice versa.

Definition 3 There is a random function $f()$, the input of which is a dataset D and the output is a D-dimensional real number vector. Thus, for any adjacent datasets D and D',

$$GS_f = MAX_{D,D'}||f(D) - f(D')|| \quad (2)$$

where GS_f is the sensitivity of the function $f()$, which characterizes the degree of the influence of noise on the query results of a dataset.

Definition 4 For the noise Lab (b) that obeys the Laplace distribution, the probability density function is

$$P(x) = (1/2b) \times \exp(-|x|/b) \quad (3)$$

where b is the scale parameter; given the dataset D, the function $f()$ is a query on D, and Δf is its sensitivity, then the random algorithm $M(D) = f(D)+Y$ provides ϵ-DP; $Y = \text{Lab}(\Delta f/\epsilon)$ is the random noise that obeys the Laplacian distribution with a scale parameter of

$\Delta f/\epsilon$.

DP protection has strict mathematical proofs to ensure its reliability, it does not need to consider an attacker's background knowledge. Compared with other privacy protections, such as k-anonymous, it has higher security[37].

3.2 Information entropy

Information is a broad concept, and it is difficult to describe it completely and accurately with a simple definition. For any probability distribution, it can be defined as a quantity called entropy. The concept of entropy originated in physics and was proposed by Clausius. Entropy represents the internal stable state of the system without external interference and is used to measure the degree of disorder in a thermodynamic system. In information theory, the uncertainty of random variables is measured by information entropy. Shannon[38] introduced information entropy and defined it as the probability of discrete random events. The more chaotic a system, the higher its information entropy. Conversely, the more orderly a system, the lower the information entropy. Therefore, information entropy can be considered as a measure of system ordering[39].

Definition 5 Suppose X is a discrete random variable, its value space table is R, and the probability density function $P_X(x) = P_r(X = x), x \in R$. For convenience, let the probability density function be $p(x)$ instead of $P_X(x)$, then the entropy of the random variable X is defined as follows:

$$H(X) = -\sum_{x \in R} p(x) \log p(x) \tag{4}$$

where the base of the logarithm is 2, and the unit of entropy is expressed in bits.

3.3 Smart contract

Smart contract[40] refers to a program fragment that can be automatically executed, which can execute the general contract conditions without the participation of a third party, and minimize the occurrence of malicious behaviors and unexpected situations caused by the participation of a third party. Specifically, smart contract is a code snippet which is deployed on blockchain, and encapsulates preset contract states, contracts response rules, presets trigger conditions, and responses actions in specific scenarios. The smart contract monitors the real-time status of a blockchain. When the parties signing the contract reach an agreement and detect that the external environment and activities meet the preset trigger conditions, the contract is automatically executed. The smart contract operation mechanism is shown in Fig. 1.

4 Differential Privacy Protection Protocol Based on Location Entropy

4.1 Selecting the optimal user to assist

In this section, to avoid the risk of privacy leakage caused by third-party participation and prevent attackers from associated analysis of location data, the privacy protection of user location data is effectively realized. First, according to the user's acceptable service quality, submit an anonymous set request $r = (id, l, k, \delta, W, W_{\min}, data)$, where id is the request user identifier, $l = (x, y)$ is the user's current location, k is the minimum degree of anonymity, δ is the maximum anonymous area radius, W is the user's reputation, $W_{\min}$ is the minimum reputation, and $data$ is the query content, as shown in Fig. 2.

When the requesting user can accept a lower service quality, the δ value is larger, if the adjacent auxiliary users meeting the conditions in the area are less than k, the DP mechanism is used to add anonymous points to construct an anonymous location

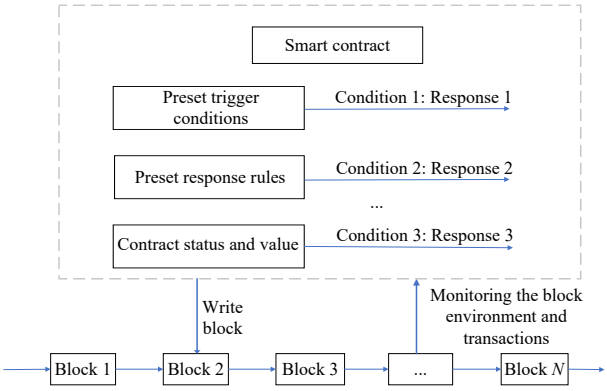

Fig. 1 Smart contract operating mechanism.

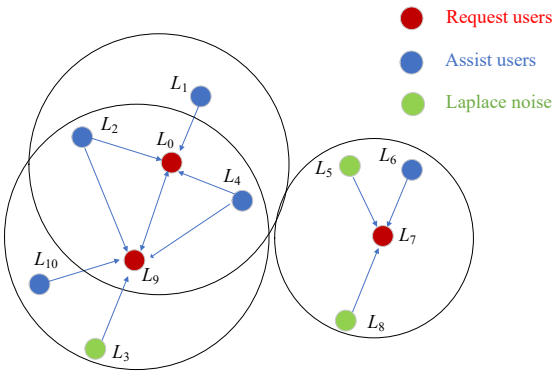

Fig. 2 Anonymous set construction method.

dataset. $C_j = \{id', l_i, data\}$ $(i = 1, 2, \ldots, k)$, where id' represents the user virtual identifier, l_i means participating in assisting the user's location, so that the user realizes the DP protection of the location data without the participation of a third-party server, and prevents the attacker's associated analysis, while ensuring the user's controllable DP, and the user's location data are traceable. To make the anonymous location of the assisting user like the real location of the requesting user, this paper uses a method of recording historical query categories to obtain similar locations. In detail, categorizing user's queries firstly, then recording the query category label once the user sends a query request. To prevent assisting users from maliciously exposing their position in the process of participating in the construction of anonymity set, a reputation W is set for each user. When the user successfully participates in an anonymous set construction, the reputation increases. If the user is assisted in maliciously revealing his position, the reputation decreases. As shown in Fig. 3, $L_i(r, s)$ is the user's location attribute, where r is the user's reputation value and s is the location similarity. When a user sends an LBS request at the location $L_0(0.2, 0.1)$, if the selected anonymous location is at the location $L_6(0.6, 0.6)$, then it is less likely to be attacked, because L_6 has a high reputation value and the two positions have highly similar query requests. When $L_2(0.2, 0)$ is selected as an anonymous location, if the attacker has some auxiliary information to attack this location, the real location may be exposed. Therefore, in this study, users with the highest reputation value and similar requests to the real location are selected as assisted users to participate in constructing an anonymous set.

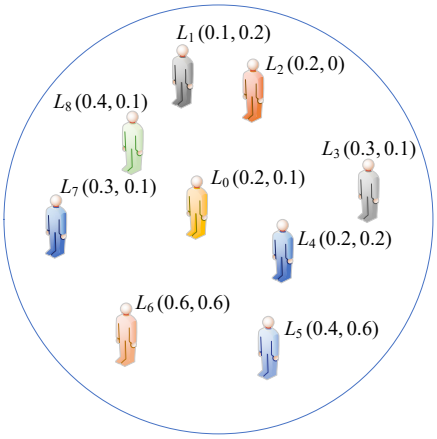

Fig. 3 Schematic of assisting user selection based on location entropy.

4.2 Anonymous set construction

When a user initiates a request for assistance, selecting k neighboring users, whose reputation scores are greater than $W_{\min}$ within δ to construct anonymity set, then, set the location similarity of the user i as P_i and the reputation as W_i, then the reliability h_i of auxiliary user can be expressed as

$$h_i = -\sum_{i=1}^{k} W_i \times P_i \times \log P_i \tag{5}$$

where W_i represents the latest reputation score of node i.

The requesting user uses the anonymity set LE value as the selection condition for users to participate in anonymous assistance. The larger the LE value, the higher the reputation of the anonymity set assistance user and the higher the similarity to the query of the requesting user. On the contrary, anonymous sets are less reliable, and it is more likely to maliciously disclose the real location of the requesting user. The main steps are as follows shown in Algorithm 1:

Step 1: Record the total amount of a certain query category of the user as E_q and the total amount of all query categories as $E_q(C_q = \sum_{i-1}^{n} E_q)$,

$$P_i = E_q / C_q \tag{6}$$

where P_i represents the probability of querying a certain category, where $i = 1, 2, \ldots, n$ (n is the total of query categories).

Step 2: Rank P_i and W_i from large to small, and then select the top k users in Q_i and W_i (total $2k$) as the candidate set C_m, where $m = 1, 2, \ldots, 2k$.

Step 3: Do m operations on C_m, each operation selects $k - 1$ users to constract anonymous set C_j $(j = 1, 2, \ldots, K)$ with request users, and then calculate h_i by Eq. (5).

Step 4: Select the anonymous set with the largest h_i and return it to the requesting user to submit a location query.

Algorithm 1 Best-assisted user selection

Input: $\delta, k, W_{\min}, data$
Output: C_j
1: $E_q \leftarrow data$;
2: $P_i = E_q / C_q$;
3: $C_j = CheckRepInqvalue\,(P_i, W, W_{\min})$;
4: **while** $j < k$ **do** $h_j = -\sum_{j=1}^{k} W_j \times P_j \times \log_2 P_j$;
5: return j;
6: **end while**
7: $C_j = CheckEntropy(h_j)$; /*Select the user with the largest entropy value as the assisting user

When the number of assisted users within the maximum anonymity radius r is less than k , the DP protection requires the addition of a noise mechanism. In this study, an anonymous point is constructed by adding noise that conforms to the Laplace distribution to achieve ϵ-DP protection, the privacy budget parameter ϵ indicates the degree of privacy protection, the smaller the value, the higher the degree of privacy protection.

Definition 6 If the probability density function distribution of a random variable is $f(x\|\mu, b) = 1/2b\times \exp(-|x-\mu|/b)$, where μ is the positional parameter, then is the Laplace distribution.

For anonymous points, the protocol in this study treats longitude and latitude independently, and generates a longitude and latitude that are indistinguishable within the privacy budget parameters, which are recorded as anonymous points $L_a(x_a, y_a)$. According to Definition 1, by setting the position parameter $\mu = 0$, the latitude and longitude of the anonymous point should meet the following conditions, as shown in Algorithm 2.

$$\begin{aligned} Pr(x_a) &= 1/2b_x \times \exp(x_a/b_x), \\ Pr(y_a) &= 1/2b_y \times \exp(y_a/b_y) \end{aligned} \tag{7}$$

$$b_x = \bar{x}/\epsilon, b_y = \bar{y}/\epsilon,$$

$$\bar{x} = 1/K\sum_{j=1}^{K} x_j,\ \bar{y} = 1/K\sum_{j=1}^{K} y_j \tag{8}$$

Step 1: Calculate $\bar{x} = average(C_{Kx})$ and $\bar{y} = average(C_{Ky})$.

Step 2: Calculate $b_x = \bar{x}/\epsilon$ and $b_y = \bar{y}/\epsilon$.

Step 3: Calculate anonymous point results $L_a = (b_x, b_y)$.

4.3 Evaluation reputation

Considering that in the process of constructing anonymity sets, not all users are honest and reliable, and there are deceptive behaviors, such as users deliberately leaking locations and spreading false query information on the network, and the current trust mechanism has problems, such as heavy reliance on trusted third parties and malicious evaluations[41, 42]. Therefore, to regulate the behavior of participants and improve the system reliability, a reputation evaluation mechanism is introduced for each participant in the blockchain network. After a transaction is completed, the behavior of each participant is judged, quantified by credibility, and stored in the blockchain data ledger. Users with high credibility can get certain rewards. The reputation mechanism is used as the system's supervision mechanism to improve blockchain network security. The reputation contract is called when the trigger rule is satisfied in the transaction upload contract. User credibility is determined from three aspects: the act of assisting user i to upload information, the rest of assisting user j in determining the behavior of i, and requesting user l evaluation of assisting users. Therefore, the comprehensive reputation score of each user is as follows:

$$\begin{aligned} W_i &= W_i^{\Delta t} + \alpha_i W_{\alpha_i}^{\Delta t} + \beta_i \sum_{j=1}^{k-1} \alpha_j W_{ij}^t + \gamma_i W_{il}^t, \\ &\alpha_i, \beta_i, \delta_j, \gamma_i > 0, \\ &\alpha_i + \beta_i + \delta_j + \gamma_i = 1, \sum_{j=1}^{k-1} \delta_j = 1 \end{aligned} \tag{9}$$

where $W_i^{\Delta t}$ represents the final reputation score of the last time, $W_{\alpha_i}^{\Delta t}$ means that the system judges its credibility after itself uploads information, α_i is the weight of the item, $\sum_{j=1}^{k-1} \alpha_j W_{ij}^t$ is the user i after participating in the assistance, the rest of the assisting user j is true or false to the information i at t judgment, assuming that there are k users participating in the assistance, β_i is the weight of the item, and δ_j means that the other assisting users j except for user i have different reputations. W_{il}^t is the evaluation of the requesting user l for assisting users, and γ_i is the proportion of the item.

4.4 Query-result optimization processing

When a user successfully constructs an anonymous set to submit a query, the user's privacy needs are met, but the privacy of the user's location data must also be ensured at the same time as the data availability. In this section, we construct a query-result graph based on a directed graph. The LSP applies the shortest path algorithm[43] idea to further filter and process the query results and finally returns the optimal query results to the requesting user (see Fig. 4), as shown in Algorithm 3.

Query result processing:

Step 1: Request user r_1 to successfully construct an

Algorithm 2 Anonymous point generation

Input: C_K, ϵ, k /*The set of assisted users who meet the conditions and the privacy budget

Output: L_a

1: $\overline{x} = average\left(C_{jx}\right), \overline{y} = average\left(C_{jy}\right)$;
2: $b_x = \overline{x}/\epsilon, b_y = \overline{y}/\epsilon$;
3: **while** $j < k - K$ **do**
$L_a = GenerationAnonymouspoint\,(b_x, b_y)$;
4: **end while**
5: $j++$;

Algorithm 3 Smart contract (reputation)

Input: $W_i^{\Delta t}, (i, j) \in k$
Output: W_i

while (k) $In \leftarrow$ Upload assistance information;
end while
if In==true **then** $W_{i\leftarrow a} > 0$, W_{ia}=$W_{i\leftarrow a}$; /*That is the grade of α versus i, same as below.
else $W_{i\leftarrow a} < 0$, W_{ia}=$W_{i\leftarrow a}$;
$W_i = W_i^{\Delta t} + \alpha_i W_{ia}$;
end if
$Sort\,(W_1, W_2, W_j)$;
$W_1 = W_{\min}$, $W_i = W_{\max}$;
$d = (2 - 0.2n)/((n-1)n)$; $\delta_j = 0.1 + (j-1)\,d$;
$W_{ij} = \delta_j$, $W_{i\leftarrow j}$;
$W_i = W_i + \beta_i W_{ij}$;
if In==true **then** $W_{ij} = W_{i\leftarrow l}$, $W_{i\leftarrow a} > 0$;
$W_i = W_i + \gamma_i W_{il}$;
else $W_{il} = W_{i\leftarrow l}$;
$W_{j\leftarrow a} < 0$;
end if
$W_i = W_i + \gamma_i W_{il}$;

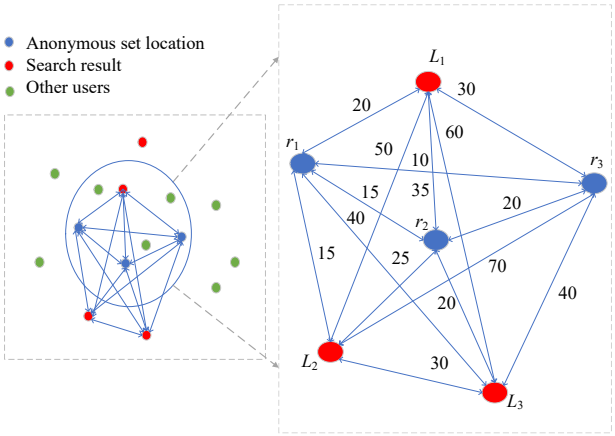

Fig. 4 Weighted directed graph of query results.

anonymous set $As = (id', L, data)$, where L is the current location of requesting user r_1 and $k-1$ assisting user geographic location sets, *data* is the content of the submitted query.

Step 2: After receiving the query-request, the server obtains the query result set $Sq = (id'_1, l_1), (id'_2, l_2), (id'_3, l_3), \ldots, (id'_k, l_k)$ according to the anonymous set. Constructing a digraph between the query result and anonymous point, the weights of the edges in digraph are the distance between users.

We find the result with the smallest sum of the distance from the anonymous set position and return it as the exact query result (Dijkstra seeks the shortest path problem), the solution process is shown in Table 1.

As we can see, we find the sum of the distances (L_1, r_1), $(L_1, r_2,)$, and (L_1, r_3) are the smallest, so L_1 can be returned as the optimal result to the requesting users r_1, r_2, and r_3.

5 Experiment and Analysis

5.1 Security analysis

The proposed DP protection protocol first considers the location similarity between the assisting and requesting users and the integrity of the assisting user during the anonymity set construction. The contract restricts the creditworthiness of the assisting user. When the user's creditworthiness is low or the similarity of the assisted user's location to that of the requesting user is low, the user will not be selected as an assisting user, which ensures that the locations of the assisting and requesting users are indistinguishable. Table 2 shows location attribute information of the requesting user $L_a(L, data, P_i, W_i)$ and the assisting user $L_b(L, data, P_i, W_i)$. The attacker cannot inference the user for each information when the user ID is not known; when the query probabilities $P_{i,b1}, P_{i,b2}, P_{i,a}$ are closer to the value, the query content data are more similar, and the user reputation is W_i, the higher value of reputation, the lower possibility of malicious disclosure of the user's location.

When a user's privacy needs are high, the DP mechanism is used to add anonymity points that conform to the Laplace distribution. DP is a privacy protection technology based on data distortion; injecting noise,

Table 1 Shortest path from each query vertex to the anonymous location.

Vertex	Requesting user		
	r_1	r_2	r_3
L_1	$L_1 \rightarrow r_1$	$L_1 \rightarrow r_2$	$L_1 \rightarrow r_2 \rightarrow r_3$, $L_1 \rightarrow r_3$
L_2	$L_2 \rightarrow r_1$	$L_2 \rightarrow r_2$	$L_2 \rightarrow r_3 \rightarrow r_3$
L_3	$L_3 \rightarrow r_2 \rightarrow r_1$	$L_3 \rightarrow r_2$	$L_3 \rightarrow r_3$

Table 2 Anonymous set information.

User	Location L	Query data	Probability P_i	Reputation
Requesting user L_a	(x, y)	$Data_1$	$P_{i,a}$	$W_{i,a}$
Anonymous user L_{b1}	(x_{b1}, y_{b1})	$Data_2$	$P_{i,b1}$	$W_{i,b1}$
Anonymous user L_{b2}	(x_{b2}, y_{b2})	$Data_3$	$P_{i,b2}$	$W_{i,b2}$
Laplace noise L_{b3}	(x_{b3}, y_{b3})	$Data_4$	$P_{i,b3}$	$W_{i,b3}$

adding, or deleting a record makes the output result indistinguishable.

(1) Let D_1 and D_2 be two datasets with N records difference, and the initial sensitivity of the dataset is D. If a data record is inserted or deleted in the dataset, the sensitivity becomes $D+1$.

(2) Let $[M(D_1)]$ and $[M(D_2)]$ be the results of datasets D_1 and D_2 after adding noise, the sensitivity of $[M(D_1)]$ is D, and the sensitivity of $[M(D_2)]$ is $D+N$.

Combined with the definition of DP: $P_r[M(D_1)]/P_r[M(D_2)] = d/(d+n) \leqslant 1 \leqslant e^{\epsilon}$. Therefore, after adding noise, the anonymous datset satisfies the ϵ-DP protection and is indistinguishable from the requesting user. The higher the similarity of location data processed by DP, the more difficult for the attacker to inference whether the user is in the dataset, so the higher the degree of privacy protection is for the user.

As the core parameter of the DP protection method, the differential privacy budget not only determines the level of DP protection, but also determines the degree of privacy leakage. In this study, when a user adds noise through the Laplace mechanism to construct an anonymous set and submits a service request, the service initiates a query in response to the request, and no longer queries the dataset, so you can make $\epsilon = 1$. According to the definition, the smaller the privacy budget , the higher the degree of privacy protection, and the lower the data availability. When $\epsilon = 1$, the user privacy protection level is the highest, but the data availability is lower. In this scheme, DP budget is determined by the number of assiting users generated, the maximum value is the number of assiting users, in brief, $1 < \epsilon < k$. And the added value of Laplace noise is dynamic which determined by the privacy budget for users, so we set a DP budget as $\epsilon = k/2$ to experiment.

5.2 Experimental results

5.2.1 Credibility value changes with the credibility

In Fig. 5, the reputation of different user behaviors changes differently. When User 1 has been honestly participating in collaboration, as the number of participants increases, the reputation value becomes higher and the score gradually increases, and vice versa. When User 2 is a malicious participant, the reputation will decrease quickly. When the reputation is 0, the user is no longer eligible to participate in assistance. User 3 has not participated in assistance; the reputation value remains unchanged. User 4 is participating in the process of collaboration, occasionally maliciously or honestly, and the cost of former is often greater. Based on the experiment, the reputation evaluation mechanism designed in this study can better constrain the user's behavior and prevent the user's location from being maliciously leaked.

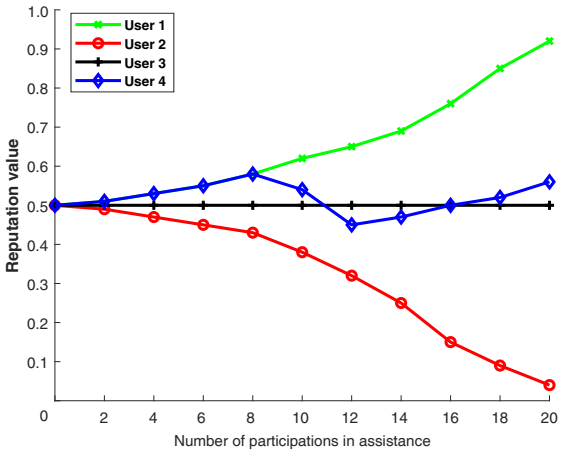

Fig. 5 Changes in reputation value.

5.2.2 Relationship between location entropy and user reputation and location similarity

Figure 6 shows the influence on LE of reputation and query probability. According to the performance of entropy, the larger the entropy, the higher the information uncertainty; the smaller the entropy, the higher the information certainty. The experimental results show that the proposed LE can be found to be consistent with changes in entropy. LE increases with the increase of user reputation and location similarity, so the indiscernibility of anonymous sets also increases. It can be verified by experiments when the entropy value is greater, the higher the location similarity between the assisting user and the requesting user, the better the

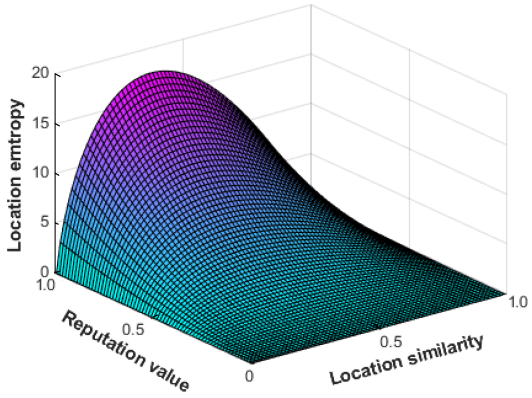

Fig. 6 Relationship among location entropy and user reputation and location similarity.

indistinguishability of the constructed anonymous set.

5.2.3 Error between the query result and the real position

Experimental setup: The datasets used in this article are all high-tech maps to capture longitude and latitude data. The experimental algorithm uses Python language and MATLAB, the language programming environment is PyCharm 2020.3.3 x64 and MATLAB R2020b. The experimental environment is configured as 3.0 GHz i5-8500, 8 GB memory, 1 TB hard disk +256 GB SSD, Win10 operating system. This experiment uses the AutoNavi Map API to simulate the submission and acquisition of user service requests, submits the query request with the real location $L_p(x, y)$ and the anonymous set C_m, and queries the content for the user $data$, using the position offset rate p to measure the quality of data service.

Submitting a query by constructing an anonymous set to obscure the real location can effectively protect user privacy, but it also causes errors in the query results. We select the best result by constructing a query-result directed graph; the selection of an anonymous set directly determines the LBS quality, so we first use experiments to measure the impact of different privacy measures on service quality,

$$p = (length_p - length_{cm})/length_p \tag{10}$$

where $length_p$ is the true distance of the results returned by the real position submission query, and $length_{cm}$ is the true distance of the results returned by the anonymous set submission query.

Experimental result: This experiment uses Python language to call Gaode Map API to simulate different anonymity and submit the query to return the best results after optimization. Finally, according to different real location service requests and anonymity set service location requests, the location offset rate is calculated using Eq. (10). As shown in Fig. 7, when maximum anonymous radius $\delta = \{200, 400, 600, \ldots, 1800, 2000\}$, when the minimum anonymity $k = \{2, 4, 6, 8, 10\}$, the effect of different k and δ values on the query results can be seen when k<6 and $\delta \leqslant 1000$, $p \leqslant 0.5$, The deviation rate of results returned after anonymity set query is small, when $6 \leqslant k \leqslant 10$ and $1000 \leqslant \delta \leqslant 2000$, $0.5 \leqslant p \leqslant 1$, the deviation rate of anonymous set query result is directly proportional to the anonymity degree of the requesting user, when the user's privacy requirements are high, the user's location privacy and service quality are guaranteed.

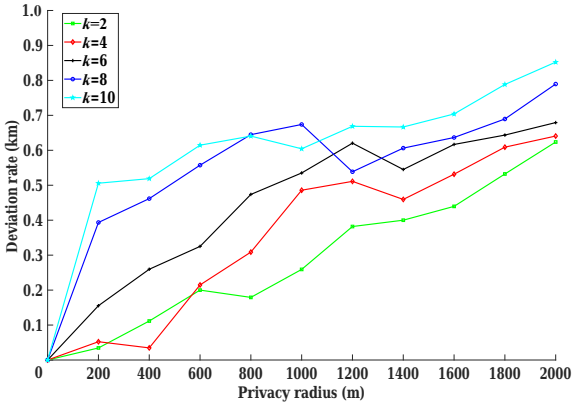

Fig. 7 Deviation rate of query results for different privacy requirements.

6 Conclusion

In this study, we design a best-assisted user selection algorithm to construct an anonymous set that effectively protects the location privacy of users. Then, we contract a smart contract to evaluate user reputation, ensuring that participants are honest. Finally, the proposed protocol uses Dijkstra's algorithm to protect the privacy of a user's location while guaranteeing data availability. The experimental results show that the proposed protocol can protect the privacy of the user's location effectively while guaranteeing data availability.

The proposed protocol is aimed at resisting attack models with background knowledge. Moreover, privacy protection is controllable by users in this protocol. It is a rigorous privacy protection protocol. The protocol mainly focuses on anonymous privacy. Additionally, DP is applied to the anonymity set construction process, and adaptive Laplace noise is added to satisfy the users' privacy requirements, which effectively protects the users' privacy. Then, the query results are optimized and the best results are returned to the requesting user.

The proposed protocol assumes that users voluntarily participate in the construction of anonymous sets as assisting users without incentives; however, in real life, most users are self-interested. A future direction is to consider incentive mechanisms that reward users' participation, enhance their motivation to participate, and consider privacy protection of location data from the perspective of publishing location data results. Additionally, trajectory privacy is an issue worthy of research and attention.

Acknowledgment

This study was supported by the National Natural

Science Foundation of China (No. 61962009), Major Scientific and Technological Special Project of Guizhou Province (No. 20183001), Science and Technology Support Plan of Guizhou Province (No. [2020] 2Y011), and the Foundation of Guizhou Provincial Key Laboratory of Public Big Data (No. 2018BDKFJJ005).

References

[1] X. Pan, Z. Huo, and X. F. Meng, *Location Big Data Privacy Management*, (in Chinese). Beijing, China: Machine Press, 2017.

[2] J. Li, X. Pei, X. J. Wang, D. Y. Yao, Y. Zhang, and Y. Yue, Transportation mode identification with GPS trajectory data and GIS information, *Tsinghua Science and Technology*, vol. 26, no. 4, pp. 403–416, 2021.

[3] M. Azrour, J. Mabrouki, A. Guezzaz, and Y. Farhaoui, New enhanced authentication protocol for internet of things, *Big Data Mining and Analytics*, vol. 4, no. 1, pp. 1–9, 2021.

[4] L. Y. Qi, C. H. Hu, X. Y. Zhang, M. R. Khosravi, S. Sharma, S. N. Pang, and T. Wang, Privacy-aware data fusion and prediction with spatial-temporal context for smart city industrial environment, *IEEE Trans. Industr. Inform.*, vol. 17, no. 6, pp. 4159–4167, 2021.

[5] Y. L. Chen, J. Sun, Y. X. Yang, T. Li, X. X. Niu, and H. Y. Zhou, PSSPR: A source location privacy protection scheme based on sector phantom routing in WSNs, *Int. J. Intell. Syst.*, vol. 37, no. 2, pp. 1204–1221, 2022.

[6] J. Mabrouki, M. Azrour, D. Dhiba, Y. Farhaoui, and S. El Hajjaji, IoT-based data logger for weather monitoring using arduino-based wireless sensor networks with remote graphical application and alerts, *Big Data Mining and Analytics*, vol. 4, no. 1, pp. 25–32, 2021.

[7] Y. Khazbak, J. Y. Fan, S. C. Zhu, and G. H. Cao, Preserving personalized location privacy in ride-hailing service, *Tsinghua Science and Technology*, vol. 25, no. 6, pp. 743–757, 2020.

[8] Y. W. Liu, A. X. Pei, F. Wang, Y. H. Yang, X. Y. Zhang, H. Wang, H. N. Dai, L. Y. Qi, and R. Ma, An attention-based category-aware GRU model for the next poi recommendation, *Int. J. Intell. Syst.*, vol. 36, no. 7, pp. 3174–3189, 2021.

[9] P. Nitu, J. Coelho, and P. Madiraju, Improvising personalized travel recommendation system with recency effects, *Big Data Mining and Analytics*, vol. 4, no. 3, pp. 139–154, 2021.

[10] R. Kumari, S. Kumar, R. C. Poonia, V. Singh, L. Raja, V. Bhatnagar, and P. Agarwal, Analysis and predictions of spread, recovery, and death caused by covid-19 in India, *Big Data Mining and Analytics*, vol. 4, no. 2, pp. 65–75, 2021.

[11] H. S. Chen, Y. P. Zhang, Y. R. Cao, and J. Xie, Security issues and defensive approaches in deep learning frameworks, *Tsinghua Science and Technology*, vol. 26, no. 6, pp. 894–905, 2021.

[12] S. Y. Xu, X. Chen, and Y. H. He, EVchain: An anonymous blockchain-based system for charging-connected electric vehicles, *Tsinghua Science and Technology*, vol. 26, no. 6, pp. 845–856, 2021.

[13] C. Dwork, Differential privacy, in *Proc. 33rd Int. Colloquium on Automata, Languages and Programming*, Venice, Italy, 2006, pp. 1–12.

[14] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor, Our data, ourselves: Privacy via distributed noise generation, in *Proc. 24th Annu. Int. Conf. on the Theory and Applications of Cryptographic Techniques*, St. Petersburg, Russia, 2006, pp. 486–503.

[15] C. Dwork and G. N. Rothblum, Concentrated differential privacy, arXiv preprint arXiv: 1603.01887, 2016.

[16] P. Austrin, Towards sharp inapproximability for any 2-CSP, in *Proc. 48th Annu. IEEE Symp. on Foundations of Computer Science* (*FOCS'07*), Providence, RI, USA, 2007, pp. 307–317.

[17] Q. Geng and P. Viswanath, The optimal noise-adding mechanism in differential privacy, *IEEE Trans. Inf. Theory*, vol. 62, no. 2, pp. 925–951, 2016.

[18] C. Li, G. Miklau, M. Hay, A. McGregor, and V. Rastogi, The matrix mechanism: Optimizing linear counting queries under differential privacy, *VLDB J.*, vol. 24, no. 6, pp. 757–781, 2015.

[19] Y. L. Chen, S. Dong, T. Li, Y. L. Wang, and H. Y. Zhou, Dynamic multi-key FHE in asymmetric key setting from LWE, *IEEE Trans. Inf. Foren. Sec.*, vol. 16, pp. 5239–5249, 2021.

[20] R. Shokri, G. Theodorakopoulos, C. Troncoso, J. P. Hubaux, and J. Y. Le Boudec, Protecting location privacy: Optimal strategy against localization attacks, in *Proc. 2012 ACM Conf. on Computer and Communications Security*, Raleigh, NC, USA, 2012, pp. 617–627.

[21] K. Chatzikokolakis, C. Palamidessi, and M. Stronati, Geo-indistinguishability: A principled approach to location privacy, in *Proc. 11th Int. Conf. on Distributed Computing and Internet Technology*, Bhubaneswar, India, 2015, pp. 49–72.

[22] R. Shokri, Privacy games: Optimal user-centric data obfuscation, *Proc. Priv. Enhanc. Technol.*, vol. 2015, no. 2, pp. 299–315, 2015.

[23] L. N. Ni, C. Li, X. Wang, H. L. Jiang, and J. G. Yu, DP-MCDBSCAN: Differential privacy preserving multi-core DBSCAN clustering for network user data, *IEEE Access*, vol. 6, pp. 21053–21063, 2018.

[24] B. Niu, Q. H. Li, X. Y. Zhu, G. H. Cao, and H. Li, Achieving k-anonymity in privacy-aware location-based services, in *Proc. IEEE Conf. on Computer Communications*, Toronto, Canada, 2014, pp. 754–762.

[25] B. Niu, Q. H. Li, X. Y. Zhu, G. H. Cao, and H. Li, Enhancing privacy through caching in location-based services, in *Proc. 2015 IEEE Conf. on Computer Communications* (*INFOCOM*), Hong Kong, China, 2015, pp. 1017–1025.

[26] H. To, K. Nguyen, and C. Shahabi, Differentially private publication of location entropy, in *Proc. 24th ACM SIGSPATIAL Int. Conf. on Advances in Geographic Information Systems*, Burlingame, CA, USA, 2016, p. 35.

[27] Y. L. Wang, G. Y. Yang, T. Li, F. Y. Li, Y. L. Tian, and X. M. Yu, Belief and fairness: A secure two-party protocol toward the view of entropy for IoT devices, *J. Netw. Comput. Appl.*, vol. 161, p. 102641, 2020.

[28] G. J. Han, H. Wang, M. Guizani, S. Chan, and W. B. Zhang, KCLP: A k-means cluster-based location privacy protection scheme in WSNs for IoT, *IEEE Wirel. Commun.*, vol. 25, no. 6, pp. 84–90, 2018.

[29] L. N. Ni, F. L. Tian, Q. H. Ni, Y. Yan, and J. Q. Zhang, An anonymous entropy-based location privacy protection scheme in mobile social networks, *EURASIP J. Wirel. Commun. Netw.*, vol. 2019, p. 93, 2019.

[30] H. Liu, X. H. Li, B. Luo, Y. W. Wang, Y. B. Ren, J. F. Ma, and H. F. Ding, Distributed k-anonymity location privacy protection scheme based on blockchain, (in Chinese), *Chin. J. Comput.*, vol. 42, no. 5, pp. 942–960, 2019.

[31] T. Li, Z. J. Wang, G. Y. Yang, Y. Cui, Y. L. Chen, and X. M. Yu, Semi-selfish mining based on hidden Markov decision process, *Int. J. Intell. Syst.*, vol. 36, no. 7, pp. 3596–3612, 2021.

[32] T. Li, Z. J. Wang, Y. L. Chen, C. M. Li, Y. L. Jia, and Y. X. Yang, Is semi-selfish mining available without being detected? *Int. J. Intell. Syst.*, doi: 10.1002/int.22656.

[33] Y. L. Wang, G. Y. Yang, A. Bracciali, H. F. Leung, H. B. Tian, L. S. Ke, and X. M. Yu, Incentive compatible and anti-compounding of wealth in proof-of-stake, *Inf. Sci.*, vol. 530, pp. 85–94, 2020.

[34] T. Li, Y. L. Chen, Y. L. Wang, Y. L. Wang, M. H. Zhao, H. J. Zhu, Y. L. Tian, X. M. Yu, and Y. X. Yang, Rational protocols and attacks in blockchain system, *Sec. Commun. Netw.*, vol. 2020, p. 8839047, 2020.

[35] X. J. Zhu, E. Ayday, and R. Vitenberg, A privacy-preserving framework for outsourcing location-based services to the cloud, *IEEE Trans. Dependable Secure Comput.*, vol. 18, no. 1, pp. 384–399, 2021.

[36] Z. Huo and X. F. Meng, A trajectory data publication method under differential privacy, (in Chinese), *Chin. J. Comput.*, vol. 41, no. 2, pp. 400–412, 2018.

[37] X. D. Bi, Y. Liang, H. Z. Shi, and H. Tian, A parameterized location privacy protection method based on two-level anonymity, (in Chinese), *J. Shandong Univ. (Nat. Sci.)*, vol. 52, no. 5, pp. 75–84, 2017.

[38] C. E. Shannon, A mathematical theory of communication, *ACM Sigmobile Mobile Comput. Commun. Rev.*, vol. 5, no. 1, pp. 3–55, 2001.

[39] Y. F. Wang, Y. L. Luo, Q. Y. Yu, Q. Q. Liu, and W. Chen, Trajectory privacy-preserving method based on information entropy suppression, (in Chinese), *J. Comput. Appl.*, vol. 38, no. 11, pp. 3252–3257, 2018.

[40] L. W. Ouyang, S. Wang, Y. Yuan, X. C. Ni, and F. Y. Wang, Smart contracts: Architecture and research progresses, (in Chinese), *Acta Automat. Sin.*, vol. 45, no. 3, pp. 445–457, 2019.

[41] F. Y. Li, D. F. Wang, Y. L. Wang, X. M. Yu, N. Wu, J. G. Yu, and H. Y. Zhou, Wireless communications and mobile computing blockchain-based trust management in distributed internet of things, *Wirel. Commun. Mobile Comput.*, vol. 2020, p. 8864533, 2020.

[42] F. Y. Li, R. Ge, H. Y. Zhou, Y. L. Wang, Z. X. Liu, and X. M. Yu, Tesia: A trusted efficient service evaluation model in internet of things based on improved aggregation signature, *Concurr. Comput.: Pract. Exp.*, doi: 10.1002/cpe.5739.

[43] B. K. Samanthula, D. Karthikeyan, B. X. Dong, and K. A. Kumari, ESPADE: An efficient and semantically secure shortest path discovery for outsourced location-based services, *Cryptography*, vol. 4, no. 4, p. 29, 2020.

Ping Guo received the BEng degree from Nanchang Jiaotong Institute, China in 2019. He is currently a master student at the School of Computer Science and Technology, Guizhou University. His research interests focus on cryptography, difference privacy, and location privacy protection.

Yuling Chen received the BEng degree from Taishan University, China in 2006, the MEng degree from Guizhou University, China in 2009, and the PhD degree from Guizhou University, China in 2021. She is currently a professor at State Key Laboratory of Public Big Data, Guizhou University. Her research interests focus on cryptography and information safety, as well as blockchain.

Baopeng Ye received the MEng degree from Liupanshui Normal University, China in 2019. He is currently working at Information Technology Innovation Service Center of Guizhou Province. His research interests focus on information security, privacy protection, and big data.

Tao Li received the BEng degree from Shandong Normal University, China in 2001, the MEng degree from Dalian University of Technology, China in 2007. He is currently a PhD candidate at the school of Computer Science and Technology, Guizhou University. His research interests focus on information security, cryptography, and blockchain technology.

Yixian Yang received the BEng degree from University of Electronic Science and Technology, China in 1983, the MEng and PhD degrees from Beijing University of Posts and Telecommunications, China in 1986 and 1988, respectively. He is currently the director of Information Security Center, Beijing University of Posts and Telecommunications. His research interests focus on coding and cryptography, information and network security, and signal and information processing.

Xiaobin Qian received the BEng and MEng degrees from Tsinghua University, China in 1996 and 2001, respectively. He is currently working at Guizhou CoVision Science & Technology Co., Ltd. His research interests focus on network security, big data, and artificial intelligence.

Xiaomei Yu received the BEng and MEng degrees from Shandong Normal University, China in 1996 and 2004, respectively, the PhD degree from Shandong Normal University, China in 2016. She is currently an associate professor of computer education at School of Information Science and Engineering, Shandong Normal University. Her research interests focus on data mining, recommended system, and big data cloud.