Personalized Location Privacy Protection Algorithm for Crowd-Sensing Networks: Postprint

Authors: Hu Min, Zhang Yan, Huang Hongcheng

Date: 2018-05-18T00:00:00+00:00

Abstract

Existing privacy protection algorithms in crowd-sensing networks adopt identical privacy protection strategies for all locations, resulting in either excessive or insufficient protection of location privacy, and the obtained sensing data suffers from low accuracy. To address this issue, a location privacy protection algorithm that satisfies users' personalized privacy and security requirements is proposed. First, based on users' historical mobility trajectories, the duration of visits, visit frequency, and regularity of visits to different locations are mined to predict the social attributes of locations for users; then, combined with the natural attributes of locations, the sensitivity level of user-location pairs is predicted; finally, by leveraging the characteristic that users have different privacy and security requirements at different locations, a dynamic privacy determination scheme is established, where users with low sensitivity are selected to participate in sensing tasks at each location, thereby ensuring that users contribute highly accurate spatio-temporally correlated sensing data under the premise of privacy security. Simulation results demonstrate that the algorithm improves the accuracy of sensing data while simultaneously enhancing the level of privacy protection.

Full Text

Personalized Location Privacy Protection Algorithm in Crowd Sensing Networks

Hu Min, Zhang Yan, Huang Hongcheng† School of Communication & Information Engineering, Chongqing University of Posts & Telecommunications, Chongqing 400065, China

Abstract

Existing privacy protection algorithms in crowd sensing networks apply uniform privacy protection strategies to all locations, resulting in either excessive or insufficient protection for certain locations and yielding low-precision sensing data. To address this issue, this paper proposes a location privacy protection algorithm that satisfies users' personalized privacy and security requirements. First, the algorithm mines users' access duration, frequency, and regularity at different locations based on their historical movement trajectories to predict the social attributes of locations to users. Then, it combines the natural attributes of locations with these social attributes to predict user-location sensitivity levels. Finally, considering that users have varying privacy security requirements at different locations, the algorithm establishes a dynamic privacy decision scheme that selects users with low sensitivity at each location to participate in sensing tasks. This ensures that users can contribute highly accurate spatiotemporally correlated sensing data while maintaining privacy security. Simulation results demonstrate that the algorithm improves both privacy protection levels and sensing data accuracy.

Keywords: location privacy protection; personalization; sensitive level; crowd sensing networks

0 Introduction

Crowd sensing refers to the formation of interactive, participatory sensing networks through people' s existing mobile devices, distributing sensing tasks to individuals or groups within the network to help professionals or the public collect data, analyze information, and share knowledge. By leveraging existing sensing devices and communication networks, crowd sensing networks can be constructed without additional deployment and maintenance of sensing equipment or transmission infrastructure, enabling large-scale, fine-grained sensing at low cost. These networks are widely applied in intelligent transportation, social networks, environmental monitoring, health monitoring, and other domains.

Since most sensing data requires associated spatiotemporal location information to be valuable, and this information directly contains users' private data while also implicitly revealing other sensitive information such as home addresses, lifestyle habits, health conditions, and social relationships, users face unprecedented privacy security risks. If users' privacy cannot be adequately protected, they may become unwilling to share their sensing data. Therefore, research on privacy protection in crowd sensing networks, particularly location privacy protection for users participating in spatiotemporal-related sensing tasks, holds significant importance.

1 Related Work

Numerous studies have investigated location privacy protection in crowd sensing networks, with location anonymization being a commonly employed method. Mehta et al. utilize dummy locations or semantically related locations around the true position to replace real locations when submitting data to servers. While this approach effectively protects location privacy, it severely impacts the quality of sensing data services and degrades user experience. Reference [6] adopts K-anonymity to protect location privacy by replacing a user' s true location with a spatial region containing K users, among whom any individual' s location is indistinguishable from the other K-1 users. However, this method introduces severe spatial distortion and fails to meet sensing data utility requirements. Yu et al. propose the l-diversity principle, ensuring that each equivalence class satisfying K-anonymity contains at least l distinct sensitive attribute values. L-diversity prevents homogeneous sensitive attribute values within an equivalence class, limiting privacy leakage risk to no more than 1/l, but remains vulnerable to similarity attacks.

Reference [10] employs differential privacy by adding random noise to data, ensuring that an adversary can obtain little more personal information than they could from a dataset without that individual' s record, thereby achieving privacy protection. However, since the original Laplacian noise used is unbounded, this renders the published data meaningless, causing privacy leakage and hindering data publication utility. To address this, reference [11] proposes a differentiated data publication algorithm with bounded Laplacian noise generation. This noise generation mechanism samples noise within an appropriate range to achieve differential privacy, significantly reducing privacy loss and improving data publication utility.

Another common location privacy protection method involves encryption techniques. Wei et al. propose a base station-based attribute encryption algorithm combined with pseudonym technology to achieve isolation between user identity and data storage. Reference [13] presents a secure framework based on homomorphic encryption that ensures users' location information is never published to anyone while still enabling the system to assign tasks to sensing users near each sensing task location. Although this method effectively protects user location privacy, the strong mobility of nodes in crowd sensing networks results in excessive key update overhead and complex key distribution processes.

Reference [14] proposes a static privacy protection mechanism employing location obfuscation and data hiding techniques, assuming all locations require the same privacy protection level. Consequently, both the average static algorithm (satisfying average privacy threshold) and max static algorithm (satisfying maximum privacy threshold) use statically fixed privacy protection parameters. While this static parameter approach provides some protection, it sacrifices data value and neglects data utility. To resolve this, reference [15] proposes an adaptive location privacy protection method that, based on static privacy pro-

tection strategies, incorporates sensing application utility and comprehensively considers the trade-off between utility and privacy. However, this method still overlooks personalized privacy requirements of different users at different locations.

Moreover, sensing scenarios with high spatiotemporal correlation requirements often demand participants to submit high-precision sensing data, yet existing location privacy protection methods provide coarse-grained sensing data to protect privacy, significantly reducing service quality. To solve this problem, this paper proposes a location privacy protection algorithm for personalized security requirements of different users (LPPA-PSRDU). Based on users' historical trajectory information, the algorithm performs real-time sensitivity calculations to provide a dynamic privacy decision scheme for sensing users. At each location, it selects users with low sensitivity to participate in sensing while providing spatiotemporally precise sensing data, thereby protecting participants' location privacy and improving sensing data service quality.

2 Personalized Privacy Protection Algorithm

The privacy protection algorithm proposed in this paper for users' personalized privacy and security requirements primarily considers that different participants have varying sensitivity levels to the same location. By evaluating users' sensitivity levels to each location in real-time through their historical trajectory records, the algorithm selects users with lower sensitivity to complete data collection at that location. This achieves the goal of satisfying all users' personalized privacy and security requirements while providing high-precision sensing data.

To reasonably evaluate a sensing user' s sensitivity to a particular location, we consider both the natural attributes inherent to the location and the social attributes that the location holds for that specific user.

Definition 1 (Natural Attribute). Natural attributes refer to the unified functional status of a location for all public users. For example, a hospital has the same inherent natural attribute for all patients and does not vary from person to person.

Definition 2 (Social Attribute). Social attributes refer to the specific characteristics a location holds for a particular user, which vary according to the user' s social relationships. For instance, if location A is Person A' s home, Person B' s friend' s house, and a place Person C occasionally passes by, then location A has different social attributes for A, B, and C.

2.1 Natural Attributes

Since most modern sensing devices are equipped with GPS positioning capabilities, we can determine the natural attribute of a location based on GPS positioning, denoted as l_{ij}^{n} representing the natural attribute of location j for user i.

Locations with different natural attributes contain different types of privacy information, and users have varying sensitivity levels to them. For example, hospitals, banks, parks, supermarkets, wilderness areas, and rivers have distinct natural attributes. For patients, hospitals have very high sensitivity as they are typically unwilling to disclose their health conditions. Similarly, privacy-related information from banking transactions requires high confidentiality. In contrast, users have lower sensitivity to supermarkets, parks, wilderness areas, or rivers because these places contain less privacy information that could harm personal interests. Therefore, we can preliminarily determine a location' s natural attribute based on GPS positioning information and assign a natural attribute sensitivity value l_{ij}^{n} according to its real-world significance.

2.2 Social Attributes

2.2.1 Visit Duration By analyzing users' movement trajectories, reference [16] found that each user' s daily average activity time follows a clear power-law distribution. Based on a user' s historical trajectory records over a period, we can calculate the average visit duration to each location, reflecting the user' s dependence on that location. If location A is user A' s home, user A will frequently appear at that location except for special circumstances such as travel or business trips. If location A has no special social significance for user B, who only occasionally passes by, the visit duration to that location will be very short within a given period. Thus, visit duration can reflect the social significance of a location to a user to some extent, thereby determining the user' s sensitivity to that location.

Considering that users are constantly moving and the social significance of locations may change (e.g., when a user moves to a new home), visit durations can be updated dynamically in real-time as movement trajectories are updated.

We define the visit duration t_{ij} of user i to location j as:

$$t_{ij} = \int_{t_1}^{t_2} f(t)\, dt$$

where t_1 and t_2 represent the start and end times of the access record, respectively, and $f(t)$ is defined as:

$$f(t) = \begin{cases} 1 & \text{if the user is at the location} \\ 0 & \text{otherwise} \end{cases}$$

To normalize t_{ij}, we use a Gaussian similarity function to obtain the relationship strength between user i and location j:

$$C_{ij} = e^{-\frac{t_{ij}^2}{2\sigma^2}}$$

where σ is a scaling parameter for the separation period.

2.2.2 Visit Frequency A user' s movement trajectory during time period (t_1, t_2) is represented as $L = \{l_1, l_2, ..., l_n\}$, where l_j denotes one of the locations visited by the user. Reference [17] indicates that location semantic information can be derived from visit frequency rankings. If a user visits a location with high frequency, it reflects the location' s importance to the user, suggesting that the location may contain more user privacy information. Therefore, we also consider visit frequency as a metric for determining user sensitivity. As user movement trajectories change continuously, visit frequencies to each location can be updated dynamically to achieve real-time prediction of user sensitivity to each location.

Visit frequency refers to the ratio of a user' s visits to a particular location to the total visits to all locations in the entire movement trajectory during a given period. We define the visit frequency f_{ij} of user i to location j as:

$$f_{ij} = \frac{N(l_j)}{N(l_i)}$$

where $N(l_j)$ represents the frequency of user i arriving at location j, and $N(l_i)$ represents the total visit frequency to all locations in user i' s entire movement trajectory.

2.2.3 Regularity of Visits To more accurately predict user sensitivity, we must consider another metric: regularity of visits. Regularity reflects whether a user' s visits to a location conform to normal patterns, thereby excluding misjudgments caused by occasional factors. Consider a scenario where a sensing user stays at a location for several consecutive days due to travel or business trips, but actually has low sensitivity to that location. The calculated visit duration would be high, leading to misjudgment. Alternatively, if a user frequently passes by a location without pattern, the calculated visit frequency would be high, also causing misjudgment. For private locations such as homes or workplaces, users typically exhibit both high visit duration and high visit frequency with regular patterns rather than occasional visits.

To calculate the regularity of a user' s visits to a location, we first compute the average separation period between the user and the location—that is, how often the user leaves the location. The average separation period A_{ij} is defined as:

$$A_{ij} = \frac{1}{n} \int_{t_1}^{t_2} \delta_{ij}(t)\, dt$$

where $\delta_{ij}(t)$ is defined as:

$$\delta_{ij}(t) = \begin{cases} 1 & \text{if the user is at the location} \\ 0 & \text{otherwise} \end{cases}$$

To ultimately reflect the regularity of a user' s visits to a location, we measure the variance of separation periods and use an irregularity metric I_{ij} to reflect the magnitude of fluctuation (regularity):

$$I_{ij} = \frac{1}{n} \sum_{X=l}^{l_{ij}} (C_{ij} - X)^2$$

where C_{ij} represents the separation period length.

2.2.4 Sensitivity Level Function To satisfy all users' personalized privacy and security requirements, we establish a real-time, dynamic privacy classification model. Considering both the natural attributes of locations and the metrics of visit duration, visit frequency, and regularity, we define a sensitivity level function to determine a user' s sensitivity to a visited location. For a given sensing task, users with low sensitivity to that location are selected to participate. For any given sensing user, different locations elicit different sensitivity levels, allowing the user to contribute sensing resources at locations with lower sensitivity. This approach protects user privacy information while maximizing the utilization of sensing resources and providing high-precision sensing data under the premise of meeting personalized privacy security requirements.

The final sensitivity level function S_{ij} is defined as:

$$S_{ij} = \alpha \cdot l_{ij}^n + \beta \cdot D_{ij} + \omega \cdot f_{ij} + \mu \cdot A_{ij}$$

where α, β, ω, and μ are parameters that adjust the weight of each metric.

2.2.5 User-Location Matrix Based on users' sensitivity values to locations, we construct matrix M, where columns represent multiple locations and rows represent multiple users. This matrix can represent the sensitivity relationships between users and locations within a certain area. For a specific location, we can select k users with lower sensitivity levels to complete sensing tasks. For a specific user, we can select m locations with lower sensitivity levels for participation, contributing sensing resources. An example matrix is shown in equation (9):

$$M = \begin{bmatrix} l_{11} & l_{12} & l_{13} & \cdots & l_{1m} \\ l_{21} & l_{22} & l_{23} & \cdots & l_{2m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ l_{n1} & l_{n2} & l_{n3} & \cdots & l_{nm} \end{bmatrix}$$

where $1 \leq i \leq n$ and $1 \leq j \leq m$.

Let u_i represent user i and l_j represent location j. The user-location corresponding sensitivity threshold matrix D can be expressed as:

$$D = \begin{bmatrix} \beta_{11} & \beta_{12} & \beta_{13} & \cdots & \beta_{1m} \\ \beta_{21} & \beta_{22} & \beta_{23} & \cdots & \beta_{2m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \beta_{n1} & \beta_{n2} & \beta_{n3} & \cdots & \beta_{nm} \end{bmatrix}$$

where $\beta_{ij} = 1 - l_{ij}$ and $0 \leq \beta_{ij} \leq 1$.

2.3 Real-Time Dynamic Privacy Classification Algorithm

Building upon the aforementioned sensitivity parameter calculation methods, we present a real-time dynamic privacy classification algorithm to satisfy each user' s privacy and security requirements at different locations. This ensures users can contribute sensing data without leaking private information while providing dynamic privacy decisions through real-time sensitivity calculations to protect user privacy and improve sensing data accuracy in sensing environments with dynamic, unpredictable attack methods.

The real-time dynamic privacy classification algorithm proceeds as follows:

a) Determine the natural attributes of certain special locations based on GPS positioning from sensing devices.

b) Obtain users' historical trajectory records, which need to be regularly updated with the latest records (weekly updates in this paper).

c) Based on users' access records, calculate visit duration, visit frequency, and regularity metrics for each location within the current period to characterize the social attributes that each location holds for each user.

d) Integrate the natural and social attributes of each location and calculate user sensitivity to the location using the sensitivity function.

e) Store user-location sensitivity values in the user-location matrix. The system dynamically sets privacy thresholds θ based on user quantity and sensing task requirements (where $0 \leq \theta \leq 1$, $\theta = 0$ represents no privacy protection, and $\theta = 1$ represents maximum privacy protection). If $\beta_{ij} \leq \theta$, the user can participate in sensing tasks at that location without privacy leakage; otherwise, the location is deemed sensitive for the user, and participation is prohibited.

Since the same location has different practical significance and sensitivity for different users, selecting users with low sensitivity ensures sensing task completion. Similarly, since the same user has different sensitivity levels at different locations, the user can contribute sensing resources at locations that won' t leak

personal privacy information. This approach both protects user privacy and maximizes sensing resource utilization.

3 Experiments

3.1 Experimental Data Preprocessing

For ease of description, we define the following terms: GPS record (P), GPS trajectory (Traj), stay point (S), and location history (LocH).

Definition 3 (GPS Record). A GPS record is a collection of GPS points $P = \{p_1, p_2, ..., p_n\}$. Each GPS point p_i contains latitude ($p_i.Lat$), longitude ($p_i.Lngt$), and timestamp ($p_i.T$).

Definition 4 (GPS Trajectory). A GPS trajectory is a curve connecting GPS points according to their time sequence. As shown in [Figure 1: see original paper], if the time interval between two consecutive GPS points exceeds a determined threshold ΔT, these two points are decomposed into two different GPS trajectories. Therefore, for $p_i \in P$, $p_{i+1}.T > p_i.T$, and $p_{i+1}.T - p_i.T < \Delta T$ $(1 \leq i < n)$.

Definition 5 (Stay Point). A stay point represents a geographic area where a user remains for a period, as shown by stay point S in [Figure 1: see original paper]. In this paper, each stay point carries specific semantic meaning, such as living or working places, visited restaurants, or tourist attractions. Stay point extraction depends on two parameters: time threshold (T_{threh}) and distance threshold (D_{threh}). Similar to points $\{p_3, p_4, p_5, p_6\}$ in [Figure 1: see original paper], a single stay point S can be considered as a set of consecutive GPS points $P = \{p_m, p_{m+1}, ..., p_n\}$. The stay point S is determined by three factors: D_{threh}, T_{threh}, and the average coordinates of these GPS points. Since S typically involves a spatial region containing multiple GPS points, we calculate the region' s average coordinates based on these points:

$$S = (S.Lat, S.Lngt, S.arvT, S.levT)$$

where $S.Lat = \frac{1}{|P|}\sum_{i=m}^{n} p_i.Lat$, $S.Lngt = \frac{1}{|P|}\sum_{i=m}^{n} p_i.Lngt$, $|P|$ represents the number of elements in set P, $S.arvT = p_m.T$ represents the user' s arrival time at stay point S, and $S.levT = p_n.T$ represents the departure time.

Definition 6 (Location History). Location history is a record of locations visited by an entity in geographic space over a period. In this paper, a person' s location history $LocH = \langle s_1 \xrightarrow{\Delta t_1} s_2 \xrightarrow{\Delta t_2} ... \xrightarrow{\Delta t_{n-1}} s_n \rangle$ represents the sequence of arrival and departure times for visited stay points, where $\Delta t_i = s_{i+1}.arvT - s_i.levT$ represents the time interval between different stay points visited by the user.

3.2 Experimental Environment

The algorithm is implemented in C++ and runs on a Windows 7 platform with an Intel(R) Core(TM) i3-2350M 2.3 GHz processor and 4 GB memory. The simulation uses real datasets collected from the GeoLife project [18], which contains data captured by GPS loggers at intervals of 2-5 seconds from April 2007 to August 2012. The dataset includes 18,670 GPS trajectory records from 182 users, comprising 24.87 million data points, making it a typical spatiotemporal dataset.

Stay Point Detection: In this experiment, we set parameters $T_{threh} = 20$ minutes and $D_{threh} = 200$ meters. If a user remains within a 200-meter region for over 20 minutes, that region is identified as a stay point (i.e., a location).

3.3 Algorithm Performance Metrics

This section introduces metrics to evaluate the proposed algorithm' s effectiveness, comparing its performance with previous schemes through privacy protection level, data completeness, and data accuracy.

3.3.1 Privacy Protection Level Participant privacy leakage probability $P_{Disclosure}$ is defined as the ratio of the number of location privacy leaks to the total number of locations requiring privacy protection:

$$P_{Disclosure} = \frac{n_{Disclosure}}{N_{Disclosure}}$$

where $n_{Disclosure}$ is the number of leaked locations and $N_{Disclosure}$ is the total number of locations requiring privacy protection. Privacy protection level $ProtectionLevel$ is defined as:

$$ProtectionLevel = 1 - P_{Disclosure}$$

3.3.2 Data Completeness An important factor affecting sensing data availability is data loss. Due to privacy concerns, some sensing data collected by users may not be uploaded to servers. Data completeness $DataCompleteness$ is defined as:

$$DataCompleteness = \frac{\sum_{i=1}^{k} d_{data}^{i}}{\sum_{i=1}^{N} D_{data}^{i}}$$

where d_{data}^{i} represents the sensing data submitted by participant i to the server, and D_{data}^{i} represents the total amount of data sensed by participant i.

3.3.3 Data Accuracy Due to privacy protection concerns, participants may submit imprecise or coarse-grained location information, leading to decreased sensing data precision and affecting data availability. We measure data accuracy using the mean absolute error between submitted sensing data locations and actual locations:

$$DataAccuracy = \frac{1}{N}\sum_{i=1}^{N}|r_i - r_i'|$$

where r_i' is the submitted sensing data location, r_i is the actual location, and N is the total number of locations.

3.4 Experimental Results Analysis

To evaluate the performance of the proposed LPPA-PSRDU algorithm, we conduct simulation comparisons under identical conditions with two static location privacy protection schemes (Avg Static and Max Static) and one adaptive privacy protection strategy (Adaptive algorithm). The algorithms are assessed from three perspectives: privacy protection level, data completeness, and data accuracy. Parameter settings for the algorithms are shown in .

TABLE:1 Algorithm Parameter Settings

Algorithm	Parameters
Avg Static	$\lambda = 0.5, \alpha = 0.3, \omega = 0.2, \mu = 0.2$
Max Static	$\lambda = 0.8, \alpha = 0.3, \omega = 0.2, \mu = 0.2$
Adaptive	$\lambda = 0.5, \alpha = 0.3, \omega = 0.2, \mu = 0.2$
LPPA-PSRDU	$\lambda = 0.5, \alpha = 0.3, \omega = 0.2, \mu = 0.2$

3.4.1 Privacy Protection Level The privacy protection levels of the four algorithms are shown in [Figure 2: see original paper]. As the privacy threshold increases, the privacy protection levels of all four algorithms rise because higher thresholds impose stronger location privacy protection. However, under the same privacy threshold conditions, the dynamic privacy protection mechanisms (Adaptive and LPPA-PSRDU) achieve significantly higher privacy protection levels than the static mechanisms (Max Static and Avg Static). Static strategies assume all participants' locations require the same protection level, using predefined fixed parameters that result in insufficient protection for some locations and excessive protection for others. Dynamic strategies enable the system to calculate the actual required privacy protection level and adjust parameters dynamically to meet diverse location-specific privacy requirements. The proposed LPPA-PSRDU algorithm achieves higher privacy protection levels than Adaptive because, although Adaptive dynamically calculates participants' privacy levels, some privacy leakage still occurs after location obfuscation. In contrast,

LPPA-PSRDU dynamically selects users with low sensitivity to participate in sensing tasks, satisfying different users' privacy needs.

3.4.2 Data Completeness Sensing data completeness is illustrated in [Figure 3: see original paper]. At low privacy thresholds, all four algorithms maintain high data completeness. At high privacy thresholds, more data is lost because higher thresholds require hiding more sensing data to achieve stronger privacy protection. Overall, the two dynamic privacy protection algorithms outperform the static ones in data completeness because static algorithms apply uniform privacy protection levels across all locations, causing excessive protection at non-sensitive locations and wasting sensing resources. Between the two dynamic algorithms, LPPA-PSRDU shows slightly lower data completeness than Adaptive because, at high privacy thresholds, Adaptive employs location obfuscation to upload sensing data, whereas LPPA-PSRDU only allows a subset of qualified participants to contribute sensing data, excluding those with excessively high sensitivity.

3.4.3 Data Accuracy The accuracy of sensing data for the four algorithms is shown in [Figure 4: see original paper]. As the privacy threshold increases, the data accuracy of Max Static, Avg Static, and Adaptive gradually decreases, while the accuracy of data submitted using LPPA-PSRDU remains close to 100%. This occurs because the other three algorithms use location obfuscation mechanisms—higher privacy thresholds require larger generalization areas, reducing data accuracy. In contrast, LPPA-PSRDU enables participants with low sensitivity to submit accurate sensing data while meeting privacy requirements. The LPPA-PSRDU mechanism demonstrates clear advantages in sensing scenarios with high spatiotemporal correlation requirements.

Compared to the Adaptive algorithm, LPPA-PSRDU sacrifices some data completeness but achieves higher privacy protection levels and data accuracy. In scenarios with many participating users and high spatiotemporal correlation requirements for sensing data, the proposed algorithm is highly significant. However, the method has limitations, such as substantial computational requirements when evaluating user sensitivity to locations. Future work will focus on rapidly and accurately assessing user sensitivity to better protect privacy.

4 Conclusion

To address severe location privacy leakage in crowd sensing networks, where existing algorithms apply uniform privacy protection intensity to all locations—causing excessive protection for some locations, insufficient protection for others, and low-precision sensing data—this paper proposes a personalized location privacy protection algorithm (LPPA-PSRDU). The algorithm calculates users' visit duration, frequency, and regularity to different locations based on historical GPS trajectory data, mines the social attributes that each location holds for different users, and combines these with location natural attributes to obtain a

user-location sensitivity level matrix. At each location, users with sensitivity levels below the privacy threshold are selected to participate in sensing tasks. The algorithm is suitable for sensing scenarios with high spatiotemporal correlation requirements, achieving high-precision sensing data while meeting users' privacy security needs at different locations. Comparative experiments with two static location privacy protection algorithms (Avg Static and Max Static) and one adaptive algorithm (Adaptive) demonstrate that LPPA-PSRDU outperforms Avg Static and Max Static in privacy protection level, data completeness, and data accuracy. Compared to Adaptive, the proposed algorithm achieves higher privacy protection levels and data precision at the cost of slightly reduced data completeness.

References

[1] Wu Yao, Zeng Juru, Peng Hui, et al. Survey of incentive mechanisms in crowd sensing [J]. Journal of Software, 2016, 27(8): 2025-2047.

[2] Sun W, Liu J. Congestion-aware communication paradigm for sustainable dense mobile crowd-sensing [J]. IEEE Communications Magazine, 2017, 55(3): 62-67.

[3] Zhang Xuejun, Gui Xiaolin, Wu Zhongdong. Survey of location service privacy protection [J]. Journal of Software, 2015, 26(9): 2373-2395.

[4] Shen H, Bai G, Yang M, et al. Protecting trajectory privacy: a user-centric analysis [J]. Journal of Network & Computer Applications, 2017, 82: 128-140.

[5] Mehta K, Liu D, Wright M. Protecting location privacy in sensor networks against a global eavesdropper [J]. IEEE Trans on Mobile Computing, 2011, 11(2): 320-336.

[6] Yang D, Fang X, Xue G. Truthful incentive mechanisms for k-anonymity location privacy [C]//Proc of INFOCOM. Turin: IEEE Infocom, 2013: 1885-1893.

[7] Ziegeldorf J H, Henze M, Bavendiek J, et al. TraceMixer: privacy-preserving crowd-sensing sans trusted third party [C]//Proc of Wireless On-demand Network Systems and Services. Jackson, WY: IEEE Press, 2017: 115-122.

[8] Yu Z T, Qian Q, Lin C Y, et al. High performance datafly based anonymity algorithm and its l-diversity [J]. International Journal of Grid and High Performance Computing, 2015, 7(3): 85-100.

[9] Zhang Yixuan, He Jingsha, Zhao Bin, et al. A privacy protection model based on game theory [J]. Chinese Journal of Computers, 2016, 39(3): 615-627.

[10] Zhang Ning, Li Ming, Lou Wenjing. Distributed data mining with differential privacy [C]//Proc of IEEE International Conference on Communications. Kyoto: IEEE Press, 2011: 1-5.

[11] Li M, Zhu L, Zhang Z, et al. Achieving differential privacy of trajectory data publishing in Participatory Sensing [J]. Information Sciences, 2017, 400-401: 1-13.

[12] Wei Z, Zhao B, Liu Y, et al. PPSense: a novel privacy-preserving system in people-centric sensing networks [C]//Proc of the 8th International ICST Conference on Communications and Networking in China. 2013: 461-467.

[13] Liu Bozhong, Chen Ling, Zhu Xingquan, et al. Protecting location privacy in spatial crowdsourcing using encrypted data [C]//Proc of the 20th International Conference on Extending Database Technology. Venice, Italy: OpenProceedings, 2017: 21-24.

[14] Krumm J. A survey of computational location privacy [J]. Personal and Ubiquitous Computing, 2009, 13(6): 391-399.

[15] Agir B, Papaioannou T G, Narendula R, et al. User-side adaptive protection of location privacy in participatory sensing [J]. Geoinformatica, 2014, 18(1): 165-191.

[16] Henderson T, Kotz D, Abyzov I. The changing usage of a mature campus-wide wireless network [J]. Computer Networks, 2008, 52(14): 2690-2712.

[17] Prabhala B, Porta T L. Spatial and temporal considerations in next place predictions [J]. Proceedings IEEE INFOCOM, 2015, 2015(9): 390-395.

[18] Zheng Y, Xie X, Ma W Y. GeoLife: a collaborative social networking service among user, location and trajectory [J]. Bulletin of the Technical Committee on Data Engineering, 2010, 33(2): 32-39.

Note: Figure translations are in progress. See original paper for figures.