IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received February 27, 2019, accepted April 2, 2019, date of publication April 15, 2019, date of current version April 25, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2911107

Enabling Smart Anonymity Scheme for Security Collaborative Enhancement in Location-Based Services

HONGCHEN WU, MINGYANG LI, AND HUAXIANG ZHANG
School of Information Science and Engineering, Shandong Normal University, Jinan 250014, China

Corresponding author: Hongchen Wu (wuhongchen@sdnu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61702312, in part by the Natural Science Foundation of Shandong Province of China under Grant ZR2017BF019, and in part by the Project of Shandong Province Higher Educational Science and Technology Program under Grant J17KB178. The work of H. WU was supported by the China Scholarship Council (CSC) under Grant 201306220132.

ABSTRACT Security enhancement is and always will be a prime concern for the deployment of point-of-interest (POI) recommendation services in mobile sensing environment. Recent tamper-proof technical protection such as strong encryption has undoubtedly become a major safeguard against threats to privacy in location-based services. Although the disclosure of location information could increase recommendation accuracy, the publication of trajectory data to untrusted entities could reveal sensitive details, e.g., daily routes, destinations, and favorite restaurants. In this paper, we propose a smart scheme named BUSA to approach the above problem by reconciling the tension between privacy protection and recommendation accuracy in location-based recommendation services. This scheme uses anonymizer agents positioned between the service-requesting users and location service providers; these agents operate by dividing the query information and using k-anonymity to enhance privacy protection. The scheme also utilizes clustering techniques to group users into clusters by learning their trajectory data and selects the spatial center cells as a cluster core and a benchmark for calculating recommendations via trust computing. An anonymizer coordination strategy is proposed to replace a low-performing anonymizer with one that provides stronger privacy protection for a recommendation service. The BUSA scheme adopts k-anonymity and clustering to protect privacy, and the calculated recommendation will be suitable for the cluster core that represents the entirety of users' location preferences. The security analysis reveals that the BUSA scheme can effectively protect privacy against fraudulent query requestors and the simulation results also indicate that it provides stronger privacy protection than its counterparts from the perspective of recommendation hit rate and the extent of disclosure.

INDEX TERMS Collaboration for security, k-anonymity, location-based services, clustering, recommendation and security, trajectory, mobile sensing.

I. INTRODUCTION

The contemporary proliferation of smart devices has contributed to development of impressive services that have by now become available, where enabling convenient and configurable computing sources, e.g., mobile devices, requires integration of artificial intelligence into the physical world [1]–[3]. The so-called Internet of Things (IOT) has provided a great opportunity for collaboration with reduced management effort and better interaction among infrastructure applications compared to those of traditional closed-source networks and vendor-specific proprietary technologies that cannot be provided at such low costs under similar circumstances [4]. In location-based services (LBS), users have to upload their location information to the LBS server and receive the nearby points of interests (POIs) with the optimum routes [5]; examples include finding the closest restaurant during a one-hour lunch break, sending a patient with acute gastritis to the best emergency hospital, driving to the office early in the morning via an alternative route to avoid a traffic jam, etc. However, simply connecting these smart devices without considering their security will expose user data to unexpected

The associate editor coordinating the review of this manuscript and approving it for publication was Mamoun Alazab.

potential risks [6]. A disclosure of sensitive information can result not only in great convenience but also in leakage of such information, leading to its unauthorized use by a third party [7]. Specifically, queries intended for the LBS server may be intercepted by an adversary that has just compromised the LBS server, and the attached personal information of the victim's only route, preferred restaurant, and even trajectory data can be inferred [8]. As a result, security issues have drawn widespread attention and need to be solved.

Many approaches to lowering the potential risk in LBS services have been proposed and mainly concentrate on privacy protection with anonymization during the information delivery process intermediated by an absolutely trusted third-party anonymizer (ATTA). The ATTA acts as an intermediate tier between the user and the LBS server by anonymizing the trajectory data. When a user of a smart device uploads his or her query to the anonymizer agent, and the data are generalized by the cloaking method, the data is made the ''same'' as that of other k-1 users. This method is called k-anonymity [9]–[11]; it confuses an online adversary and makes it difficult to distinguish a user from at least $k - 1$ other users. The location service provider (LSP) receives the cloaked query, generates the candidate POIs and returns them to the user [12]. Nevertheless, the ATTA architecture still suffers from two significant flaws, as discussed in IOT literature.

First, the generated POIs aim to serve the k users but not the anonymized user. The value of k indicates the overall gaps among the k users, and if this value is very high, the POIs may not be accurate. Suppose that user Nancy is searching for a nearby restaurant; the agent anonymizes her trajectory data together with that of 99 other users in a small cloaking region, or with five other users in a large cloaking region. Obtaining a desirable service is always a major concern from the user's perspective. Excessive anonymization would lead to a too low personalization quality.

Second, another barely explored issue is whether the anonymizer agent is trustworthy when compared with the LBS. If an adversary can hack an LBS, then it may also be able to hack an anonymizing agent. Although the failure of a single agent may not corrupt the entire scheme, without a trust computing methodology, an LBS may be easily subverted by a fraudulent agent that sends most of the query requests. Moreover, hardly any studies have focused on the theoretical selection of an agent as a candidate backup agent to replace a fraudulent agent within a limited reaction time. The new randomly selected backup agent may not provide a proper personalized service due to the cold-start problem, resulting in a loss of users' confidence.

The above two weaknesses both lower the quality of the location recommendation service in device collaboration of the IoT. To the best of our knowledge of the state-of-the-art literature, hardly any existing studies have reconciled the tension between security and service quality in the context of location-based recommendation. To fill this gap, the main contributions of this paper are as follows:

1) We present a scheme called BUSA that applies a modified k-anonymity to protect user's trajectory data during the user's query to the LBS. This scheme deploys multiple agents for anonymization between the user and the LBS with a proper number of agents and a proper value of k, depending on the size of cloaking on the map. The queried location information is divided into segments that are sent to several different agents for anonymization. In this case, a single agent will not know the entire private data.
2) A clustering technique is applied to select the spatial center containing the core users as the cloaking region. The concept of core users represents a major trend of location preferences and a generalization of blurred privacy of the entire set of clustered users. This is a novel way of reconciling the tension between the degree of anonymity and the quality of recommendation service.
3) An anonymity coordination strategy is proposed to detect the failure of an anonymizer agent that could have been hacked or could be providing bad recommendations. This strategy picks an alternative competent agent with a high rank in trust-computing values to anonymize the users' location information while providing accurate recommendations.

We introduce the related work in Section II. An overview of the architecture of the BUSA scheme is provided in Section III, including k-anonymity and the descriptions of the clustering method and the anonymizer coordination strategy. An overview of the BUSA scheme of security and service follows in Section IV, and a detailed experiment to compare the BUSA scheme to its counterparts is proposed in Section V, which also contains the evaluation. Finally, we conclude this paper and outline directions for future research in Section VI.

II. RELATED STUDIES

In this section, the current state of the security-service tension is reviewed from the following perspectives: (A) trajectory privacy protection techniques; and (B) location-based recommendations. Gaps are identified in these studies, and we subsequently discuss how to reconcile the tension.

A. TRAJECTORY PRIVACY PROTECTION TECHNIQUES

An untrustworthy LSP may disclose trajectory data to a third party without authorization and leak private data. Many studies have sought to protect trajectory data in an edge network [13], a fog environment [14], a cloud server [15], etc., and a commonly applied method is to blur the location by k-anonymity applied to trajectory data to prevent an adversary from knowing the exact data. Zhang *et al.* designed auction-based mechanisms to achieve a high satisfaction ratio of privacy protection attained with k-anonymity [16]. Niu *et al.* proposed a dummy location selection algorithm to achieve k-anonymity for users of an LBS by adopting an entropy metric to spread the dummies and enlarge the cloaking region [17]. To protect users' personal data from unauthorized tracking by untrusted or malicious LBS servers, Sun *et al.* [18] analyzed a dummy location selection (DLS) algorithm and designed an attack algorithm for testing the emerging IoT security in terms of attaining a lower probability of revealing the user's location and improving computational cost and efficiency. In [19], Tong *et al.* studied the privacy protection of users' location information in the scheduling of ridesharing services and proposed a scheduling protocol to achieve joint differential privacy with reasonable efficiency. Wernke *et al.* systematically assessed the applicability and effectiveness of location privacy approaches and summarized existing approaches by different protection goals and the ability to resist attacks [20]. Zhao *et al.* proposed an ILLIA scheme with a credibility-based k-anonymity to defend against location injection attacks without requiring advance knowledge of how fake locations were manipulated [21]. In summary, researchers have applied a variety of improved scholarly techniques to protect users' data in location-based services.

III. LOCATION-BASED RECOMMENDATION

In contrast, overprotection will lead to a low quality of location services. Specifically, location-based recommendations have to integrate users' trajectory data to provide suggestions of closest restaurants [22], best routing [23], etc. With increasing use of Web services in the IoT, the subject of helping users select services that match users' needs by analyzing their personal data has become popular among researchers. Zheng *et al.* proposed a user collaboration mechanism for past Web service data collection from different service users, and designed a collaborative filtering approach to predict valuable Web services [24]. In [25], Wang *et al.* analyzed the differential problem of client location changes, and proposed an integrated Quality-of-Service (QoS) prediction approach that unified the modeling of multidimensional QoS data via multilinear algebra-based concept of a tensor and enabled efficient Web service recommendation for mobile clients via tensor decomposition and reconstruction optimization algorithms. Zhou *et al.* introduced an approach to identifying and recommending scientific workflows for reuse and repurposing with a layer hierarchy, which adopted a graph skeleton-based clustering technique to cluster the core workflows [26]. In [27], the researcher also retrieved and recommended subchains of possible service invocations by leveraging the semantic similarity between the name and textual description of parameters, where a network model was constructed to represent possible invocations between operations; the results proved that the researcher's approach could solve the problem of a geospatial Web service. An automated filtering mechanism capable of categorizing members within a group based on their response patterns was proposed in [28] by Thampi; the mechanism clustered user posts into groups based on stylistic, thematic, and emotional aspects. Luo *et al.* proposed a dynamic prediction approach to selecting an optional service in cloud computing, combining fuzzy neural networks and adaptive dynamic programming [29].

The above studies were informative and achieved the goals of protecting the users' privacy or providing recommendations with proper solutions and suitable methodologies. To the best of our knowledge, studies that reconcile the tension between privacy protection and privacy requesting remain rare. Our research aims to better define privacy management for integrating a proper degree of personal information [30], [31], and intends to apply clustering, anonymity, and security analysis to provide recommendations based on a medium volume of disclosure.

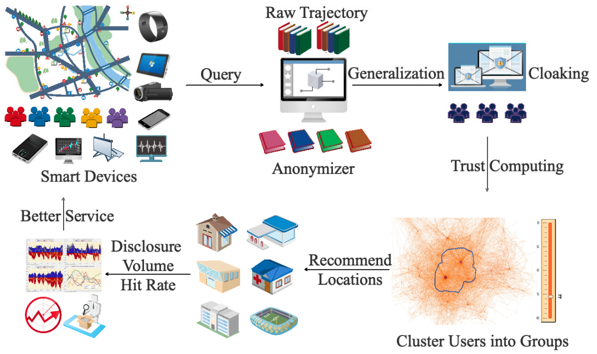

FIGURE 1. Workflow of our proposed BUSA scheme.

IV. MODEL IMPLEMENTATION

As shown in Fig. 1, the proposed BUSA scheme first sends users' queries to anonymizers for dividing them into information segments; afterwards, the raw trajectory data will be generalized by k-anonymity with the cloaking technique. Additionally, trust computing can help cluster users into groups, and the core of each cluster is utilized to generate location recommendations. We use the adversary test and the hit rate to demonstrate that the BUSA scheme can provide better recommendations and secure users' trajectory data while maximizing user satisfaction and minimizing the privacy concerns. This section introduces the preliminaries of the basic concepts involved in the BUSA scheme, such as a service level agreement (SLA) and a privacy level agreement (PLA). The concepts of SLA and PLA were proposed by Zhu *et al.* [32] and inspired us to extend the researchers' work on the foundation of location-based services. In what follows, we describe the preliminaries.

A. PRELIMINARIES OF SLA AND PLA

An SLA is a negotiated agreement reached by at least two parties, usually between a service-requesting user (SRU) and an LSP. It describes the details of a service contract, where the service includes the recommended items, e.g., the best route home, or a nearby restaurant between two or more parties, in which one is the customer and the others are service providers. In short, it is a part of the service contract, where a service is formally defined. In trajectory data, an SLA specifies the estimated time of reaching the destination, the reasons for this recommendation, the modality of reaching the destination, the levels of availability, and the validity duration. The content of the SLA directly influences how likely a user will be to accept the recommendation from the provider.

In location-based recommendations, a PLA is an agreement to describe the level of privacy protection, and a higher level indicates a greater difference between the original information and the anonymized information. The anonymity includes the information generalization on the location server. A too high level of PLA will hide too much personal information, causing the returned recommendation to be too far away from the users' personal interests. Too low a level of PLA cannot generalize the sensitive information against an adversarial agent. Investigating the proper level of PLA, i.e., the proper level of anonymization, is a major objective.

B. PRELIMINARIES OF TRUST COMPUTING IN TRAJECTORIES

A trust computing value (TCV) represents a measure of trustworthiness of each agent assigned by users and varying depending on the specific context, e.g., an SRU will send another location query with more disclosed private data to an anonymizer agent (AA) if it receives a satisfactory location recommendation. A TCV also stands for the strength of reliance of the information requestor as to how much private data he or she can disclose, or a global perception of an entity's trustworthiness determined through the information exchange process. The value of TCV is directly proportional to the benefit provided by a specific entity and equals the weight value between an SRU and an AA. We consider the trust computing in the users' trajectory data, which means considering the factors of both location and time. The value of TCV is updated over time because an SRU could gain TCV when he or she receives a greater benefit at time t_x or lose TCV due to an improper request for private data at time t_y.

We assume that the role of the trust computing auditor is assigned to the agent invitation unit (AIU). An AIU consists of a datacenter that stores the copied queries and anonymized information. In location-based recommendations, an AIU is capable of the following duties: 1) recording the signed SRUs, LSPs, and AAs in a location network; 2) dividing the query information into information segments and sending each segment to a proper AA for anonymization; 3) receiving the general feedback of SRUs, providing timely updates in the determination of whether an AA provides competent anonymization and service, and updating the value of computed trust from various types of SRU feedback, e.g., clicks on the recommended items, positive comments, etc.; and 4) clustering the AAs according to their competence and usage, real-time monitoring of their performance, and replacing a bad AA with an alternative AA. The value of trust $P_{ij}(k)$ at time t_k on aspect m is defined as follows:

$$P_{ij}(k) = \left\{P_{ij}^1(k), P_{ij}^2(k), \ldots, P_{ij}^m(k) \,|t_k\right\} \tag{1}$$

where i is the index of an SRU, j is the index of an AA, and k is the number of instances of interaction between SRU$_i$ and AA$_j$. The value of $P_{ij}(k)$ will be updated using the time decay parameter $\theta(t_k - t_l)$ that records an SRU's reliance trend from time t_k to t_l, and t_l is a previous record. Every t_k represents an interaction that has occurred. After s instances of interactions, the value of trust $P_{ij}(k)$ should be determined as

$$P_{ij}(k) = \frac{\sum_{k\le s} P_{ij}^m(k) \times \theta(t_k - t_l)}{\sum_{k\le s} \theta(t_k - t_l)} \tag{2}$$

where $P_{ij}^m(k)$ represents the value after k instances of interactions on the reliance level of SRU$_i$ on AA$_j$.

C. PRELIMINARIES OF ANONYMIZER COORDINATION STRATEGY

In this paper, we consider the anonymizer coordination strategy in two cases:

1) A positive updating coordination (PUC) stands for an SRU's awareness of experiencing more potential benefits than risks. It could result from a good location recommendation or long-term usage without experiencing security threats. At this time, the updates of the value of trust $P_{ij}(k)$ are fixed with a coordination parameter $\propto$ ($\propto < 1$):

$$P_{ij}(k) = \frac{\sum_{k\le s} P_{ij}^m(k) \times \theta(t_k - t_l)}{\sum_{k\le s} \theta(t_k - t_l)} + (1-\propto) \times \frac{\sum_{k\le s} P_{ij}^m(r) \times \theta(t_r - t_k)}{\sum_{k\le s} \theta(t_r - t_k)} \times r_{ik} \times s_{ik} + b_{im} \tag{3}$$

where r_{ik} represents that SRU$_i$ is satisfied with the location recommendation provided by AA$_j$ at time t_i, s_{ik} represents that SRU$_i$ does not experience a privacy threat during the time interval from t_k to t_r ($r > k$), and b_{im} is the item bias from SRU$_i$ on query m.

2) A negative updating coordination (NUC) stands for an SRU's awareness of experiencing more potential risks than benefits. It could result from a bad location recommendation or the perception of a security threat. At this time, the updates of the value of trust $N_{ij}(k)$ are fixed with a coordination parameter β ($\beta < 1$):

$$N_{ij}(k) = \frac{\sum_{k\le s} N_{ij}^m(k) \times \theta(t_k - t_l)}{\sum_{k\le s} \theta(t_k - t_l)} + (1-\beta) \times \frac{\sum_{k\le s} N_{ij}^m(r) \times \theta(t_r - t_k)}{\sum_{k\le s} \theta(t_r - t_k)} \times r_{ik} \times s_{ik} + b_{im} \tag{4}$$

where r_{ik} represents that SRU$_i$ is dissatisfied with the location recommendation provided by AA$_j$ at time t_i, s_{ik} represents that SRU$_i$ perceives a privacy threat at time t_k ($r > k$), and b_{im} is the item bias from SRU$_i$ on query m. We rank the combined value of PUC and NUC in the descending order for all anonymization agents and place the high-ranking $R_{ij}(k)$ of a candidate anonymizer as a replacement once a failure has been detected:

$$R_{ij}(k) = P_{ij}(k) - N_{ij}(k) \times (1 - \mu) \ll S \tag{5}$$

where S is the threshold of the prescribed minimum of the trust-computing value.

D. PRELIMINARIES OF SRU CLUSTERING

Extending our previous work [33], we cluster all SRUs into several groups according to similar attributes. In this paper, in the scenario of location recommendation services, SRUs are clustered into one group if they are currently located closest to each other among other clustering results, and the recommended items will be more likely to satisfy the SRU in this group. An SRU's mobility generally exhibits certain regularity, such as following a fixed route from home to work, and the recommended item, e.g., the closest restaurant, can always be a good recommendation of a lunch location because such an item's location is also fixed. As a result, an SRU's trajectory J can be represented as a set of discrete 2-tuples of locations $L\{l_1, l_2, \cdots, l_n\}$ at points of time $T\{t_1, t_2, \cdots, t_n\}$:

$$J = \{(l_1, t_1), (l_2, t_2), \ldots, (l_n, t_n) \,|m\} \tag{6}$$

and each tuple $\left(l_x, t_y\right)$ represents this SRU's time-specific location (x, y), where x and y represent its geographic latitude and longitude, respectively. The value of m indicates a specific query to the LSP, e.g., a route home, a restaurant route, etc., and locations $L\{l_1, l_2, \cdots, l_n\}$ are the set of points passed to reach m. Suppose that the whole group of SRUs is a set of m entities, denoted by $(J_1, J_2, \cdots, J_m)$. Each entity J_x is expressed as J_x $(J_{x1}, J_{x2}, J_{x3}, \cdots, J_{xn})$, where J_{xj} $(j = 1, 2.., n)$ stands for a historical record consisting of the number of times SRU x has followed trajectory j. The center of each cluster represents a core user. An SRU is included in a cluster if its trajectory has properties that are most similar to those of the cluster's core. The SRUs may be similar to each other, and the similarities are represented by m-1 edges; hence, the total number of edges is $E = m \times (m-1)/2$. The distance $D = \sum_{u=1}^{p} \sum_{v \neq u} (E_u, E_v)/2$ measures the differences between the SRUs, and corresponds to how unlikely they are to choose the same routes, where p is the total number of SRUs, and routine (E_u, E_v) stands for the rating differences between two SRUs, e.g., E_u and E_v:

$$routine\,(E_u, E_v) = \sqrt{\sum\nolimits_{i=1}^{n} \left(\hat{r}_{u,i,m} - \hat{r}_{v,i,m}\right)^2} \tag{7}$$

where m is the number of queries. An SRU E_x is considered abnormal if routine $(E_x, E_y) > D/T$, which indicates that its values are too different from those of the other SRUs.

We denote the number of clusters by c. For x SRUs, $c = 1$ if all users are in the same cluster, and $c = x$ if each user belongs to one individual cluster. Hence, the value of c varies from 1 to x. Another parameter, namely, the minimum distance variance in n folds denoted by MDV_n, is introduced. The users' datasets are randomly divided into 10 folds, numbered from 1 to 10. For each turn, one fold is regarded as a testing set, while the remaining nine folds are used as a training set, and n is the fold index:

$$MDV_n = \sum\nolimits_{j=1}^{c} \sum\nolimits_{y \epsilon C_j} routine\,(y, U) \tag{8}$$

where k is the total number of clusters, y is the user set in cluster C_j, and m is the core user. Each time the value of c changes, MDV_n is computed accordingly. The minimum value of MDV_n is recorded, and the corresponding value of c, denoted by c_n, is regarded as the number of clusters. The corresponding value of c_n is the final value. The clustering algorithm is performed accordingly with the final value of c_n.

E. PRELIMINARIES OF K-ANONYMITY

One major issue in anonymity is avoiding too much generalization of private data, thus causing too little information being usable for recommendations. In location-based anonymization, the process of anonymization consists of two parts:

1) Dividing the SRU's query information into several information segments, and distributing them to different AAs. Given this manipulation, no AA can obtain the entire query. The dividing process separates the identifiable attributes and recommendation-relevant attributes. An identifiable attribute, e.g., name, SSN, home location, driver's license number, phone number, personal webpage, etc., can be used for inferring a specific individual, and recommendation-relevant attributes, e.g., current location, preferred restaurant, preferred transportation, mobile information, office hours, etc., are used to generate recommendations. The recommendation-relevant attributes belong to query m from SRU$_x$, represented by $R_{xm} = \{r_1, r_2, \cdots, r_n|m\}$. An information segment is denoted by $\bar{R_l}$, where $\bigcup_{j=1}^{n} \bar{R_l} = R_{xm}$, and $\bar{R_l} \neq \emptyset$. An information segment $\bar{R_l}$ will be sent to an AA that has received the most queries in category m for anonymization. Information is segmented according to the following method.

Algorithm 1 Query-Dividing Algorithm

```
Input: SRU$_x$'s query M and its attributes
Output: A set of information segments $\bar{R_l}$ for anonymization
1: Remove the identifiable attributes
   $\bar{R_{xm}}$, where $\bar{R_{xm}} \bigcup R_{xm} = $ M
2: Divide the recommendation-purpose attributes
   $R_{xm}\{r_1, r_2, \ldots, r_n\}$
3: Calculate the AA receiving the most queries in category M.
4: Match AAs and $\bar{R_l}$
5: return pairs {AA, $\bar{R_l}$}
```

2) Each AA forms a cloaking region and selects k cells to perform spatial k-anonymity. An example of a cloaking region in the BUSA scheme is shown in Fig. 2. A cluster of SRUs is plotted in one cloaking region, which is a grid structure regarded as the map. The cluster core corresponds to the highest query probability and can be regarded as the

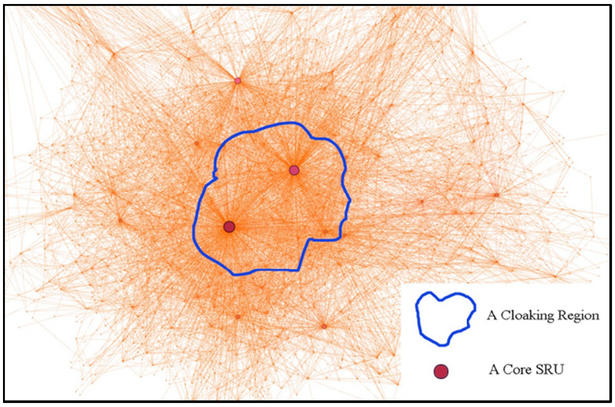

FIGURE 2. Spatial cell selection in the BUSA scheme for forming a cloaking region.

Algorithm 2 Cloaking Region for k-Anonymity

```
Input: A set of information segment pairs $\{AA, \bar{R}_l\}$
Output: A set of anonymized information segments $\tilde{R}_l$ and a cloaking region
for each AA, plot the SRU points over an n × n grid structure according to $\bar{R}_l$
    In category M, plot the POIs over the same n × n grid structure
    Calculate the cluster core according to $\bar{R}_l$
    Coordinate the POIs, $\bar{R}_l$, and the core location on the grid under min MDVn
    Construct the weight graph G (V, E, routines)
    Select N cells around the core with the lowest differences within min $(|r_i| \times |r_m|)$ i, m $\in (1, l)$
    region = N cells, where core attributes are $\left\{\check{r}_1, \check{r}_2, \ldots, \check{r}_c, \ldots, \check{r}_n\right\}$ $c \in (1, n)$
    for each Rs point $\in \bar{R}_l$ do
        region + = Rs,
        $\bar{R}_l \{r_1, r_2, \ldots, r_c, .., r_n\} \underset{\text{anonymized}}{\longrightarrow} \tilde{R}_l \left\{r_1, r_2, \ldots, \check{r}_c, .., r_n\right\}$
    end for
return region, $\tilde{R}_l$
```

hotspot closest to the POIs. Although a very small number of SRUs that are located far away from the core may not be satisfied with the POIs, we still guarantee that the overall satisfaction of most SRUs in this cluster with the POI is high. In this paper, an SRU's acceptance of the recommendation is regarded as "satisfaction" if the user, e.g., clicks on the suggested term, spends a long time reading the suggestion, etc. As a result, the overall satisfaction Rs can be computed as

$$R_s = Average \sum\nolimits_{i=1}^{k} A_{rate} \times routine\,(y, U) + b_c \tag{9}$$

where A_{rate} is the percentage of clicked recommendations multiplied by the total number of recommendations. Furthermore, we also specify a threshold h for identifying a too low performance of an AA. Once $h \gg R_s$, e.g., $h = 10\ R_s$, the current AA is replaced with a backup AA in case its data are intercepted by an adversary. During the process of replacement, the SRU stops delivering the queries to the bad AA and instead recalculates the nearest core of the cluster, and picks an alternative AA that is closest to the core using Equation (7). Algorithm 2 demonstrates the process of k-anonymity and the adjustment of the cloaking region; the whole process is intended to coordinate the core's overall fit to most of the SRUs' POIs. For each $\bar{R}_l$, the anonymization process replaces a random attribute r_c with a corresponding attribute $\check{r}_c$ of the core, and the anonymized information is regarded as $\tilde{R}_l$.

1) As demonstrated in **Algorithm 2**, the core of this cluster is regarded as the cloak.

F. OVERVIEW OF THE SCHEME'S SECURITY AND SERVICE

In this section, we briefly analyze the security and service performance of our proposed scheme and determine if it can reconcile the tension between privacy and recommendation quality. The security of the proposed scheme is analyzed to assess if it can protect against an adversary in referring a specific SRU. When an SRU makes a location query request, its query will have the identifiable information removed. Then, the rest of the query information will be divided into information segments, and sent to an anonymizer for generalization. As each anonymizer can only receive a piece of information, inferring the original SRU becomes very difficult. The only information an adversary can obtain is that of a core, which is a generalization of a cluster of SRUs. The results sent from the anonymizer to the user are symmetrically encrypted with a secret known only to the user himself, so an adversary cannot decrypt the results.

On the other hand, the robustness of the proposed scheme against fraudulent SRUs is also enhanced. Some restaurants may deliberately hire SRUs to pick the POIs closer to its location, and in doing so will increase the chances of recommending the restaurant. This fraudulent behavior will not succeed in switching the POI in our scheme. Suppose that n fraudulent SRUs have deliberately clicked on the recommendation of lunch in position d, which is far away from the core and the POIs. According to Equations (6) – (8), if the value of n is very high, the n fraudulent SRUs will be clustered by themselves, and the recommended lunch place may indeed be calculated as their intended restaurant, but the recommendation receivers can only be themselves; if the value of n is less than the number of regular SRUs, their fraudulent clicks will not be strong enough to affect the calculated results. The AAs are assigned with typical kinds of queries, so the recommendations are generated with high accuracy.

V. EXPERIMENTAL RESULTS

In this section, we perform the experiment and evaluate the efficiency and effectiveness of the proposed scheme. The performance is appraised based on the recommendation click rate and the dummy adversary attack.

A. SIMULATION SETUP

We hired more than 860 local mobile SRUs, who were employed long-term residents, and could send queries by uploading their personal information. The trajectory data were collected for one week, and we created their integrated daily routes by observing locations they stayed at every 30 minutes. More than 2000 POIs were generated, including cinemas, restaurants, parks, plazas, etc. The recommendations were calculated and ranked according to the closest location or the highest number of clicks. The experiment was performed on a PC with 16 GB of RAM and an Intel(R) Core(TM) i7-7500U CPU operating at 2.70-2.90 GHz; the proposed scheme was implemented with MyEclipse and the Java Development Kit. Performance was measured by the SRUs' click behaviors, disclosure volume, and privacy awareness.

B. RECOMMENDATION PERFORMANCE

A recommendation r is generated according to an SRU's query information, and an added justification will increase the possibility of accepting the recommendation. The samples of generated recommendations include a target location and justifications S $\{s_1, s_2, \cdots, s_n\}$, varying according to results:

1) XXX hotel, low price/not far away/good window view/quiet location/free breakfast…
2) XXX restaurant, suits your favorite taste/offers a coupon/visit it soon/few people waiting in line…
3) XXX cinema, upcoming movie/festival deal/famous actors/your friends have all watched/…
4) XXX plaza, reduced price/fresh vegetables/…

The justifications originate from other SRUs' comments made after visiting. We regard the following two cases as an SRU accepting a recommendation, called a recommendation hit:

1) The user has clicked the suggested recommendation or has looked at the recommendation for more than one minute.
2) One of the user's next observed locations matches the recommended place.

We selected the CSKA [9] scheme and the CaDSA [34] scheme as the baseline algorithm. The CKSA scheme adopts spatial k-anonymity to better protect location privacy. The CaDSA scheme is a cache-based solution for protecting privacy that also utilizes the entropy metric and maximizes the dummy users' contribution. In [9], the authors proposed the CKSA scheme and used multilevel caching to reduce the risk of exposure of user information to untrusted entities. One advantage of CKSA is the application of the Markov model to predict the users' next query location as they move, which could enhance location privacy and increase the cache hits. In [34], the authors proposed CaDSA to investigate the proper amount of caching that could be used to improve privacy protection and designed two novel caching-aware dummy selection algorithms for privacy enhancement by maximizing both the privacy of the current query and the dummies' contribution. Both schemes are interesting approaches that inspired us to go further by considering the aspect of recommendation. However, when taking collaboration into account, an emergent tension between recommendation accuracy and privacy protection has to be reconciled.

C. RESULTS

The effectiveness of the proposed scheme was analyzed by varying the cluster count c of SRUs and the anonymity degree k, and considering the performance of the recommendation hit rate, and the SRUs' disclosure volume. If $c = 58$ and $k = 30$, the SRUs are sending the most queries in the range of several individual radiuses. The experiment was performed for seven days; the recommendations were generated at the start of the second day, and 97 SRUs with first-day routes significantly different from their second-day routes were removed. The rest of the SRUs sent queries with at least their current location to obtain the recommendation; in return, the recommendations were generated from one of the AAs for 2 hours (at 11:00 a.m., 13:00 p.m., 15:00 p.m., and 17:00 p.m.) from 10:00 a.m. to 18:00 p.m., and each gap between observed locations was at least 100 meters. During the experiment, 763 users were clustered into 58 groups, and 1309 recommendations of where to go were provided. Additionally, the total set of disclosures included 3082 private items; 730 trust relations were constructed, and 13 AAs were replaced when the hit rate was too low.

As shown in Table 1, the replacement of AAs only occurred on the second and third days and did not occur on the following four days. On the first two days, the SRUs mostly disclosed only their basic locations, as we requested (averaging 1.20 disclosures for each SRU); however, for the remainder of the four days, their tendency to disclose information increased (averaging 11.57 disclosures for each SRU). We received more disclosures as the input of the recommendation provider, so the hit rate values on the six days were 17.27%, 20.61%, 24.19%, 32.81%, 38.02%, and 42.17%, due to growth of available information.

D. COMPARISON

We compared our proposed scheme with the CSKA and CaDSA in terms of the hit rate and recommendation accuracy to illustrate the performance of our scheme. The performance results initially stabilized on the 6th day; hence, we applied the respective methods together with BUSA on the 7th day. In CSKA and CaDSA, we considered their cache hit rates to be equivalent to our recommendation hit rate. As shown in Fig. 3(A), the hit rates of the three methods generally decreased with the increased value of anonymity degree k. For most values of k, the hit rate for BUSA was higher than those for CSKA and CaDSA; however, for $k = 30$, BUSA was still outperformed CaDSA, while CSKA was the best performer. On the other hand, as shown in Fig. 3(B), the willingness of self-disclosure of the SRUs was the highest for BUSA, while the value for CSKA was higher than that of CaDSA. A higher value of k indicates a greater degree of anonymity

TABLE 1. Hit rate, disclosure volume, and anonymizer backups during seven days.

Time	Disclosure	Hit Rate (%)	AA Backups++
2.11	2	12.76	2
2.13	1	11.09	3
2.15	3	13.58	1
2.17	4	14.2	3
3.11	3	14.57	2
3.13	2	13.07	0
3.15	4	15.69	1
3.17	7	17.81	1
4.11	4	19.24	0
4.13	7	18.17	0
4.15	8	20.05	0
4.17	9	23.98	0
5.11	11	26.69	0
5.13	8	27.31	0
5.15	10	25.43	0
5.17	12	29.53	0
6.11	10	30.18	0
6.13	10	29.34	0
6.15	12	32.67	0
6.17	14	37.24	0
7.11	10	40.13	0
7.13	11	35.67	0
7.15	13	38.61	0
7.17	16	46.28	0

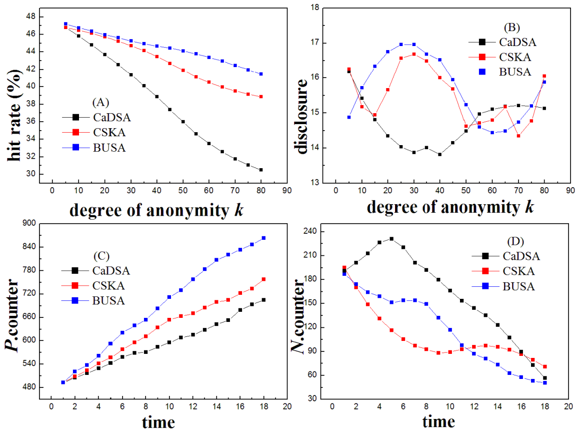

FIGURE 3. Comparison of the hit rate (A) and disclosure volume (B) among the three schemes for various values of the degree of anonymity *k*. SRUs' evaluation of the potential benefit and risk at runtime are shown in (C) and (D).

and knowing a lesser amount of personal information. In general, BUSA achieved the highest average hit rate compared to those of CSKA and CaDSA (35.01% > 31.79% > 26.36%) and the highest average disclosure volume each day (8.11 > 6.73 > 3.89) from the SRUs. We believe that the baseline scheme anonymized the SRUs' data more than we did. To investigate what led to these results, on the seventh day, we required the SRUs to give feedback in comments when they were using one of the three schemes (BUSA.N = 254, CaDSA.N = 254, CSKA.N = 255). The comments could be positive, e.g., "I'd like to give more support to the service, so I want to disclose more information" or negative, e.g., "The queries contain too much private data". Positive comments were selected for forwarding to other SRUs as justifications. We used the natural language processing techniques to determine if each comment was positive or negative. Two counters P and N were initialized to 0 to record the total counts of positive and negative comments on each recommended location. Whenever a positive (negative) word was detected, counter P.counter (N.counter) was incremented by 1, indicating that there were more benefits (risks) perceived by an SRU than concerns with the service-query exchange. When an SRU disclosed a private item, the respective P.counter was also incremented by 1; e.g., the comment "This restaurant sells 'spicy food' that really suites my taste and is so 'close to my house'" results in P.counter being incremented by 2.

In Fig. 3(C), the SRUs' awareness of the potential benefits rose, and the BUSA method ranked the highest in average benefits, while CSKA ranked higher than CaDSA (5.39 > 3.19 > 2.07). We believe that BUSA collected more private information with a proper degree of anonymity to provide good recommendations and successfully eased the effect of the data sparsity problem with a low rate of negative comments. Positive comments indicate that the SRUs are satisfied with the recommendation, as well as privacy protection, and can disclose more information to obtain even better recommendations, which becomes a positive feedback loop. We regarded the negative comments as demonstrating the SRUs' awareness of potential risk. As shown in Fig. 3(D), the SRUs' risk awareness generally decreased in three conditions, which supported the idea that the schemes provided good suggestions of what items to view and helped the SRUs gain trust based on the recommendations' benefits. Over an entire day, the average risks the BUSA method achieved were lower than those of CSKA and CaDSA (3.85 < 4.03 < 6.72). In the beginning, BUSA collected the SRUs' data so that its N. counter was higher than that of CSKA, but later the risks declined. Although the cloaking region generalized the SRUs' data, the recommendations could still be generated according to each cluster's core, so the recommendations still fit most of the SRUs' interests and thus resulted in a greater reduction of risk awareness in BUSA. As a result, we believe that our proposed scheme has achieved the research goal of reconciling the tension between privacy and personalization.

VI. CONCLUSIONS AND DIRECTIONS FOR FUTURE RESEARCH

This paper developed a scheme called BUSA to reconcile the tension between privacy protection and requesting privacy in location-based recommendations. This scheme used an anonymizing agent between service-requesting users and location service providers to perform dividing query information and k-anonymity to enhance privacy protection. It first utilized clustering techniques to group users into clusters by learning their trajectory data and then selected k cells as a cluster core using trust computing, regarded as the benchmark for calculating recommendations. The proposed anonymizer agent updates and coordination can help improve the service and anonymization performance. The BUSA method adopts the proposed k-anonymity to protect privacy, and the calculated recommendations will fit the core, which represents the location preference of all users. The security analysis revealed that the BUSA scheme could effectively protect privacy and resist fraudulent query requestors, and the simulation results also proved that it exhibited stronger privacy protection than those of its counterparts from the perspective of recommendation hit rate and disclosure volume.

While applying clustering techniques, we only utilized users' trust computing to form the whole cluster's major trend as to location preference while ignoring the potential value obtainable from friends. Another aspect we did not consider was the users' interests at different times, e.g., preferring work-related information during office hours, while preferring entertainment-related suggestions outside of work; furthermore, we did not consider the privacy details from the perspective of life-long problems, e.g., decoy routes, and energy consumption. In future studies, we would further investigate various topics using multilayers. Additionally, when a bad news event occurs in the entire scheme, relieving users' concerns and enacting an emergency recovery plan is also an essential issue that merits investigation.

REFERENCES

[1] Z. Sheng, C. Mahapatra, C. Zhu, and V. C. M. Leung, "Recent advances in industrial wireless sensor networks toward efficient management in IoT," *IEEE Access*, vol. 3, pp. 622–637, Jun. 2015.

[2] J. Wen, Z. Zhou, Z. Shi, J. Wang, Y. Duan, and Y. Zhang, "Crossing scientific workflow fragments discovery through activity abstraction in smart campus," *IEEE Access*, vol. 6, pp. 40530–40546, 2018.

[3] Z. Sheng, H. Wang, C. Yin, X. Hu, S. Yang, and V. C. M. Leung, "Lightweight management of resource-constrained sensor devices in Internet of Things," *IEEE Internet Things J.*, vol. 2, no. 5, pp. 402–411, Oct. 2015.

[4] Z. Sheng, S. Yang, Y. Yu, A. Vasilakos, J. McCann, and K. Leung, "A survey on the IETF protocol suite for the Internet of Things: Standards, challenges, and opportunities," *IEEE Wireless Commun.*, vol. 20, no. 6, pp. 91–98, Dec. 2013.

[5] T. Peng, Q. Liu, D. Meng, and G. Wang, "Collaborative trajectory privacy preserving scheme in location-based services," *Inf. Sci.*, vol. 387, pp. 165–179, May 2017.

[6] S. Zhang, K.-K. R. Choo, Q. Liu, and G. Wang, "Enhancing privacy through uniform grid and caching in location-based services," *Future Gener. Comput. Syst.*, vol. 86, pp. 881–892, Sep. 2018.

[7] J. Zhou, Z. Cao, X. Dong, N. Xiong, and A. V. Vasilakos, "4S: A secure and privacy-preserving key management scheme for cloud-assisted wireless body area network in m-healthcare social networks," *Inf. Sci.*, vol. 314, pp. 255–276, Sep. 2015.

[8] X. Yang, R. Lu, J. Shao, X. Tang, and A. Ghorbani, "Achieving efficient and privacy-preserving multi-domain big data deduplication in cloud," *IEEE Trans. Services Comput.*, to be published.

[9] S. Zhang, X. Li, Z. Tan, T. Peng, and G. Wang, "A caching and spatial K-anonymity driven privacy enhancement scheme in continuous location-based services," *Future Gener. Comput. Syst.*, vol. 94, pp. 40–50, May 2019.

[10] B. Gedik and L. Liu, "Protecting location privacy with personalized k-anonymity: Architecture and algorithms," *IEEE Trans. Mobile Comput.*, vol. 7, no. 1, pp. 1–18, Jan. 2008.

[11] G. Ghinita, K. Zhao, D. Papadias, and P. Kalnis, "A reciprocal framework for spatial k-anonymity," *Inf. Syst.*, vol. 35, no. 3, pp. 299–314, 2010.

[12] Z. Chi, Y. Wang, Y. Huang, and X. Tong, "The novel location privacy-preserving CKD for mobile crowdsourcing systems," *IEEE Access*, vol. 6, pp. 5678–5687, 2017.

[13] L. Qi, X. Zhang, W. Dou, C. Hu, C. Yang, and J. Chen, "A two-stage locality-sensitive hashing based approach for privacy-preserving mobile service recommendation in cross-platform edge environment," *Future Gener. Comput. Syst.*, vol. 88, pp. 636–643, Nov. 2018.

[14] T. Wang *et al.*, "Fog-based storage technology to fight with cyber threat," *Future Gener. Comput. Syst.*, vol. 83, pp. 208–218, Jun. 2018.

[15] D. Liao, G. Sun, H. Li, H. Yu, and V. Chang, "The framework and algorithm for preserving user trajectory while using location-based services in IoT-cloud systems," *Cluster Comput.*, vol. 20, no. 3, pp. 2283–2297, 2017.

[16] Y. Zhang, W. Tong, and S. Zhong, "On designing satisfaction-ratio-aware truthful incentive mechanisms for k-anonymity location privacy," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 11, pp. 2528–2541, Nov. 2016.

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

[18] G. Sun *et al.*, "Efficient location privacy algorithm for Internet of Things (IoT) services and applications," *J. Netw. Comput. Appl.*, vol. 89, pp. 3–13, Jul. 2017.

[19] W. Tong, J. Hua, and S. Zhong, "A jointly differentially private scheduling protocol for ridesharing services," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 10, pp. 2444–2456, Oct. 2017.

[20] M. Wernke, P. Skvortsov, F. Dürr, and K. Rothermel, "A classification of location privacy attacks and approaches," *Pers. Ubiquitous Comput.*, vol. 18, no. 1, pp. 163–175, 2014.

[21] P. Zhao, J. Li, F. Zeng, F. Xiao, C. Wang, and H. Jiang, "ILLIA: Enabling k-anonymity-based privacy preserving against location injection attacks in continuous LBS queries," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 1033–1042, Apr. 2018.

[22] H. Gao, J. Tang, X. Hu, and H. Liu, "Content-aware point of interest recommendation on location-based social networks," in *Proc. AAAI*, 2015, pp. 1721–1727.

[23] C.-Y. Tsai and B.-H. Lai, "A location-item-time sequential pattern mining algorithm for route recommendation," *Knowl.-Based Syst.*, vol. 73, pp. 97–110, Jan. 2015.

[24] Z. Zheng, H. Ma, M. R. Lyu, and I. King, "QoS-aware Web service recommendation by collaborative filtering," *IEEE Trans. Services Comput.*, vol. 4, no. 2, pp. 140–152, Apr./Jun. 2011.

[25] S. Wang, Y. Ma, B. Cheng, F. Yang, and R. Chang, "Multi-dimensional QoS prediction for service recommendations," *IEEE Trans. Services Comput.*, vol. 12, no. 1, pp. 47–57, Jan./Feb. 2019.

[26] Z. Zhou, Z. Cheng, L. Zhang, W. Gaaloul, and K. Ning, "Scientific workflow clustering and recommendation leveraging layer hierarchical analysis," *IEEE Trans. Services Comput.*, vol. 11, no. 1, pp. 169–183, Jan. 2018.

[27] X. Wang, Z. Cheng, Z. Zhou, K. Ning, and L. J. Zhang, "Geospatial Web service sub-chain ranking and recommendation," in *Proc. IEEE Int. Conf. Services Comput.*, Jun. 2014, pp. 91–98.

[28] A. U. and S. M. Thampi, "Linguistic feature based filtering mechanism for recommending posts in a social networking group," *IEEE Access*, vol. 6, pp. 4470–4484, 2018.

[29] X. Luo, Y. Lv, R. Li, and Y. Chen, "Web service QoS prediction based on adaptive dynamic programming using fuzzy neural networks for cloud services," *IEEE Access*, vol. 3, pp. 2260–2269, 2015.

[30] H. Wu, B. P. Knijnenburg, and A. Kobsa, "Improving the prediction of users' disclosure behavior by making them disclose more predictably?" in *Proc. Symp. Usable Privacy Secur. (SOUPS)*, 2014, pp. 1–7.

[31] H. Wu, H. Zhang, L. Cui, and X. Wang, "CEPTM: A cross-edge model for diverse personalization service and topic migration in MEC," *Wireless Commun. Mobile Comput.*, vol. 2018, Aug. 2018, Art. no. 8056195.

[32] C. Zhu, V. C. M. Leung, L. Shu, and E. Ngai, "Green Internet of Things for smart world," *IEEE Access*, vol. 3, pp. 2151–2162, 2015.

[33] H. Wu, X. Wang, Z. Peng, and Q. Li, "Div-clustering: Exploring active users for social collaborative recommendation," *J. Netw. Comput. Appl.*, vol. 36, no. 6, pp. 1640–1642, 2013.

[34] B. Niu, Q. Li, X. Zhu, G. Cao, and H. Li, "Enhancing privacy through caching in location-based services," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Apr. 2015, pp. 1017–1025.

HONGCHEN WU received the Ph.D. degree in computer science and technology from Shandong University, in 2016. He studied at the University of California at Irvine, U.S.A. as a joint Ph.D. student supported by the CSC for two years, since 2013. He is currently a Lecturer with the School of Information Science and Engineering, Shandong Normal University, China. His research interests include machine learning, network security, and data management.

MINGYANG LI received the B.S. degree in computer science and technology from the Harbin University of Science and Technology, in 2016. He is currently pursuing the degree with the School of the School of Information Science and Engineering, Shandong Normal University, China. He was with Shunfeng Technology as a Big Data Engineer. His research interests include recommender systems, network security, and data mining.

HUAXIANG ZHANG received the Ph.D. degree from Shanghai Jiaotong University, in 2004. He is currently a Professor with the School of Information Science and Engineering, Shandong Normal University, China. He has authored more than 180 journals and conference papers. His research interests include machine learning, pattern recognition, evolutionary computation, and web information processing.