Hindawi
Security and Communication Networks
Volume 2017, Article ID 4842694, 11 pages
https://doi.org/10.1155/2017/4842694

Research Article

An Efficient Context-Aware Privacy Preserving Approach for Smartphones

Lichen Zhang,[1,2] Yingshu Li,[3] Liang Wang,[1,2] Junling Lu,[1,2] Peng Li,[1,2] and Xiaoming Wang[1,2]

[1]*Ministry of Education Key Laboratory for Modern Teaching Technology, Shaanxi Normal University, Xi'an 710119, China*
[2]*School of Computer Science, Shaanxi Normal University, Xi'an 710119, China*
[3]*Department of Computer Science, Georgia State University, Atlanta, GA 30303, USA*

Correspondence should be addressed to Xiaoming Wang; wangxm@snnu.edu.cn

Received 11 March 2017; Accepted 12 April 2017; Published 27 April 2017

Academic Editor: Qing Yang

Copyright © 2017 Lichen Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the proliferation of smartphones and the usage of the smartphone apps, privacy preservation has become an important issue. The existing privacy preservation approaches for smartphones usually have less efficiency due to the absent consideration of the active defense policies and temporal correlations between contexts related to users. In this paper, through modeling the temporal correlations among contexts, we formalize the privacy preservation problem to an optimization problem and prove its correctness and the optimality through theoretical analysis. To further speed up the running time, we transform the original optimization problem to an approximate optimal problem, a linear programming problem. By resolving the linear programming problem, an efficient context-aware privacy preserving algorithm (CAPP) is designed, which adopts active defense policy and decides how to release the current context of a user to maximize the level of quality of service (QoS) of context-aware apps with privacy preservation. The conducted extensive simulations on real dataset demonstrate the improved performance of CAPP over other traditional approaches.

1. Introduction

Nowadays, smartphones have been greatly proliferated and smartphone applications (apps) have been widely developed. Specifically, context-aware apps greatly facilitate people as context-aware personalized services related to people' contexts have been provided. In fact, a variety of sensors (e.g., GPS, microphone, accelerometers, magnetometer, light, and proximity) embedded in smartphones have the capability to measure the surroundings and the status related to the smartphone owner and then provide related data to context-aware apps. These sensory data can be exploited to infer the context or the status about a user. For example, the location information of a user can be reported by GPS data, the transportation state (e.g., walking, running, or standing) can be evaluated by the accelerometers, and the voice and scene can be recorded by microphone and camera, respectively. Furthermore, the inferred context can be further analyzed by context-aware apps for providing context-aware personalized services. There exist a variety of context-aware apps, of which GeoReminder can notify a user when she/he enters particular locations, HealthMonitor can record the amount of exercise of a user in each day, and PhoneWise can smartly mute the phone.

While people's experience and convenience are enhanced by context-aware apps, they raise serious privacy issues [1–3]. Specifically, those untrusted context-aware apps may infer the sensitive context related information about a user and then disclose it to a third party for commercial or malicious intent, thus disclosing user's privacy [4]. In fact, due to the convenient services and functionalities provided by context-aware apps, most users would not refuse to allow these apps to access these related sensory data. Therefore, an increasing demand arises for reducing the risk of context-privacy disclosure while providing the context related services.

However, context-privacy preservation for smartphones is not an easy task because there exist high temporal correlations among human contexts and behaviors in daily life, and these temporal correlations can be used by adversaries to infer the hidden sensitive information. In fact, temporal correlations among human contexts can be modeled well with a Markov chain [5, 6]. By using the knowledge of the temporal correlations between contexts and the current context that a user dwells in, the probability that the user being in any context in the past or in future can be inferred. Thus, the naive approach, in which all the sensitive contexts are simply hidden or suppressed while leaving the others released, would fail to protect user sensitive context due to the absent consideration of the temporal correlations between user contexts.

To cope with the temporal correlations between contexts, Götz et al. [7] proposed MaskIt, in which not only sensitive contexts but also some nonsensitive contexts may be suppressed to decrease the temporal correlations between contexts. Evidently, since more contexts are hidden in MaskIt, the level of quality of services (QoS) provided by context-aware smartphone apps degrades. In fact, the hiding-sensitive policy adopts passive defense, which unavoidably discloses some knowledge to adversaries. For example, an adversary is sure that the released contexts are always real no matter whether the hiding ones are sensitive or not. Recently, a few active defense policies are proposed [8–10]. FakeMask, proposed in [8], first introduces a deception policy with the consideration of decreasing the temporal correlations between contexts. In FakeMask, the released contexts may be not real but still have some meaning (i.e., from the history, the user may have a probability being in that context at that time) to confuse the adversaries. With such a deception policy, the released number of real contexts increases greatly and then leads to a better service quality for users. However, in FakeMask, the brute-force search for the optimal solution consumes huge computation resources, thus restricting its applications on smartphones. Therefore, it is necessary and important to propose an efficient lightweight privacy preservation approach with the temporal correlations between user contexts taken into consideration.

In this paper, we first model the temporal correlations between user contexts with a heterogeneous Markov model and then formalize the context-privacy problem for smartphones to an optimization problem followed with correctness proof. Then, in order to speed up the running time, we further transform the original optimization problem to a near optimal problem, a linear programming problem. Moreover, by resolving the linear programming problem, we design an efficient context-aware privacy preserving algorithm (CAPP), which adopts active defense policy, and can decide how to release the current context of a user to maximize the level of quality of service (QoS) of context-aware apps with privacy preservation. Finally, we conduct extensive simulations to evaluate the algorithm performance, and the simulation results demonstrate the effectiveness and efficiency of the proposed algorithm. In summary, the main contributions of this paper are threefold. First, we formalize the context-privacy problem with the consideration of existence of temporal correlations between user contexts to an efficient optimization problem and prove its correctness and the optimality. Second, to speed up the running time further, we transform the original optimization problem to an approximate optimal problem, a linear programming problem. By resolving the linear programming problem, an efficient context-aware privacy preserving algorithm (CAPP) is designed, which adopts active defense policy and decides how to release the current context of the user to maximize the level of quality of service (QoS) of context-aware apps with privacy preservation. Finally, we conduct extensive evaluations on real smartphone context traces to demonstrate the effectiveness and efficiency of the proposed CAPP compared with the traditional approaches.

The rest of the paper is organized as follows. Section 2 introduces the related works. Section 3 presents the models and preliminaries, followed by the problem formulation and the proposed privacy preserving algorithm in Section 4. Section 5 illustrates the performance evaluation. Finally, Section 6 concludes the paper.

2. Related Works

With the rapidly growing popularity of smartphones as well as popular mobile social applications, various kinds of mobile smartphone apps are developed to provide context-aware services for users. Meanwhile, individual privacy issues on smartphones are increasingly receiving attentions due to the risk of disclosure of user's privacy sensitive information. Various approaches have been proposed to protect users' sensitive information in location-based services (LBSs) and participatory sensing applications [11]. In fact, most previous privacy protection techniques focus on the static scenarios [12–19], in which the instant sensitive location information is protected without consideration of temporal correlations among locations.

The hiding or deception policies are first used in location privacy preserving approaches in [14, 16], in which the current location information of a person may be hidden or a fake location is released to replace the real one if the current location information is sensitive and should not be accessed by untrusted apps. Among the techniques, spatial cloaking and anonymization are widely adopted [20–22], in which the identity of a user who issues a query specifying his/her location is hidden by replacing that user's exact location with a broader region containing at least k users. However, these techniques do not protect privacy against adversaries who have the knowledge of the temporal correlations between contexts. Moreover, the anonymity-based approaches do not readily imply privacy sometimes. For example, if all the k users are in the same sensitive region, an adversary would know the fact.

There have been several popular works of privacy protection against adversaries who are aware of the temporal correlations between contexts [7–9, 23, 24]. The work in [23] considers that an adversary can adopt a linear interpolation to infer the supposedly hidden locations from prior-released locations of a user, in which some zones containing multiple sensitive locations are created in order to increase uncertainty

that the user dwells at one of the sensitive locations. Due to the suppression of sensitive locations and the uncertainty of zones, this approach greatly reduces privacy disclosure compared with the simple hiding-sensitive policy.

MaskIt [7] is the first approach to preserve privacy against the adversaries who know the temporal correlations between the contexts of user. In MaskIt, a user's contexts and their temporal correlations are modeled with a time-heterogeneous Markov chain, which can be also observed by an adversary. By hiding most sensitive contexts and partial nonsensitive ones, MaskIt can increase the difficulty of inferring the hidden sensitive context by adversaries and thus could achieve a better privacy and utility tradeoff. As aforementioned, the number of suppressed contexts is much greater than that in the simple hiding-sensitive approach, leading to a degraded utility and functionality.

The work in [24] considers the interaction between a user and an adversary as well as the temporal correlations between contexts. Unlike MaskIt, in [24], a user controls the granularity of the released contexts, and an adversary has limited capability which means the adversary can only obtain a subset of the user's contexts as the goal of attacking and then actively adjusts his/her future strategies based on the attacking results. In this approach, the interactive competition between the user and the adversary is formalized as a stochastic game, and its Nash Equilibrium point is then obtained. Since the released contexts are some granularity of the truth, the adversary can only gain partial contexts, thus decreasing the privacy disclosure to some degree. On the other hand, since the deception policy is not applied, the obtained contexts by the adversary are still approximately consistent with the truth, which also could be used by the adversary to infer the real sensitive contexts.

A number of privacy preservation techniques have been proposed by using access control techniques [25–27], in which the smartphone resources are controlled by the user-defined access control policies. BlurSense, presented in [25], is an efficient tool that implements a context-aware reference monitor to control all the access on the resources. By using BlurSense, a smartphone user is provided with an interface to define flexible access control policies for all the embedded sensors, which are monitored and controlled by reference monitors for achieving a fine-grained access control.

Besides the aforementioned mechanisms, a variety of privacy preservation schemes have been introduced in other application scenarios like data collection [11, 28, 29], medical care [30], influence maximization [31, 32], collaborative decision-making [33], and others [18, 34–36].

To the best of our knowledge, our approach is the first work to provide an efficient optimal approach in which the deception policy is introduced with privacy preservation on smartphones while considering the temporal correlations between user contexts. In the proposed approach, a Markov chain is used to model the contexts of a user and the temporal correlations between user contexts. Then, with the Markov model, the context-privacy problem for smartphones is formalized to an optimization problem and its correctness and the optimality are proved. To further speed up the computation, a linear programming problem is obtained to look for an efficient feasible solution. By resolving the linear programming problem, a near optimal context-aware privacy preserving algorithm (CAPP) is proposed, which is designed to accelerate the computation through local optimization at any time with user-defined privacy preservation.

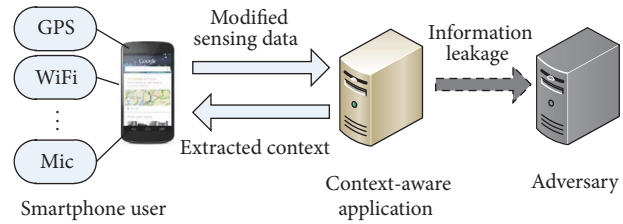


Figure 1: A mobile phone context sensing system [7, 8].

3. Models and Preliminaries

3.1. Models and Assumptions

3.1.1. System Model. We illustrate a smartphone context sensing system in Figure 1, where the privacy preserving system protects a user's privacy context from those untrusted smartphone apps. In Figure 1, the raw sensory data are first collected by smartphone sensors and filtered by the privacy preserving system, which in turn transmits the processed sensory data to those untrusted context-aware apps. Thus, the privacy preserving system served as a middleware in the system, and then the untrusted context-aware apps could not access the raw sensory data and could only obtain the released sensory data from the privacy preserving system. In the process of handling the sensory data, the privacy preserving system infers the related context from the collected sensory data by using the model about the temporal correlations between user context and then releases the filtered sensory data with privacy preservation. Based on the released sensory data from the privacy preserving system, the context about the user could be reasoned and the context-aware services are accordingly provided to the user by the context-aware apps with the capability of obeying the user's privacy protection policy.

User's context can be inferred from sensory data. That is, at any time the privacy preserving system can obtain user's context according to the collected sensory data. So, in the following we use context to represent the related sensory data for ease of illustration. In this paper, we adopt periodic discrete time as in [7, 8, 24]. At any discrete time period t, a user's context c_t can be inferred and then handled by the privacy preserving system, and then the result context o_t is released to the context-aware apps with privacy preservation. To preserve user's privacy, the output o_t from the privacy preserving system falls in two different forms, real or fake. The real ($o_t = c_t$) means the raw sensory data related to the real context c_t is released to the context-aware apps. On the contrary, a fake context means the context o_t inferred from the released sensory data is not the original context c_t at time t. Based on the user's predefined privacy parameter, the privacy preserving system makes a decision to release the real sensory data or a fake one with the goal that the expectation

of the released real contexts is maximized while guaranteeing the privacy preservation.

Unlike the "release or suppress" paradigm in [7], the privacy preserving system in this paper introduces the "release or deceive" paradigm in [8] to increase the number of releasing real contexts while guaranteeing user's privacy. Compared with the traditional schemes, such as MaskIt [7] and FakeMask [8], our novel approach is optimal under the above system model through theoretical analysis and could substantially improve the number of released real contexts while preserving privacy.

3.1.2. Context Model and Markov Chain. As aforementioned, the periodic discrete time is adopted, so we try to model a user's contexts over a period of discrete time (e.g., a day, a week). All the possible contexts of a user in a period of time are represented by a finite set $C = \{c_1, \ldots, c_N\}$, in which N represents the number of discrete times in one period of time. As in [7, 24], we adopt a time-heterogenous Markov chain to capture the temporal correlations between contexts of a user. A time-heterogenous Markov process is denoted by $Z = \{Z_1, Z_2, \ldots\}$, in which Z_t represents the context of the user at discrete time t. Due to the cyclic nature of time, we infer that $Z_{N+m} = Z_m$ for any integer m. The independence property of the time-heterogenous Markov process states that

$$P\left\{Z_{t+1} = j \mid Z_1, \ldots, Z_t = i\right\} = P\left\{Z_{t+1} = j \mid Z_t = i\right\}, \tag{1}$$

where $P\{Z_{t+1} = j \mid Z_t = i\}$ is the probability that the process enters state j at time $t+1$ with the condition that the process was in state i at time t, also denoted by $P^t_{i,j}$.

3.1.3. Adversary Model. To make our approach more robust, we assume adversaries could obtain the knowledge of the Markov chain, in which the temporal correlations between the contexts of a user through observing the output sequence of the sensory data are modeled. By using the Markov chain Z and the distribution of the initial contexts of a user, an adversary could conclude the prior belief about the user being in any context c at time t, denoted by probability $P\{Z_t = c\}$. Furthermore, through the observation of the previously released contexts of the user, the adversary can apply the Bayesian reasoning to obtain their posterior belief about the user being in a context. That is, the posterior belief, denoted by $P\{Z_t = c \mid \vec{o}\}$, can be inferred by conditioning the observed output sequence $\vec{o}$ from the privacy preserving system. The goal of an adversary is to increase the posterior belief about the user being in a sensitive context and try to break the user's privacy protection policy. Note that the posterior belief is usually greater than the corresponding prior belief due to the fact that more knowledge about the posterior belief is obtained.

3.2. Preliminaries about Context Reasoning

3.2.1. Hidden Markov Chain. Let $Z = \{Z_1, Z_2, \ldots\}$ be a Markov chain with transition probabilities $P^t_{i,j}$, where $P^t_{i,j}$ is the probability that the process enters context j at time $t+1$ with the condition that the process was in context i at time t. Suppose that a novel context is emitted each time the Markov chain enters a context, and there exists a finite set of emitted contexts. Specifically, if the Markov chain enters context i at time t, then, independently of previous contexts and emitted contexts, the present context emitted is o with probability $p(o \mid i, t)$ with $\sum_o p(o \mid i, t) = 1$, where o is the emitted context observed by adversaries. Thus, the output contexts also construct a process $\{O_1, O_2, \ldots\}$, where O_t represents the emitted context variable at time t. Formally, we have $p(o \mid i, t) = P\{O_t = o \mid Z_t = i\}$. Since the inside process Z is hidden from the observers and can only be reasoned through the emitted context, the process Z is called a hidden Markov chain.

3.2.2. Reasoning on Hidden Markov Chain. Consider a hidden Markov chain Z, with each random variable Z_t taking a value in the set of contexts C at time t. As aforementioned, the hidden Markov chain can model the temporal correlations between contexts of a user and can also be obtained by adversaries through the output contexts. In the following, we illustrate how the adversaries infer the hidden context from the output context sequence. Note that the actual released contents are sensory data, which can be inferred by adversaries to obtain the related context. Supposing that an adversary knows the hidden Markov chain Z and the initial probability P^0_j, where P^0_j is the probability that the user is in context j at the beginning time, the adversary could apply the Bayesian reasoning to obtain the prior belief that the user enters any context at any time.

Proposition 1. *The prior belief of an adversary (who knows a user's hidden Markov chain Z and the initial probability P^0_j) about the user being in context k at time t is equal to*

$$P\left\{Z_t = i\right\} = \sum_{i_0} \cdots \sum_{i_{t-1}} P^0_{i_0} \cdot P^1_{i_0,i_1} \cdots P^t_{i_{t-1},i_t}, \tag{2}$$

where $i_0, i_1, \ldots, i_t \in C$, with $i_0 = j$ being the beginning context and $i_t = k$ being the context at time t.

It is worth mentioning that, whatever policies are applied and whatever the output context is, if an adversary guesses that the user is in a sensitive context $s \in C$ at time t, the probability that the guess result is true is at least $P\{Z_t = s\}$ because this probability can be computed by using (2). Moreover, since more information (i.e., the inferred context from the released sensory data) can be observed by an adversary, the guess probability can be larger than the prior belief. That is, an adversary could infer the present context with the knowledge of the prior-released context sequence and the related Markov model.

For a hidden Markov chain Z, each context has a distribution over possible outputs at any time. The output context at time t is a random variable O_t. We define the emission matrix B whose element is equal to

$$b^t_{c,c'} = P\left\{O_t = c' \mid Z_t = c\right\}, \tag{3}$$

where $b^t_{c,c'}$ denotes the probability of releasing the context c' at time t with the condition that the context is c at time t.

From (3), we knows that $b^t_{c,c}$ is the probability of releasing the real context and $b^t_{c,c'}$ is the probability of releasing a fake context c' where $c' \neq c$. Note that if we let $c' = \phi$ denote nothing is released and there is no fake output, the above policy is just MaskIt in [7]. Furthermore, in our general policy, since the output context c' may belong to possible contexts C, it could confuse the adversaries and then allows the privacy preserving system to release more real contexts with the same user predefined privacy.

For a user, the context that the user dwells at any time is hidden from the adversaries. Suppose at time t that the hidden context takes a value from Z_t and the emitted context takes a value from O_t. The adversaries could only infer the hidden context of a user based on the observation of the emitted contexts. Furthermore, the emitted context is determined according to the emission probability. For a given output sequence $\vec{o} = o_1, \ldots$ released context from the privacy preserving system, an adversary could obtain the conditional probability (posterior probability) that at time t the hidden context was c by

$$P\left\{Z_t = c \mid \vec{o}\right\} = \frac{P\left\{Z_t = c, o_1, \ldots, o_{t-1}\right\} \cdot P\left\{o_t, \ldots, o_T \mid Z_t = c\right\}}{P\left\{\vec{o}\right\}}. \quad (4)$$

For the detailed process of the above conditional probability, please refer to [7].

4. Problem Formulation and Our Approach

4.1. Problem Formulation. We adopt the definition of privacy in [7], in which a user declares a subset of contexts $S \subset C$ as private sensitive contexts and also claims a privacy preservation parameter δ with $\delta \in [0, 1]$. Informally, we declare that a released context sequence $\vec{o}$ preserves privacy if the adversary cannot learn much about the user being in a sensitive context from the released context sequence $\vec{o}$. That is, for all sensitive contexts and all times, the posterior belief about the user being in a sensitive context cannot be larger than the prior belief plus a predefined privacy parameter δ. Formally, we have the following δ-privacy definition.

Definition 2 (see [7]). We claim that a system preserves δ-privacy against an adversary if, for all possible outputs $\vec{o}$, all times t, and all sensitive contexts $s \in S$, the following inequation holds:

$$P\left\{Z_t = s \mid \vec{o}\right\} - P\left\{Z_t = s\right\} \leq \delta. \quad (5)$$

Note that the δ-privacy definition guarantees that an adversary cannot learn too much about the user being in a sensitive context even though the adversary has an access to the output sequence of the system and also knows the Markov chain of the temporal correlations between the user's contexts.

The goal of a privacy preserving system is to release as many real contexts as possible, while satisfying the δ-privacy constraint. Specifically, a privacy preserving system tries to obtain an emission matrix B, which preserves user's privacy (i.e., (5) holds), and maximizes the utility of the system. Formally, the utility of a privacy preserving system is defined as follows.

Definition 3. We say that the utility of a system is the expectation of the number of the released real contexts; that is,

$$u\left(\mathscr{A}\right) = \sum_{\vec{o}} P\left\{\vec{o}\right\} \cdot \left|\left\{i \mid o_i = c_i\right\}\right| = \sum_{t \in [T], c \in C} P\left\{Z_t = c\right\} \cdot b^t_{c,c}, \quad (6)$$

where $b^t_{c,c}$ is the probability of releasing the real context c at time t, $P\{Z_t = c\}$ is the prior belief that the user is in context c at time t, and $[T]$ is the set of all possible discrete times in a period of time.

Therefore, the objective of a privacy preserving system is obtaining an emission matrix B, which tries to maximize the utility with the privacy preservation.

Götz et al. [7] proposed a method, in which all possible emission probabilities are brute-force-searched to find one that maximizes the utility while preserving δ-privacy. Moreover, in the process of trying each emission matrix in [7], the posterior belief has to be computed. However, the attempts on all possible emission probabilities on all possible output context sequences to resolve the solution would consume huge computation resources, thus leading to less feasible resource-constrained smartphones and even PCs.

To cope with the issue of the huge computation consumption in the above approach, in this section, we design an efficient privacy preserving approach, in which the emission matrix can be obtained in an efficient way. We first present some propositions to illustrate our privacy preserving approach and then describe our privacy preserving algorithm.

To make the privacy preservation problem easier, we first assume that there exist no temporal correlations between user contexts. Under this assumption, to preserve δ-privacy, the system should only guarantee that, at any time for any sensitive context, its posterior belief under any possible observation is not larger than δ plus its prior belief.

Proposition 4. *Under the assumption that there exist no temporal correlations between the adjacent contexts, a system $\mathscr{A}$ preserves δ-privacy against an adversary if, for any possible released context $o \in C$ and for any possible sensitive context $s \in S$ at any time t, the following inequation holds:*

$$\frac{b^t_{s,o} \cdot P^t_s}{\sum_{c \in C} b^t_{c,o} \cdot P^t_c} \leq \delta + P^t_s, \quad (7)$$

where $b^t_{s,o}$ is the emission probability of releasing context o at time t under the condition that the real context is s at time t and P^t_s and P^t_c are the prior beliefs that the context is, respectively, s and c at time t with $P^t_s > 0$ and $P^t_c > 0$.

The above proposition is evident since it needs no consideration of the temporal correlations between the adjacent contexts. Moreover, there always exists a feasible solution to (7). Specifically, whatever the current context is at any time t, a system preserves δ-privacy if the emission probability of releasing a context c equals its prior belief P_c^t. Formally, if, for any context $c, c' \in C$, we let the emission probability $b_{c,c'}^t = P_{c'}^t$, the following inequation holds:

$$\frac{b_{s,o}^t \cdot P_s^t}{\sum_{c \in C} b_{c,o}^t \cdot P_c^t} = \frac{P_o^t \cdot P_s^t}{\sum_{c \in C} P_o^t \cdot P_c^t} = P_s^t. \tag{8}$$

However, by knowing the posterior belief of a context c at time t (denoted by $Q_{c,o}^t$) and also knowing the context transition probability of entering a sensitive context s at the next time $t+1$ (denoted by $P_{c,s}^t$), an adversary could obtain the posterior belief of a user being in sensitive context s at time $t+1$ with probability $Q_{c,o}^t P_{c,s}^t$. In fact, if $Q_{c,o}^t P_{c,s}^t > P_s^{t+1} + \delta$, the δ-privacy will be broken. Therefore, in order to preserve δ-privacy, for any output context o at time t and for any possible sensitive context s at time $t+1$, the following inequation should hold:

$$\sum_{c \in C} Q_{c,o}^t \cdot P_{c,s}^t \le P_s^{t+1} + \delta. \tag{9}$$

Motivated by the above analysis, to preserve δ-privacy in the existence of temporal correlations between user contexts, the δ-privacy preserving problem is formulated as follows.

Proposition 5. *Under the existence of temporal correlations between the contexts, a system preserves δ-privacy if the emission probability is resolved from the following optimization problem:*

$$\begin{aligned}
\max \quad & \sum_{c \in C, t \in T} b_{c,c}^t \cdot P_c^t \\
\text{s.t.} \quad & (1)\ Q_{s,o}^t \le \delta + P_s^t, \\
& \qquad \forall s \in S,\ o \in C,\ t \in T \\
& (2)\ \sum_{c \in C} Q_{c,o}^t \cdot P_{c,s}^t \le P_s^{t+1} + \delta, \\
& \qquad \forall s \in S,\ o \in C,\ t \in T \\
& (3)\ \sum_{c \in C} Q_{c,o}^t \cdot \widehat{P}_{s,c}^{t-1} \le P_s^{t-1} + \delta, \\
& \qquad \forall s \in S,\ o \in C,\ t \in T \\
& (4)\ b_{c,c'}^t \in [0,1], \\
& \qquad \forall c, c' \in C,\ t \in T,
\end{aligned} \tag{10}$$

where $b_{c,c'}^t$ is the emission probability of releasing context c' at time t under the condition that the user is in context c at time t, P_c^t is the prior belief of a user being in context c at time t, $Q_{c,o}^t$ is the posterior belief that a user is in context c on the output context o at time t, $\widehat{P}_{s,c}^{t-1} = P_{s,c}^{t-1} / \sum_{c' \in C} P_{s,c'}^{t-1}$ is the normalized probability of $P_{s,c}^{t-1}$, $P_{c,s}^t$ is the transition probability from context c at time t to context s at time $t+1$, and P_s^{t+1} and P_s^{t-1} are the prior probabilities of a user being in context s at time $t+1$ and time $t-1$, respectively.

Proof. As mentioned in Proposition 4, under the assumption that there exist no temporal correlations between contexts, the solution to Constraint (1) in (10) preserves δ-privacy at time t. On the contrary, due to the existence of temporal correlations between contexts, the above solution may break δ-privacy at time $t-1$ or $t+1$. In fact, Constraint (2) in (10) guarantees that the posterior belief of a user being in sensitive context s at time $t+1$, caused by any released context at time t, will be not larger than its prior belief plus δ. Similarly, Constraint (3) in (10) guarantees that the posterior belief of a user being in sensitive context s at time $t-1$, only caused by any released context at time t, will be not larger than its prior belief plus δ.

Therefore, for any time t the solution (i.e., the emission probability at time t) to (10) satisfies the statement that an adversary, based on all possible output contexts at time t, cannot infer that the user is in a sensitive context at times $t-1$, t, and $t+1$ with a probability larger than its prior belief plus δ. In other words, an adversary cannot infer that the user is in a sensitive context at time t with a probability larger than its prior belief plus δ under all the possible released context at time t, $t-1$, and $t+1$. Thus, based on the transitivity, under the observation of any possible released context sequence, the above solution preserves δ-privacy. □

We have to mention that the condition in the posterior probability in (10) is the context at time t while, in the definition of the δ-privacy, the condition is the context sequence $\vec{o}$. Thus, the computing of (10) is much more efficient than that from the definition of the δ-privacy if a brute-force search is used. Furthermore, the above solution to (10) is also optimal. The proof is evident, because any possible solution must satisfy the above 4 constraints which are the necessary and sufficient conditions. However, (10) is not a linear programming problem due to the fact there exist multiple multiplications on different variables. In order to speed up the running time, we then propose an efficient approach which formulates the above optimization problem to a near optimal problem.

Theorem 6. *Under the existence of temporal correlations between user contexts, a system preserves δ-privacy if the emission probability at any time t is resolved from the following linear programming problem:*

$$\begin{aligned}
\max \quad & \sum_{c \in C} b_{c,c}^t \cdot P_c^t \\
\text{s.t.} \quad & (1)\ Q_{s,o}^t \le \delta + P_s^t, \\
& \qquad \forall s \in S,\ o \in C, \\
& (2)\ \sum_{c \in C} Q_{c,o}^t \cdot P_{c,s}^t \le P_s^{t+1} + \delta, \\
& \qquad \forall s \in S,\ o \in C,
\end{aligned}$$

Input: Markov chain M, sensitive contexts S, privacy threshold δ
Output: Emission probabilities $B = (b_{i,j}^t)$ with $i, j \in C,\ t \in T$
(1) **for all** $c \in C,\ t \in T$ **do**
(2) compute P_c^t by Eq. (2);
(3) **end for**
(4) **for all** $t = 1, \ldots, T$ **do**
(5) construct a linear programming problem as Eq. (11);
(6) compute $b_{c,c'}^t$, where $c, c' \in C$.
(7) **end for**

Algorithm 1: Emission probability generation.

$$
\begin{aligned}
&(3)\ \sum_{c \in C} Q_{c,o}^t \cdot \widehat{P}_{s,c}^{t-1} \le P_s^{t-1} + \delta, \\
&\qquad\qquad \forall s \in S,\ o \in C, \\
&(4)\ b_{c,c'}^t \in [0,1], \quad \forall c, c' \in C.
\end{aligned}
\tag{11}
$$

The proof is evident because, at any given time t, (11) achieves local optimization solution and guarantees δ-privacy at that time through the above 4 constraints. That is, the solution at any given time does not affect the solutions in the future. We have to mention that (11) is not optimal to δ-privacy problem. There exist some assignments of emission probabilities under which the result in (11) at some given time may not be maximized but leads to the global optimization value in (10). The reason lies in the relation between the local optimization problem and the global one. Detailedly, if we decrease the emission probability at some given time, then a lower posterior probability is achieved which means less posterior belief. Based on less posterior belief, an adversary could infer current and future context with less correctness. Thus, we could increase the emission probability at next time to release more real contexts while still guaranteeing the predefined δ-privacy. Although (11) is not optimal, it is a linear programming problem; thus we can resolve it efficiently by using the existing methods such as the simplex method. To make it better, the above linear programming problem can also be resolved in advance to reduce the computation consumption. It needs to mention that the computing process of (11) at time t requires the solution results of (11) at other times prior to t due to the fact the posterior probability at time t is related to the emission probabilities at time prior to t. That is, in order to compute the solution to (11) at time t, we should compute the optimal solution to (11) at times prior to t first. Thus, it requires that the process of solution to (11) in ascending order of time t.

4.2. The Proposed Approach. According to Theorem 6, we propose our efficient context-aware privacy preserving approach, called CAPP. Algorithm 1 generates the emission probabilities according to the user's Markov model M, sensitive contexts S, and privacy parameter δ. Note that M is learned from historical observations at the training phase of Markov chain.

Input: context c at time t, emission probability $b_{i,j}^t \in B$
Output: output context o at time t
(1) **for all** $c' \in C$ **and** $b_{c,c'}^t > 0$ **do**
(2) **return** c' with probability $b_{c,c'}^t$
(3) **end for**

Algorithm 2: An efficient checking decision algorithm (CAPP).

Based on the generated emission probabilities, Algorithm 2 decides how to release the context of a user with δ-privacy preservation.

It is worth mentioning that even if an adversary had known the Markov model and even the related emission probability matrices, he/she cannot infer the original context with a large probability from the output context sequence of CAPP. The main reason lies in the fact that the constraint of δ-privacy guarantees that an adversary cannot learn too much about the user being in a sensitive context.

5. Evaluation

5.1. Settings. We implement our context-aware privacy preserving algorithm (called CAPP) and compare it with traditional privacy approaches, such as MaskSensitive, MaskIt (using the hybrid check) [7], and EfficientFake [8]. MaskSensitive is a naive approach, in which all sensitive contexts are hidden or suppressed while releasing all nonsensitive ones. All the simulations are conducted in the platform MATLAB 8.4, which runs on the Windows 8.1 operating system with the hardware of Intel Core 1.80 GHz CPU and 8 GB memories.

In this paper, the dataset used in the simulation is from real human traces: Reality Mining dataset, in which fine-grained mobility data of 100 students and staff at MIT over the 2004-2005 academic year are contained [37]. In Reality Mining dataset, the GPS location contexts of each user are, respectively, obtained from the cell towers in the trace through the public cell ID database (e.g., Google location API). We consider 91 users who have at least 1 month of data, in which the total length is 11,091 days. The average, minimum, and maximum trace length per user are 122 days, 30 days, and 269 days, respectively. The average, minimum, and maximum number of distinct locations per user are 19, 7, and 40, respectively.

To obtain a Markov chain for each user, we train on the first half of the user's trace with the remaining half being used for evaluation. Note that, during the collection of the trace of the user, δ-privacy may not be guaranteed due to lack of the prior belief and the emission probabilities. Upon obtaining the solution to (11), we can guarantee the δ-privacy preservation for the user.

For the simulation parameters, we choose the privacy parameter $\delta = 0.1$. It is worth mentioning that the larger the privacy parameter δ is, the lower the user's privacy protection level is and thus the more the real sensory data is released. There are two different ways of choosing sensitive contexts. Unless stated, for each user, we choose uniformly at random sensitive contexts for each user, named "random as sensitive." Alternatively, for each user, we choose the location with the highest prior probability as the home of the user and choose it as sensitive, named "home as sensitive."

As aforementioned, the utility of a privacy preserving approach is the expectation of the number of the released real contexts, so we use the normalized utility as the measurement which is defined as the fraction of the released real contexts. We should note that a higher utility of an approach means a higher quality of service is provided by context-aware apps. Similarly, we measure privacy breaches as the number of the sensitive contexts in the user's context sequence that are breached divided by the length of the user's context sequence. Note that, from the definition, the three approaches CAPP, EfficientFake, and MaskIt always guarantee no privacy breaches. MaskSensitive probably cannot guarantee the δ-privacy due to the absent consideration of the existence of temporal correlations between user contexts.

5.2. Results. First, we compare the privacy breaches of CAPP with other approaches in the following two scenarios. In one scenario, we choose three contexts for each user at random as sensitive, and, in the other, we choose the home of each user as sensitive. Note that the home of a user has the highest prior belief, which means the user spends most of his/her time at home compared to that at other locations.

Figures 2 and 3 report the average fractions of released and suppressed contexts by various algorithms in the above two scenarios, respectively. From the figures, we observe that MaskSensitive suppresses all the sensitive contexts in both scenarios. Although all sensitive contexts are not released in MaskSensitive, an adversary who knows the Markov chain of contexts can infer about 40–60% sensitive contexts from the suppressed ones in the two scenarios. The main reason lies in that the temporal correlation between contexts discloses enough information to an adversary and then makes an adversary infer a larger posterior belief which may exceed the corresponding prior belief by the privacy parameter δ. On the contrary, the other three approaches such as CAPP, EfficientFake, and MaskIt guarantee δ-privacy through. For CAPP, EfficientFake, and MaskIt, we can see that some sensitive contexts as well as some nonsensitive ones are suppressed and released. Furthermore, the average fraction of the released real contexts by CAPP is larger than that of MaskSensitive, MaskIt, and EfficientFake. From the figures, we see that MaskIt sacrifices less than 20% of the utility of

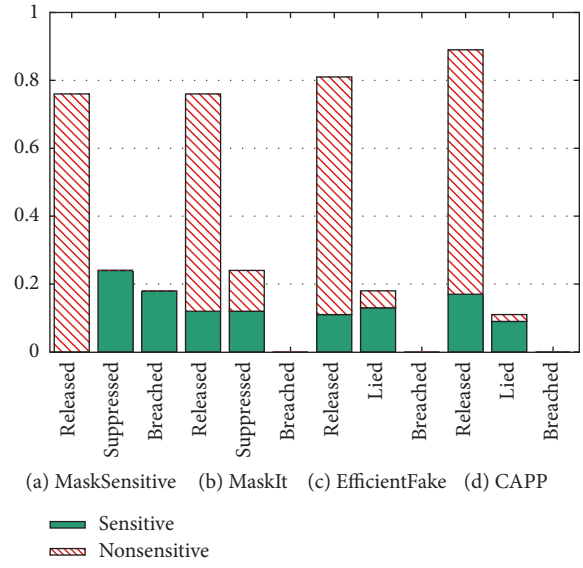


Figure 2: Privacy breach comparison (home as sensitive).

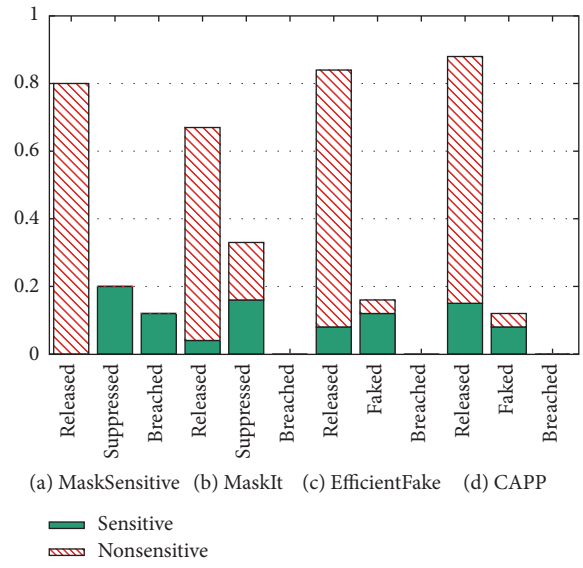


Figure 3: Privacy breach comparison (random as sensitive).

MaskSensitive to guarantee privacy. However, both EfficientFake and CAPP increase near 20% of the utility compared to MaskSensitive while guaranteeing privacy. The main reason is that the introduced deception policy makes an adversary difficult to infer the posterior belief and then allows releasing more real contexts. Although both EfficientFake and CAPP are formalized to linear programming problems, our CAPP performs better than EfficientFake in the aspect of average utility in both scenarios. The main reason is twofold. The first is that the goal in EfficientFake is to maximize the emission probability only while in CAPP the goal is to maximize the utility value at a given time. The second is that solution space on EfficientFake is greatly decreased. Specifically, in EfficientFake, the form of the emission probability matrix is decreased to a vector, which decreases the accuracy of the

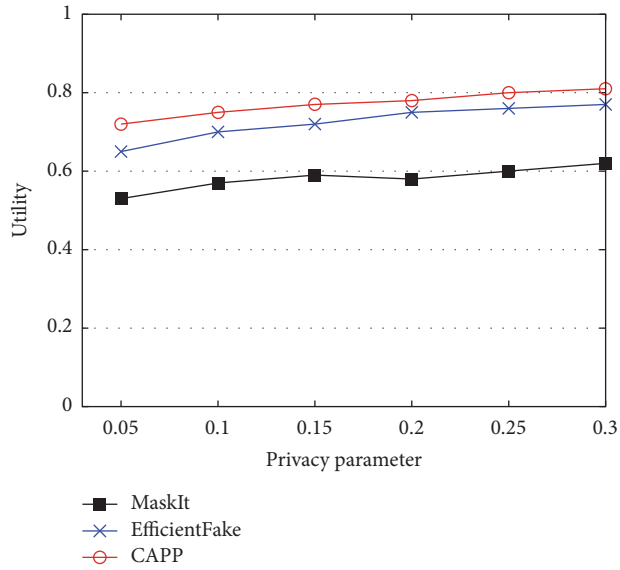


Figure 4: Privacy-utility tradeoff (home as sensitive).

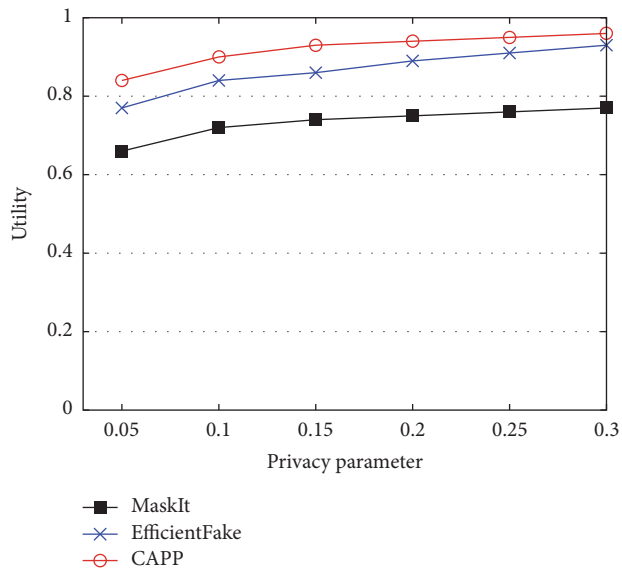


Figure 5: Privacy-utility tradeoff (random as sensitive).

solution greatly in EfficientFake, leading to less utility than CAPP. On the contrary, in CAPP, the solution space is not shrunk, and we can obtain a better optimization solution.

We then compare the utility of our CAPP with other approaches under different privacy parameters which varies from 0.05 to 0.3. Similar to the former experiments, we choose different sensitive contexts in the experiments: the sensitive context for a user is chosen to be the user's home, and the other is chosen at random. We expect the utility to increase with the decrease of the privacy requirement. As we can see from Figures 4 and 5, the utility increases slowly as δ increases in both scenarios. Furthermore, we can see that, at the same privacy parameter δ, each approach performs better in the second scenario where random context is chosen as sensitive than that in the first scenario where home is sensitive. Since, in the first scenario in Figure 4, the location for each person with the highest prior belief is chosen as sensitive context, the number of sensitive contexts is larger than that in the second scenario in Figure 5 where a context is randomly chosen as sensitive. To guarantee the same δ-privacy, CAPP and EfficientFake should disguise more contexts by releasing more fake contexts in the first scenario. But, compared with other approaches, our CAPP achieves the best due to its fine approximation to the optimal optimization of the problem.

6. Conclusions

In this paper, we address the context-aware privacy preserving problem for smartphones. We formalize the context-privacy preservation problem to an optimization problem and prove the correctness and the optimality of our formulation through theoretical analysis. In order to speed up the computing further, we propose an efficient near optimal approach in which a linear programming problem is formulated. By resolving the linear programming problem, an efficient context-aware privacy preserving algorithm (CAPP) is proposed. Through the extensive experimental evaluations on real mobility trace, we demonstrate that our proposed CAPP achieves much more utility than the traditional approaches while guaranteeing the user's δ-privacy policy. One interesting future work is to determine an online context releasing decision algorithm which could make quicker and more efficient decisions only based on the present context of the user with privacy preservation. Since this paper concerns the privacy preservation for a single user, another future work is to propose a privacy preservation approach with the consideration of interactions among users since there exists group mobility in humans.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work is partly supported by the National Natural Science Foundation of China (nos. 61402273, 61373083, and 61601273), the NSF of USA (no. CNS-1252292), the Fundamental Research Funds for the Central Universities of China (nos. GK201603115 and GK201703061), and the Program of Shaanxi Science and Technology Innovation Team of China (no. 2014KTC-18).

References

[1] P. Corcoran, "Privacy in the age of the smartphone," *IEEE Potentials*, vol. 35, no. 5, pp. 30–35, 2016.

[2] J. Tsai, PG. Kelley, LF. Cranor, and N. Sadeh, "Location sharing technologies: privacy risks and controls," *I/S: A Journal of Law and Policy for the Information Society*, vol. 6, no. 2, pp. 119–317, 2010.

[3] X. Zheng, Z. Cai, J. Li, and H. Gao, "Location-privacy-aware review publication mechanism for local business service systems," in *Proceedings of the 36th Annual IEEE International Conference on Computer Communications (INFOCOM '17)*, Atlanta, Ga, USA, 2017.

[4] W. Enck, P. Gilbert, B.-G. Chun et al., "Taint droid: an information flow tracking system for real-time privacy monitoring on smartphones," *Communications of the ACM*, vol. 57, no. 3, pp. 99–106, 2014.

[5] E. Kim, S. Helal, and D. Cook, "Human activity recognition and pattern discovery," *IEEE Pervasive Computing*, vol. 9, no. 1, pp. 48–53, 2010.

[6] A. Mannini and A. M. Sabatini, "Accelerometry-based classification of human activities using Markov modeling," *Computational Intelligence and Neuroscience*, vol. 2011, Article ID 647858, 10 pages, 2011.

[7] M. Götz, S. Nath, and J. Gehrke, "MaskIt: Privately releasing user context streams for personalized mobile applications," in *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD '12)*, pp. 289–300, USA, May 2012.

[8] L. Zhang, Z. Cai, and X. Wang, "FakeMask: A Novel Privacy Preserving Approach for Smartphones," *IEEE Transactions on Network and Service Management*, vol. 13, no. 2, pp. 335–348, 2016.

[9] W. Wang and Q. Zhang, "Privacy Preservation for Context Sensing on Smartphone," *IEEE/ACM Transactions on Networking*, 2016.

[10] Z. He, Z. Cai, J. Yu, X. Wang, Y. Sun, and Y. Li, "Cost-efficient strategies for restraining rumor spreading in mobile social networks," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 3, pp. 2789–2800, 2017.

[11] L. Zhang, X. Wang, J. Lu, P. Li, and Z. Cai, "An efficient privacy preserving data aggregation approach for mobile sensing," *Security and Communication Networks*, vol. 9, 3844, no. 16, p. 3853, 2016.

[12] C. S. Jensen, H. Lu, and M. L. Yiu, "Location privacy techniques in client-server architectures," in *Privacy in Location-Based Applications, Lecture Notes in Computer Science*, vol. 5599, pp. 31–58, 2009.

[13] M. Gruteser and D. Grunwald, "Anonymous usage of location-based services through spatial and temporal cloaking," in *Proceedings of the 1st International Conference on Mobile Systems, Applications and Services (MobiSys '03)*, pp. 31–42, San Francisco, Calif, USA, May 2003.

[14] A. R. Beresford, A. Rice, N. Skehin, and R. Sohan, "MockDroid: trading privacy for application functionality on smartphones," in *Proceedings of the 12th International Workshop on Mobile Computing Systems and Applications (HotMobile '11)*, pp. 49–54, ACM, Phoenix, Ariz, USA, March 2011.

[15] P. Hornyack, S. Han, J. Jung, S. Schechter, and D. Wetherall, "These aren't the droids you're looking for: retrofitting android to protect data from imperious applications," in *Proceedings of the 18th ACM Conference on Computer and Communications Security (CCS '11)*, pp. 639–651, Chicago, Illinois, USA, October 2011.

[16] K. Fawaz and K. G. Shin, "Location privacy protection for smartphone users," in *Proceedings of the 21st ACM Conference on Computer and Communications Security (CCS '14)*, pp. 239–250, USA, November 2014.

[17] X. Wang, Y. Mu, and R. Chen, "One-round privacy-preserving meeting location determination for smartphone applications," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 8, pp. 1712–1721, 2016.

[18] X. Zheng, Z. Cai, J. Yu, C. Wang, and Y. Li, "Follow but no track: privacy preserved profile publishing in cyber-physical social systems," *IEEE Internet of Things Journal*, 2017.

[19] X. Li, J. Yang, Z. Sun, and J. Zhang, "Differential privacy for edge weights in social networks," *Security and Communication Networks*, vol. 2017, Article ID 4267921, 10 pages, 2017.

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

[21] C. Reynold, Y. Zhang, E. Bertino, and S. Prabhakar, "Preserving user location privacy in mobile data management infrastructures," in *Proceedings of the 6th international conference on Privacy Enhancing Technologies (PET '06)*, G. Danezis and P. Golle, Eds., vol. 4258 of *Lecture Notes in Computer Science*, pp. 393–412, Springer, Cambridge, UK, 2006.

[22] K. Vu, R. Zheng, and J. Gao, "Efficient algorithms for K-anonymous location privacy in participatory sensing," in *Proceedings of the 31st Annual IEEE International Conference on Computer Communications (INFOCOM '12)*, pp. 2399–2407, Orlando, Fla, USA, March 2012.

[23] M. Gruteser and X. Liu, "Protecting privacy in continuous location-tracking applications," *IEEE Security & Privacy*, vol. 2, no. 2, pp. 28–34, 2004.

[24] W. Wang and Q. Zhang, "A stochastic game for privacy preserving context sensing on mobile phone," in *Proceedings of the 33rd Annual IEEE International Conference on Computer Communications (INFOCOM '14)*, pp. 2328–2336, can, May 2014.

[25] J. Cappos, L. Wang, R. Weiss, Y. Yang, and Y. Zhuang, "BlurSense: dynamic fine-grained access control for smartphone privacy," in *Proceedings of the 9th IEEE Sensors Applications Symposium (SAS '14)*, pp. 329–332, Queenstown, New Zealand, February 2014.

[26] L. Zhang, X. Wang, W. Dou, and X. Zhao, "Secure verifiable active access control for medical sensor networks," *Chinese Journal of Electronics*, vol. 21, no. 3, pp. 555–558, 2012.

[27] J. Abdella, M. Özuysal, and E. Tomur, "CA-ARBAC: privacy preserving using context-aware role-based access control on Android permission system," *Security and Communication Networks*, vol. 9, no. 18, pp. 5977–5995, 2016.

[28] F. Rahman, D. Williams, S. I. Ahamed, J.-J. Yang, and Q. Wang, "PriDaC: privacy preserving data collection in sensor enabled RFID based healthcare services," in *Proceedings of the IEEE 15th International Symposium on High-Assurance Systems Engineering (HASE '14)*, pp. 236–242, USA, January 2014.

[29] Z. Cai, Z. He, X. Guan, and Y. Li, "Collective data-sanitization for preventing sensitive information inference attacks in social networks," *IEEE Transactions on Dependable and Secure Computing*, 2016.

[30] H. Zhu, X. Liu, R. Lu, and H. Li, "Efficient and privacy-preserving online medical pre-diagnosis framework using nonlinear svm," *IEEE Journal of Biomedical and Health Informatics*, 2016.

[31] M. Han, J. Li, Z. Cai, and Q. Han, "Privacy reserved influence maximization in gps-enabled cyber-physical and online social networks," in *Proceedings of the IEEE International Conference on Social Computing and Networking (SocialCom '16)*, pp. 284–292, Atlanta, Ga, USA, October 2016.

[32] M. Han, Q. Han, L. Li, J. Li, and Y. Li, "Maximizing influence in sensed heterogenous social network with privacy preservation," *International Journal of Sensor Networks*, pp. 1–11, 2017.

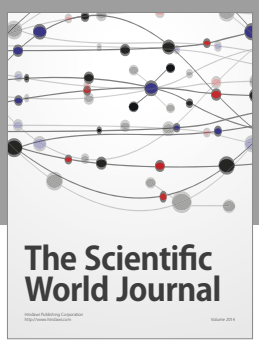
[33] I. Bilogrevic, M. Jadliwala, V. Joneja, K. Kalkan, J.-P. Hubaux, and I. Aad, "Privacy-preserving optimal meeting location determination on mobile devices," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 7, pp. 1141–1156, 2014.

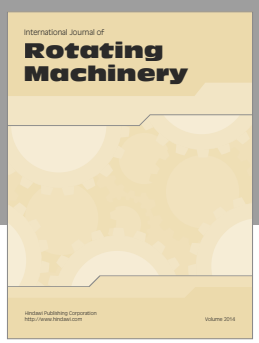
[34] X. Wang, Y. Lin, Y. Zhao, L. Zhang, J. Liang, and Z. Cai, "A novel approach for inhibiting misinformation propagation in human mobile opportunistic networks," *Peer-to-Peer Networking and Applications*, vol. 20, no. 2, pp. 337–394, 2016.

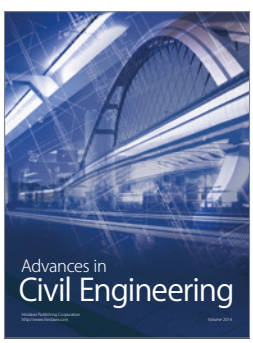
[35] S. Gisdakis, V. Manolopoulos, S. Tao, A. Rusu, and P. Papadimitratos, "Secure and privacy-preserving smartphone-based traffic information systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 3, pp. 1428–1438, 2015.

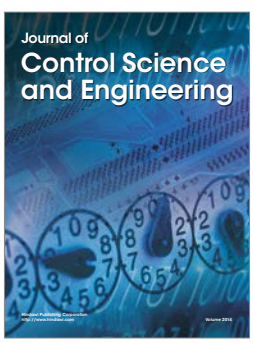
[36] K. Grover, A. Lim, S. Lee, and Q. Yang, "Privacy-enabled probabilistic verification in broadcast authentication for vehicular networks," *Ad-Hoc and Sensor Wireless Networks*, vol. 32, no. 3-4, pp. 239–274, 2016.

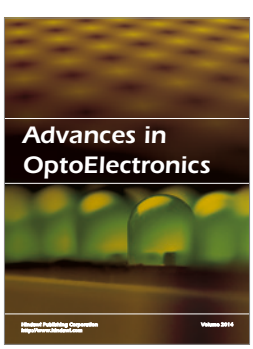
[37] N. Eagle and A. Pentland, "Reality mining: sensing complex social systems," *Personal and Ubiquitous Computing*, vol. 10, no. 4, pp. 255–268, 2006.

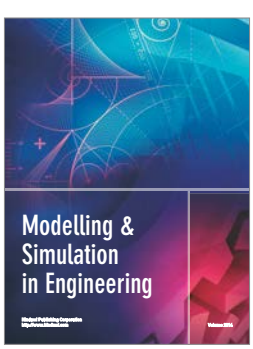


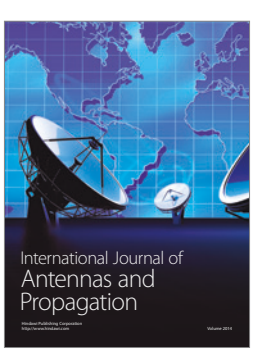


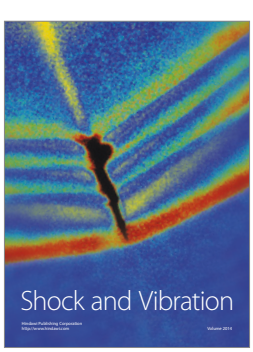


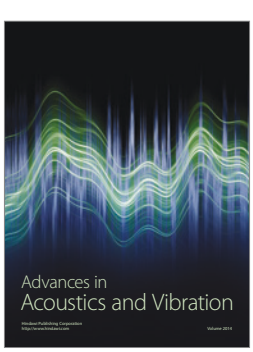












Hindawi

Submit your manuscripts at
https://www.hindawi.com