Towards Long-Term Quality-of-Protection in Mobile Networks: A Context-Aware Perspective

Wei Wang, Student Member, IEEE, Qian Zhang, Fellow, IEEE

Abstract

Sensor-equipped smartphones as well as wearable devices have undoubtedly become the predominant source of user-generated data in mobile networks. The proliferation of user-generated data has created a plethora of opportunities for personalized services based on the states of the users and their surrounding environments. Those personalized services, although improving users’ perceived quality-of-experience (QoE), have raised severe privacy concerns, as most of these applications aggressively collect users’ personal data without providing clear statements on the usage and disclosure policies of such sensitive information. In order to sustain personalized services with long-term privacy preservation, disruptive paradigms are required. We envision that context awareness is a key pillar to providing long-term Quality-of-Protection (QoP) for individual privacy. In particular, users transit between different contexts, including mobility modes and social activities, and these contexts are temporally or logically correlated, which can be leveraged by adversaries to compromise users’ privacy. In addition, users may have different QoP preferences in different contexts. With these salient features in mind, this article investigates context-aware QoP mechanism designs for personalized services in mobile networks. We discuss possible attacks and propose corresponding countermeasures. In particular, we develop a QoP framework that exploits context awareness to achieve better tradeoffs between service quality and privacy protection in long-term services. Finally, we provide some implications for future context-aware QoP mechanism designs by conducting a case study on smartphone traces.

W. Wang is with the Fok Ying Tung Research Institute, Hong Kong University of Science and Technology. E-mail: gswwang@cse.ust.hk.

Q. Zhang is with the Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Hong Kong. E-mail: qianzh@cse.ust.hk.

May 28, 2015 DRAFT
Index Terms

Quality-of-Protection (QoP), Context Awareness, Privacy

I. INTRODUCTION

Due to continuing advances in low-power sensors and actuators, we are entering an era of rapid expansion in wearable and pervasive computing. Nowadays, it is common for people to carry smart devices, such as sensor-equipped smartphones, smart watches, and healthcare devices. Those devices periodically measure users’ physical activities (such as Fitbit) and physiology (such as Samsung Simband with electrocardiogram (ECG) and photoplethysmogram (PPG) sensors embedded), making it possible to continuously track users’ contexts including mobility modes, social activities, and health conditions [1]. These devices are normally capable of wireless connectivity, which enables them to upload their sensor readings to their service providers. Examples of these service providers include healthcare providers that own electronic medical record systems for user’s data collection and analysis, activity trackers that keep track of users’ trajectories or mobility modes, and environment-based applications that offer personalized services based on the operating conditions of users and their surrounding environments.

On the one hand, these services provided along with smart devices truly reap the benefits of context awareness to improve user’s Quality-of-Experience (QoE) [2]. QoE is a holistic evaluation that measures a user’s experiences with a service focusing on the entire service experience [3]. The context-aware applications optimizes QoE by providing services tailored to users’ contexts. On the other hand, these services require users to continuously upload their personal data, which has imposed wide privacy concerns. Such privacy threats come from the fact that many service providers aggressively collect users’ data without providing clear statements about how to use the data and whom the data will be shared with. A recent survey [4] studied thirty popular Android applications that have access to user’s location, camera, microphone data, and found that fifteen of them sent users’ information to remote advertisement or analytics servers. Being aware of such risks, users may be reluctant to upload their data to service providers. However, this would also prohibit personalized services based on users’ data. Consequently, novel privacy preservation paradigms for personalized services in mobile networks are imperative to allow users to enjoy these services with guaranteed Quality-of-Protection (QoP).
Mobile user and service provider communicate via mobile networks.

Mobile user encounters a series of different contexts over time.

Service provider offers personalized services based on users’ sensing data.

Fig. 1. Long-term services and QoP in mobile networks.

As smart devices continuously collect users’ data, long-term QoP is envisioned to guarantee users’ privacy against untrusted service providers. Long-term data collection and service delivery typically follow the architecture presented in Fig 1, where users transit between different contexts, and continuously upload their sensor data to service providers, who extract users’ contexts and offer context-related services. To provide long-term protection, several salient features about user’s QoP preferences and behaviors need to be considered. First, users are sensitive to certain contexts and are unwilling to disclose them to the service providers. Take the context transition scenario in Fig. 1 as an example. The user is sensitive about the contexts “in meeting” and “in hospital”, while considering “walking” and “in café” to be insensitive. Second, a user’s private preferences vary across different contexts. For example, a user may consider the call logs during meetings as private information, while being sensitive to location information when she is at home. In addition, as users’ behaviors follow certain patterns, contexts are temporally correlated. The connections between contexts can be exploited by adversaries to launch new attacks.

The goal of this article is to first investigate each of the aforementioned features of long-term QoP for personalized services, and call attention to a clean-slate redesign of QoP frameworks for long-term personalized services. We start at a deep dive at what potential threats arise with these long-term services, and then propose corresponding countermeasures. Specifically, we shed light on two types of adversaries – active adversaries who continuously launch real-time attacks and passive adversaries who passively collect users’ data. Different attacking strategies are discussed.
Fig. 2. Context acquisition from sensor data.

We review the pros and cons of existing QoP mechanisms under these long-term attacks, and propose a context-aware QoP framework that exploits the features of long-term services. We also take a case study on real context traces of 94 smartphone users. Merits of the context-aware framework are verified, and implications about future QoP mechanism designs for long-term services are provided.

II. EXPLOITING CONTEXT AWARENESS IN MOBILE NETWORKS: LONG-TERM SERVICES AND PRIVACY THREATS

A. Long-Term Services in Mobile Networks

The contexts of a user refer to her state or surrounding environmental conditions [1]. A user encounters a series of contexts and transits between various contexts. The gamut of contexts is a set of interrelated user states and environmental conditions that are logically and temporally correlated. As illustrated in Fig. 1, the mobile user usually goes to a café (the second context in Fig. 1) after leaving home (the first context in Fig. 1) – this temporal correlation is the user’s habit that can be learned from the user’s historical data. The mobile user walks on the street (the third context in Fig. 1) after having coffee in the café – in this case the context “walking” has logical relation with the context “in café” as a user needs to walk out of the café after having coffee.

A range of sensors equipped on smart devices have been effectively used to infer user’s contexts including location from GPS or wireless signals (Wi-Fi or cellular signals), mobility
modes from accelerometer, and social activities from a combination of multiple sensors. Fig. 2 summarizes how sensors equipped on commercial off-the-shelf smart devices are used to extract different types of contexts. Consequently, a large bundle of personalized services are emerging by utilizing user-generated sensor data. Fig. 1 illustrates how personalized services are delivered to users in mobile networks. Smart devices carried by a mobile user continuously acquire readings from equipped sensors, and periodically upload the data to corresponding service providers via the user’s smartphone. After receiving the user-generated sensor data, service providers perform a series of actions to process the data, including context acquisition, pre-processing, storing, and reasoning within the application boundaries [1]. The target of these actions is to “understand” the current state of the user and environmental conditions, based on which service providers deliver personalized services to the user. These personalized services can be categorized into:

- **Passive services.** This type of services passively tracks the user’s activities or conditions for offline analysis. Examples are healthcare monitoring and activity tracking applications such as Apple Healthbook and FitBit.

- **Active services.** This type of services recognizes the user’s contexts in real time and autonomously take actions based on the current context. For example, an iOS application named *AutoSilent* automatically mutes the phone when the user is in a meeting.

### B. Privacy Threats in Context-Aware Perspective

#### QoP preferences in long-term services.** Nowadays, location privacy has already been widely concerned by both users and governments, as location traces can be used to infer many individual behaviors and preferences that user do not want to disclose [5]. Context information, which provides more profound implications about users’ behaviors and preferences, enlarges the range of sensitive information exposed to service providers. Today, many service providers aggressively collect much more personal data than what is required to support their functionalities. Users, on the other hand, are sensitive about some contexts and corresponding sensor data whose disclosure is undesired by the users. Users’ QoP preferences can be categorized into the following two levels.

- **Context sensitivity.** A user normally has a sensitivity preference on contexts. First, a user is sensitive to a subset of contexts. As illustrated in Fig. 1, a user might not want the contexts “in meeting” and “in hospital” to be learned by service providers but she is willing
to disclose the contexts “in café” and “walking”. In addition, a user has different levels of sensitivities. For example, a user is normally more reluctant to disclose the context “in hospital” than disclosing the context “in meeting”.

- **Data sensitivity.** The sensitivity level of sensor data is context-specific. For instance, a user may consider the call logs during meetings as private information, while the user is sensitive to location information when she is at home.

**Context-aware adversaries and attacking strategies.** Adversaries include untrusted service providers and other entities who are interested in user’s private information and acquire users' data by trading with the service providers. Context awareness is a double-edged sword – it enables service providers to offer personalized services that substantially improve user’s perceived experience, while it also adds extra shots for adversaries to compromise users’ privacy. Different from “single-shot” scenarios as considered in many location and participatory sensing privacy preservation studies, contexts are temporally and logically correlated [6]. Adversaries that are aware of this unique feature can compromise the user’s privacy by exploiting such correlations [7]. Specifically, adversaries can infer the presence of sensitive contexts based on their previous observations on non-sensitive contexts. Context-aware adversaries can be categorized into passive and active adversaries, as elaborated below.

- **Passive adversary.** The adversary aggressively collects the user’s sensing data for offline analysis. The target of passive attack is to obtain the user’s personal information and preferences, which are considered to be of great value.

- **Active adversary.** The adversary launches real-time attacks based on the user’s instant context information. The adversary may push unwanted advertisements or context-based spams/scams to the user, or even make the user a victim of blackmail or physical violence. Different from the passive attack, the active attack has instant impact on the user and thus is observable to the user.

Both types of adversaries can launch the following two types of attacks.

- **Static attack.** The static attack follows pre-defined rules to acquire user’s data and infer user’s private information. The adversary may adopt certain models, such as *Bayesian Networks* (BN) or *Hidden Markov Model* (HMM), to make inference about user’s sensitive information. The major limitation of the static attack is that it largely depends on adversary’s
Fig. 3. Illustration of context dynamics. The arrows refer to the possible context transitions.

background knowledge, and cannot update strategies according to the user’s actions.

- **Adaptive attack.** The adaptive attack evolves over time by continuously tuning attacking strategies or parameters. The adaptation is made based on new observations during the long-term services. The adversary learns more accurate patterns of the user based on newly acquired data released by the user, and adjusts its attacking strategies accordingly.

We illustrate how these two attacks work by using a concrete example of a user’s possible daily activities as illustrated in Fig. 3. An advertiser may want to push context-related ads through smartphone apps to the user. The advertiser cannot directly infer the context at 9 a.m. as the related data is suppressed, while after observing the user’s non-sensitive context “driving” at 10 a.m., the advertiser can infer backward that the user’s context at 9 a.m. is “in hospital”. Then, the advertiser links the sensor data at 9 a.m. to the context “in hospital”, and use this observation to update the inference model.

C. Achieving QoP: Models and Challenges

Plenty of QoP mechanisms have been proposed to thwart privacy breach in sharing user-generated content, including user’s location information in location-based services (LBSs) [8], [9], and sensor data in participatory sensing [10], [11]. These proposals employ several QoP models that can be categorized into the following three classes.

- **Anonymization and generalization models.** The crux of anonymization and generalization models is to replace a specific value of user’s attributes (such as sensor data) with a coarser range to provide privacy protection. In LBS systems, the location coordinates of
multiple users are transformed into the same geographical region to hide a user’s specific location spot. In participatory sensing systems, sensor data from multiple users can also be protected by releasing data of coarser granularity. The main advantage of anonymization and generalization models is that the generalized range truly includes the real value and thus there is no falsified data. Whereas, their limitations are also obvious. These techniques normally require a trusted third party to serve as the anonymizer. Moreover, these techniques work well under certain assumptions about adversaries’ prior knowledge, while the users’ privacy can be jeopardized in the presence of stronger adversaries, such as adversaries knowing the anonymization algorithm.

- **Randomization-based models.** Randomization-based models include random perturbation and differential privacy techniques. Random perturbation transforms the original data by replacing a subset of the data points with randomly selected values, which can be generated by adding random noise to the original data or executing randomized algorithms. Differential privacy is achieved by injecting randomness into each user’s data so that aggregated results are insensitive to the change of any single user’s data. A major appealing feature of differential privacy is that it makes the worst case guarantee, that is, even if the adversaries know the data of all the individuals except the target individual, the adversaries are still uncertain about the data of the target individual. Though randomization-based models can defend a wider range of attacks than anonymization and generalization models, the loss of data quality and truthfulness caused by randomness is the main shortcoming.

- **Cryptographic techniques.** End-to-end encryptions are widely adopted to secure the transmissions of users’ reported data. These approaches mainly protect users’ privacy from external adversaries (such as eavesdroppers), while internal adversaries (such as compromised or untrusted service providers) can decrypt the ciphertext to obtain users’ data. Secure multi-party computation (SMC) allows multiple peers to jointly compute a function over their inputs without revealing inputs to each other. However, SMC approaches are normally limited to specific computations and incurs substantial communication or computation overhead.

**Challenges in long-term services.** Existing studies leverage the above techniques to enable privacy preservation in LBSs or participatory sensing applications. Most of them are designed
for protecting data privacy in single-shot scenarios [10], [12], or focus on static attacks [13], while the user’s two-level sensitivities and the context-aware adversaries have not yet been fully explored. To effectively preserve users’ privacy in long-term services, we require a disruptive framework that considers temporal correlations and user dynamics. In particular, the framework should be carefully designed to guarantee user’s QoP requirements (context sensitivity and data sensitivity) against context-aware adversaries (passive and active adversaries) and different attacking strategies (static and adaptive attacks). To this end, we propose a QoP framework for long-term services and shed light on how context awareness is exploited to satisfy user’s QoP requirements under different types of attacks.

III. CONTEXT-AWARE QoP FOR LONG-TERM SERVICES

A. Framework Overview

We consider a scenario where the user’s smart devices (such as smart watch, FitBit) connect to her smartphone, which uploads sensor data to the untrusted service providers. Fig. 4 outlines the work flow of the proposed context-aware QoP framework, which is implemented as a middleware [14] on the smartphone to control the data flow from sensors to upper layer applications and their service providers.

The crux of the framework is adaptively controlling “what” (data transformation) to release and “when” (data release control) to release data to service providers based on the user’s profile (behavior model and QoP preferences) and feedback (perceived QoE and context information) from service providers or the user.

Concretely, this framework employs three functional primitives as follows.

- Sensitivity evaluation. The function of sensitivity evaluation is to estimate the amount of privacy loss incurred by releasing current data. Sensitivity evaluation consists of two aspects: direct privacy leakage of current context and indirect privacy leakage of future contexts. The privacy leakage of current context can be estimated based on the user’s QoP preference profile, which specifies the user’s sensitivity of each context (context sensitivity) and data attribute (data sensitivity). Furthermore, to take into account the impact of releasing current data on future private contexts and other private information, the sensitivity evaluation considers the temporal and logical correlations among contexts, which can be learned from
the user’s behavior model or historical traces. As such, the framework jointly considers the context correlations and the user’s QoP preferences via the sensitivity evaluation component.

- **Adaptive data transformation.** The adaptive data transformation component converts the raw sensor readings to private forms by employing certain privacy preservation techniques such as anonymization or randomization-based approaches. To combat context-aware adversaries, the data transformation strategies are tuned based on the updated sensitivity evaluation as well as service quality and context information feedback from the user and service providers. In particular, when adversaries are active, attacking information (such as context-related spams/scams delivered by adversaries) is also included as an additional dimension of feedback for data transformation adaptation.

- **Data release control.** The data release control component decides when to release the transformed data to service providers based on the results of adaptive data transformation. The aim of the data release control component is to find an optimal releasing period that preserves most of the perceived service quality given the QoP requirements.
In the following two sections, we shed light on how to implement the above framework to thwart attacks from active and passive adversaries, respectively.

**B. Providing QoP Against Active Adversary**

Note that the static attack can be considered to be a special case of adaptive attacks (by fixing the adversaries’ choices). Without loss of generality, we focus on the adaptive attacks where adversaries intelligently adjust their strategies based on their new observations on the user’s data. To combat against adaptive attacks, the framework observes the feedback from service providers (context information), the user (perceived service quality), or even the adversaries (such as spams/scams delivered by active adversaries), and updates the releasing granularity accordingly.

We first model the user behavior based on historical data or public profiles. Previous work has shown that a user’s transitions between various contexts can be captured by a Markov chain. The sensitivity of each context is estimated by jointly considering its own sensitivity specified by the user and its correlations with sensitive contexts. The amount of privacy loss of releasing a context can be quantified by the weighted sum of the difference of adversaries’ prior and post inference probabilities on each sensitive context, where the weight of each context is its sensitivity specified by the user.

Since the user’s context is considered to keep changing over time and both the user and the adversary make different actions at different times, the interactions between the user and the adversary is in a stochastic setting and can be formulated as a competitive MDP, which is a dynamic game with probabilistic transitions played in a sequence of stages, where each player receives a stage payoff based on players actions and current system state and attempts to maximize its expected sum of discounted payoffs. The user’s context is included in the system state as the user’s action depends on its observation of contexts. In the case of active adversaries, previous attack results should also be included in the system state. As the adversary’s strategy is not known by the user, the user can only conjecture the adversary’s strategy from previous attack results, which are assumed to be observable to the user as online attacks have instant impact on the user. After observing the system state at each stage, both the user and the adversary decide their actions for the current stage. The user controls the granularity of the released sensing data to protect its context privacy while preserving the quality of context-based services.
the other hand, the adversary selects a proper subset of sensing data to infer user’s contexts. The payoff function of the user is defined to be the quality of the context-based service with weighted penalty on privacy loss. The user’s utility is to the expected sum of discounted stage payoffs. Optimal strategies can be obtained at the Nash Equilibrium (NE) point of the game. As such, at each stage, the QoP framework observes the current system state and performs data transformation according to the optimal strategies.

C. Providing QoP Against Passive Adversary

Different from active adversaries, attacks by passive adversaries are not observable to the user. Thus, we adopt differential privacy as the data transformation technique, which makes no assumptions about adversaries’ background knowledge. As discussed earlier, the user’s data sensitivity varies over time due to context transitions. In our framework, we jointly adjust the releasing period and data transformation to minimize the overall data quality loss incurred by data transformation.

The core idea is to use the data quality loss to guide the releasing period selection and enforce differential privacy by injecting randomness into each selection. As we consider passive services, the sensor data does not require to be uploaded in real time. We assume that the maximal delay for data uploading is $T$ time slots, and the data within the $T$ time slots can be divided into $K$ consecutive periods for data transformation and uploading. To this end, the framework first lists all possible periods, and computes the minimal data quality loss in each period that is required to guarantee the user’s privacy requirements. Then, the framework randomly selects each period to ensure differential privacy. In order to guarantee differential privacy, the framework conforms to the exponential mechanism by selecting periods of less quality loss with exponentially greater probabilities to ensure differential privacy. Interested readers can refer [15] for detailed information about the exponential mechanism.

D. Enabling QoP middleware on Smartphones

The proposed QoP framework serves as a middleware that sanitizes raw sensor data and releases the sanitized data to upper layer applications. We can leverage the sandbox mechanism in today’s smartphone platforms, such as Android and iOS, to implement such a middleware. A sandbox typically provides a tightly controlled set of resources for applications to run in
and prohibits these applications from accessing resources outside the sandbox. Therefore, we can build a sandbox to confine all untrusted applications, and use the sandbox to sanitize raw sensor data according to certain privacy preserving mechanisms before providing the data to applications.

IV. Case Study on Smartphone Traces

In this section, we evaluate the proposed framework using the Reality Mining dataset\(^1\), which was collected by the MIT Media Laboratory from September 2004 to June 2005. The Reality Mining dataset records the continuous activities of 94 students and staff at MIT equipped with Nokia 6600 smartphones, which are pre-installed with several pieces of software that collects data about call logs, Bluetooth devices in proximity of approximately five meters, location at granularity of cell tower, application usage, transportation model (including driving, walking, stationary), and so on. Based on the traces, we train a Markov chain for each user. For each user, a certain percentage of contexts are selected as sensitive contexts.

A. Performance Under Active Adversaries

**Baselines.** We compare our framework with fixed and single-shot mechanisms. The fixed mechanism draws an action that uniformly sets the granularity of each sensor. And the single-shot mechanism adopts the optimal policy that maximizes the user's current payoff without considering future impact and the temporal correlations among contexts.

The average sums of discounted payoff of all users are reported in Fig. 5. The gap between the sums of discounted payoff obtained by adopting the proposed and single-shot approaches to zero when the percentage of sensitive contexts goes down. This observation can provide some guidance for the context privacy preserving schemes that for the users with a small faction of sensitive contexts, the impact of current actions on the future payoff can be neglected so as to design more efficient algorithm.

B. Performance Under Passive Adversaries

**Baselines.** In the case of passive adversaries, the fixed mechanism treats each context as a period, and the single-shot mechanism directly applies data transformation to the whole \(T\) time

\(^1\)http://realitycommons.media.mit.edu/realitymining.html
Fig. 5. Sum of discounted payoff vs. percentage of sensitive context.

Fig. 6. Relative error vs. number of time slots.

To measure the data quality loss, we use relative error, which is defined by sum of differences between raw data and released data divided by number of time slots. Fig. 6 shows that the relative error decreases when number of time slots increases. The error of the proposed framework stays lower than those of the other mechanisms in all cases, which implies the merits of the adaptive period control in the proposed framework.
V. CONCLUDING REMARKS

This article has envisioned the crucial role of context awareness in achieving QoP in long-term services. Instead of focusing on protecting the user-generated sensor data in each context independently, context awareness exploits the correlations among different contexts and facilitates adaptive QoP paradigms. Through careful investigation of possible adversaries in long-term services, we have presented a context-aware framework and demonstrated its merits using real smartphone traces. Some observations in the smartphone case study can provide some implications for future designs of context-aware QoP provisions.

ACKNOWLEDGEMENT

The research was support in part by grants from 973 project 2013CB329006, China NSFC under Grant 61173156, RGC under the contracts CERG 622613, 16212714, HKUST6/CRF/12R, and M-HKUST609/13, as well as the grant from Huawei-HKUST joint lab.

REFERENCES


Wei Wang (S’10) is currently a Research Assistant Professor in Fok Ying Tung Research Institute, Hong Kong University of Science and Technology (HKUST). He received his Ph.D. degree in Department of Computer Science and Engineering from HKUST, where his Ph.D. advisor is Prof. Qian Zhang. Before he joined HKUST, he received his bachelor degree in Electronics and Information Engineering from Huazhong University of Science and Technology, Hubei, China, in June 2010. His research interests include PHY/MAC in Wi-Fi networks, privacy and fault management in wireless networks.

Qian Zhang (M’00-SM’04-F’12) joined Hong Kong University of Science and Technology in Sept. 2005 where she is a full Professor in the Department of Computer Science and Engineering. Before that, she was in Microsoft Research Asia, Beijing, from July 1999, where she was the research manager of the Wireless and Networking Group. Dr. Zhang has published about 300 refereed papers in international leading journals and key conferences in the areas of wireless/Internet multimedia networking, wireless communications and networking, wireless sensor networks, and overlay networking. She is a Fellow of IEEE for “contribution to the mobility and spectrum management of wireless networks and mobile communications”. Dr. Zhang has received MIT TR100 (MIT Technology Review) worlds top young innovator award. She also received the Best Asia Pacific (AP) Young Researcher Award elected by IEEE Communication Society in year 2004. Her current research is on cognitive and cooperative networks, dynamic spectrum access and management, as well as wireless sensor networks. Dr. Zhang received the B.S., M.S., and Ph.D. degrees from Wuhan University, China, in 1994, 1996, and 1999, respectively, all in computer science.