The Accuracy-Privacy Trade-off of Mobile Crowdsensing

Mohammad Abu Alsheikh, Yutao Jiao, Dusit Niyato, Ping Wang, Derek Leong, and Zhu Han

The authors address the contradicting incentives of privacy preservation by crowdsensing users and accuracy maximization and collection of true data by service providers. They define the individual contributions of crowdsensing users based on the accuracy in data analytics achieved by the service provider from buying their data. They propose a truthful mechanism for achieving high service accuracy while protecting privacy based on user preferences.

ABSTRACT

Mobile crowdsensing has emerged as an efficient sensing paradigm that combines the crowd intelligence and the sensing power of mobile devices, such as mobile phones and Internet of Things gadgets. This article addresses the contradicting incentives of privacy preservation by crowdsensing users, and accuracy maximization and collection of true data by service providers. We first define the individual contributions of crowdsensing users based on the accuracy in data analytics achieved by the service provider from buying their data. We then propose a truthful mechanism for achieving high service accuracy while protecting privacy based on user preferences. The users are incentivized to provide true data by being paid based on their individual contribution to the overall service accuracy. Moreover, we propose a coalition strategy that allows users to cooperate in providing their data under one identity, increasing their anonymity privacy protection, and sharing the resulting payoff. Finally, we outline important open research directions in mobile and people-centric crowdsensing.

INTRODUCTION

The proliferation of mobile devices with built-in sensors has made mobile crowdsensing an efficient sensing paradigm, especially in people-centric and Internet of Things (IoT) services. Crowdsensing users collect sensing data using their personal mobile devices (e.g., mobile phones and IoT gadgets). However, the development of crowdsensing services is impeded by many challenges, especially criticism of the privacy protection of crowdsensing users. Service providers require true data, which is a key factor in optimizing data originated services [1]. This introduces contradicting incentives of maximizing the privacy protection of users and the prediction accuracy of service providers. Most of the existing incentive models in the literature are monetary motivated with sole profit maximization objective (e.g., [2–4]), while the privacy incentive of users is neglected. Therefore, conventional monetary-based incentive models are inapplicable in privacy preserving crowdsensing systems, and new privacy-aware incentive models are required. Several major questions related to developing privacy-aware incentive models in mobile crowdsensing arise. First, how does the crowdsensing service define the contributions and payoff allocations of users with varying privacy levels? Second, do crowdsensing coalitions change the attained privacy of the cooperative users? Third, how do cooperative users divide the coalition payoff among themselves?

This article provides answers to the aforementioned questions by presenting a novel incentive framework for privacy preservation and accuracy maximization in mobile crowdsensing. Sensing users select their preferred data anonymization levels without knowing the privacy preferences of other users. The data anonymization is inversely proportional to the accuracy of data analytics of the service provider. Accordingly, users are paid based on their marginal contributions to service accuracy. Users can be also penalized with a negative payoff if they cause marginal harm to the service accuracy (e.g., an outlier providing misleading data). Moreover, a set of k cooperative users can jointly work by forming a crowdsensing coalition, increasing the anonymity privacy protection measured by the k-anonymity metric. The total coalition payoff is then divided among the cooperative users based on their marginal contributions to the coalition's data quality. Our experiments on a real-world dataset of a crowdsensing activity recognition system show that the payoff allocation of a particular user does not directly depend on the contributed data size but on the data quality. Likewise, the payoff allocation is found to decrease as the privacy level increases.

The rest of this article is organized as follows. We first present an overview of mobile crowdsensing in people-centric and IoT services, and review some related incentive mechanisms. Next, we discuss the privacy preservation in mobile crowdsensing. Then we propose an incentive framework for privacy preservation and accuracy maximization in crowdsensing services. After that, we present numerical experiments based on a real-world crowdsensing dataset. Finally, we outline some interesting research directions and conclude the article.

MOBILE CROWDSENSING AND IoT

This section first gives an overview of mobile crowdsensing in IoT and then reviews some monetary incentive mechanisms in mobile crowdsensing.
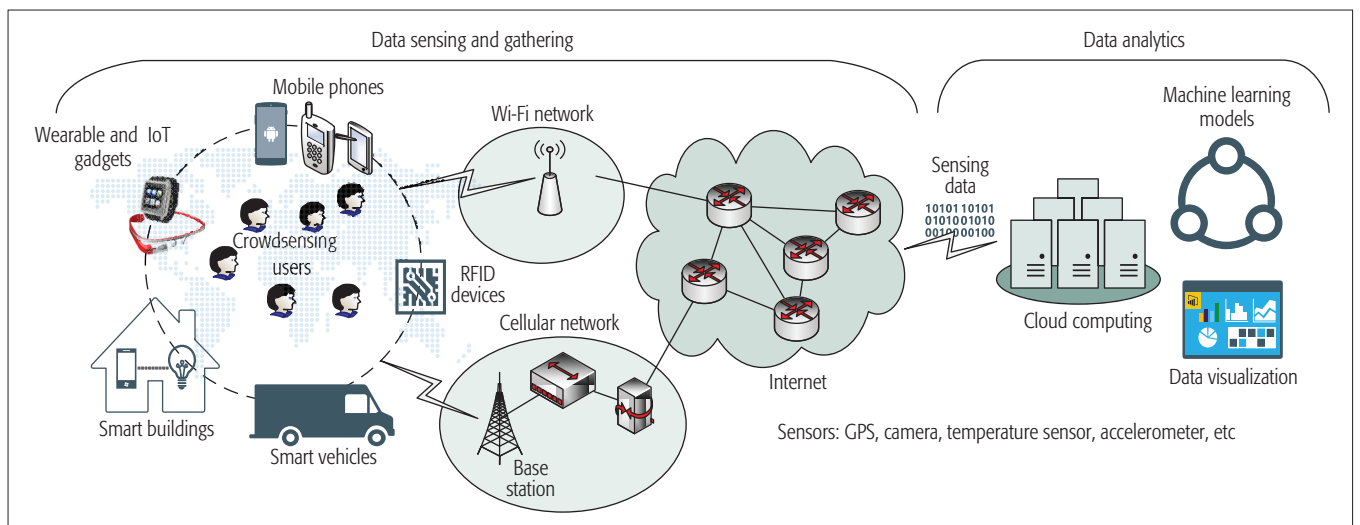
Mohammad Abu Alsheikh, Yutao Jiao, Dusit Niyato, and Ping Wang are with Nanyang Technological University;
Derek Leong is with the Institute for Infocomm Research; Zhu Han is with the University of Houston.

Figure 1. System model of mobile crowdsensing.

ARCHITECTURES AND DATA MANAGEMENT

In mobile crowdsensing, mobile devices and human intelligence are jointly adopted for collecting sensing data regardless of geographic separation among users and service providers. As shown in Fig. 1, the design of mobile crowdsensing services includes the following stages.

Data Sensing and Gathering: Crowdsensing users sense and collect data using mobile devices including phones, wearable devices, and in-vehicle sensing devices. Users can also annotate the sensory data with subjective observations and reports such as their emotions and surrounding events. The data is sent to the cloud server through various types of networks including cellular and Wi-Fi networks.

Data Analytics: After receiving the raw data from the users, cloud computing can be used to store and process the large-scale data. Data analytics (e.g., machine learning methods) are typically applied to extract useful information and make effective predictions. Services also support data visualization, generate reports, and provide platforms to share the outcomes with other collaborative entities, such as social networking services.

APPLICATIONS

Mobile crowdsensing has become an efficient sensing paradigm in people-centric and IoT services. People-centric services contain sensing, computing, and communication components that aim to assist human life. The following are some pertinent crowdsensing applications.

Traffic Monitoring: Mobile Millennium[1] is a traffic crowdsensing service. Millennium collects geolocation data from taxi drivers. It also assimilates other data obtained in real time from radars, loop detectors and historical databases. The traffic information can be accessed by drivers for accurate real-time traffic conditions (e.g., traffic congestion points).

Wi-Fi Sharing: WiFi-Scout[2] is a crowdsensing service for sharing reviews and connection quality of Wi-Fi hotspots. Users can easily search for free and paid Wi-Fi hotspots covering the locations that they will be visiting. Users are also rewarded based on their compliance and review quality.

Healthcare: PatientsLikeMe[3] is a healthcare crowdsensing service that collects health data from patients. The patients provide their experience on medication, supplements, or devices. PatientsLikeMe also sells the collected data to pharmaceutical companies in order to improve and develop effective medication and healthcare equipment.

MONETARY INCENTIVE MODELS

Mobile crowdsensing should incorporate efficient incentive mechanisms to attract and retain enough crowdsensing users. In [5], the authors compared the resulting data quality and user compliance of three incentive schemes. The *uniform* scheme pays a user at a fixed rate of 4 cents per completed task. The *variable* scheme selects the payoff in the range of 2 to 12 cents based on the required task and user performance. Finally, the *hidden* scheme includes a lottery factor in defining the payoff values where the users are not informed of the expected payoff before completing the task. The study showed that the variable scheme reduces the total cost by 50 percent compared to the uniform scheme for the same completion rate and performance. The hidden scheme is found to be the least effective incentive scheme.

Next, we review monetary incentive mechanisms for mobile crowdsensing with an emphasis on reverse auction mechanisms [6] as they fit well and are commonly applied for mobile crowdsensing with multiple users. As shown in Fig. 2, a typical reverse auction framework occurs between the crowdsensing users and service. The crowdsensing users compete among themselves to perform the sensing task. The service provider first announces the description of the crowdsensing tasks to potential mobile users. Users are rational entities and will set their bids based on the cost of the crowdsensing task. In order to maximize the utility of the crowdsensing service, the auction system determines the task assignment and payoff of each user including both selected and rejected bids. For example, the crowdsensing tasks are assigned to the winning users with the lowest bids to perform the crowdsensing tasks and submit the data to the service. The service provider will provide the agreed payoff to the winning users. Table 1 provides a summary of the monetary incentive

Model	Main entities	Payoff scheme	Maximization objective	Solution properties
Bayesian auction [7]	Multiple risk-neutral users	Threshold winner payoff	The target tracking accuracy	Bayesian Nash equilibrium (profitable and individual rational)
Sealed-bid auction [2]	Fixed budget with risk-neutral users	Threshold winner payoff	The service utility (more user and less payoff)	Profitable and individual rational
Stackelberg competition [2]	A leader (service) and followers (users)	Threshold winner payoff	The service utility	Nash equilibrium (profitable and individual rational)
Vickrey auction [3]	Multiple risk-neutral users	Contribution-dependent payoff	Data integrity	Profitable and truthful
All-pay auction [4]	Risk-averse and risk-neutral users	All-pay contribution-dependent payoff	The service utility	Nash equilibrium (profitable and individual rational)

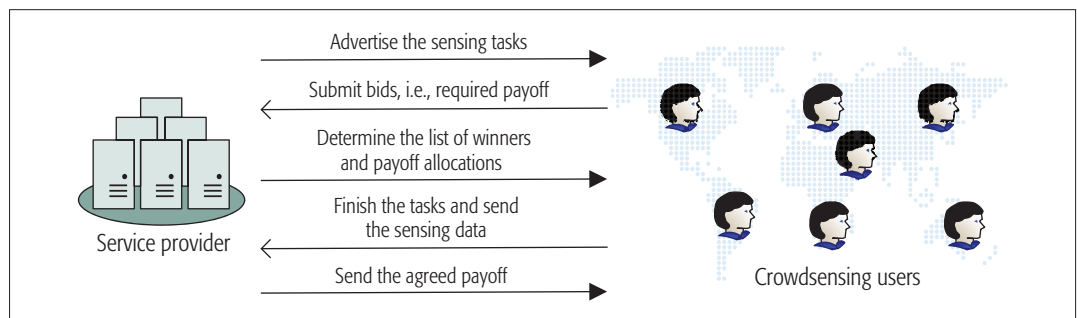Table 1. Summary of the monetary incentive models in mobile crowdsensing.

Figure 2. Crowdsensing incentive mechanism as a reverse auction.

models reviewed in this section. From the table, "risk-neutral" means that the user is unaware of the loss of its payoff, for example, when choosing between guaranteed $5 and conditioned $10 payoffs. A "profitable" solution guarantees a nonnegative utility for the service provider. An "individual rational" solution guarantees a nonnegative utility for each user. A "truthful" solution guarantees that users cannot increase their payoff by submitting misleading bids for the crowdsensing task. Therefore, a truthful incentive mechanism provides a dominant strategy for rational users in bidding their true cost of performing the crowdsensing task.

We divide the incentive schemes into two main categories of *threshold winner* and *contribution-dependent* payoffs.

Threshold Winner Payoff: In this payoff scheme, only the winning users will be paid for performing the sensing task, and there is no payoff allocation for rejected users. For example, the authors in [7] presented a Bayesian reverse auction model for target tracking with crowdsourcing, assuming that the value estimate of a user can be drawn from a continuous probability distribution. The residual energy of the mobile devices has an impact on the prior distribution of the user bids. The objective of this model is maximizing the total target tracking utility of the service by solving the multiple-choice knapsack problem. Likewise, the authors in [2] proposed two complementary payoff scenarios of *user-centric* and *platform-centric* schemes. In the user-centric scheme, the service defines the payoff using a reverse auction by following the steps shown in Fig. 2. In the platform-centric scheme, the crowdsensing problem is formulated as a Stackelberg game. The Nash equilibrium is solved using backward induction and found to be unique. A major limitation of [2, 7] is assuming a known prior distribution of user bids. In the real world, users can collude and submit misleading bids to increase their own payoff. This problem is solved in contribution-dependent payoff schemes, as discussed next.

Contribution-Dependent Payoff: A practical incentive mechanism requires all participants to be truthful. One principal way to achieve truthful user interaction is by choosing an appropriate pricing scheme where the payoff allocations of participants are not solely defined by their bids. The authors in [3] applied the Vickrey-Clarke-Groves (VCG) reverse auction with the objective of minimizing the sum of payoff values to crowdsensing users. A user is paid based on the difference between the sum of costs with and without that particular user. Reporting truthful bids is a weakly-dominant strategy in the VCG auction. The authors in [4] modeled the mobile crowdsensing problem as an all-pay auction where the crowdsensing users are not required to submit their bids at the beginning of the auction. Instead, the payoff is calculated based on the user contributions after completing the sensing tasks. The users with the highest contribution receive a payoff, while the rest of the users do not receive any payoff allocation.

PRIVACY PRESERVATION IN MOBILE CROWDSENSING

Even though most of the existing works in the literature focus on monetary incentive models to achieve the maximum possible payoff allocation, privacy preservation is still a top priority for

crowdsensing users. In this section, we first discuss the data anonymization properties that can be used to measure the privacy protection. Then we discuss the challenges of privacy preservation in mobile crowdsensing.

Privacy Properties and Data Anonymization

Mobile crowdsensing comes with challenging privacy issues. In particular, crowdsensing users are typically concerned that their personal information can be leaked from the collected data. Personal information of users can be categorized into three main classes:

- *Explicit identifiers* are the data attributes that directly reveal the user identity (e.g., full name and social security number).
- *Non-explicit identifiers* can be combined with background knowledge to reveal the user identity (e.g., zip code and birth date).
- *Sensitive attributes* can be utilized to extract private information about the user (e.g., real-time activity tracking using accelerometer data) [8].

Explicit identifiers should be completely removed before trading the crowdsensing data among businesses. To protect the non-explicit identifiers and sensitive attributes, data anonymization methods can be applied to sensing data.

Privacy is defined by the information gain of an adversary. The following syntactic privacy properties can be used to define of the privacy protection requirements.

k-anonymity [9]: This property is developed to guarantee that a data sample of a particular user in public datasets cannot be re-identified by potential intruders. Specifically, for a crowdsensing service to possess the *k*-anonymity property, each user should not be distinguishable from at least $k - 1$ other users. For example, a user should be unidentifiable by combining the available gender and birth date crowdsensing data. This can be achieved by transformation techniques, such as identity generalization and suppression, to reduce the granularity of the data. For example, the birth dates can be replaced by date ranges instead of the exact values.

l-diversity [10]: The *k*-anonymity does not work well if the sensitive data attributes lack diversity. For example, if a few users of a healthcare crowdsensing service used a particular zip code and are infected by a disease, background knowledge can be used to reveal the health privacy of a user which is known by the adversary to use that zip code. In order to avoid this privacy threat, the *l*-diversity property requires that each equivalence class has at least *l* "well-represented" values. An equivalence class is a set of data samples with the same anonymized data attributes.

t-closeness [11]: The *t*-closeness property requires the distribution of sensitive values within each equivalence class to be "close" to their distribution in the entire original dataset. *t*-closeness is an extension of the *l*-diversity model as it takes the distribution of sensitive values into account. *t*-closeness can be achieved by adding random noise to sensitive data attributes. For example, adding Gaussian noise to accelerometer data can restrict the tracking of particular activities.

Challenges of Privacy Preservation in Mobile Crowdsensing

The authors in [12] reviewed the privacy threats and protection methods during the task management in mobile crowdsensing. A taxonomy of privacy methods was provided including pseudonyms, connection anonymization, and spatial cloaking. The authors also highlighted the challenging process of defining the user contribution in incentive-based task assignment. The authors in [13] discussed the privacy and data integrity of mobile crowdsensing. The privacy is observed to be user-dependent.

Achieving the syntactic privacy properties can reduce the accuracy of data analytics algorithms. Applying strict data anonymization to all users results in poor accuracy of the data analytics. Instead, users can be given the choice of setting their preferable data anonymization level such that reliable users receive high payoff allocation. The trade-off between privacy preservation and accuracy maximization should also be taken into consideration, which is the main objective of the next section.

Incentive Mechanism for Privacy Preserving Crowdsensing

In this section, we introduce a privacy preserving incentive framework for mobile crowdsensing where participating users can protect their private data by data anonymization. The level of data protection will accordingly be used to set the resulting payoff allocation such that the users have an incentive to provide their true data. We first present the system model and major entities. Then we discuss the proposed incentive framework, which is intended to maximize the accuracy of data analytics while preserving the privacy of crowdsensing users.

System Model

As shown in Fig. 3, the crowdsensing system under consideration consists of the following three main entities.

Crowdsensing **users** are the participants who collect sensing data using their personal mobile devices (e.g., mobile phones and IoT gadgets). The contribution of a particular user to the crowdsensing community is defined based on the quality of the sensing data from the data analytics perspective. A user with positive contribution to the sensing process is considered *pivotal*. Users can apply data anonymization (e.g., adding noise to the sensing data) to protect their privacy and personally identifying information. Additionally, crowdsensing coalitions can be built as an efficient scheme for achieving *k*-anonymity protection, where *k* is the number of cooperative users in the coalition.

A **service provider** buys data from the crowdsensing users through a mediator, applies data analytics, and delivers a service to a set of customers. The provider makes a profit by charging the customers a subscription fee.

A **mediator** is the auction management entity that controls the exchange of data between the crowdsensing users and the service provider. Moreover, the mediator divides the payoff

Applying a strict data anonymization to all users results in poor accuracy of the data analytics. Instead, the users can be given the choice of setting their preferable data anonymization level such that reliable users receive high payoff allocation. The trade-off between privacy preservation and accuracy maximization should also be taken into consideration.
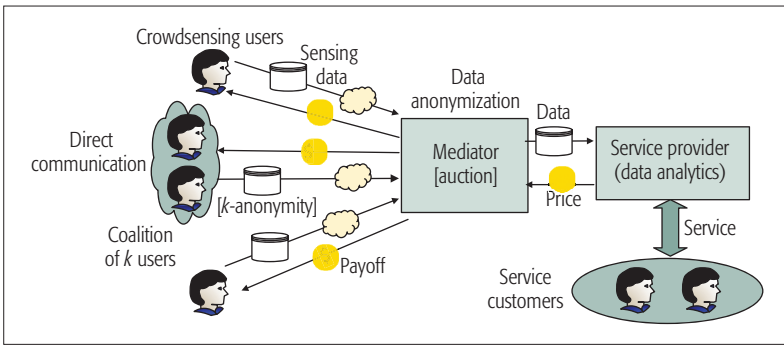
Figure 3. System model of the privacy preserving crowdsensing framework supporting both data anonymization and identity generalization through crowdsensing coalition formulation. Cooperative users are connected using device-to-device (D2D) communication.

received from the service provider among the crowdsensing users based on their contributions to the crowdsensing system.

We next discuss the privacy preserving model through which the crowdsensing users can sell data to the service provider and receive a payoff according to their individual contributions, as illustrated in Fig. 3. First, we define the individual contributions and resulting payoffs of the users from the data analytics perspective. Second, we develop a privacy preserving mechanism that gives the users an incentive for contributing their true data with the lowest possible data anonymization level. Third, we consider the case where users can form a crowdsensing coalition for identity generalization, and we present a fair payoff allocation among cooperative users.

DATA ANALYTICS

Crowdsensing data $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{L}$ usually includes tuples of sensing feature set $\mathbf{x}_i \in \mathbb{R}^M$ and a class label $y_i \in \mathbb{R}$, where L is the number of data tuples and M is the number of data attributes. The feature set $\mathbf{x}_i$ includes the sensing data, such as images in vision services and geographic coordinates in transport services. The class label y_i contains human input and is only available in supervised data analytics (e.g., specifying accident events in transport services). After collecting sufficient data, the service provider applies data analytics methods such as deep learning [14] to build data originated services. For example, transport services can provide accurate prediction of vehicle arrival times and road congestion. We denote the accuracy function of the data analytics model trained using dataset $\mathcal{D}$ as $f(\mathcal{D})$. $f(\mathcal{D})$ measures the performance of the service in providing accurate prediction of the ground truth.

INCENTIVE MECHANISM DESIGN

We consider a set of N users who are connected to a privacy preserving crowdsensing service. Each user n generates true sensing data $\mathcal{D}_n$ and selects a data anonymization level p_n. The data anonymization can be performed by adding random noise to the true data $\mathbf{x}_i$ subject to p_n, for example, Gaussian noise $\mathcal{N}(0, p_n \mathbf{I}_M)$ with zero mean and a variance of p_n, where $\mathbf{I}_M$ is the identity matrix of size M. Each user submits its anonymized data $\tilde{\mathcal{D}}_n$ and data anonymization level p_n to the crowdsensing mediator, without knowing the preferences of other users. The full anonymized dataset

$$\tilde{\mathcal{D}}n = \bigcup_{1 \le n \le N} \tilde{\mathcal{D}}_n$$

and data anonymization preferences $\mathcal{P} = p_1, ..., p_n\}$ are collected by the mediator from all users. According to the VCG auction [6], the mediator calculates the payoff of user n as follows:

$$F_n = f(\tilde{\mathcal{D}}_n) - f(\tilde{\mathcal{D}}_{-n}), \tag{1}$$

where $\tilde{\mathcal{D}}_{-n}$ is the anonymized data after excluding the data of user n. The following three cases for the payoff function exist:

• If $F_n > 0$, the user will receive a positive payoff allocation of F_n as its data contribution increases the accuracy. These users are called pivotal.
• If $F_n = 0$, the user does not change the crowdsensing choice or the service accuracy. Such users receive zero payoff and can be advised to decrease their data anonymization level.
• If $F_n < 0$, the user has a negative contribution (e.g., excessive data anonymization) and will accordingly be penalized with a negative payoff. The data collected from such users should not be used in the data analytics.

Sending true data to the service provider is a weakly dominant strategy under the VCG rules regardless of the data anonymization levels of the other users.

CROWDSENSING COALITION

A set of k users can cooperate to form a crowdsensing coalition, denoted by $\mathcal{K}$, which increases the privacy level by providing the data of the cooperative users under one generalization identity and achieving k-anonymity privacy protection. Those k users must be connected using D2D communication without traversing the service provider. The generalization identity guarantees that a data sample cannot be used to identify its source from the k cooperative users. $\mathcal{K}$ is a virtual alliance of users who work collectively and are seen as one sensing entity by the service provider. Specifically, the service provider cannot identify the source of data samples as a particular data sample can relate to any of the k cooperative users. The payoff of the coalition is

$$F_{\mathcal{K}} = f(\tilde{\mathcal{D}}) - f(\tilde{\mathcal{D}}_{-\mathcal{K}}), \tag{2}$$

where $\tilde{\mathcal{D}}_{-\mathcal{K}}$ is the anonymized data after excluding the data from all users in the coalition $\mathcal{K}$. Solution concepts from cooperative game theory, such as the Shapley value and Nash bargaining solution [15], can be applied to share the resulting payoffs among the cooperative users in the coalition $\mathcal{K}$. From the Shapley value, the payoff allocation (i.e., monetary payment), of each user is defined based on its contribution to the coalition.

NUMERICAL RESULTS

In this section, we present numerical experiments to evaluate the performance of the proposed privacy preserving framework.

SYSTEM SETUP

In this section, we use a real-world dataset [8] of a crowdsensing activity recognition system of six activities including walking, jogging, upstairs, downstairs, sitting, and standing. The dataset

includes L = 1,098,207 samples of accelerometer data that were collected by N = 36 users. The mobile devices sampled at a rate of 20 Hz resulted in M = 120 data features of framed 3-axial acceleration. We assume that the service provider uses deep learning [14] to develop the prediction service. The service provider buys the crowdsensing data from the users through the auction mediator and sells an activity tracking service to customers.

We assume that users 2 and 3 protect their sensitive activities by adding varied levels of Gaussian noise $\mathcal{N}(0, p_n\mathbf{I}_M)$ to the acceleration data. Accordingly, users 2 and 3 acquire the t-closeness property, where t is equal to the variance of the added noise p_n. The payoff of each user is defined based on the payoff rule in Eq. 1. Moreover, users 2 and 3 can collaborate in the crowdsensing coalition $\mathcal{K}$ to acquire the k-anonymity protection, where k = 2 for two cooperative users. The coalition's total payoff is defined based on the payoff rule in Eq. 2, while the payoff sharing among users 2 and 3 is defined according to the Shapley value.

USER CONTRIBUTIONS AND PIVOTAL USERS

Figure 4 shows the contributed data rates from each user and the resulting service accuracy $f(\cdot)$ by training a deep learning model on the data of each user separately. Two key results can be noted. First, the data rate varies among different users. However, there is no correlation between the service accuracy from the data analytics perspective and the contributed data rate from the sensing perspective. The service accuracy depends on the quality of the used mobile device, the user's performance during task execution, and data annotation. For example, user 1 contributes more data than user 2, while the accuracy resulting from the data of user 1 is lower than that of user 2. Second, users 3 and 6 are pivotal, and they score the highest standalone accuracy values of 68.3 and 68.1 percent, respectively. The standalone accuracy for the rest of the users is less than 64 percent. The pivotal users are important to the service provider to ensure high service accuracy.

THE IMPACT OF PRIVACY ON ACCURACY

In Fig. 5a, we consider the impact of the data anonymization level on the accuracy of the crowdsensing service. Several important results are observed. First, there is an inverse relationship between the prediction accuracy and the data anonymization level. The maximum service accuracy of $f(\mathcal{D})$ = 92.5 percent is achieved when all users provide true data samples without any anonymization. This maximum value decreases as user 3 increases the level of data anonymization. A high level of data anonymization can be required by the users to protect their privacy. Second, the service provider has an incentive to reject users with high data anonymization levels. For example, the service will reject user 3 when its data anonymization level is greater than 8, which is labeled "critical point 1" in Fig. 5a. This is due to the resulting harm to the overall system accuracy. Third, the prediction accuracy decreases as more users adopt the data anonymization scheme. For example, the accuracy is negatively affected when both users 2 and 3 apply the data anonymization compared to the case of user 3 only. According-
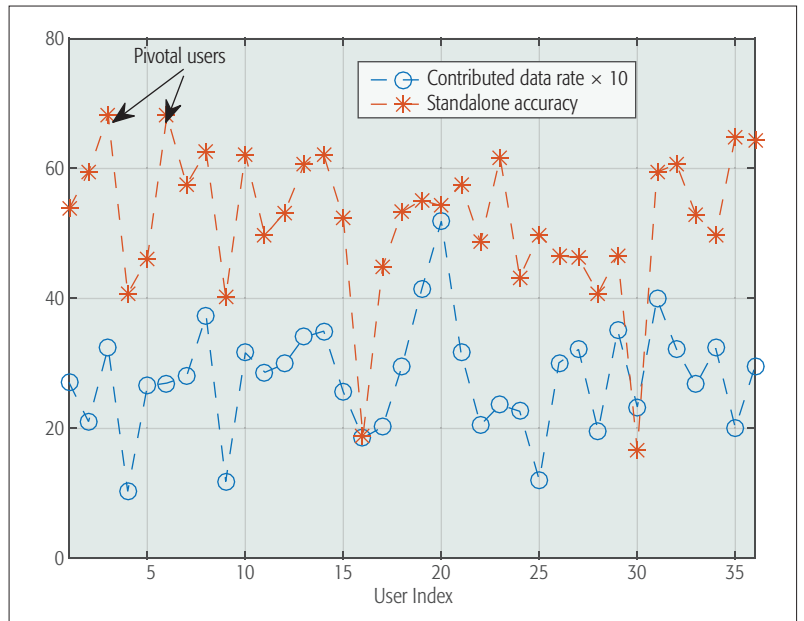

Figure 4. User contribution to the crowdsensing service.

ly, the crowdsensing system has an incentive for reducing the number of users applying the data anonymization scheme. As presented next, this can be achieved by increasing the payoff allocation of users who provide their true data.

PAYOFF ALLOCATION

Figure 5b shows the payoff allocation of users 2 and 3 under the varied data anonymization levels. First, the payoff allocation of any user decreases as its data anonymization level increases. For a high data anonymization level which is equal to or greater than the over anonymization levels specified in Fig. 5b, users will be penalized by receiving negative payoff. Second, pivotal users receive a higher payoff compared to normal and low-performing users; for example, the payoff of user 3 is greater than that of user 2. For the crowdsensing coalition case, the payoff allocation to the cooperative users is found using the Shapley value, which reflects the individual contribution of each user. The cooperative users receive not only the same payoff in both the crowdsensing coalition and the standalone cases, but also a higher level of the k-anonymity privacy protection.

FUTURE DIRECTIONS

Based on the proposed incentive framework, the following open research directions can be further pursued.

COOPERATION AND COMPETITION AMONG SERVICE PROVIDERS

To collect high-quality data, service providers may cooperate or compete with each other to attract and retain crowdsensing users. With cooperation, service providers collude to set payoff strategies that maximize their profit as a cooperative coalition. In the competitive scenario, service providers can apply non-cooperative game and Nash equilibrium solutions for the service's subscription fee and crowdsensing data's prices. The strategic interaction among providers can also benefit the users in making higher revenues.
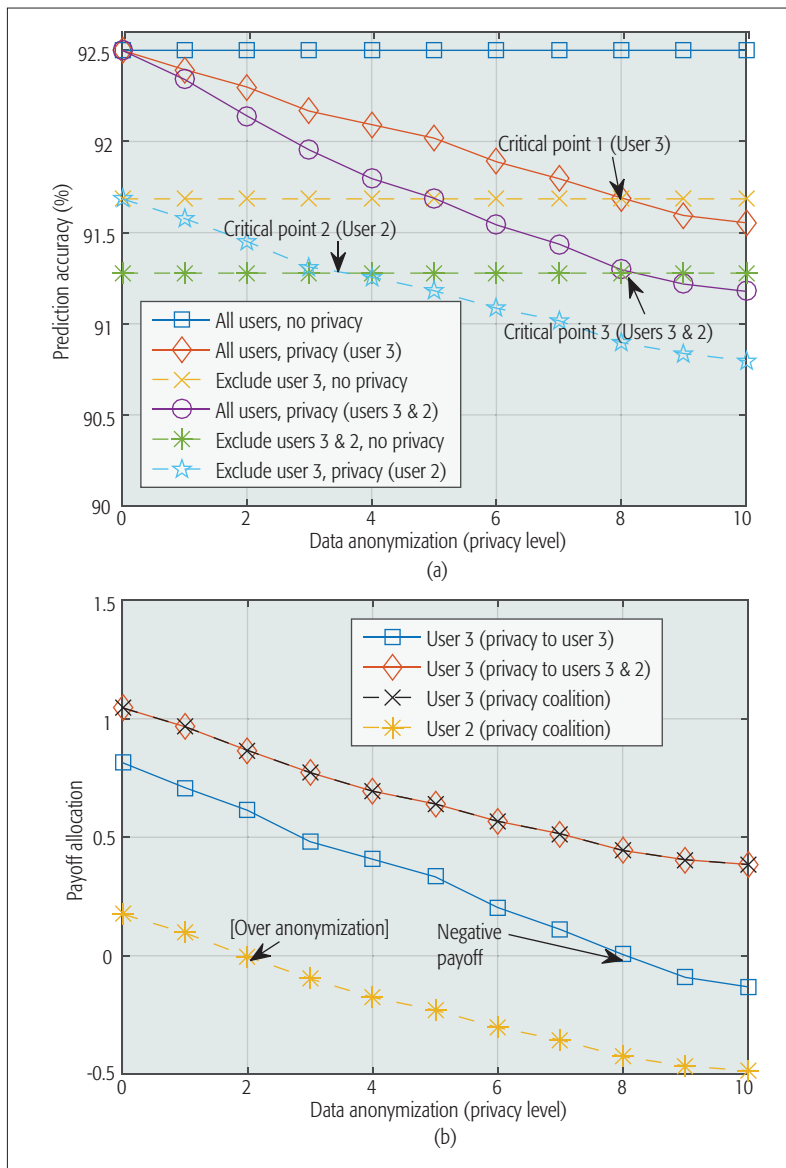
Figure 5. Performance of the proposed privacy preserving framework under varied privacy levels: a) the resulting accuracy of the deep learning service trained on the crowdsensing data; b) the payoff allocation of Users 2 and 3. The privacy level is equal to the variance of the added Gaussian noise.

INCENTIVE MECHANISM DESIGN FOR FOG COMPUTING

Analyzing crowdsensing data can be computationally expensive. Fog computing provides a solution by allowing partial data processing at the mobile devices owned by users. In such a design, users are paid not only for sensing data, but also for available computing power. Incentive mechanisms are required to attract large contributions from users as fog nodes. Likewise, mobile devices come with varying hardware resources; methods for defining the user contributions in fog computing are also required.

DYNAMIC AND HETEROGENEOUS CROWDSENSING

Crowdsensing users can be heterogeneous in term of the sensing precision and technical experience. Thus, the service provider has an incentive of attracting powerful users by increasing their payoff allocations, and the incentive mechanism has to optimize these payoff values. Additionally,

users asynchronously join and leave the crowdsensing system. Stochastic optimization methods (e.g., Markov decision processes) can be formulated to determine the optimal payoff rates over time, for example, to attract users during the off-peak times.

CONCLUSION

Privacy awareness has the potential to significantly boost the performance of mobile crowdsensing, attracting more sensing users, and enabling the protection of privileged information. This article has presented an incentive mechanism for privacy preservation and accuracy maximization in mobile crowdsensing. It has been shown that the coalition strategy can be used by users to send their data under one generalized identity, increase the *k*-anonymity privacy protection, and share the resulting payoffs among cooperative users based on their individual sensing contribution. The proposed incentive framework has been evaluated using a real-world crowdsensing dataset. Finally, open research directions have been presented.

REFERENCES

[1] J. Vaidya, C. W. Clifton, and Y. M. Zhu, "Privacy Preserving Data Mining," *Springer Science & Business Media*, vol. 19, 2006.
[2] D. Yang et al., "Crowdsourcing to Smartphones: Incentive Mechanism Design for Mobile Phone Sensing," *Proc. 18th ACM Annual Int'l. Conf. Mobile Computing and Networking*, 2012, pp. 173–84.
[3] J. Xu, J. Xiang, and D. Yang, "Incentive Mechanisms for Time Window Dependent Tasks in Mobile Crowdsensing," *IEEE Trans. Wireless Commun.*, vol. 14, Nov. 2015, pp. 6353–64.
[4] T. Luo et al., "Incentive Mechanism Design for Crowdsourcing: An All-Pay Auction approach," *ACM Trans. Intelligent Systems and Technology*, vol. 7, Apr. 2016, pp. 35:1–35:26.
[5] M. Musthag et al., "Exploring Micro-Incentive Strategies for Participant Compensation in Highburden Studies," *Proc. 13th ACM Int'l. Conf. Ubiquitous Computing*, 2011, pp. 435–44.
[6] P. Klemperer, *Auctions: Theory and Practice*, Princeton Univ. Press, 2004.
[7] N. Cao, S. Brahma, and P. K. Varshney, "Target Tracking via Crowdsourcing: A Mechanism Design Approach," *IEEE Trans. Signal Processing*, vol. 63, Feb. 2015, pp. 1464–76.
[8] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Activity Recognition Using Cell Phone Accelerometers," *ACM SigKDD Explorations Newsletter*, vol. 12, Mar. 2011, pp. 74–82.
[9] L. Sweeney, "k-anonymity: A Model for Protecting Privacy," *Int'l. J. Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, Oct. 2002, pp. 557–70.
[10] A. Machanavajjhala et al., "l-diversity: Privacy beyond k-anonymity," *ACM Trans. Knowledge Discovery from Data*, vol. 1, Mar. 2007, pp. 1–52.
[11] N. Li, T. Li, and S. Venkatasubramanian, "t-closeness: Privacy beyond k-anonymity and l-diversity," *Proc. 23rd IEEE Int'l. ConfF Data Engineering*, 2007, pp. 106–15.
[12] L. Pournajaf et al., "Participant Privacy in MobileCrowd Sensing Task Management: A Survey of Methods and Challenges," *ACM SIGMOD Record*, vol. 44, Dec. 2016, pp. 23–34.
[13] R. K. Ganti, F. Ye, and H. Lei, "Mobile Crowdsensing: Current State and Future Challenges.," *IEEE Commun. Mag.*, vol. 49, no. 11, Nov. 2011, pp. 32–39.
[14] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," *Nature*, vol. 521, May 2015, pp. 436–44.
[15] B. Peleg and P. Sudhôlter, "Introduction to the Theory of Cooperative Games," *Springer Science & Business Media*, vol. 34. 2007.

Biographies

Mohammad Abu Alsheikh [S'14] (mohammad027@e.ntu.edu.sg) received his B.Eng. degree in computer systems engineering from Birzeit University, Palestine, in 2011. Between 2010 and 2012, he was a software engineer working on developing robust web services, Ajax-based web components, and smartphone applications. He is currently a Ph.D. candidate in the School of Computer Science and Engineering, Nanyang Technological University, Singapore. His research interests include machine learning in big data analytics, mobile sensing technologies, and sensor-based activity recognition.

Yutao Jiao (yjiao001@ntu.edu.sg) is currently a Ph.D. student in the School of Computer Science and Engineering, Nanyang Technological University. He received a B.S. degree from the College of Communications Engineering, Nanjing, China, in 2013. His research interests include resource management in the Internet of Things and economics of big data.

Dusit Niyato [M'09, SM'15, F'17] (dniyato@ntu.edu.sg) is currently an associate professor in the School of Computer Science and Engineering, Nanyang Technological University. He received his B.Eng. from King Mongkuts Institute of Technology Ladkrabang (KMITL), Thailand, in 1999 and his Ph.D. in electrical and computer engineering from the University of Manitoba, Canada, in 2008. His research interests are in the area of energy harvesting for wireless communication, the Internet of Things, and sensor networks.

Ping Wang [M'08, SM'15] (wangping@ntu.edu.sg) received her Ph.D. degree in electrical engineering from the University of Waterloo, Canada, in 2008. Currently she is an associate professor in the School of Computer Science and Engineering, Nanyang Technological University. Her current research interests include resource allocation in multimedia wireless networks, cloud computing, and smart grid. She was a corecipient of the Best Paper Award from the IEEE Wireless Communications and Networking Conference 2012 and IEEE ICC 2007.

Derek Leong (dleong@i2r.a-star.edu.sg) received his B.S. degree in electrical and computer engineering from Carnegie Mellon University in 2005, and his M.S. and Ph.D. degrees in electrical engineering from the California Institute of Technology in 2008 and 2013, respectively. He is a scientist with the Smart Energy and Environment cluster at the Institute for Infocomm Research (I^{2}R), A*STAR, Singapore. His research and development interests include distributed systems, sensor networks, smart cities, and the Internet of Things.

Zhu Han [S'01, M'04, SM'09, F'14] (zhan2@uh.edu) received his B.S. degree in electronic engineering from Tsinghua University in 1997, and his M.S. and Ph.D. degrees in electrical engineering from the University of Maryland, College Park, in 1999 and 2003, respectively. From 2000 to 2002, he was an R&D engineer of JDSU, Germantown, Maryland. From 2003 to 2006, he was a research associate at the University of Maryland. From 2006 to 2008, he was an assistant professor at Boise State University, Idaho. Currently, he is a professor in the Electrical and Computer Engineering Department as well as the Computer Science Department at the University of Houston, Texas. His research interests include wireless resource allocation and management, wireless communications and networking, game theory, wireless multimedia, security, and smart grid communication. He received an NSF Career Award in 2010, the Fred W. Ellersick Prize of the IEEE Communication Society in 2011, the EURASIP Best Paper Award for the *Journal on Advances in Signal Processing* in 2015, several best paper awards at IEEE conferences, and is currently an IEEE Communications Society Distinguished Lecturer.