A Secure Mobile Crowdsensing Game With Deep Reinforcement Learning

Liang Xiao, *Senior Member, IEEE*, Yanda Li, Guoan Han, *Student Member, IEEE*, Huaiyu Dai, *Fellow, IEEE*, and H. Vincent Poor, *Fellow, IEEE*

***Abstract*— Mobile crowdsensing (MCS) is vulnerable to faked sensing attacks, as selfish smartphone users sometimes provide faked sensing results to the MCS server to save their sensing costs and avoid privacy leakage. In this paper, the interactions between an MCS server and a number of smartphone users are formulated as a Stackelberg game, in which the server as the leader first determines and broadcasts its payment policy for each sensing accuracy. Each user as a follower chooses the sensing effort and thus the sensing accuracy afterward to receive the payment based on the payment policy and the sensing accuracy estimated by the server. The Stackelberg equilibria of the secure MCS game are presented, disclosing conditions to motivate accurate sensing. Without knowing the smartphone sensing models in a dynamic version of the MCS game, an MCS system can apply deep Q-network (DQN), which is a deep reinforcement learning technique combining reinforcement learning and deep learning techniques, to derive the optimal MCS policy against faked sensing attacks. The DQN-based MCS system uses a deep convolutional neural network to accelerate the learning process with a high-dimensional state space and action set, and thus improve the MCS performance against selfish users. Simulation results show that the proposed MCS system stimulates high-quality sensing services and suppresses faked sensing attacks, compared with a Q-learning-based MCS system.**

***Index Terms*— Mobile crowdsensing, game theory, deep reinforcement learning, faked sensing attacks, deep Q-networks.**

I. Introduction

With the ubiquity of mobile devices such as smartphones and tablets that are equipped with multiple sensors including accelerometers and global positioning systems, mobile crowdsensing (MCS) provides location-based services to improve social welfare. In particular, an MCS platform or server recruits mobile users to monitor the surrounding features and offers crowdsensing applications in environmental, network and traffic monitoring [1]–[3]. With the rapid development of smart & programmable radio devices, a user obtains powerful control over his or her radio device as well as the embedded sensors, and thus accurately decides the sensing effort or quality. As a selfish smartphone user chooses his or her sensing effort to maximize the individual payoff, a mobile crowdsensing system has to stimulate users to contribute accurate sensing reports and suppress faked sensing attacks. Otherwise, if knowing that cheating in an MCS duty incurs no punishment, selfish smartphone users are motivated to launch faked sensing attacks, in which low-quality or even faked sensing reports are sent to an MCS server with a mobile application such as Fake GPS in [4] to save sensing effort and avoid the possible leak of user privacy.

Game theory is an important means of formulating the MCS process and provides approaches such as auctions, pricing and reputation-based mechanisms to motivate users to contribute to MCS services [5]–[12]. For instance, an auction-based MCS solution proposed in [5] pays the user who bids the lowest price in the auction to save costs. We note that the utility of an MCS server not only relies on the payment to the serving users, but also depends on their locations, sensing efforts, and sensors' qualities. Therefore, an MCS server can improve its sensing performance by evaluating the sensing quality and only recruiting the smartphones that provide accurate reports [13]–[15]. However, to the best of our knowledge, the game theoretic study of MCS against faked sensing attacks is still an open problem.

In this paper, we formulate the interactions between a number of self-interested smartphone users and an MCS server as a secure mobile crowdsensing game to address faked sensing attacks. In this game, the server applies a classification algorithm, such as the blind image quality assessment algorithm in [16], or a majority voting algorithm in [17] to evaluate the accuracy of each sensing report and thus estimate the sensing effort of the corresponding user. Each user is paid according to his or her sensing accuracy and cheating users are punished with zero payment to suppress the incentives of faked sensing attacks. The goal of this work is to suppress the attack motivation of selfish users who aim to improve their own utilities instead of malicious attackers.

As shown in Fig. 1, the server as the leader of a Stackelberg game first determines and broadcasts its payment policy, and

Manuscript received March 2, 2017; revised June 4, 2017 and July 12, 2017; accepted July 13, 2017. Date of publication August 9, 2017; date of current version November 20, 2017. This work was supported in part by the National Natural Science Foundation of China under Grant 61671396, in part by the U.S. National Science Foundation under Grant ECCS-1307949, Grant EARS-1444009, and Grant CMMI-1435778, and in part by the U.S. Army Research Office under Grant W911NF-17-1-0087 and Grant W911NF-16-1-0448. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Guofei Gu. *(Corresponding author: Liang Xiao.)*

L. Xiao is with the School of Data and Computer Science, Sun Yat-sen University, Guangdong Key Laboratory of Big Data Analysis and Processing, Guangzhou 510006, China (e-mail: xiaoliang3@mail.sysu.edu.cn).

Y. Li and G. Han are with the Department of Communication Engineering, Xiamen University, Xiamen 361005, China (e-mail: lxiao@xmu.edu.cn).

H. Dai is with the Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC 27616 USA (e-mail: huaiyu_dai@ncsu.edu).

H. V. Poor is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: poor@princeton.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIFS.2017.2737968

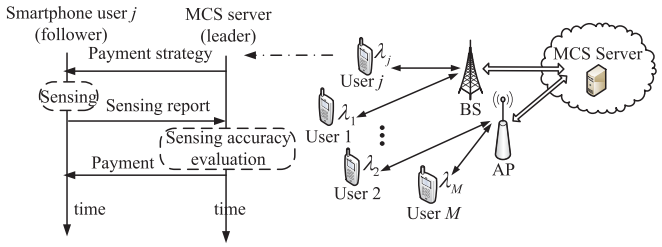


Fig. 1. Illustration of a mobile crowdsensing system, where λ_j represents the sensing importance of user j in the MCS application.

each user as a follower chooses his or her sensing effort afterwards. The Stackelberg equilibria (SEs) of the secure MCS game are derived and the conditions under which each SE exists are provided to reveal the impact of the sensing costs, the contribution factor of each user in the MCS application, the density of active smartphone users, and the sensing evaluation accuracy of the server. A higher payment stimulates more smartphone users to contribute to the mobile crowdsensing application and suppresses faked sensing attacks. However, the over-payment leads to over-sensing and sometimes results in network congestion, thus decreasing the utility of the server. Therefore, the MCS server has to provide a tradeoff between the sensing cost and the sensing accuracy.

The payment of the MCS server provides motivation for the sensing efforts of the mobile users in the future and thus the payment process can be formulated as a finite Markov decision process (MDP). In dynamic environments where the sensing models of active smartphones are not readily available, an MCS server can apply Q-learning, a model-free reinforcement learning method, which can achieve the optimal policy via trials in the finite Markov decision process, to derive the optimal payment policy without being aware of the sensing model. More specifically, the MCS server with Q-learning decides the payment strategy based on the observed state of the previous sensing report quality and payment policy and a quality function or Q-function that describes the discounted long-term reward for each state-action pair.

However, the Q-learning-based MCS system has a slow learning rate in the presence of a large state space due to the curse of high dimensionality. This problem can be addressed by the deep Q-network (DQN), a deep reinforcement learning technique developed by Google DeepMind in 2014 for video games [18]. More specifically, we propose a DQN-based MCS payment strategy that exploits a deep convolutional neural network (CNN) to compress the learning state space and estimate the Q-value for each payment value. This MCS payment scheme combines deep learning and Q-learning to accelerate the learning rate to obtain an optimal payment policy and thus enhance the sensing performance against faked sensing attacks.

The main contributions of this paper are summarized below:

(1) We formulate a secure MCS game and derive its SEs, disclosing conditions that allow an MCS system to avoid both under-sensing and over-sensing and suppress faked sensing attacks.

(2) We propose a DQN-based MCS strategy that does not require knowledge of the sensing model or the active user density in the dynamic secure MCS game. This MCS system facilitates the learning process to achieve an optimal payment policy, and thus improves the sensing performance against faked sensing attacks compared with the Q-learning based strategy.

The remainder of the paper is organized as follows. We review related work in Section II, and present the MCS system model in Section III. We formulate the secure MCS game and present the SE of the game in Section IV. We design a DQN-based MCS scheme for the dynamic MCS game in Section V. We provide simulation results in Section VI and conclude in Section VII.

II. Related Work

An auction-based MCS system proposed in [5] motivates smartphone users to participate in the MCS tasks at the Stackelberg equilibrium. The location-aware auction developed in [7] stimulates users to submit bids that reflect the actual costs of performing sensing tasks and join MCS applications. The auction-based incentive mechanism designed in [8] motivates users to declare actual costs to reduce the compensation cost of the server. An all-pay auction based MCS system developed in [9] motivates risk-averse users to participate without knowing the marginal cost of the users. The online auction with budget constraints proposed in [10] applies the greedy task allocation strategy to achieve high energy efficiency with good fairness among users who arrive sequentially and randomly. However, it is challenging for the MCS server to accurately estimate the sensing models of smartphone users and avoid cheating attacks.

Reputation and trust are used to stimulate users to participate in MCS in [11]. The social norm system designed in [12] integrates the payment and reputation mechanism into crowdsoucing websites. The crowdsensing platform developed in [19] applies a polynomial-time greedy algorithm in the sensing task allocation to achieve both fairness and energy efficiency among participatory smartphones. The credit-based MCS scheme proposed in [20] provides both single-report and multiple-report sensing tasks and protects user privacy. The data aggregation system designed in [21] analyzes empirical data and builds a reputation mechanism to protect user privacy and suppress false data attacks.

Finally, payment and reward are also widely used to stimulate mobile crowdsensing. For instance, a coalition game based MCS setting formulated in [22] provides payment to stimulate users and improves social welfare. A reward based MCS system proposed in [23] motivates smartphone users to collaborate in both data acquisition and distributed computing. In addition, incentive and punishment mechanisms can suppress faked sensing attacks in spectrum sensing. For example, the spectrum sharing and punishment strategy proposed in [24] can suppress cooperative sensing attacks by secondary users. The spectrum sensing incentive scheme proposed in [25] adjusts a penalty factor for access collision to reduce the missed detection rate. A mobile sensing server that applies data mining and learning algorithms to evaluate fake sensing reports can suppress the motivation to cheat [26].

However, because of evaluation error, it is still challenging for the server to stimulate accurate reports without

TABLE I
SUMMARY OF SYMBOLS AND NOTATION

Notation	Definition
M	Number of active smartphone users
$x_j^{(k)}$	Sensing effort/accuracy of user j at time k
$\mathbf{y}^{(k)}$	Payment policy of the server at time k
L	Number of evaluation accuracy levels
p_j	User mobility index of user j
$C_j^{(i)}$	Sensing cost of user j with sensing accuracy level i
$G^{(i)}$	System gain from a report of accuracy i
λ_j	Contribution coefficient of user j
$\mathbf{P}$	Feasible payment policy set
$u_{j/s}$	Utility of user j/server
$\hat{N}_i^{(k)}$	Estimated number of reports of accuracy level i at time k
$\mathbf{s}^{(k)}$	System state at time k
$\boldsymbol{\varphi}^{(k)}$	State sequence
θ	CNN weights
B	Size of the minibatch for the CNN
W	Experience size in the CNN input sequence

knowing the sensing model of users. We have proposed a Q-learning based MCS system to suppress faked sensing attacks in [27], where we have studied the equilibria of a secure MCS game. Compared with our previous work, the new contributions of this paper include: (1) we provide the SEs for a generic case with multiple smartphone users and multiple sensing accuracy levels, instead of the special case with a single user and one sensing accuracy level as in [27]; and (2) we propose a DQN-based MCS system to improve the performance of the Q-learning based MCS system in [27].

III. System Model

An MCS server aims to recruit some of the M smartphone users located in an area of interest to gather sensing data and thus establish an MCS application, as shown in Fig. 1. The MCS server first chooses its payment policy and broadcasts a recruiting message that lists the payment for each sensing accuracy level. According to the MCS policy, each selfish and rational smartphone user then chooses the sensing effort, such as the sensing time and power, which in turn determines the sensing accuracy.

For simplicity, let x_j denote both the actual sensing accuracy level and the sensing effort of smartphone j. The sensing accuracy is quantized into $L+1$ levels, with $x_j \in A = \{-1, 0, \cdots, L\}$, where $x_j = 0$ represents faked sensing and $x_j = L$ corresponds to the most accurate sensing. The sensing effort of user j increases with x_j; and if $x_j = -1$, the user does not participate in the MCS application. In a mobile network, users sometimes leave the area of interest during an MCS game, and new mobile users sometimes enter the network. The mobility index of user j denoted by p_j is the probability that user j leaves or re-enters the network during the sensing task, which depends on the mobility speed of the user.

Let $C_j^{(i)}$ be the nonnegative cost of user j with sensing accuracy level i, and $G^{(i)}$ denote the benefit to the server from a sensing report of accuracy level i. The system gain set is denoted by $\mathbf{G} = \left[G^{(i)}\right]_{-1 \le i \le L}$. As shown in Fig. 1, the contribution coefficient of user j, denoted by λ_j, represents the importance of the sensing data from the user and reflects the dynamism of the contribution of the user due to the difference in the sensing location and response time. If the contribution coefficients are the same for all users, i.e., $\lambda_j \equiv \lambda_i$, $\forall i,\ j$, the server obtains a higher benefit from a user exerting more sensing effort, i.e., $0 < G^{(h)} < G^{(g)}$, $\forall g > h > 0$. Therefore, the gain that the server obtains from a sensing report sent by user j is $\lambda_j G^{(x_j)}$.

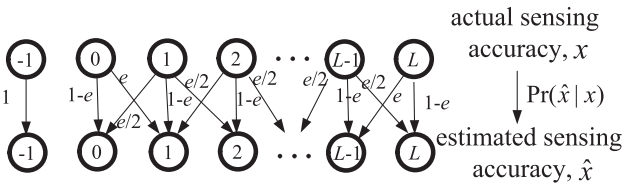


Fig. 2. Sensing evaluation model of the MCS server given by (1), in which $\Pr(\hat{x}|x)$ is the probability that a sensing report of accuracy level x is taken as level $\hat{x}$, with x and $\hat{x} \in \{-1, 0, \cdots, L\}$.

The sensing effort of a mobile user has an impact on the contribution of another user. For instance, if user A has already sent an accurate sensing report to the MCS server, the second report sent by user B will have a much lower impact and thus the server is motivated to give user B a lower payment, even if user B applies the same sensing effort. In addition, due to the transmission cost, a smartphone has to expend effort to send a faked report, i.e., $C_j^{(0)} > 0$, and clearly $G^{(-1)} = C_j^{(-1)} = 0$, $\forall j$.

By applying an advanced and efficient sensing quality evaluation algorithm, such as the classification algorithm in [16] and [28]–[30], and the majority voting algorithm in [17], the server estimates the sensing effort of each user. For example, in cases where the sensed quantity is an image, the server can use the blind image quality assessment algorithm in [16] to extract image features of interest and thus evaluate the quality of the sensing report. For some environmental monitoring applications in which a single report cannot be accurately evaluated, the server has to recruit more users and evaluate the sensing quality based on data truthfulness [31].

In the sensing evaluation model of the server, $\Pr(\hat{x}|x)$ is the probability that the server takes the actual sensing accuracy level x as $\hat{x}$. It is assumed that the server can receive the sensing reports under a good radio channel conditions, i.e., $\Pr(-1|-1) = 1$. An example of the probability mass function of the sensing evaluation model as shown in Fig. 2 is given by

$$\Pr(\hat{x}|x) = \begin{cases} 1-e, & \text{if } \ \hat{x} = x \ge 0 \\ 1, & (x, \hat{x}) = (-1, -1) \\ e, & (x, \hat{x}) = (0, 1) \ \text{ or } (L, L-1) \\ e/2, & \hat{x} = x \pm 1, 0 < x < L \\ 0, & \text{otherwise,} \end{cases} \tag{1}$$

where e denotes the evaluation error rate of the server, which depends on the classification algorithm used by the server and the overall quality of the sensing reports.

In this MCS system, the server sometimes fails to detect a faked sensing report, i.e., $\Pr(1|0) = e > 0$, and makes mistakes in evaluating the sensing accuracy, e.g., $\Pr(L|L-1) = e$.

The probability that a report with sensing accuracy l is taken as $l \pm 1$ (if applicable) is $e/2$, and this work assumes an accurate sensing evaluation with $e < 1/2$. The evaluation error e depends on the number of sensing accuracy levels L. For simplicity, both L and e are assumed to be known and fixed in the following analysis.

As an MCS server sometimes takes an accurate report as a fake one by mistake, the server cannot simply remove the "cheating" user, which suppresses the sensing motivation of the users and thus decreases the long-term utility of the server. Note that although Eq. (1) is adopted in the simulations presented in Section VI, the framework presented in this work is not restricted to the sensing evaluation error model in (1). More specifically, the analytical results in the next section are independent of (1) and the proposed learning based MCS schemes in Section V do not rely on (1) either. For ease to reference, we summarize the commonly used notation in Table I.

IV. Secure Mobile Crowdsensing Game

In the Stackelberg MCS game, the MCS server first determines and broadcasts the payment policy and the M smartphone users choose their sensing efforts $x_{1\le j\le M} \in A$ afterwards. The payment policy is denoted by $\mathbf{y} = [y_i]_{-1\le i\le L}$, where $y_i \in \mathbf{P}_i$ is the payment for the sensing report of accuracy level i, and $\mathbf{P}_i$ contains H nonzero feasible payment levels. A rational server pays only for useful sensing reports $\big($i.e., $y_{-1} = y_0 = 0\big)$, and pays user j according to the estimated sensing accuracy $\hat{x}_j$. For simplicity, we omit $y_{-1} = y_0 = 0$ in $\mathbf{y}$ if no confusion results.

Let $f(x)$ denote the estimated sensing accuracy of the report with actual accuracy level x at the server. If user j sends a sensing report of accuracy x_j, the utility of user j denoted as u_j, is given by

$$u_j(x_j, \mathbf{y}) = \mathbf{y}\big(f(x_j)\big) - C_j^{(x_j)}, \tag{2}$$

where $\mathbf{y}\big(f(x_j)\big)$ is the payment for the report of user j. Based on the system gain and the payment for user j, the utility of the server denoted by u_s is given by

$$u_s(x_j, \mathbf{y}) = \lambda_j G_j^{(x_j)} - \mathbf{y}\big(f(x_j)\big). \tag{3}$$

As $y_{-1} = y_0 = 0$ and $C_j^{(-1)} = 0 < C_j^{(0)}$, by (2), user j receives a lower utility by sending a faked report than taking no response, i.e., $u_j(-1, \mathbf{y}) > u_j(0, \mathbf{y})$, under accurate sensing evaluation. Otherwise, if $\Pr\big(f(x) = x\big)$ is small, a user is either over-paid or under-paid, with the expected payment over all possible sensing evaluation results given by $\sum_{l=-1}^{L} \mathbf{y}(l)\Pr(l|x)$. Therefore, by (2) and (3), the expected utility of user j (or the server) denoted by U_j (or U_s) is given by

$$U_j(x_j, \mathbf{y}) = \mathbb{E}_{\Pr(f(x_j)|x_j)}[u_j(x_j, \mathbf{y})] = \sum_{l=-1}^{L} \mathbf{y}(l)\Pr(l|x_j) - C_j^{(x_j)} \tag{4}$$

$$\begin{aligned} U_s(x_j, \mathbf{y}) &= \mathbb{E}_{\Pr(f(x_j)|x_j)}[u_s(x_j, \mathbf{y})] \\ &= \lambda_j G^{(x_j)} - \sum_{l=-1}^{L} \mathbf{y}(l)\Pr(l|x_j). \end{aligned} \tag{5}$$

With $\mathbf{x} = [x_j]_{1\le j\le M}$, the expected overall utility of the server from the M users is given by

$$\begin{aligned} U_s(\mathbf{x}, \mathbf{y}) &= \sum_{m=1}^{M} U_s(x_m, \mathbf{y}) \\ &= \sum_{m=1}^{M} \left(\lambda_m G^{(x_m)} - \sum_{l=-1}^{L} \mathbf{y}(l)\Pr(l|x_m) \right). \end{aligned} \tag{6}$$

In summary, we have formulated a secure mobile crowdsensing game denoted by $\mathbb{G}$, in which the server as the leader of the Stackelberg game first decides $\mathbf{y} \in \mathbf{P}^L$ to maximize U_s, and each user as a follower chooses his or her sensing effort or accuracy according to the MCS strategy.

An SE strategy of the secure MCS game, denoted by $(\mathbf{x}^* = [x_j^*]_{1\le j\le M}, \mathbf{y}^*)$, consists of the best response of each player in the Stackelberg game. By definition, we have

$$x_j^*(\mathbf{y}) = \arg\max_{x\in A} U_j(x, \mathbf{y}), \quad 1 \le j \le M \tag{7}$$

$$\mathbf{y}^* = \arg\max_{\mathbf{y}\geqslant 0} U_s\big(\mathbf{x}^*(\mathbf{y}), \mathbf{y}\big). \tag{8}$$

We first consider the SE of the MCS game with a single positive sensing accuracy ($L = 1$), in which user j applies his or her full sensing effort with $x_j = 1$, sends a faked report with $x_j = 0$ and gives no response with $x_j = -1$. As $y^*(-1) = y^*(0) = 0$, we have $\mathbf{y}^* = [0, 0, y^*]$.

Proposition 1: In the secure MCS game $\mathbb{G}$ with $L = 1$ positive sensing levels, user j is motivated to apply full sensing effort, i.e. $x_j^* = 1$, if $e < 1/2$ and

$$y^* \ge \max\left(\frac{C_j^{(1)}}{1-e}, \frac{C_j^{(1)} - C_j^{(0)}}{1-2e} \right), \tag{9}$$

and is motivated not to respond, i.e. $x_j^* = -1$, if

$$y^* \le \min\left(\frac{C_j^{(0)}}{e}, \frac{C_j^{(1)}}{1-e} \right). \tag{10}$$

Proof: See Appendix A. □

The optimal payment y^* increases with the sensing evaluation error to make up for the underpay loss of the user. On the other hand, as the offered payment has to exceed the sensing cost to stimulate a user to take the sensing task, the probability that a sensing task fails due to insufficient sensing results increases with the sensing costs of the smartphones in the area. If the payment for an accurate report is too low compared with that given in (9) and (10), a user will benefit from cheating and thus is motivated to send faked sensing reports.

In image-centric MCS applications such as a public information sharing system in [32], a server that applies a classification algorithm such as that in [16] to evaluate sensing reports chooses the most accurate report (e.g., the photo with the highest resolution) in the same location and discards the other reports. In this case, the server aims to motivate the smartphone with the best sensing condition to apply its full sensing effort while suppressing the other smartphones to save costs. On letting $\mathbf{1}$ be the all-1 vector and $\mathbf{I}_M(j)$ be the M-dimension indication vector with the i-th element equal 0 if $i \ne j$ and 1 if $i = j$, we have the following result:

Theorem 2: The secure MCS game $\mathbb{G}$ with $L = 1$ positive sensing levels and M users has an SE given by

$$(\boldsymbol{x}^*, \boldsymbol{y}^*) = \left(-\boldsymbol{1}_M + 2\mathbf{I}_M(j^*), \left[0, 0, \max\left(\frac{C_{j^*}^{(1)}}{1-e}, \frac{C_{j^*}^{(1)} - C_{j^*}^{(0)}}{1-2e} \right) \right] \right), \tag{11}$$

if

$$\min_{j \neq j^*} \left(\frac{C_j^{(1)}}{\lambda_j(1-e)}, \frac{C_j^{(0)}}{\lambda_j e} \right) \geq \frac{G^{(1)}}{1-e} \geq \max\left(\frac{C_{j^*}^{(1)}}{\lambda_{j^*}(1-e)}, \frac{C_{j^*}^{(1)} - C_{j^*}^{(0)}}{\lambda_{j^*}(1-2e)} \right) \tag{12}$$

with

$$j^* = \arg \min_{1 \leq j \leq M} \max\left(\frac{C_j^{(1)}}{\lambda_j(1-e)}, \frac{C_j^{(1)} - C_j^{(0)}}{\lambda_j(1-2e)} \right). \tag{13}$$

Proof: See Appendix B. □

Remark: By (13), user j^* has the lowest sensing cost and payment expectation. If the offered payment equals his or her expectation and is less than the expectation of the other users as indicated by (12), the server obtains a single accurate sensing report. A user expects a lower utility by cheating alone, and thus is never motivated to send a faked sensing report, unless he or she knows for sure that a large number of users simultaneously launch faked sensing attacks, which is unlikely to happen in a large-scale wireless network.

If the server can benefit from all the sensing reports in the MCS application such as in the pollution monitoring application in [33], all the smartphone users are motivated to participate in the MCS activity, and we have the following result:

Theorem 3: The MCS game $\mathbb{G}$ with $L = 1$ positive sensing levels and M users has an SE given by

$$(\boldsymbol{x}^*, \boldsymbol{y}^*) = \left(\boldsymbol{1}_M, \left[0, 0, \max_{1 \leq j \leq M} \max\left(\frac{C_j^{(1)}}{1-e}, \frac{C_j^{(1)} - C_j^{(0)}}{1-2e} \right) \right] \right), \tag{14}$$

if

$$\frac{G^{(1)}}{1-e} \geq \max_{1 \leq j \leq M} \max\left(\frac{C_j^{(1)}}{\lambda_j(1-e)}, \frac{C_j^{(1)} - C_j^{(0)}}{\lambda_j(1-2e)} \right). \tag{15}$$

Proof: The proof is similar to that of Theorem 2. □

Remark: In this case, the server offers a high payment to motivate all the users to apply full sensing effort. As shown in Fig. 3, both the average utility of the users and the utility of the server increase with the number of users, because the server can obtain more sensing reports of high quality.

If there are $L = 2$ non-zero sensing accuracy levels, we have $\mathbf{y}^* = [0, 0, y_1^*, y_2^*]$. In this case, user j uses full sensing effort if $x_j = 2$, applies coarse sensing effort if $x_j = 1$, sends a faked report if $x_j = 0$, and does not respond if $x_j = -1$.

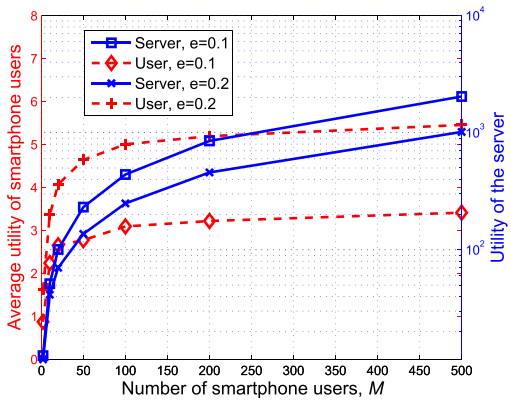


Fig. 3. Performance of the MCS game $\mathbb{G}$ as given by Theorem 2, with $L = 1$, $\lambda_j \sim \mathrm{U}(0, 5, 1)$ and the system gains $\mathbf{G} = [0, -5, 20]$.

Proposition 4: In the MCS game $\mathbb{G}$ with $L = 2$ positive sensing levels, user j is motivated to use full sensing effort, i.e. $x_j^* = 2$, if

$$\begin{cases} y_2^* > \max\left(\dfrac{C_j^{(2)} - C_j^{(0)}}{1-e}, \dfrac{(1-e)C_j^{(2)} - eC_j^{(1)}}{1-2e+0.5e^2} \right) & \text{(16a)} \\ \dfrac{1-1.5e}{1-2e} y_2^* - \dfrac{C_j^{(2)} - C_j^{(1)}}{1-2e} > y_1^* > \dfrac{C_j^{(2)} - (1-e)y_2^*}{e}; & \text{(16b)} \end{cases}$$

use coarse sensing effort, i.e. $x_j^* = 1$, if

$$\begin{cases} y_1^* > \max\left(\dfrac{(2-2e)C_j^{(1)} - eC_j^{(2)} - (2-3e)C_j^{(0)}}{2-6e+4e^2}, \dfrac{(2-2e)C_j^{(1)} - eC_j^{(2)}}{e^2-4e+2} \right) & \text{(17a)} \\ \dfrac{(1-3e)y_1^* + C_j^{(2)} - C_j^{(1)}}{1-2e} > y_2^* > \max\left(\dfrac{2C_j^{(1)} - (2-2e)y_1^*}{e}, \dfrac{2C_j^{(1)} - 2C_j^{(0)} - (2-4e)y_1^*}{e} \right); & \text{(17b)} \end{cases}$$

and keep silent, i.e. $x_j^* = -1$, if

$$\begin{cases} y_1^* < \dfrac{C_j^{(0)}}{e} & \text{(18a)} \\ y_2^* < \min\left(\dfrac{C_j^{(1)} - (1-e)y_1^*}{0.5e}, \dfrac{C_j^{(2)} - ey_1^*}{1-e} \right). & \text{(18b)} \end{cases}$$

Proof: The proof is similar to the proof of Proposition 1. □

Remark: To stimulate accurate sensing, the server has to offer a higher payment for the high-quality sensing report y_2^* given by (16a) than that for the low-quality report y_1^* given by (16b). If y_1^* is very high, i.e., (17a), user j is motivated to provide a low-quality sensing report. If the sensing cost exceeds the offered payment in (18), a user does not take the MCS task.

If an MCS application requires only one accurate sensing report from the area of interest, the smartphone user with the

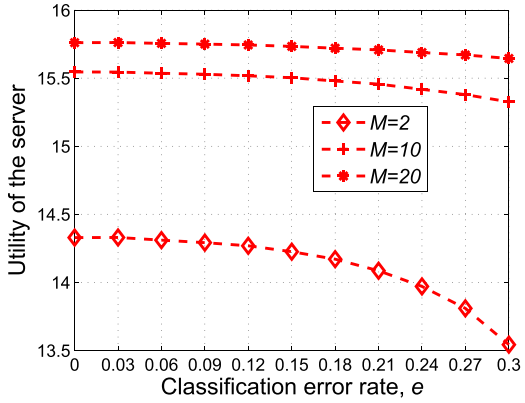


Fig. 4. Performance of the secure MCS game $\mathbb{G}$ with M users, $L = 2$, $\mathbf{G} = [0, -5, 5, 20]$, and $\lambda_j \sim \mathrm{U}(0.8, 1)$, as given by Theorem 5.

lowest sensing cost is motivated to apply his or her full sensing effort and all the others have to keep silent to avoid over-sensing, yielding the following result:

Theorem 5: The secure MCS game $\mathbb{G}$ with $L = 2$ positive sensing levels and M users has an SE $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ given by

$$\boldsymbol{x}^* = -\boldsymbol{I}_M + 3\mathbf{I}_M(j^*) \tag{19}$$

$$\boldsymbol{y}^* = \left[0, 0, \frac{C_{j^*}^{(2)} - (1-e)y_2^*}{e}, \max\left(\frac{C_{j^*}^{(2)} - C_{j^*}^{(0)}}{1-e}, \frac{(1-e)C_{j^*}^{(2)} - eC_{j^*}^{(1)}}{1-2e+0.5e^2}\right)\right], \tag{20}$$

if

$$\begin{cases} \min\left(\lambda_{j^*}\dfrac{(1-e)G^{(2)} - eG^{(1)}}{1-2e+0.5e^2}, \dfrac{C_{j'}^{(1)} - (1-e)y_1^*}{0.5e}, \right. & \text{(21a)} \\ \left. \dfrac{C_{j'}^{(2)} - ey_1^*}{1-e}\right) > \max\left(\dfrac{C_{j^*}^{(2)} - C_{j^*}^{(0)}}{1-e}, \dfrac{(1-e)C_{j^*}^{(2)} - eC_{j^*}^{(1)}}{1-2e+0.5e^2}\right) & \\ \min\left(\lambda_{j^*}\dfrac{0.5eG^{(2)} - (1-e)G^{(1)}}{-1/e+2-0.5e}, C_{j'}^{(0)}\right) > C_{j^*}^{(2)} - (1-e)y_2^*, & \text{(21b)} \end{cases}$$

where

$$j^* = \arg\min_{1\le j\le M} \max\left(\frac{C_j^{(2)} - C_j^{(0)}}{1-e}, \frac{(1-e)C_j^{(2)} - eC_j^{(1)}}{1-2e+0.5e^2}\right), \tag{22}$$

$$j' = \arg\min_{1\le j\le M, j\ne j^*} C_j^{(0)}. \tag{23}$$

Proof: See Appendix C. □

Remark: As indicated by Theorem 5, if condition (21) holds, the smartphone with the best sensing condition, i.e., user j^*, sends an accurate sensing report to receive a payment y_2^* given by (20) that is higher than the payment given by (11) for the case with (12) in Theorem 1. In addition, as shown in Fig. 4, the utility of the server increases with the sensing evaluation accuracy (i.e., $1 - e$) and the number of users M.

If the MCS application requires multiple high-quality reports to evaluate the data truthfulness, e.g., the sensing task as described in [31], the server has to pay more for each accurate sensing report compared with the case in Theorem 5, yielding the following result:

Theorem 6: The MCS game $\mathbb{G}$ with $L = 2$ positive sensing levels and M users has an SE $(\boldsymbol{2}_M, \boldsymbol{y}^*)$ given by

$$\boldsymbol{y}^* = \left[0, 0, \max_{1\le j\le M} \frac{C_j^{(2)} - (1-e)y_2^*}{e}, \max_{1\le j\le M} \max\left(\frac{C_j^{(2)} - C_j^{(0)}}{1-e}, \frac{(1-e)C_j^{(2)} - eC_j^{(1)}}{1-2e+0.5e^2}\right)\right], \tag{24}$$

if

$$\frac{(1-e)G^{(2)} - eG^{(1)}}{1-2e+0.5e^2} \ge \max_{1\le j\le M} \max\left(\frac{C_j^{(2)} - C_j^{(0)}}{\lambda_j(1-e)}, \frac{(1-e)C_j^{(2)} - eC_j^{(1)}}{\lambda_j\left(1-2e+0.5e^2\right)}\right). \tag{25}$$

Proof: The proof is similar to that of Theorem 5. □

Similarly to Theorem 3, the server provides a high payment given by (24) for all the users that provide high-quality sensing reports to compensate for their sensing efforts, if (25) holds.

In a generic MCS game with M users and L accuracy levels, the SE is denoted by $(\mathbf{L}_M, \mathbf{y}^*)$. By definition, we have $\forall 1 \le j \le M$

$$U_s(\mathbf{L}_M, \mathbf{y}^*) \ge U_s([x_j]_{1\le j\le M}, \mathbf{y}^*), \ \forall -1 \le x_j \le L \tag{26}$$

$$U_j(L, \mathbf{y}^*) \ge \max_{-1\le i\le L-1} U_j(i, \mathbf{y}^*). \tag{27}$$

If the benefit obtained from accurate sensing reports is large, the MCS server offers a high payment to stimulate each user to use full-sensing effort. On the other hand, the utility of the server decreases with increasing payment level. The optimal payment policy given by the SE in Eqs. (26) and (27) depends on the maximum sensing cost of the smartphones.

The server can estimate the sensing parameters such as λ_j and C_j via surveys and the sensing history, and use them to design optimal payment strategies according to the SE in a static game. However, a server usually has difficulty estimating the sensing model in a realistic MCS application, and thus can use reinforcement learning techniques to design the payment policy to motivate accurate sensing reports without knowing the user sensing model. This approach will be discussed in the following section.

V. Dynamic MCS Game

The repeated interactions between an MCS server and M smartphone users can be formulated as a dynamic MCS game. In general, a high payment for an accurate sensing report decreases the server's immediate utility, but it also stimulates more users to participate in the MCS task in the future. An over-paid MCS task tends to result in over-sensing and thus decreases the server's long-term utility. On the other hand, an under-paid policy usually causes under-sensing or even a failed sensing task. In addition, a mobile user usually chooses its sensing effort based on the MCS payment history,

because an under-paid sensing experience indicates a server with inaccurate sensing evaluation. Therefore, the payment policy of an MCS server impacts the future sensing results and the future reward. The MCS payment decision in the dynamic game can be formulated as a Markov decision process and reinforcement learning algorithms can be used to achieve an optimal policy.

A. Q-Learning Based MCS Payment

The server can apply Q-learning, a model-free reinforcement learning technique to derive an optimal MCS policy without knowing the sensing model of the active smartphones for the MDP. It is assumed that the server applies high quality data mining algorithms such as the expectation-maximization algorithm in [28] to evaluate each sensing report and thus estimate the sensing effort of each user with high accuracy. In this paper, we assume that the quality evaluation algorithm is effective and thus the sensing accuracy evaluation is accurate, with $x_j^{(k)} \approx \hat{x}_j^{(k)}$ in most cases. According to the estimate sensing efforts of M users at time k, denoted by $\hat{x}_{1\le j\le M}^{(k)}$, the server obtains the estimated number of reports of accuracy level i, denoted by $\hat{N}_i^{(k)}$ and given by

$$\hat{N}_i^{(k)} = \sum_{j=1}^{M} \mathrm{I}(\hat{x}_j^{(k)} = i), \quad 0 \le i \le L, \tag{28}$$

where $\mathrm{I}(\cdot)$ is the indicator function taking the value 1 if its argument is true and 0 otherwise.

The Q-learning based payment is chosen based on the system state, denoted by $\mathbf{s}^{(k)}$, which consists of the previous sensing quality and the payment policy, i.e., $\mathbf{s}^{(k)} = \left[\hat{N}_{0\le i\le L}^{(k-1)}, \mathbf{y}^{(k-1)}\right]$. For simplicity, the feasible payments for each accuracy level are quantized into H levels, with $y_i^{(k)} \in \mathbf{P}_i$, where $\mathbf{P}_i$ contains H feasible payments. The server pays user j according to the current sensing evaluation result, i.e., $\mathbf{y}^{(k)}(\hat{x}_j^{(k)})$. As the total payment to the M users is $\sum_{i=1}^{L} \hat{N}_i^{(k)} \mathbf{y}^{(k)}(i)$, by (6), the utility of the server observed at time k can be written as

$$u_s^{(k)} = \sum_{i=0}^{L}\sum_{j=1}^{M} \lambda_j \mathrm{I}\left(\hat{x}_j^{(k)} = i\right) G^{(i)} - \hat{N}_i^{(k)} \mathbf{y}^{(k)}(i). \tag{29}$$

The Q-learning based MCS payment strategy as shown in Algorithm 1 depends on the Q function denoted by $Q(\mathbf{s}, \mathbf{y})$; that is the expected long-term discounted utility for the state-action pair $(\mathbf{s}, \mathbf{y})$ is updated according to the iterative Bellman equation as follows:

$$Q(\mathbf{s}, \mathbf{y}) \leftarrow (1-\alpha) Q(\mathbf{s}, \mathbf{y}) + \alpha \left(u_s(\mathbf{s}, \mathbf{y}) + \gamma V(\mathbf{s}')\right), \tag{30}$$

$$V(\mathbf{s}') \leftarrow \max_{\mathbf{y}' \in \mathbf{P}^L} Q(\mathbf{s}', \mathbf{y}'), \tag{31}$$

where $\mathbf{s}'$ is the new state from $\mathbf{s}$ with action $\mathbf{y}$, the value function $V(\cdot)$ provides the highest value of the Q function, γ is the discount factor indicating the myopic view of the server regarding the future reward, and $\alpha \in (0, 1]$ is the learning rate of the current experience $\mathbf{s}$-$\mathbf{y}$-$\mathbf{s}'$.

According to the Q-function values of the actions and the current system state $\mathbf{s}^{(k)}$, the MCS server chooses the payment strategy with the ϵ-greedy algorithm to avoid staying in the local maxima. More specifically, the "optimal" payment vector $\mathbf{y}^* = \arg\max_{\mathbf{y}' \in \mathbf{P}^L} Q(\mathbf{s}^{(k)}, \mathbf{y}')$ is chosen with a high probability $1-\epsilon$, and the other payment strategies are randomly chosen with a very small probability, as summarized in Algorithm 1. The convergence time of the Q-learning based MCS strategy depends on the size of the state space, i.e., M^{L+1}, which increases with the number of active users in the area of interest. Therefore, the Q-learning based MCS system has to address the curse of dimensionality for applications involving a large number of smartphone users.

B. DQN-Based MCS Payment

The DQN-based MCS strategy uses a convolutional deep neural network to accelerate the convergence rate of Algorithm 1. More specifically, the DQN-based MCS system also updates a quality function Q for each action-state pair in the dynamic MCS game, given by definition as

$$Q(\mathbf{s}, \mathbf{y}) = \mathbb{E}_{\mathbf{s}'}\left[u_s + \gamma \max_{\mathbf{y}' \in \mathbf{P}^L} Q(\mathbf{s}', \mathbf{y}') | \mathbf{s}, \mathbf{y}\right]. \tag{32}$$

The deep Q-network technique accelerates the learning speed of Q-learning by using a nonlinear neural network function approximator to estimate the values of the Q-function. More specifically, our proposed DQN-based MCS system as shown in Fig. 5 estimates the Q-value in (32) for each action $\mathbf{y}$ using a convolutional deep neural network, which consists of two convolutional (Conv) layers and two fully connected (FC) layers. The first convolutional layer consists of 20 filters each with size 4×4 and stride 1, and applies the rectified linear unit (ReLU) given by [18] as the activation function. The second convolutional layer involves 40 filters each with size 2×2 and stride 1, and uses the same nonlinear rectifier. The first FC layer uses 320 rectified linear units, while the second FC layer has H^L units for each payment policy $\mathbf{y}$, where H is the number of feasible payments and L is the number of positive sensing levels. The filter weights of the four layers in the CNN at time k are denoted by $\theta^{(k)}$, with CNN parameters as shown in Table II.

As shown in Fig. 5, the input to the CNN in the DQN-based MCS system consists of the current and the previous $W = 12$ system states, i.e., $\boldsymbol{\varphi}^{(k)} = (\mathbf{s}^{(k-W)}, \mathbf{s}^{(k-W+1)}, \cdots, \mathbf{s}^{(k)})$, which are then reshaped into an 8×8 matrix. The H outputs at the CNN are the estimated Q values of each action $Q(\boldsymbol{\varphi}^{(k)}, \mathbf{y}^{(k)}|\theta^{(k)})$ for a given system state sequence $\left[\mathbf{s}^{(k-w)}\right]_{0\le w\le W}$. According to the experience replay, $\theta^{(k)}$ is updated at each time.

The experience replay uses the memory as shown in Fig. 5 to store the experiences. The experience that the MCS server obtained at time k is denoted by $\mathbf{e}^{(k)} = (\boldsymbol{\varphi}^{(k)}, \mathbf{y}^{(k)}, u_s^{(k)}, \boldsymbol{\varphi}^{(k+1)})$, and the memory pool is $\mathcal{D} = \{\mathbf{e}^{(1)}, \cdots, \mathbf{e}^{(k)}\}$. According to the experience replay technique, the server randomly chooses an experience from $\mathcal{D}$ to update the CNN weights $\theta^{(k)}$. The stochastic gradient algorithm is applied with learning rate η, which is the

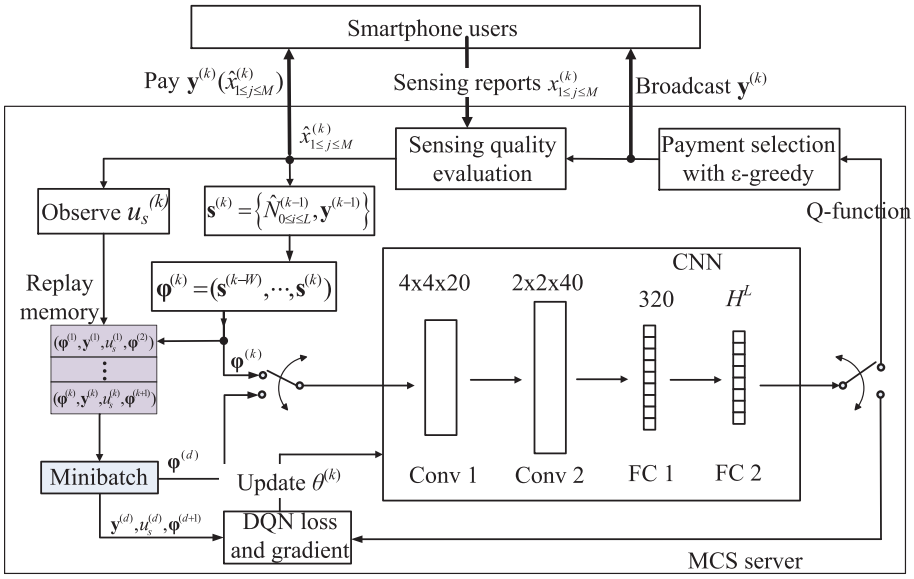


Fig. 5. Illustration of the DQN-based MCS system.

TABLE II
ARCHITECTURE PARAMETERS OF THE CNN

Layer	Input	Filter size	Stride	Number of filters	Activation	Output
Conv 1	8×8	4×4	1	20	ReLU	$5\times 5\times 20$
Conv 2	$5\times 5\times 20$	2×2	1	40	ReLU	$4\times 4\times 40$
FC 1	640			320	ReLU	320
FC 2	320			H^L	ReLU	H^L

Algorithm 1 Q-Learning Based MCS Payment Strategy

Initialize δ, α, $\mathbf{P}$, $\hat{N}_i=0$, $\forall 0\le i\le L$, $Q(\mathbf{s},\mathbf{y})=0$, and $V(\mathbf{s})=0$, $\forall \mathbf{s},\mathbf{y}$
for $k=1,2,3,\cdots$ **do**
  $\mathbf{s}^{(k)}=\left[\hat{N}^{(k-1)}_{-1\le i\le L},\mathbf{y}^{(k-1)}\right]$
  Select and perform $\mathbf{y}^{(k)}\in\mathbf{P}^L$ via the ϵ-greedy algorithm
  Broadcast the recruiting message with $\mathbf{y}^{(k)}$
  while receiving a sensing report from user j **do**
    Apply the sensing evaluation algorithm to obtain $\hat{x}_j^{(k)}$
    Pay user j with $\mathbf{y}^{(k)}(\hat{x}_j^{(k)})$
  end while
  Obtain $u_s^{(k)}$
  for $0\le i\le L$ **do**
    Calculate $\hat{N}_i^{(k)}$ via (28)
  end for
  Update $Q(\mathbf{s}^{(k)},\mathbf{y}^{(k)})$ via (30)
  Update $V(\mathbf{s}^{(k)})$ via (31)
end for

step size of the gradient descent iterations in the training process and represents the speed of the CNN weight updates. The mean squared-error of the target values is minimized over minibatches, for the loss function chosen from [18] as

$$L(\theta^{(k)})=\mathbb{E}_{\boldsymbol{\varphi},\mathbf{y},u_s,\boldsymbol{\varphi}}\left[\left(u_s+\gamma\max_{\mathbf{y}'\in\mathbf{P}^L}Q(\mathbf{s}',\mathbf{y}';\theta^{(k-1)})-Q\left(\mathbf{s},\mathbf{y};\theta^{(k)}\right)\right)^2\right]. \quad (33)$$

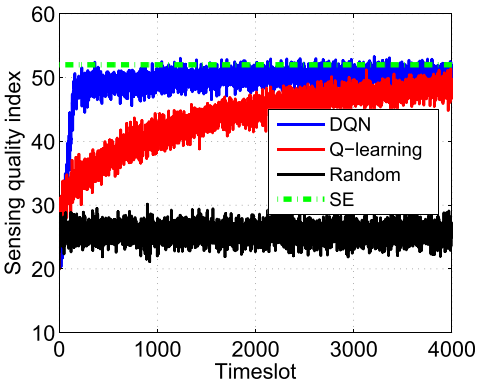
Thus

$$\nabla_{\theta^{(k)}}L(\theta^{(k)})=-\mathbb{E}_{\boldsymbol{\varphi},\mathbf{y},u_s,\boldsymbol{\varphi}}\left[\left(u_s+\gamma\max_{\mathbf{y}'\in\mathbf{P}^L}Q\left(\mathbf{s}',\mathbf{y}';\theta^{(k-1)}\right)-Q\left(\mathbf{s},\mathbf{y};\theta^{(k)}\right)\right)\nabla_{\theta^{(k)}}Q\left(\mathbf{s},\mathbf{y};\theta^{(k)}\right)\right], \quad (34)$$

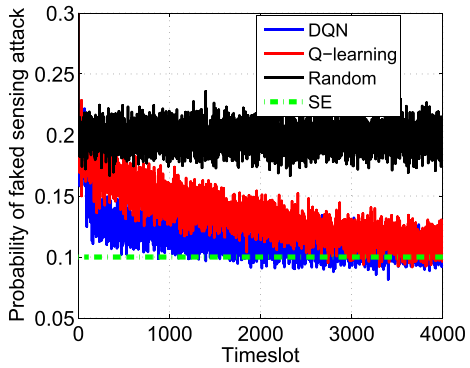
where $\mathbf{s}'$ is the next state after $\mathbf{s}$ by taking action $\mathbf{y}$. This process repeats B times at each time slot. In each update, $\theta^{(k)}$ is chosen according to the randomly selected experiences from $\mathcal{D}$, as summarized in Algorithm 2.

VI. Simulation Results

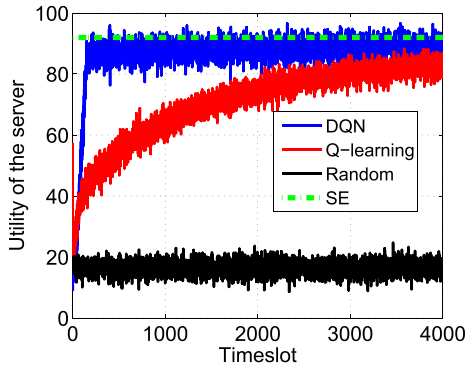
Simulations have been performed to evaluate the performance of the dynamic MCS game, with $M=5$, $L=1$, $H=11$, $\mathbf{G}=[0,-1,9]$, $C_j^{(-1)}=0$, $C_j^{(0)}\sim \mathrm{U}(0,1)$, $C_j^{(1)}\sim \mathrm{U}(3,5)$ and $\lambda_j\sim \mathrm{U}(0.8,1)$, $\forall 1\le j\le M$, if not specified otherwise, where $\mathrm{U}(a,b)$ denotes the uniform distribution



(a)



(b)



(c)

Fig. 6. Performance of the dynamic MCS game, with $M = 60$, $L = 1$, $e = 0.1$, $\mathbf{G} = [0, -1, 9]$, $C_j^{(-1)} = 0$, $C_j^{(0)} \sim \mathrm{U}(0, 1)$, $C_j^{(1)} \sim \mathrm{U}(3, 5)$, and $\lambda_j \sim \mathrm{U}(0.8, 1)$, $\forall 1 \leq j \leq M$. (a) Sensing quality index. (b) Attack probability. (c) Utility of the server.

between a and b. In the simulations, user j leaves or re-enters the network or the area of interest in each time slot with a probability $p_j = 0.1$, with $1 \leq j \leq M$. The parameters of the Q-learning algorithm are set as $\alpha = 0.2$ and $\delta = 0.7$.

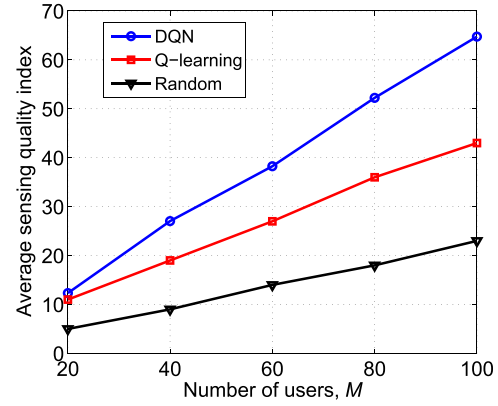
As shown in Fig. 6, the Q-learning based payment strategy achieves an optimal policy in this dynamic MCS game after convergence, which matches the theoretical results of the SE given by Theorem 2. The DQN-based MCS strategy achieves an optimal policy even faster and takes much less time to converge to the theoretical result given by Theorem 2. More specifically, the DQN-based MCS strategy outperforms the Q-learning based strategy, which in turn exceeds the random payment policy, with a higher sensing quality, lower attack probability and thus higher utility. For instance, the faked sensing probability of the game with the DQN-based MCS strategy decreases from 0.2 at the beginning of the game to 0.1 at the 2000-th time slot, while the Q-learning based strategy takes 3500 time slots to achieve the same performance. Therefore, the utility of the server increases from 18 to 88 at time slot 200, which is 96% and 340% higher than the Q-learning based MCS and the random strategy, respectively.

As shown in Fig. 7, the overall sensing quality index of the MCS system increases with the number of users, as the server obtains more accurate sensing reports and fewer faked sensing reports. For example, our proposed DQN-based MCS system increases the utility from 28 for the case with $M = 20$ users to 132 with $M = 100$ users, and u_s with $M = 100$ is 65% and 371% higher than that of Q-learning and the random strategy, respectively.

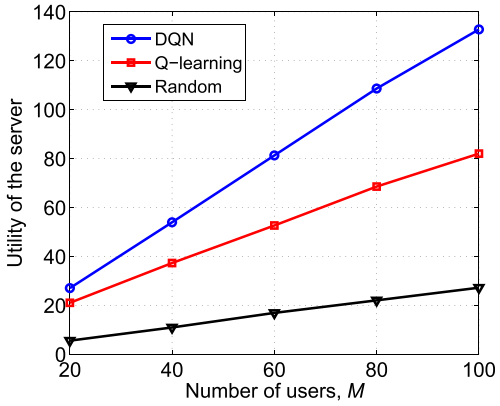
The performance of the dynamic game under user mobility is presented in Fig. 8, in which each user leaves or re-joins the network with probability $p_j = 0.1$ and $\forall 1 \leq j \leq 60$ at every 100 time slots, with $L = 2$, $\mathbf{G} = [0, -1, 9, 12]$, and $C_j^{(2)} \sim \mathrm{U}(9, 10)$, $\forall 1 \leq j \leq 60$. The Q-learning based system fails to converge, while the DQN-based MCS system still converges and provides accurate sensing. For instance,

Algorithm 2 DQN-Based MCS Payment Strategy

Initialize δ, α, $\mathbf{P}$, $\mathbf{y} = \mathbf{1}$, $B = 16$, $W = 12$, $\mathcal{D} = \emptyset$, and $\hat{N}_i = 0$, $\forall 0 \leq i \leq L$
Initialize the Q-network with random weights θ
for $k = 1, 2, 3, \ldots$ **do**
  $\mathbf{s}^{(k)} = \left[\hat{N}_{-1 \leq i \leq L}^{(k-1)}, \mathbf{y}^{(k-1)}\right]$
  if $k \leq W$ **then**
    Select $\mathbf{y}^{(k)} \in \mathbf{P}^L$ at random
  else
    Input $\boldsymbol{\varphi}^{(k)} = \left(\mathbf{s}^{(k-W)}, \mathbf{s}^{(k-W+1)}, \cdots, \mathbf{s}^{(k)}\right)$ to the CNN as shown in Fig. 5, with weights θ
    Obtain the CNN output $Q(p)$, $\forall 1 \leq p \leq H^L$
    Select $\mathbf{y}^{(k)}$ via the ϵ-greedy algorithm
  end if
  Broadcast the recruiting message with $\mathbf{y}^{(k)}$
  while receiving sensing report from user j **do**
    Apply the sensing evaluation algorithm to obtain $\hat{x}_j^{(k)}$
    Pay user j with $\mathbf{y}^{(k)}(\hat{x}_j^{(k)})$
  end while
  Obtain $u_s^{(k)}$
  for $0 \leq i \leq L$ **do**
    Calculate $\hat{N}_i^{(k)}$ via (28)
  end for
  $\mathcal{D} \leftarrow \mathcal{D} + \left\{\mathbf{s}^{(k)}, \mathbf{y}^{(k)}, u_s^{(k)}, \mathbf{s}^{(k+1)}\right\}$
  for $d = 1, 2, \cdots, B$ **do**
    Select $\left(\mathbf{s}^{(d)}, \mathbf{y}^{(d)}, u_s^{(d)}, \mathbf{s}^{(d+1)}\right)$ from $\mathcal{D}$ at random
    $R^{(d)} \leftarrow r_s^{(d)} + \delta \max_{\mathbf{y}'} Q\left(\mathbf{s}^{(d+1)}, \mathbf{y}'; \theta\right)$
  end for
  Calculate $\theta^{(k)}$ via (34)
  Update the CNN weights with $\theta^{(k)}$
end for



(a)



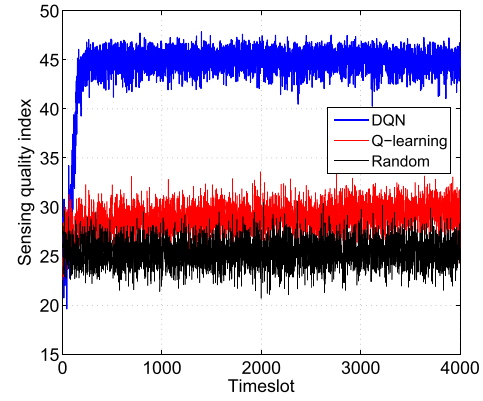
(b)

Fig. 7. Performance of the dynamic MCS game, with $L = 1$, $e = 0.1$, $\mathbf{G} = [0, -1, 9]$, $C_j^{(-1)} = 0$, $C_j^{(0)} \sim \mathrm{U}(0, 1)$, $C_j^{(1)} \sim \mathrm{U}(3, 5)$, and $\lambda_j \sim \mathrm{U}(0.8, 1)$, $\forall 1 \leq j \leq M$. (a) Average sensing quality index. (b) Average utility of the server.

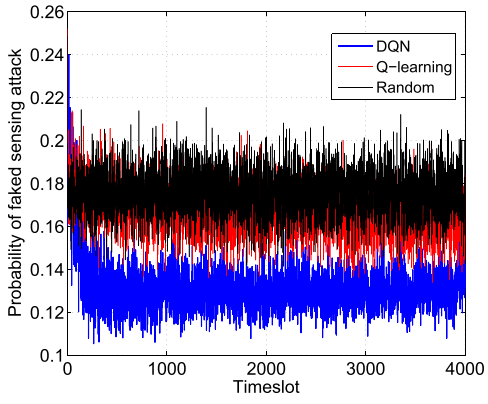
the sensing quality index and the utility of the server based on the DQN-based strategy converges at 200 time slots to 43 and 80, which are 54% and 225% higher than those of Q-learning, respectively. In addition, the faked attack rate of the DQN-based MCS system decreases by 35% to 0.13 after 200 time slots, while that of the Q-learning based strategy is 31% higher at that time.

VII. Conclusion

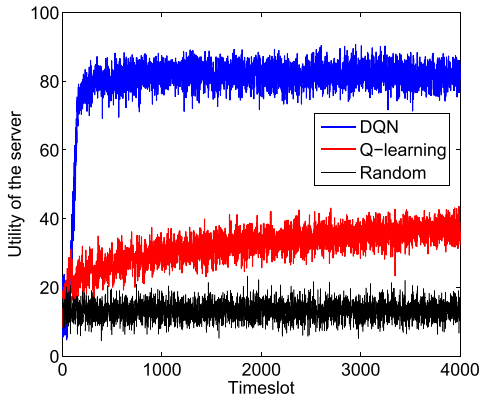
We have formulated a secure mobile crowdsensing game and provided the SEs of the static game, showing that an MCS system that pays each user according to the estimated sensing accuracy can suppress faked sensing attacks, even in the presence of sensing accuracy evaluation error. In the dynamic secure MCS game, a DQN-based MCS system has been proposed to achieve the optimal payment policy without knowing the sensing model of the users, which promotes user participation in MCS and avoids both under-sensing and over-sensing. Simulation results show that the proposed MCS system exceeds both the Q-learning strategy and the random payment strategy with a higher sensing quality, lower attack rate and higher utility of the server. In our future work, we plan to focus on implementing a realistic MCS application for traffic information collection based on the proposed payment strategy and extracting the behavior features of cheating users to achieve better performance against faked sensing attacks.



(a)



(b)



(c)

Fig. 8. Performance of the dynamic MCS game with user mobility, with $M = 60$, $L = 2$, $e = 0.1$, $\mathbf{G} = [0, -1, 9, 12]$, $p_j = 0.1$, $C_j^{(-1)} = 0$, $C_j^{(0)} \sim \mathrm{U}(0, 1)$, $C_j^{(1)} \sim \mathrm{U}(3, 5)$, $C_j^{(2)} \sim \mathrm{U}(9, 10)$ and $\lambda_j \sim \mathrm{U}(0.8, 1)$, $\forall 1 \leq j \leq M$. (a) Sensing quality index. (b) Attack probability. (c) Utility of the server.

Appendix A
Proof of Proposition 1

As $L = 1$, Eq. (1) becomes

$$\Pr(\hat{x}|x) = \begin{cases} 1 - e, & \text{if } (x, \hat{x}) = (0, 0) \text{ or } (1, 1) \\ e, & (x, \hat{x}) = (0, 1) \text{ or } (1, 0) \\ 1, & (x, \hat{x}) = (-1, -1) \\ 0, & \text{otherwise.} \end{cases} \tag{35}$$

As $e < 1/2$, by (4) and (35), we have

$$U_j(1, y^*) = (1-e)y^* - C_j^{(1)} \geq 0 = U_j(-1, y^*), \quad (36)$$

if

$$y^* \geq \frac{C_j^{(1)}}{1-e}, \quad \forall 1 \leq j \leq M; \quad (37)$$

and

$$U_j(1, y^*) = (1-e)y^* - C_j^{(1)} \geq ey^* - C_j^{(0)} = U_j(0, y^*), \quad (38)$$

if

$$y^* \geq \frac{C_j^{(1)} - C_j^{(0)}}{1-2e}, \quad \forall 1 \leq j \leq M. \quad (39)$$

By combining (37) and (39), we have

$$U_j(1, y^*) \geq \max\left(U_j(-1, y^*), U_j(0, y^*)\right), \quad (40)$$

if

$$y^* \geq \max\left(\frac{C_j^{(1)}}{1-e}, \frac{C_j^{(1)} - C_j^{(0)}}{1-2e}\right). \quad (41)$$

Thus (7) holds for $x_j^* = 1$ if (9) holds.

Similarly, by (4) and (35), we have

$$U_j(-1, y^*) = 0 \geq (1-e)y^* - C_j^{(1)} = U_j(1, y^*), \quad (42)$$

if

$$y^* \leq \frac{C_j^{(1)}}{1-e}, \quad \forall 1 \leq j \leq M; \quad (43)$$

and

$$U_j(-1, y^*) = 0 \geq ey^* - C_j^{(0)} = U_j(0, y^*), \quad (44)$$

if

$$y^* \leq \frac{C_j^{(0)}}{e}. \quad (45)$$

Thus, (7) holds for $x_j^* = -1$ if (10) holds.

Appendix B
Proof of Theorem 1

By (9), we have $\sum_m (x_m^*) > 0$, if

$$y \geq \min_{1 \leq j \leq M} \max_{x_j^* = 1} \left(\frac{C_j^{(1)}}{1-e}, \frac{C_j^{(1)} - C_j^{(0)}}{1-2e}\right). \quad (46)$$

Therefore, by (5), if

$$\max\left(\frac{C_{j^*}^{(1)}}{1-e}, \frac{C_{j^*}^{(1)} - C_{j^*}^{(0)}}{1-2e}\right) \leq y \leq \frac{\lambda_{j^*} G^{(1)}}{1-e}, \quad (47)$$

we have

$$U_s\left(x_{j^*}^*, \mathbf{y}\right) = (1-e)y^* - \lambda_{j^*} G^{(1)} \geq 0 = U_s\left(x_{j^*}^*, \mathbf{0}\right). \quad (48)$$

By (10), if

$$\min_{j \neq j^*} \left(\frac{C_j^{(1)}}{1-e}, \frac{C_j^{(0)}}{e}\right) \geq \frac{\lambda_j G^{(1)}}{1-e}, \quad (49)$$

we have $x_{-j^*}^* = -1$. Thus, if (12) holds, we have (7) for $\mathbf{x}^* = \mathbf{I}_M(j^*)$.

By (6), U_s decreases monotonically with y, indicating that y^* is the smallest positive solution to $\sum_m (x_m^*) > 0$. By (47)-(49), we see that (7) and (8) hold, indicating that (11) provides the SE of the game.

Appendix C
Proof of Theorem 3

Similarly to the proof of Theorem 1, by (16), we have $x_{j^*}^* = 2$, if

$$\begin{cases} y_2^* > \max\left(\dfrac{C_{j^*}^{(2)} - C_{j^*}^{(0)}}{1-e}, \dfrac{(1-e)C_{j^*}^{(2)} - eC_{j^*}^{(1)}}{1-2e+0.5e^2}\right) & (50a) \\ \dfrac{1-1.5e}{1-2e} y_2^* - \dfrac{C_{j^*}^{(2)} - C_{j^*}^{(1)}}{1-2e} > y_1^* > \dfrac{C_{j^*}^{(2)} - (1-e)y_2^*}{e}; & (50b) \end{cases}$$

and $x_{-j^*}^* = -1$, if

$$\begin{cases} y_1^* < \dfrac{C_{-j^*}^{(0)}}{e} & (51a) \\ y_2^* < \min\left(\dfrac{C_{-j^*}^{(1)} - (1-e)y_1^*}{0.5e}, \dfrac{C_{-j^*}^{(2)} - ey_1^*}{1-e}\right), & (51b) \end{cases}$$

where j^* is given by (22).

In addition, by Eq. (6), we have $U_s(2, \mathbf{y}^*) \geq \max\left(U_s(-1, \mathbf{y}^*), U_s(0, \mathbf{y}^*), U_s(1, \mathbf{y}^*)\right)$, if

$$\begin{cases} \lambda_{j^*} G^{(2)} \geq ey_1^* + (1-e)y_2^* & (52a) \\ \lambda_{j^*}(G^{(2)} - G^{(1)}) > (1-1.5e)y_2^* - (1-2e)y_1^* & (52b) \\ \lambda_{j^*} G^{(1)} < (1-e)y_1^* + 0.5ey_2^*, & (52c) \end{cases}$$

yielding

$$y_2^* < \frac{\lambda_{j^*}\left((1-e)G^{(2)} - eG^{(1)}\right)}{1-2e+0.5e^2}, \quad (53)$$

$$y_1^* < \frac{\lambda_{j^*}\left(0.5eG^{(2)} - (1-e)G^{(1)}\right)}{-1+2e-0.5e^2}. \quad (54)$$

Therefore, by (50a) and (53), we have

$$\frac{\lambda_{j^*}\left((1-e)G^{(2)} - eG^{(1)}\right)}{1-2e+0.5e^2} > \max_{x_{j^*}^* = 2} \left(\frac{C_{j^*}^{(2)} - C_{j^*}^{(0)}}{1-e}, \frac{(1-e)C_{j^*}^{(2)} - eC_{j^*}^{(1)}}{0.5e^2 - 2e + 1}\right). \quad (55)$$

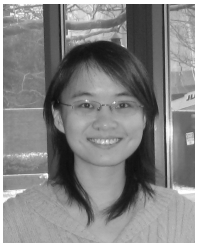
According to Eqs. (50), (51), (54) and (55), we have (21). Therefore, by (50) and (51), we have (19) and (20) after simplification.

References

[1] J. Mineraud, F. Lancerin, S. Balasubramaniam, M. Conti, and S. Tarkoma, "You are AIRing too much: Assessing the privacy of users in crowdsourcing environmental data," in *Proc. IEEE Int. Conf. Trust, Secur. Privacy Comput. Commun.*, Helsinki, Finland, Aug. 2015, pp. 523–530.

[2] R. K. Ganti, F. Ye, and H. Lei, "Mobile crowdsensing: Current state and future challenges," *IEEE Commun. Mag.*, vol. 49, no. 11, pp. 32–39, Nov. 2011.

[3] X. Hu, X. Li, E. Ngai, V. Leung, and P. Kruchten, "Multidimensional context-aware social network architecture for mobile crowdsensing," *IEEE Commun. Mag.*, vol. 52, no. 6, pp. 78–87, Jun. 2014.
[4] S. Saroiu and A. Wolman, "Enabling new mobile applications with location proofs," in *Proc. ACM Wksp. Mobile Comput. Syst. Appl.*, Santa Cruz, CA, USA, Feb. 2009, p. 3.
[5] D. Yang, G. Xue, X. Fang, and J. Tang, "Crowdsourcing to smartphones: Incentive mechanism design for mobile phone sensing," in *Proc. ACM Annu. Int. Conf. Mobile Comput. Netw.*, Istanbul, Turkey, Aug. 2012, pp. 173–183.
[6] A. K. Chorppath and T. Alpcan, "Trading privacy with incentives in mobile commerce: A game theoretic approach," *Pervasive Mobile Comput.*, vol. 9, no. 4, pp. 598–612, Aug. 2013.
[7] Z. Feng, Y. Zhu, Q. Zhang, L. M. Ni, and A. V. Vasilakos, "TRAC: Truthful auction for location-aware collaborative sensing in mobile crowdsourcing," in *Proc. IEEE Int. Conf. Comput. Commun. (INFOCOM)*, Toronto, ON, Canada, Apr. 2014, pp. 1231–1239.
[8] I. Koutsopoulos, "Optimal incentive-driven design of participatory sensing systems," in *Proc. IEEE Int. Conf. Comput. Commun. (INFOCOM)*, Turin, Italy, Apr. 2013, pp. 1402–1410.
[9] T. Luo, H.-P. Tan, and L. Xia, "Profit-maximizing incentive for participatory sensing," in *Proc. IEEE Int. Conf. Comput. Commun. (INFOCOM)*, Toronto, ON, Canada, Apr. 2014, pp. 127–135.
[10] D. Zhao, X. Li, and H. Ma, "How to crowdsource tasks truthfully without sacrificing utility: Online incentive mechanisms with budget constraint," in *Proc. IEEE Int. Conf. Comput. Commun. (INFOCOM)*, Toronto, ON, Canada, Apr. 2014, pp. 1213–1221.
[11] X. O. Wang, W. Cheng, P. Mohapatra, and T. Abdelzaher, "ARTSense: Anonymous reputation and trust in participatory sensing," in *Proc. IEEE Int. Conf. Comput. Commun. (INFOCOM)*, Turin, Italy, Apr. 2013, pp. 2517–2525.
[12] Y. Zhang and M. van der Schaar, "Reputation-based incentive protocols in crowdsourcing applications," in *Proc. IEEE Int. Conf. Comput. Commun. (INFOCOM)*, Orlando, FL, USA, Mar. 2012, pp. 2140–2148.
[13] C. Junghans, M. Karnstedt, and M. Gertz, "Quality-driven resource-adaptive data stream mining?" *ACM SIGKDD Explorations Newslett.*, vol. 13, no. 1, pp. 72–82, Jun. 2011.
[14] I. Boutsis and V. Kalogeraki, "Mobile stream sampling under time constraints," in *Proc. IEEE Int. Conf. Mobile Data Manage.*, vol. 1. Milan, Italy, Jun. 2013, pp. 227–236.
[15] G. Tychogiorgos and C. Bisdikian, "Selecting relevant sensor providers for meeting 'your' quality information needs," in *Proc. IEEE Int. Conf. Mobile Data Manage.*, vol. 1. Lulea, Sweden, Jun. 2011, pp. 200–205.
[16] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3339–3352, Aug. 2012.
[17] K. L. Huang, S. S. Kanhere, and W. Hu, "On the need for a reputation system in mobile phone based sensing," *Ad Hoc Netw.*, vol. 12, pp. 130–149, Jan. 2014.
[18] V. Mnih *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, pp. 529–533, Feb. 2015.
[19] Q. Zhao, Y. Zhu, H. Zhu, J. Cao, G. Xue, and B. Li, "Fair energy-efficient sensing task allocation in participatory sensing with smartphones," in *Proc. IEEE Int. Conf. Comput. Commun. (INFOCOM)*, Toronto, ON, Canada, Apr./May 2014, pp. 1366–1374.
[20] Q. Li and G. Cao, "Providing efficient privacy-aware incentives for mobile sensing," in *Proc. IEEE Int. Conf. Distrib. Comput. Syst.*, Madrid, Spain, Jun. 2014, pp. 208–217.
[21] Y. Zhang, Q. Chen, and S. Zhong, "Privacy-preserving data aggregation in mobile phone sensing," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 5, pp. 980–992, May 2016.
[22] B. Di, T. Wang, L. Song, and Z. Han, "Incentive mechanism for collaborative smartphone sensing using overlapping coalition formation games," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Atlanta, GA, USA, Dec. 2013, pp. 1705–1710.
[23] L. Duan, T. Kubo, K. Sugiyama, J. Huang, T. Hasegawa, and J. Walrand, "Motivating smartphone collaboration in data acquisition and distributed computing," *IEEE Trans. Mobile Comput.*, vol. 13, no. 10, pp. 2320–2333, Oct. 2014.
[24] L. Duan, A. W. Min, J. Huang, and K. G. Shin, "Attack prevention for collaborative spectrum sensing in cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 9, pp. 1658–1665, Oct. 2012.
[25] W. Wang, L. Chen, K. G. Shin, and L. Duan, "Thwarting intelligent malicious behaviors in cooperative spectrum sensing," *IEEE Trans. Mobile Comput.*, vol. 14, no. 11, pp. 2392–2405, Nov. 2015.
[26] Q. Li, Y. Li, J. Gao, B. Zhao, W. Fan, and J. Han, "Resolving conflicts in heterogeneous data by truth discovery and source reliability estimation," in *Proc. ACM Special Interest Group Manage. Data (SIGMOD)*, Snowbird, UT, USA, Jun. 2014, pp. 1187–1198.
[27] L. Xiao, J. Liu, Q. Li, and H. V. Poor, "Secure mobile crowdsensing game," in *Proc. IEEE Int. Conf. Commun. (ICC)*, London, U.K., Jun. 2015, pp. 7157–7162.
[28] D. Peng, F. Wu, and G. Chen, "Pay as how well you do: A quality based incentive mechanism for crowdsensing," in *Proc. ACM Int. Symp. Mobile Ad Hoc Netw. Comput. (MobiHoc)*, Hangzhou, China, Jun. 2015, pp. 177–186.
[29] P. Zhang, W. Zhou, L. Wu, and H. Li, "SOM: Semantic obviousness metric for image quality assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 2394–2402.
[30] J. A. D. Santos *et al.*, "Descriptor correlation analysis for remote sensing image multi-scale classification," in *Proc. IEEE Int. Conf. Pattern Recognit. (ICPR)*, Tsukuba, Tokyo, Nov. 2012, pp. 3078–3081.
[31] H. Jin, L. Su, D. Chen, K. Nahrstedt, and J. Xu, "Quality of information aware incentive mechanisms for mobile crowd sensing systems," in *Proc. ACM Int. Symp. Mobile Ad Hoc Netw. Comput. (MobiHoc)*, Hangzhou, China, Jun. 2015, pp. 167–176.
[32] B. Guo, H. Chen, Z. Yu, X. Xie, S. Huangfu, and D. Zhang, "FlierMeet: A mobile crowdsensing system for cross-space public information reposting, tagging, and sharing," *IEEE Trans. Mobile Comput.*, vol. 14, no. 10, pp. 2020–2033, Oct. 2015.
[33] X. Liu, Z. Song, E. Ngai, J. Ma, and W. Wang, "PM2:5 Monitoring using images from smartphones in participatory sensing," in *Proc. IEEE Int. Conf. Comput. Commun. (INFOCOM)*, Hong Kong, Apr./May 2015, pp. 630–635.



Liang Xiao (M'09–SM'13) received the B.S. degree in communication engineering from the Nanjing University of Posts and Telecommunications, China, in 2000, the M.S. degree in electrical engineering from Tsinghua University, China, in 2003, and the Ph.D. degree in electrical engineering from Rutgers University, New Brunswick, NJ, USA, in 2009. She was a Visiting Professor with Princeton University, Virginia Tech, and the University of Maryland, College Park. She is currently a Professor with the School of Data and Computer Science, Sun Yat-sen University, Guangzhou, and also with the Department of Communication Engineering, Xiamen University, Xiamen, China. Her research interests include wireless security, smart grids, and wireless communications. She was a recipient of the Best Paper Award at the 2016 IEEE INFOCOM Bigsecurity WS. She has served in several editorial roles, including as an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY and *IET Communications*.

Yanda Li received the B.S. degree in communication engineering from Xiamen University, Xiamen, China, in 2015, where he is currently pursuing the M.S. degree with the Department of Communication Engineering. His research interests include network security and wireless communications.

Guoan Han (S'16) received the B.S. degree in communication engineering from Southwest Jiaotong University, Chengdu, China, in 2015. He is currently pursuing the M.S. degree with the Department of Communication Engineering, Xiamen University, China. His research interests include network security and wireless communications.

Huaiyu Dai (F'17) received the B.E. and M.S. degrees in electrical engineering from Tsinghua University, Beijing, China, in 1996 and 1998, respectively, and the Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ, USA, in 2002. He was with Bell Labs, Lucent Technologies, Holmdel, NJ, USA, in summer 2000, and with AT&T Labs-Research, Middletown, NJ, USA, in summer 2001. He is currently a Professor of electrical and computer engineering with North Carolina State University, Raleigh, NC, USA. His research interests are in the general areas of communication systems and networks, advanced signal processing for digital communications, and communication theory and information theory. His current research focuses on networked information processing and crosslayer design in wireless networks, cognitive radio networks, network security, and associated information-theoretic and computation-theoretic analysis.

He has served as an Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS, *IEEE Transactions on Signal Processing*, and *IEEE Transactions on Wireless Commuications*. He is currently an Area Editor in charge of wireless communications of the IEEE TRANSACTIONS ON COMMUNICATIONS. He co-edited two special issues of EURASIP journals on distributed signal processing techniques for wireless sensor networks, and on multiuser information theory and related applications, respectively. He co-chaired the Signal Processing for Communications Symposium of the IEEE Globecom 2013, the Communications Theory Symposium of the IEEE ICC 2014, and the Wireless Communications Symposium of the IEEE Globecom 2014.

H. Vincent Poor (S'72–M'77–SM'82–F'87) received the Ph.D. degree in EECS from Princeton University in 1977. From 1977 to 1990, he was on the faculty of the University of Illinois at Urbana–Champaign. Since 1990, he has been on the faculty at Princeton, where he is currently the Michael Henry Strater University Professor of Electrical Engineering. From 2006 to 2016, he served as the Dean of Princeton's School of Engineering and Applied Science. His research interests are in the areas of information theory and signal processing, and their applications in wireless networks and related fields. Among his publications in these areas is the recent book *Information Theoretic Security and Privacy of Information Systems* (Cambridge University Press, 2017).

Dr. Poor is a member of the National Academy of Engineering and the National Academy of Sciences, and a foreign member of the Royal Society. He received the Technical Achievement and Society Awards of the IEEE Signal Processing Society in 2007 and 2011, respectively. Recent recognition of his work includes the 2017 IEEE Alexander Graham Bell Medal, and a D.Sc. *honoris causa* from Syracuse University in 2017. In 1990, he served as the President of the IEEE Information Theory Society, and in 2004–2007 as the Editor-in-Chief of the IEEE TRANSACTIONS ON INFORMATION THEORY.