Truthful Mobile Crowdsensing for Strategic Users With Private Data Quality

Xiaowen Gong, *Member, IEEE*, and Ness B. Shroff, *Fellow, IEEE*

***Abstract*— Mobile crowdsensing has found a variety of applications (e.g., spectrum sensing, environmental monitoring) by leveraging the "wisdom" of a potentially large crowd of mobile users. An important metric of a crowdsensing task is data accuracy, which relies on the data quality of the participating users' data (e.g., users' received SNRs for measuring a transmitter's transmit signal strength). However, the quality of a user can be its private information (which, e.g., may depend on the user's location) that it can manipulate to its own advantage, which can mislead the crowdsensing requester about the knowledge of the data's accuracy. This issue is exacerbated by the fact that the user can also manipulate its effort made in the crowdsensing task, which is a hidden action that could result in the requester having incorrect knowledge of the data's accuracy. In this paper, we devise truthful crowdsensing mechanisms for Quality and Effort Elicitation (QEE), which incentivize strategic users to truthfully reveal their private quality and truthfully make efforts as desired by the requester. The QEE mechanisms achieve the truthful design by overcoming the intricate dependency of a user's data on its private quality and hidden effort. Under the QEE mechanisms, we show that the crowdsensing requester's optimal (RO) effort assignment assigns effort only to the best user that has the smallest "virtual valuation", which depends on the user's quality and the quality's distribution. We also show that, as the number of users increases, the performance gap between the RO effort assignment and the socially optimal effort assignment decreases, and converges to 0 asymptotically. We further discuss some extensions of the QEE mechanisms. Simulation results demonstrate the truthfulness of the QEE mechanisms and the system efficiency of the RO effort assignment.**

***Index Terms*— Crowdsensing, truthful incentive mechanism, data quality.**

I. Introduction

MOBILE crowdsensing has found a variety of applications, such as spectrum sensing [2], and environmental monitoring (e.g., air quality [3], noise level [4], weather conditions [5] like temperature, humidity, and wind speed). In principle, crowdsensing leverages the "wisdom" of a potentially large crowd of mobile users for a crowdsensing task. The primary advantage of crowdsensing lies in that it exploits the diversity of inherently inaccurate data from many users by aggregating the data sensed by the crowd, such that the data accuracy after aggregation can be substantially enhanced. With enormous opportunities brought by big data, crowdsensing serves as an important foundation for big data learning tools to harness the power of big data in a wide range of application domains.

To fully exploit the potential of crowdsensing, it is important to assign crowdsensing tasks to users based on the quality of their data. The quality of a user's data captures the accuracy of the data relative to the actual ground truth of the event of interest, and it generally varies for different users depending on a user's specific situation (e.g., location, surrounding). For example, if the crowdsensing task is to measure the transmit signal strength of a transmitter, then the SNR received by a user from the transmitter determines the quality of the user's data, and users generally can receive distinct SNR values depending on their locations. A user can learn the quality of its data by estimation, based on its specific situation or using some history data. For example, a user can estimate the SNR received from a transmitter based on its distance from the transmitter. However, the quality of a user's data can be its private information, which is unknown to and cannot be verified by the crowdsensing requester. For example, as a user's received SNR from a transmitter depends on its location which is its private information, the requester cannot verify what is the actual received SNR of that user. As a result, a strategic user may have incentive to manipulate its quality revealed to the requester so as to gain an advantage. For example, a user with a low quality may pretend to have a high quality in the hope of receiving a high reward for contributing high quality data to the task. To the best of our knowledge, this paper is the first to design truthful mechanisms to elicit private quality from strategic users.

In addition to the quality, the effort exerted by a user in the crowdsensing task also affects the accuracy of the user's data. A user can improve its data's accuracy by making a greater effort to complete the task. For example, to measure the signal strength of a transmitter, a user can take the average of more samples of the received signal in order to combat noise. However, a user's effort can also be its hidden action that cannot be observed by the requester. Due to the inaccurate nature of the data, a strategic user may report some arbitrary data to the requester without making any effort in the task, while the requester is not able to verify how much effort was actually made.

Manuscript received March 24, 2018; revised March 10, 2019; accepted July 30, 2019; approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor W. Lou. Date of publication August 29, 2019; date of current version October 15, 2019. The work of X. Gong was supported by his startup fund provided by Auburn University. This article was presented in part at the International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt) in 2017. *(Corresponding author: Xiaowen Gong.)*

X. Gong is with the Department of Electrical and Computer Engineering, Auburn University, Auburn, AL 36830 USA (e-mail: xgong@auburn.edu).

N. B. Shroff is with the Department of Electrical and Computer Engineering and also with the Department of Computer Science and Engineering, The Ohio State University, Columbus, OH 43210 USA (e-mail: shroff.11@osu.edu).

Digital Object Identifier 10.1109/TNET.2019.2934026

1063-6692 © 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

Authorized licensed use limited to: The Ohio State University. Downloaded on May 05,2023 at 23:56:56 UTC from IEEE Xplore. Restrictions apply.

In the presence of strategic users with private quality and hidden efforts, our goal is to incentivize users to truthfully reveal their quality, and make efforts as desired by the crowdsensing requester. Such a truthful mechanism is desirable as it eliminates the possibility of manipulation, which would encourage users to participate in crowdsensing. More importantly, the truthful elicitation of quality and effort ensures that the requester can obtain the correct knowledge of the data's accuracy after aggregating the requested data, which is a critical metric of the crowdsensing task. This is in contrast to the situation of private cost, where manipulating the cost cannot mislead the requester about the data accuracy.

The jointly truthful elicitation of quality and effort calls for new designs that are significantly different from existing mechanisms. First, a user's payoff as a function of its private quality has a structure essentially different from that of its private cost. As a result, existing designs for cost elicitation cannot work for quality elicitation. Second, due to the intricate dependency of a user's data on its private quality and hidden effort, the joint elicitation of quality and effort needs to overcome the intricate coupling between the elicitation of quality and the elicitation of effort.

Given a truthful mechanism that can elicit quality and effort from users, an important question for the requester is to determine how much effort to assign to the users based on their quality, in order to maximize the requester's payoff. This involves the tradeoff between assigning more effort to improve the data's accuracy, and assigning less effort to reduce the reward paid to the users to compensate their costs.

The main contributions of this paper can be summarized as follows.

- We devise truthful crowdsensing mechanisms for Quality and Effort Elicitation (QEE), with general effort assignment functions, which incentivize strategic users to truthfully reveal their private quality and truthfully make efforts as desired by the crowdsensing requester. The QEE mechanisms achieve the truthful design by overcoming the intricate dependency of a user's data on its private quality and hidden effort.
- Under the QEE mechanisms, we characterize the crowdsensing requester's optimal (RO) effort assignment (under some condition) that maximizes the requester's expected payoff based on the distribution information of users' quality. We show that the RO effort assignment assigns effort only to the best user that has the smallest virtual valuation, which depends on the user's quality and the quality's distribution.
- For the RO effort assignment, we show that the expected requester's payoff and the social welfare both increase as the number of users increases, or the cost decreases. We also show that the performance gap of the RO effort assignment from the SO social welfare decreases as the number of users increases, and converges to 0 asymptotically. We show via numerical results that the users' payoffs attained by the RO effort assignment can decrease when the number of users increases.

The rest of this paper is organized as follows. Section II discusses related work. In Section III, we describe the system model of crowdsensing with private data quality and formulate the problems of truthful mechanism design. In Section IV, we devise truthful mechanisms for Quality and Effort Elicitation (QEE) for continuous-valued data. In Section V, we characterize the optimal effort assignment under the QEE mechanisms and devise truthful mechanism, and analyze their performance and system efficiency. Simulation results are presented in Section VII. Section VIII concludes this paper and discusses future work.

II. Related Work

Truthful Crowdsensing With Private Cost

Crowdsensing has recently attracted a lot of research interests [6]–[9]. There have been many mechanisms to incentivize users to truthfully reveal their costs in crowdsensing [10]–[13]. The cost is considered to be a strategic user's private information that it would not be willing to reveal truthfully to the user's advantage without appropriate incentive. Departing from these works, we study the setting where the quality of a user's data contributed to the crowdsensing task is the user's private information that it can manipulate. A user's payoff as a function of its private quality has an essentially different structure as that of its private cost. As a result, existing designs for cost elicitation (such as the classical VCG auction and the characterization of truthful mechanisms [14, Theorem 9.36]) cannot work for quality elicitation, so that new designs are needed. Furthermore, this paper aims at joint elicitation of quality and effort. The intricate dependency of a user's data on its private quality and hidden effort results in the intricate coupling between the elicitation of quality and the elicitation of effort, which needs to be overcome.

Mechanism Design for Hidden Actions

Mechanism design for hidden actions has been well studied in the economics literature [15], which is concerned with strategic agents that can take hidden actions that are not desired by a principal who employs the agents to work on a task. There are a few recent studies that have investigated this problem in the context of crowdsourcing [13], [16]. Cai *et al.* [16] have proposed truthful mechanisms to incentivize users to make efforts as desired in statistical estimation. However, most of the work on truthful elicitation of effort do not take into consideration agents' other possible private information (e.g., quality, cost). In a recent work [13], Luo *et al.* have made progress in this direction by providing mechanisms that not only elicit desired efforts from users but also truthful revelation of their private costs and data. This paper is different from these works as we aim to jointly elicit users' desired effort and true quality. Due to the intricate coupling between the elicitation of quality and the elicitation of effort, existing mechanism designs cannot handle the problem studied in this paper.

Quality-Aware Crowdsensing

The quality information of users is important for allocating crowdsensing tasks to the users, and has been studied in

Authorized licensed use limited to: The Ohio State University. Downloaded on May 05,2023 at 23:56:56 UTC from IEEE Xplore. Restrictions apply.

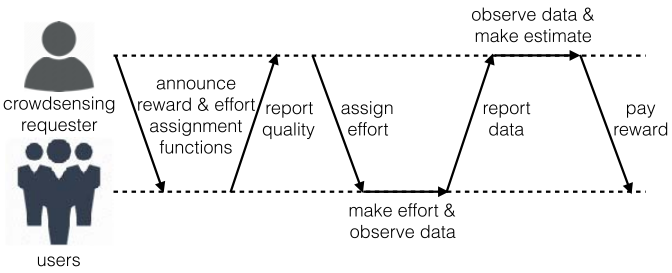

Fig. 1. Structure and procedure of the crowdsensing system.

a few works [7], [17]–[20]. One interesting line of work in this direction has focused on learning the quality of users, e.g., by exploiting the correlation of their data for the same tasks [17], [18], or allocating tasks on the fly [19]. This paper focuses on the setting where quality is a user's private information that is unknown to the requester. To the best of our knowledge, this paper is the first to design truthful mechanisms to elicit private quality from strategic users.

III. System Model and Problem Formulation

We consider a crowdsensing requester (also referred to as user 0) recruiting a set of users $\mathcal{N} \triangleq \{1, \cdots, N\}$ to work on a crowdsensing task. For convenience, let $\mathcal{N}^+ \triangleq \mathcal{N} \cup \{0\}$. The structure and procedure of the crowdsensing system is illustrated in Fig. 1 and described in detail as follows.

A. Crowdsensing With Private User Quality

Data Observation: The crowdsensing task aims to observe and estimate an unknown and random event of interest X. We consider continuous-valued data such that $X \in \mathbb{R}$ (e.g., the signal strength of a transmitter). The interested event X follows an arbitrary prior distribution which is known to the requester. Each user $i \in \mathcal{N}^+$ (i.e., including the requester) obtains random data D_i after working on the task, which is equal to X corrupted by an independent additive noise W_i, i.e.,

$$D_i \triangleq X + W_i \tag{1}$$

where

$$E[W_i] = 0, \ \text{Var}(W_i) = \frac{q_i}{e_i}. \tag{2}$$

Here the mean of W_i is assumed to be 0 without loss of generality (WLOG), and the distribution of W_i can be arbitrary. The accuracy of D_i is quantified by the variance of W_i, which is equal to the ratio of the quality q_i of user i and the effort e_i exerted by user i in the task. For ease of exposition, we assume that users' quality is within the range of $[\underline{q}, \bar{q}]$, which is known to the requester.

Worker Quality: Given the effort e_i, the quality $q_i > 0$ determines the variance $\frac{q_i}{e_i}$ of the difference $D_i - X$ (i.e., the noise W_i) which quantifies how accurate D_i is. The quality q_i is an intrinsic coefficient that captures user i's capability for the task. Note that a smaller q_i means a higher quality. The quality generally varies for different users (e.g., users can receive distinct SNRs from a transmitter based on their locations). We assume that each user $i \in \mathcal{N}^+$ knows its quality q_i (e.g., by estimating the received SNR based on the user's distance from a transmitter). However, the quality of each user $i \in \mathcal{N}$ is unknown to the requester (e.g., the received SNR from a transmitter depends on the user's location which is its private information).

Worker Effort: The effort $e_i \geq 0$ represents how much work user i devotes to the task. For example, the effort can be (approximately) the number of samples of the received signal from a transmitter taken by the user. Given the quality q_i, a higher effort e_i means a smaller variance $\frac{q_i}{e_i}$ and thus a higher accuracy of D_i. We should note that it is reasonable to model the variance of noise W_i as the function $\frac{q_i}{e_i}$ which is inversely proportional to the effort e_i (e.g., as in [13]): if each unit of user i's effort e_i is a sample of the observed event taken by user i, then when user i makes k units effort (i.e., ke_i) by taking k i.i.d. samples, it is clear that the variance of the average of the k samples is exactly $\frac{q_i}{ke_i}$. We assume that each user i can fully control its effort e_i, but it cannot be observed by the requester.

Task Assignment: The requester allocates the crowdsensing task to the users by assigning an effort e_i' that it desires each user i to exert in the task, based on the quality of the users. To this end, each user i reports its quality q_i' to the requester. Since the true quality q_i is user i's private information, it may manipulate the reported quality q_i' to its own advantage such that $q_i' \neq q_i$. Based on the quality reported by all the users, the requester determines the effort e_i' assigned to each user i according to some effort assignment function

$$e_i'(\boldsymbol{q}') \tag{3}$$

and notifies user i of its assigned effort e_i'. The effort assignment function $e_i'(\boldsymbol{q}')$ is pre-defined by the requester and announced to all the users before they report their quality. The effort assigned to a user generally varies for different users due to the diversity of their quality. Intuitively, a user with a higher quality would be assigned a larger effort. Note that in general the assigned effort e_i' is not only dependent on the quality q_i' reported by user i but also on the quality $\boldsymbol{q}'_{-i}$ reported by the other users. After being assigned effort e_i' to, each user i works on the task by making actual effort e_i. Since e_i is a hidden action of user i, it may manipulate it against the assigned effort e_i' to its own advantage such that $e_i \neq e_i'$. After obtaining data d_i from the task, which is a sample realization of the random data D_i, each user i reports data d_i' to the requester. We assume that each user i truthfully reports its data d_i (we will discuss in Remark 3 that this is a reasonable assumption).

Data Aggregation: After collecting all the data $\boldsymbol{d}$ reported by the users, the requester makes an estimate x^* of the interested event X based on $\boldsymbol{d}$. It uses the estimator that achieves the minimum mean squared error (MMSE), which is a common metric for statistical estimation [21] , i.e.,

$$x^* \triangleq \arg\min_{x'} E_{X|\boldsymbol{d}(\boldsymbol{q}',\boldsymbol{e}')}\left[(x' - X)^2\right]. \tag{4}$$

Here the estimation is based on the posteriori distribution $X|\boldsymbol{d}$ of X, which depends on the quality $\boldsymbol{q}'$ and efforts $\boldsymbol{e}'$ of the

Authorized licensed use limited to: The Ohio State University. Downloaded on May 05,2023 at 23:56:56 UTC from IEEE Xplore. Restrictions apply.

TABLE I
SUMMARY OF MAIN NOTATION

Symbol	Meaning
X	variable of interest (ground truth)
d_i	data of worker i
q_i	true quality of worker i
q_i'	reported quality of worker i
e_i	true effort of worker i
e_i'	effort assigned to worker i
x^*	optimal estimate of interested variable X
r_i	reward paid to worker i
c_i	cost coefficient of worker i
u_i	payoff of worker i
u_0	payoff of the requester

users. Then the utility of crowdsensing is represented by the estimation loss l, which is quantified by the MMSE:

$$l(\boldsymbol{d}, \boldsymbol{q}', \boldsymbol{e}') \triangleq E_{X|\boldsymbol{d}(\boldsymbol{q}', \boldsymbol{e}')} \left[(x^* - X)^2\right]. \tag{5}$$

Reward Payment.:* On the other hand, the requester pays a reward r_i to each user i for its contribution to the task, according to a certain reward function:

$$r_i(d_0, d_i, \boldsymbol{q}', e_i'). \tag{6}$$

Note that the reward r_i depends on the data d_0 observed by the requester itself. The reward function is also pre-defined by the requester, and announced to all the users before they report their quality (together with the effort assignment function $e_i'(\boldsymbol{q}')$). Note that the reward function can only depend on the information that the requester knows, i.e., d_0, $\boldsymbol{d}$, $\boldsymbol{q}'$, and $\boldsymbol{e}'$.

B. Mechanism Design Objective

Based on the crowdsensing system described above, each user i's payoff u_i is the reward r_i paid by the requester minus its cost in the task, given by,

$$u_i(d_0, d_i, \boldsymbol{q}', e_i) \triangleq r_i(d_0, d_i, \boldsymbol{q}', e_i') - c_i e_i. \tag{7}$$

Here the cost coefficient c_i represents how much resource (e.g., sensing time, energy) is consumed by user i for each unit of effort e_i devoted to the crowdsensing task. Therefore, as a linear function of e_i, the total cost $c_i e_i$ represents the total amount of resource consumed by user i in the task. Note that the weight of the total cost $c_i e_i$ relative to the reward r_i in (7) can be integrated into the cost c_i. We assume that all users have the same cost coefficient c (i.e., $c_i = c$, $\forall i$) and it is known to the requester. This assumption is reasonable when c is common knowledge that is determined by the market price (e.g., the same price is paid for each sensing sample of each user).

For the convenience of readers, we summarize the main notation used in this paper in Table I.

The requester's payoff u_0 is the crowdsensing's utility (i.e., the negative of the estimation loss l) minus the total reward paid to the users, i.e.,

$$u_0(x, \boldsymbol{d}, \boldsymbol{q}', \boldsymbol{e}') \triangleq -l(\boldsymbol{d}, \boldsymbol{q}', \boldsymbol{e}') - \sum_{i \in \mathcal{N}} r_i(d_0, d_i, \boldsymbol{q}', e_i') \tag{8}$$

As the users have private quality and make hidden efforts, if any user manipulates its reported quality or actual effort, then the estimator found by (4) would be different from the correct estimator, i.e.,

$$x^* \neq \arg\min_{x'} E_{X|\boldsymbol{d}(\boldsymbol{q}, \boldsymbol{e})} \left[(x' - X)^2\right].$$

More importantly, the estimation loss found by (5) would be different from the correct one, i.e.,

$$l(\boldsymbol{d}, \boldsymbol{q}', \boldsymbol{e}') \neq \min_{x'} E_{X|\boldsymbol{d}(\boldsymbol{q}, \boldsymbol{e})} \left[(x' - X)^2\right].$$

This means that manipulation would lead to incorrect knowledge of the requester about the estimation loss! This is undesirable since the data's accuracy is a critical performance metric that needs to be ensured to meet some requirement (e.g., a tolerance threshold for acceptable data accuracy). Note that this issue does not arise in the setting where users have private costs only, since manipulating the costs can affect only the crowdsensing's utility and the reward payment but cannot affect the requester's knowledge of the data accuracy. Furthermore, the possibility of manipulation could result in concerns that discourage users to participate in crowdsensing. Thus motivated, we aim to design a mechanism, which is a pair of an effort assignment function $e_i'(\boldsymbol{q}')$ and a reward function $r_i(d_0, d_i, \boldsymbol{q}', e_i')$, that can achieve the property of incentive compatibility as stated below.

Definition 1: A mechanism is *dominant incentive-compatible* (DIC) if, given any set of quality reported by the other users, the optimal strategy of each user i for maximizing its expected payoff is to truthfully report its quality and make the effort desired by the requester, i.e.,

$$E_{D_0, D_i(q_i, e_i)} \left[u_i(D_0, D_i, q_i, \boldsymbol{q}'_{-i}, e_i')\right] \geq E_{D_0, D_i(q_i, e_i)} \left[u_i(D_0, D_i, q_i', \boldsymbol{q}'_{-i}, e_i)\right], \forall (q_i', e_i), \forall \boldsymbol{q}'_{-i}.$$

Another natural and desirable property we aim to achieve is that each user's expected reward should at least compensate its cost (i.e., its expected payoff is nonnegative), since otherwise the user would not participate in crowdsensing for a payoff of 0. This property of individual rationality is stated as follows.

Definition 2: A mechanism is *individually rational* (IR) if for each user i, given that it truthfully reports its quality and makes the effort desired by the requester, its expected payoff is nonnegative, i.e.,

$$E_{D_0, D_i(q_i, e_i)} \left[u_i(D_0, D_i, q_i, \boldsymbol{q}'_{-i}, e_i')\right] \geq 0, \ \forall \boldsymbol{q}'_{-i}.$$

IV. TRUTHFUL QUALITY AND EFFORT ELICITATION FOR CROWDSENSING

In this section, we design truthful crowdsensing mechanisms that achieve the DIC and IR properties.

We first present the QEE mechanisms as follows.

Definition 3: A Quality and Effort Elicitation (QEE) mechanism is a pair of any effort assignment function $e_i'(\boldsymbol{q}')$ that satisfies the condition in (9) and the reward function

Authorized licensed use limited to: The Ohio State University. Downloaded on May 05,2023 at 23:56:56 UTC from IEEE Xplore. Restrictions apply.

$r_i(d_0, d_i, \boldsymbol{q}', e_i')$ given by (10) based on that $e_i'(\boldsymbol{q}')$:

$$e_i'(q_i', \boldsymbol{q}_{-i}') \geq e_i'(q_i'', \boldsymbol{q}_{-i}'), \quad \forall q_i' \leq q_i'', \ \forall \boldsymbol{q}_{-i}' \tag{9}$$

$$r_i(d_0, d_i, \boldsymbol{q}', e_i') = c\int_{q_i'}^{\bar{q}} \frac{e_i'(q, \boldsymbol{q}_{-i}')}{q} dq + 2ce_i'(\boldsymbol{q}') - \frac{ce_i'^2(\boldsymbol{q}')}{q_i'}\left[(d_0 - d_i)^2 - \frac{q_0}{e_0}\right]. \tag{10}$$

The condition in (9) captures general effort assignment functions where the effort assigned to each user increases as its quality improves, given any quality of the other users. Intuitively, these effort assignment functions are natural and desirable for system efficiency. In the following, we will show the main ideas of the design of the QEE mechanisms (9) and (10). The rationale behind the design will be explained in Remark 1.

We show how the QEE mechanisms achieve the DIC property in the following three steps. We first show how each user's expected payoff depends on its true quality and actual effort (Lemma 1). Then we show that if the elicitation of quality is achieved, the elicitation of effort is also achieved (Lemma 2). Last, given that the actual effort is optimized, we show that the elicitation of quality is achieved (Lemma 3). As a result, the DIC property is achieved (Theorem 1).

We first show that each user's expected payoff can be expressed as a function of its true quality and actual effort using (10).

Lemma 1: Under the QEE mechanisms, given that any user i reports quality q_i' and its data d_i and makes effort e_i, its expected payoff is given by

$$E_{D_0, D_i(q_i, e_i)}[u_i(D_0, D_i, \boldsymbol{q}', e_i)] = c\int_{q_i'}^{\bar{q}} \frac{e_i'(q, \boldsymbol{q}_{-i}')}{q} dq + 2ce_i'(\boldsymbol{q}') - \frac{ce_i'^2(\boldsymbol{q}')q_i}{e_i q_i'} - ce_i. \tag{11}$$

Then we show that, as user i can only affect its payoff in (11) via its reported quality q_i' and actual effort e_i, its optimal actual effort can be found as a function of q_i' using (11).

Lemma 2: Under the QEE mechanisms, given that any user i reports quality q_i' and its data d_i, its optimal actual effort is given by

$$e_i = \sqrt{\frac{q_i}{q_i'}} e_i'(\boldsymbol{q}'). \tag{12}$$

Note that the optimal actual effort e_i in (12) is equal to the desired effort e_i' when the reported quality q_i' is equal to the true quality q_i. It means that if the elicitation of quality is achieved, then the elicitation of effort is also achieved. This is a key property we need to overcome the intricate coupling between the elicitation of quality and the elicitation of effort, and achieve the elicitation of effort. Using Lemma 2, given that user i reports its data d_i and makes the optimal effort as in (12), we can express its payoff in (11) as

$$c\int_{q_i'}^{\bar{q}} \frac{e_i'(q, \boldsymbol{q}_{-i}')}{q} dq + 2ce_i'(\boldsymbol{q}') - 2c\sqrt{\frac{q_i}{q_i'}} e_i'(\boldsymbol{q}'), \tag{13}$$

by substituting (12) into (11).

Next we show that, since user i can only affect its own payoff in (13) via its reported quality q_i', its optimal q_i' is its true quality q_i, under the general condition (9) on the effort assignment function $e_i'(\boldsymbol{q}')$.

Lemma 3: Under the QEE mechanisms, given that any user i reports its data d_i and makes its optimal actual effort as (12), its optimal reported quality is its true quality $q_i' = q_i$.

Note that the optimal reported quality q_i' is always equal to the true quality q_i and is independent of e_i'. Therefore, this property achieves the elicitation of quality.

Using Lemmas 1, 2, and 3, we can show that the QEE mechanisms achieve the DIC property as in the next theorem. Given that user i reports its data d_i, makes the optimal effort $e_i = e_i'$, and reports the optimal quality $q_i' = q_i$, its payoff in (13) is given by

$$c\int_{q_i}^{\bar{q}} \frac{e_i'(q, \boldsymbol{q}_{-i}')}{q} dq. \tag{14}$$

It follows that the IR property is also achieved since (38) is no less than 0 due to the fact that $e_i'(\boldsymbol{q}') \geq 0$, $\forall \boldsymbol{q}'$.

Theorem 1: The QEE mechanisms are DIC and IR.

Remark 1: We explain the rationale behind the truthful design of the QEE mechanisms as follows. To incentivize each user i to report the true quality $q_i' = q_i$ and make the desired effort $e_i = e_i'$, the reward function r_i must depend on the true quality q_i and the actual effort e_i, since otherwise user i can manipulate q_i' and e_i without regard to q_i and e_i'. Since r_i can only be defined using the information known to the requester (i.e., d_0, d_i, $\boldsymbol{q}'$, and e_i'), and the variance of the noise W_i in data d_i is determined by the true quality q_i and the actual effort e_i, we can design r_i as a function of the squared error $(d_0 - d_i)^2$ such that the expected reward is a function of the variance and thus depends on q_i and e_i (as in Lemma 1 and (11)). Therefore, the expected reward depends only on q_i', e_i', q_i, and e_i (as in (11)). Then we can design the reward function such that the optimal actual effort e_i that maximizes user i's expected payoff is equal to e_i' when $q_i' = q_i$ (as in Lemma 2 and (12)). Given that the actual effort is optimized, the payoff only depends on q_i', e_i', and q_i (as in (13)). Next we further design the reward function such that, under the general condition on e_i' (as in (9)), the optimal reported quality q_i' that maximizes the payoff is always equal to q_i and independent of e_i' (as in Lemma 3).

Remark 2: We can see from (38) that, given the effort assignment function $e_i'(\boldsymbol{q}')$ and the users' quality q_i and $\boldsymbol{q}_{-i}'$, the user's payoff increases as the upper bound $\bar{q}$ of users' quality increases, and thus the requester's payoff decreases as $\bar{q}$ increases. This means that the requester can pay less "information rent" [22] by knowing more information (i.e., having less uncertainty) about users' quality with a smaller upper bound $\bar{q}$. In the extreme case where the requester knows that all users have the same quality (i.e., $q_i = \bar{q}$, $\forall i$), all users have 0 payoffs while the requester receives all the surplus of users' efforts, which means that the mechanism is fully "efficient" for the requester's interest.

Remark 3: We should note that, for the QEE mechanisms, it is difficult for a user to misreport its data to improve its payoff. This is because the user needs to know the requester's

Authorized licensed use limited to: The Ohio State University. Downloaded on May 05,2023 at 23:56:56 UTC from IEEE Xplore. Restrictions apply.

quality q_0 and effort e_0 (which determines the distribution of d_0). However, this is the requester's the private information that the user would not know. Therefore, it is reasonable to assume that users truthfully report their data (as we do in this paper).

V. Optimal Effort Assignment for Truthful Crowdsensing

We have shown that the DIC and IR properties can be achieved by all the QEE mechanisms which have general effort assignment functions that satisfy condition (9). We will now show how the requester can find the optimal effort assignment that maximizes its payoff, based on the distribution information of users' quality. Because of the DIC property, we assume that $\boldsymbol{q}' = \boldsymbol{q}$ and $\boldsymbol{e} = \boldsymbol{e}'$ in this section, and thus, for brevity, we use $\boldsymbol{q}$ and $\boldsymbol{e}$ instead of $\boldsymbol{q}'$ and $\boldsymbol{e}'$, respectively. For ease of analysis, we assume that the interested event X follows a normal prior distribution $\mathcal{N}(0,1)$ (i.e., with mean 0 and variance 1). We further assume that users' quality follow independent and identical uniform distributions over an interval $[\underline{q}, \bar{q}]$, which is known to the requester.

Definition 4: The *crowdsensing requester's optimal (RO) effort assignment* $\boldsymbol{e}^*(\boldsymbol{q})$ is the effort assignment function $\boldsymbol{e}(\boldsymbol{q})$ that maximizes the requester's expected payoff (8) among all the QEE mechanisms, i.e.,

$$\{\boldsymbol{e}^*(\boldsymbol{q}), \forall \boldsymbol{q}\} \triangleq \arg \max_{\{\boldsymbol{e}(\boldsymbol{q}), \forall \boldsymbol{q}\}} E_{X, \boldsymbol{D}(\boldsymbol{Q}, \boldsymbol{e})}[u_0(X, \boldsymbol{D}, \boldsymbol{Q}, \boldsymbol{e})]. \quad (15)$$

Then the optimal effort assignment can be characterized as follows.

Theorem 2: When $c \geq 1/(9\underline{q})$, the requester's optimal effort assignment (15) is given by

$$e_i^*(\boldsymbol{q}) = \begin{cases} \max\left\{ q_i \left(\frac{1}{\sqrt{\alpha(q_i)}} - 1 \right), 0 \right\}, i = \arg\min_j \alpha(q_j) \\ 0, \qquad \text{otherwise}, \end{cases} \quad (16)$$

$\forall \boldsymbol{q}$, where

$$\alpha(q_i) \triangleq c \left(\frac{F(q_i)}{f(q_i)} + q_i \right), \ \forall i \quad (17)$$

and $f(q)$ and $F(q)$ denote the probability density function (PDF) and cumulative density function (CDF) of each user's quality, respectively.

In the rest of this section, we will assume that the condition $c \geq 1/(9\underline{q})$ in Theorem 2 is satisfied so that the RO effort assignment is given in (16). This is because the characterization of the RO effort assignment under this condition and the corresponding performance analysis provide useful insights on the impact of system parameters on the performance and system efficiency. Furthermore, when this condition is not satisfied, we can still use the effort assignment in (16), and the corresponding performance can be lower bounded by that when the condition holds.

Remark 4: Theorem 2 shows that the optimal effort assignment assigns effort to only one user, and furthermore, this effort depends only on that user's quality (and its distribution) but is independent of the other users' quality. Therefore, the optimal effort assignment appears to be "single-sensing" rather than "crowdsensing". However, we should note that in fact it exploits the diversity gain of multiple users' quality, since only the "best"[1] user i that has the smallest $\alpha(q_i)$ is assigned to effort. This "single-sensing" observation is essentially because the cost functions are linear, so that the marginal gain of the requester's payoff by increasing the best user's effort is always greater than by increasing any other user's effort. It is in contrast to data estimation with no data cost [23], for which it is usually optimal to use data from multiple sources rather than only from the one source with the best quality. One attractive implication of this "single-sensing" observation is that it simplifies the implementation of crowdsensing: the requester needs to collect data only from the best user rather than a potentially large number of users.

For convenience, let q_1^* denote the quality of the best user for the RO effort assignment and $e_1^*(\boldsymbol{q})$ the RO effort assigned to the best user.

Remark 5: Theorem 2 shows that the best user is the user i with the smallest "virtual valuation" $\alpha(q_i)$ rather than the highest quality q_i, where each user i's virtual valuation depends on not only its quality q_i but also the quality's distribution $F(q_i)$ and $f(q_i)$. This implies that the range of a user's possible quality, represented by $\Delta q \triangleq \bar{q} - \underline{q}$, affects its effort assignment: given users' quality, when users have a smaller quality range, more effort is assigned to the best user due to a smaller virtual valuation. This is intuitive because a larger quality range incurs a higher payment to the user in order to truthfully elicit quality. In the special case of $\Delta q = 0$, a user's virtual valuation is equal to its quality. The concept of virtual valuation was introduced by Myerson [14] and is in the same spirit as the result here.

Remark 6: According to condition (9), the RO effort $e_1^*(\boldsymbol{q})$ assigned to the best user in (16) increases when its quality q_1^* improves. Note that no effort is assigned if its quality is too low (i.e., $\alpha(q_1^*) \geq 1$). This is because a higher quality improves the marginal utility of crowdsensing by making more effort, and thus assigns more effort to the best user. On the other hand, we can observe from (16) and (17) that $e_1^*(\boldsymbol{q})$ decreases as the cost c increases. This is due to that a larger c incurs a larger payment to compensate the higher marginal cost, which results in less effort assigned.

Next we analyze the impact of system parameters on the performance of the RO effort assignment and its system efficiency.

Proposition 1: The expected RO payoff $E_{\boldsymbol{Q}}[u_0(\boldsymbol{e}^*(\boldsymbol{Q}))]$ attained by the RO effort assignment increases as the number of users N increases, or the cost c decreases.

Remark 7: Proposition 1 shows that the requester benefits from a greater diversity gain in users' quality. This is because when there are more users, the quality of the best user is likely to be higher, which improves the crowdsensing's utility. On the other hand, a larger cost c increases the total cost and thus reduces the payoff.

[1] If there are multiple "best" users, only one of them is selected by breaking the tie randomly.

Authorized licensed use limited to: The Ohio State University. Downloaded on May 05,2023 at 23:56:56 UTC from IEEE Xplore. Restrictions apply.

The system efficiency of an effort assignment function $\boldsymbol{e}(\boldsymbol{q})$ is quantified by the social welfare v it attains, which is the crowdsensing's utility (i.e., the expected estimation loss l) minus the total cost of all users, i.e.,

$$v(\boldsymbol{e}(\boldsymbol{q})) \triangleq -E_{\boldsymbol{D}(\boldsymbol{q},\boldsymbol{e})}[l(\boldsymbol{D},\boldsymbol{q},\boldsymbol{e})] - \sum_{i\in\mathcal{N}} c_i e_i. \tag{18}$$

For the interest of system efficiency, it is desirable to achieve the optimal social welfare.

Definition 5: The *socially optimal (SO) effort assignment* $\boldsymbol{e}^{so}(\boldsymbol{q})$ is the effort assignment function $\boldsymbol{e}(\boldsymbol{q})$ that maximizes the social welfare, i.e.,

$$\boldsymbol{e}^{so}(\boldsymbol{q}) \triangleq \max_{\boldsymbol{e}(\boldsymbol{q})} v(\boldsymbol{e}(\boldsymbol{q})). \tag{19}$$

The socially optimal effort assignment can be characterized as follows.

Proposition 2: The socially optimal effort assignment is given by

$$e_i^{so}(\boldsymbol{q}) = \begin{cases} \max\left\{q_i\left(\dfrac{1}{\sqrt{cq_i}}-1\right),0\right\}, & i=\arg\min_j q_j \\ 0, & \text{otherwise.} \end{cases} \tag{20}$$

Remark 8: Proposition 2 shows that the SO effort assignment assigns effort only to the "best" user i that has the highest quality q_i. Comparing (16) and (20), we can see that the SO effort assignment is only different from the RO effort assignment in that it selects the best user and assigns the effort to it based on the highest quality rather than the smallest virtual valuation among the users. Since it can be easily seen that each user i's virtual valuation $\alpha(q_i)$ is no less than its quality q_i, the RO effort assigned to the best user is less than the SO effort assigned. This is because, although assigning more effort can improve the social welfare, it would result in a too higher payment. As a result, the RO effort assignment is not socially optimal, and the gap is essentially due to the asymmetry of users' quality information between the users and the requester.

For convenience, let q_1^{so} denote the quality of the best user for the SO effort assignment and $e_1^{so}(\boldsymbol{q})$ the SO effort assigned to the best user.

Proposition 3: The expected SO social welfare $E_{\boldsymbol{Q}}[v(e_1^{so}(\boldsymbol{Q}))]$ and social welfare $E_{\boldsymbol{Q}}[v(e_1^{*}(\boldsymbol{Q}))]$ attained by the RO effort assignment increases as the number of users N increases, or the cost c decreases.

Similar to Proposition 1, Proposition 3 shows that the social welfare also benefits from a greater diversity gain in users' quality: when there are more users, the quality of the best user is likely to improve.

Proposition 4: The gap between the expected social welfare of the RO effort assignment and the SO effort assignment $E_{\boldsymbol{Q}}[v(e_1^{so}(\boldsymbol{Q}))] - E_{\boldsymbol{Q}}[v(e_1^{*}(\boldsymbol{Q}))]$ decreases as the number of users N increases, and converges to 0 as N goes to infinity.

Remark 9: Proposition 4 shows that the performance gap between the RO effort assignment and the SO effort assignment decreases to 0 asymptotically as the number of users increases. This is because the gap between the RO effort and the SO effort assigned to the best user decreases when its quality improves, and that the crowdsensing's utility and thus the social welfare is a concave function of the effort from the best user.

VI. Extensions

In this section, we extend the QEE mechanisms to the situations where 1) workers have quadratic cost functions; 2) the requester lacks reference data for the interested variable.

A. Quadratic Cost Function

In this subsection, we design the QEE mechanisms for quadratic cost functions. We consider that each worker i's cost is a quadratic function of its effort e_i quantified by a cost coefficient c_i, i.e.,

$$\frac{1}{2}c_i e_i^2.$$

For ease of exposition, we assume that $c_i = c$, $\forall i$. We present the QEE mechanisms as follows.

Definition 6: For the quadratic cost functions, a Quality, Effort, and Data Elicitation (QEE) mechanism is any effort assignment function $e_i'(\boldsymbol{q}')$ that satisfies condition (9) and the reward function $r_i(d_0, d_i, \boldsymbol{q}', e_i')$ given by (21) based on that $e_i'(\boldsymbol{q}')$:

$$r_i(d_0,d_i,\boldsymbol{q}',e_i') = c\int_{q_i'}^{\bar{q}} \frac{e_i'^2(q,\boldsymbol{q}'_{-i})}{q}dq + \frac{3}{2}ce_i'^2(\boldsymbol{q}') - \frac{ce_i'^3(\boldsymbol{q}')}{q_i'}\left[(d_0-d_i)^2 - \frac{q_0}{e_0}\right]. \tag{21}$$

We can show that they achieve the DIC and IR properties.

Theorem 3: For the quadratic cost functions, the QEE mechanisms are DIC and IR.

The design idea and rationale of the QEE mechanisms for the quadratic cost functions are similar to those for the linear cost functions in Section IV.

Remark 10: We can see from the QEE mechanisms for the linear and quadratic cost functions that their design idea and rationale (as discussed in Section IV-A) can be used for more general cost functions, such as any n-th order cost functions $c_i e_i^n$. We can also see that workers can have diverse cost functions with different parameters and forms (e.g., linear or quadratic), in which case the reward functions of the truthful design will have different parameters and forms for different workers accordingly.

B. No Reference Data From the Requester

In the previous sections, we have assumed that the requester itself can work on the task and obtains data d_0 with quality q_0 and effort $e_0 = 1$, which are (certainly) known by the requester. The reference data d_0 and its quality q_0 are necessary information needed to achieve the truthfulness of the QEE mechanism. If the requester cannot work on the task (e.g., when it is too far away from the location of interest), we can modify the QEE mechanism to deal with this situation, described as follows.

Authorized licensed use limited to: The Ohio State University. Downloaded on May 05,2023 at 23:56:56 UTC from IEEE Xplore. Restrictions apply.

For each worker i, we pick any other worker $j \neq i$ as a *reference* worker, and define the reward function r_i as

$$r_i(\boldsymbol{q}', e_i', d_i', d_j') \tag{22}$$

where d_j' is the data reported by worker j. We are interested in a mechanism under which truthful behavior of all workers is a *Nash equilibrium*, defined as follows.

Definition 7: A mechanism achieves truthful strategies of all workers as a *Nash equilibrium* (NE) if, for each worker i, given that all other workers $j \neq i$, $\forall j$ truthfully report their quality and data, and make the effort desired by the requester, the optimal strategy of worker i for maximizing its expected payoff is also to the truthful strategy, i.e.,

$$E_{D_j, D_i(q_i, e_i)}\left[u_i(q_i, \boldsymbol{q}_{-i}, e_i', e_j', D_i, D_j)\right] \geq E_{D_j, D_i(q_i, e_i)}\left[u_i(q_i', \boldsymbol{q}_{-i}, e_i, e_j', D_i, D_j)\right], \forall (q_i', e_i), \forall \boldsymbol{q}_{-i}.$$

To deal with the lack of reference data d_0 from the requester, we modify the reward function of the QEE mechanism given in (10) by replacing d_0 with the reported data d_j' of worker i's reference worker, worker j, and replacing q_0 with worker j's quality q_j', i.e.,

$$r_i(\boldsymbol{q}', e_i', e_j', d_i, d_j) = c\int_{q_i'}^{\bar{q}} \frac{e_i'(q, \boldsymbol{q}_{-i}')}{q} dq + 2ce_i'(\boldsymbol{q}') - \frac{ce_i'^2(\boldsymbol{q}')}{q_i'}\left[(d_j - d_i)^2 - \frac{q_j'}{e_j'}\right]. \tag{23}$$

To guarantee that each worker i working on the task (i.e., $e_i > 0$) has a reference worker $j \neq i$ also working on the task (i.e., $e_j > 0$), we need to restrict the assignment function $\boldsymbol{e}'$ such that there are either at least two workers or no worker working on the task. The conditions (9) of the QEE mechanism remain the same. We can show that the modified QEE mechanism can achieve an NE where all workers behave truthfully, and also the IR property. The proof follows from the same argument as that of Theorem 1.

VII. Simulation Results

In this section, we evaluate the properties of the QEE mechanisms and its performance with the RO effort assignment using simulation results.

A. User's Payoff

To illustrate the DIC and IR properties of the QEE mechanisms, we compare a user's expected payoff when it truthfully reports its quality and makes its effort with that when it untruthfully reports its quality and/or makes its effort. We use the RO effort assignment $e_i^*(\boldsymbol{q})$ in (16) for the QEE mechanism. We set the default parameters as follows[2]: $n = 2$, $c = 0.3$, $\mu_q \triangleq (\bar{q} + \underline{q})/2 = 2$, $\Delta q = 3$, $q_1 = 1.2$, $q_2 = 2.75$.

Fig. 2 illustrates user 1's expected payoff as it reports varying quality q_1' and makes no effort, or truthful effort $e_1^*(q_1', q_2)$, or optimal effort $\sqrt{\frac{q_1}{q_1'}}e_1^*(q_1', q_2)$ (as in (12)), compared to when it truthfully reports its quality and makes its effort. Fig. 2 illustrates user 1's expected payoff as it makes varying effort e_1 and reports its actual quality q_1, or the highest quality $\underline{q}$, compared to when it truthfully reports its quality and makes its effort.. We can see that the user's payoff when its behavior is untruthful is always no greater than when truthful. Furthermore, the user's payoff gap due to untruthfulness increases when it is more untruthful (i.e., the gap between the reported quality and actual quality, or the desired effort and the actual effort is larger). This confirms that the DIC property is achieved by the QEE mechanism so that users have incentive to behave truthfully. We also observe from Figs. 2-3 that the user's payoff is always greater than 0. This confirms that the IR property is achieved by the QEE mechanism.

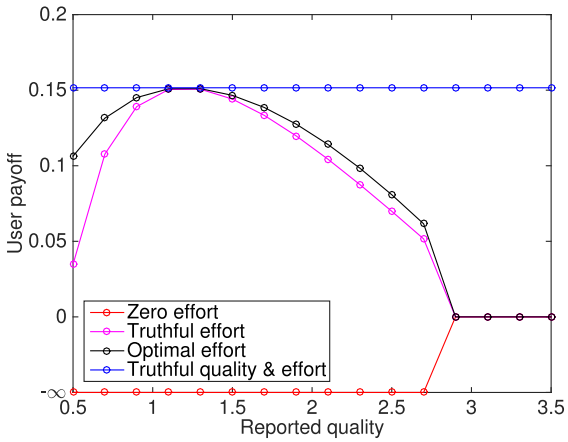

Fig. 2. Impact of reported quality q_1'.

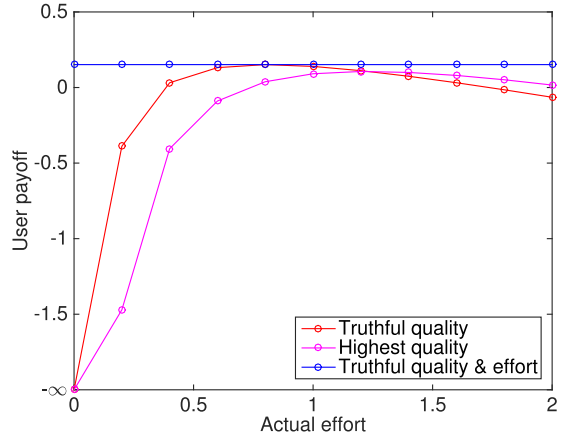

Fig. 3. Impact of actual effort e_1.

B. Requester's Payoff

To illustrate the system efficiency of the RO effort assignment, we compare the expected requester's payoff (CP) and total users' payoff (SP) attained by the RO effort assignment (CP-RO, SP-RO) with the expected social welfare (SW) attained by the SO effort assignment (SW-SO), and the expected SW attained by the RO effort assignment (SW-RO). Note that SP-RO is represented by the gap between CP-RO and SW-RO curves in figures. We set the default parameters as

[2]It is WLOG to consider 2 users only as the RO effort assignment assigns effort to the best user only based on the best user's quality.

Authorized licensed use limited to: The Ohio State University. Downloaded on May 05,2023 at 23:56:56 UTC from IEEE Xplore. Restrictions apply.

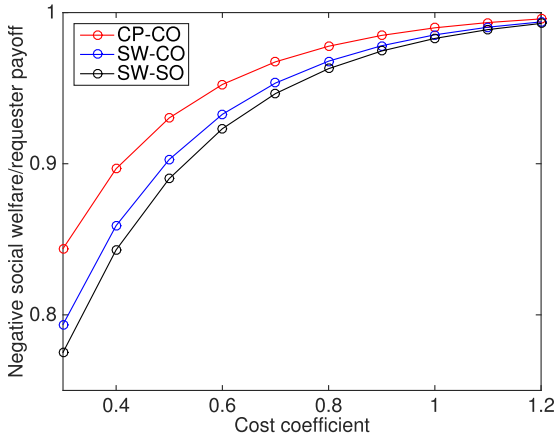

Fig. 4. Impact of cost c.

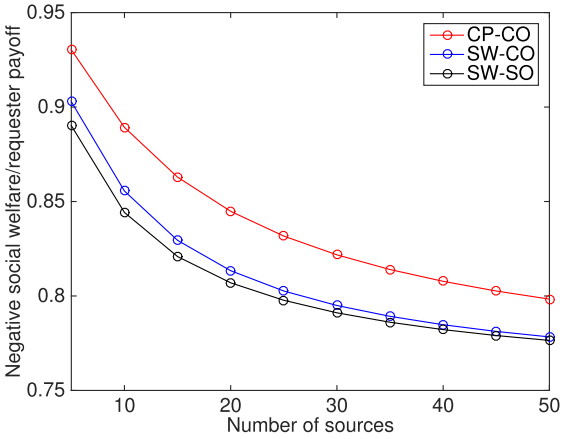

Fig. 6. Impact of the number of users N.

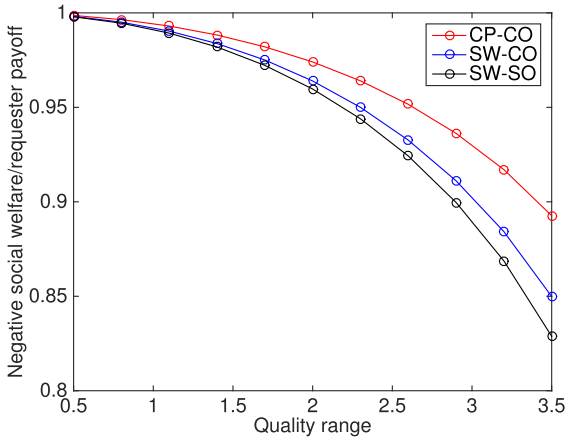

Fig. 5. Impact of quality range Δq.

follows: $N = 5$, $c = 0.5$, $\mu_q \triangleq (\bar{q} + \underline{q})/2 = 2$, $\Delta q = 3$. It can be verified that all the data points presented in the figures of this section satisfy the optimality condition $c \geq 1/(9\underline{q})$ in Theorem 2 for the RO effort assignment. For convenience, we illustrate the negative of social welfare or the requester's payoff in all figures.

Fig. 4 illustrates the impact of the cost c on the performance of CP-RO, SW-RO, and SW-SO. We observe that all the three curves and SP-RO increase as c increases, which is because higher cost results in less effort and low performance. We also observe that the gap between SW-RO and SW-SO is small when c is large. This is because a large c results in little effort so that the social welfare is close to the lower bound in which no effort is made, and thus the gap in the social welfare is also small.

Fig. 5 illustrates the impact of the quality range Δq on the performance. We observe that all the three curves are decreasing in Δq. This shows that CP and SW are concave functions of the quality, so that the increase of CP or SW at high quality is larger than the decrease of CP or SW at low quality when Δq is large. We also observe that SP-RO increases as Δq increases. This is partly due to that a larger range of possible quality would require a higher truth-eliciting reward according to (38). We further observe that the gap between SW-RO and SW-SO is small when Δq is small. This is partly because the gap between the RO and SO effort assignments is decreasing in Δq according to (16) and (20).

Fig. 6 illustrates the impact of the number of users N on the performance. We observe that all the three curves are decreasing in N, which is because they benefit from a greater diversity gain in users' quality when there are more users. We also observe that the gap between SW-RO and SW-SO is decreasing and converging to 0 when N increases, which confirms our result in Proposition 4. It is interesting to observe that, while all of CP-RO, SW-RO, and SW-SO increases as N increases, SP-RO can decrease when N increases (e.g., as SP-RO is smaller at $N = 50$ than at $N = 20$). This can be understood by examining the best user's payoff given by

$$c\int_{q_1^*}^{\bar{q}} \frac{e_1^*(q, \boldsymbol{q}_{-i})}{q} dq$$

according to (38). We can see that when N increases, q_1^* is likely to decrease which would increase the above integral. However, $\boldsymbol{q}_{-i}$ is also likely to increase as N increases, which decreases $e_1^*(q, \boldsymbol{q}_{-i})$, and this effect can outweigh the increase of q_1^* such that the integral decreases. This means that while the requester benefits from more users, the users can experience a loss due to the "competition" among the users.

VIII. Conclusion and Future Work

In this paper, we have devised the QEE mechanisms for crowdsensing, to incentivize strategic users to truthfully reveal their private quality and make efforts as desired by the crowdsensing requester. The QEE mechanisms have achieved the truthful design by overcoming the intricate coupling between the elicitation of quality and the elicitation of effort. Under the QEE mechanisms, we have analyzed the performance and system efficiency of the requester's optimal effort assignment.

For future work, one interesting direction is to consider users that have no knowledge of their quality. In this case, the requester needs to learn the quality of strategic users which may not truthfully provide data to the requester for the purpose of learning. In this paper, we have focused on the

Authorized licensed use limited to: The Ohio State University. Downloaded on May 05,2023 at 23:56:56 UTC from IEEE Xplore. Restrictions apply.

truthful quality and effort elicitation under the assumptions that users truthfully report their data and their cost is known to the requester. The truthful mechanism design when users' data and/or costs are also private information is still an open problem and will be studied in our future work.

APPENDIX

We use $f_Y(y)$ and $F_Y(y)$ to denote the probability density function (PDF) and cumulative density function (CDF) of a random variable Y, respectively.

Proof of Lemma 1

Comparing (10) and (11), it suffices to show that

$$E_{D_0,D_i(q_i,e_i)}\left[(D_0-D_i)^2\right]=\frac{q_0}{e_0}+\frac{q_i}{e_i}.$$

This follows from that

$$\begin{aligned}&E_{D_0,D_i(q_i,e_i)}\left[(D_0-D_i)^2\right]\\&=E_{X,W_0,W_i(q_i,e_i)}\left[\left((X+W_0)-(X+W_i)\right)^2\right]\\&=E_{W_0,W_i(q_i,e_i)}\left[(W_0+W_i)^2\right]\\&=\frac{q_0}{e_0}+\frac{q_i}{e_i},\end{aligned}$$

where the last equality follows from the fact that the variance of the sum of independent Gaussian random variables is equal to the sum of their variances.

Proof of Lemma 2

For brevity, we can define

$$\bar{u}_i(\boldsymbol{q}',q_i,e_i',e_i)\triangleq E_{D_0,D_i(q_i,e_i)}[u_i(D_0,D_i,\boldsymbol{q}',e_i')]\tag{24}$$

according to (11). Using (11) and (24), we have

$$\frac{\partial^2\bar{u}_i(\boldsymbol{q}',q_i,e_i',e_i)}{\partial e_i^2}=-\frac{2ce_i'^2(\boldsymbol{q}')q_i}{e_i^3q_i'}\le 0.$$

Hence the optimal actual effort e_i must satisfy

$$\frac{\partial\bar{u}_i(\boldsymbol{q}',q_i,e_i',e_i)}{\partial e_i}=\frac{ce_i'^2(\boldsymbol{q}')q_i}{e_i^2q_i'}-c=0.$$

Solving the above equation for e_i yields (12).

Proof of Lemma 3

For brevity, we can define

$$\hat{u}_i(\boldsymbol{q}',q_i,e_i')\triangleq\bar{u}_i(\boldsymbol{q}',q_i,e_i',\sqrt{\frac{q_i}{q_i'}}e_i')\tag{25}$$

according to (13) and (24). For convenience, we write $\hat{u}_i(\boldsymbol{q}',q_i,e_i')$ as $\hat{u}_i(q_i',\boldsymbol{q}'_{-i},q_i,e_i')$. According to (25), it suffices to show that $\hat{u}_i(q_i,\boldsymbol{q}'_{-i},q_i,e_i')\ge\hat{u}_i(q,\boldsymbol{q}'_{-i},q_i,e_i')$, $\forall q\ne q_i$.

Let $q_i'=a>q_i$. Using (13) and (25), we have

$$\begin{aligned}&\hat{u}_i(q_i,\boldsymbol{q}'_{-i},q_i,e_i')-\hat{u}_i(a,\boldsymbol{q}'_{-i},q_i,e_i')\\&=c\int_{q_i}^{\bar{q}}\frac{e_i'(q,\boldsymbol{q}'_{-i})}{q}dq+2ce_i'(q_i,\boldsymbol{q}'_{-i})-2c\sqrt{\frac{q_i}{q_i}}e_i'(q_i,\boldsymbol{q}'_{-i})\\&\quad-\left(c\int_a^{\bar{q}}\frac{e_i'(q,\boldsymbol{q}'_{-i})}{q}dq+2ce_i'(a,\boldsymbol{q}'_{-i})-2c\sqrt{\frac{q_i}{a}}e_i'(a,\boldsymbol{q}'_{-i})\right)\\&=c\int_{q_i}^{a}\frac{e_i'(q,\boldsymbol{q}'_{-i})}{q}dq-2ce_i'(a,\boldsymbol{q}'_{-i})\left(1-\sqrt{\frac{q_i}{a}}\right)\\&\ge ce_i'(a,\boldsymbol{q}'_{-i})\int_{q_i}^{a}\frac{1}{q}dq-2ce_i'(a,\boldsymbol{q}'_{-i})\left(1-\sqrt{\frac{q_i}{a}}\right)\\&=ce_i'(a,\boldsymbol{q}'_{-i})\left[\ln\frac{a}{q_i}-2\left(1-\sqrt{\frac{q_i}{a}}\right)\right]\ge 0\end{aligned}$$

where the first inequality follows from condition (9) and the second inequality follows from the following lemma.

Lemma 4: $g(y)\triangleq\ln y-2(1-\sqrt{\frac{1}{y}})\ge 0$, $\forall y\ge 1$.

Proof: For any $y\ge 1$, we have

$$\frac{\partial g(y)}{\partial y}=\frac{1}{y}-\frac{1}{y\sqrt{y}}=\frac{1}{y}\left(1-\frac{1}{\sqrt{y}}\right)\ge 0.$$

Then since we observe that $g(1)=0$, we have $g(y)\ge 0$, $\forall y\ge 1$. □

Let $q_i'=b<q_i$. Using (13) and (25), we have

$$\begin{aligned}&\hat{u}_i(b,\boldsymbol{q}'_{-i},q_i,e_i')-\hat{u}_i(q_i,\boldsymbol{q}'_{-i},q_i,e_i')\\&=c\int_b^{\bar{q}}\frac{e_i'(q,\boldsymbol{q}'_{-i})}{q}dq+2ce_i'(b,\boldsymbol{q}'_{-i})-2c\sqrt{\frac{q_i}{b}}e_i'(b,\boldsymbol{q}'_{-i})\\&\quad-\left(c\int_{q_i}^{\bar{q}}\frac{e_i'(q,\boldsymbol{q}'_{-i})}{q}dq+2ce_i'(q_i,\boldsymbol{q}'_{-i})-2c\sqrt{\frac{q_i}{q_i}}e_i'(q_i,\boldsymbol{q}'_{-i})\right)\\&=c\int_b^{q_i}\frac{e_i'(q,\boldsymbol{q}'_{-i})}{q}dq+2ce_i'(b,\boldsymbol{q}'_{-i})\left(1-\sqrt{\frac{q_i}{b}}\right)\\&\le ce_i'(b,\boldsymbol{q}'_{-i})\int_b^{q_i}\frac{1}{q}dq+2ce_i'(a,\boldsymbol{q}'_{-i})\left(1-\sqrt{\frac{q_i}{b}}\right)\\&=ce_i'(b,\boldsymbol{q}'_{-i})\left[\ln\frac{q_i}{b}+2\left(1-\sqrt{\frac{q_i}{b}}\right)\right]\le 0\end{aligned}$$

where the first inequality follows from condition (9) and the second inequality follows from the following lemma.

Lemma 5: $g(y)\triangleq\ln y+2(1-\sqrt{y})\le 0$, $\forall y\ge 1$.

Proof: For any $y\ge 1$, we have

$$\frac{\partial g(y)}{\partial y}=\frac{1}{y}-\frac{1}{\sqrt{y}}=\frac{1}{\sqrt{y}}\left(\frac{1}{\sqrt{y}}-1\right)\le 0.$$

Then since we observe that $g(1)=0$, we have $g(y)\le 0$, $\forall y\ge 1$. □

Proof of Theorem 1

As the IR property has been proved using (38), we only show that the DIC property is achieved. Choose and fix any (q_i',e_i). It follows from Lemma 2 that

$$\bar{u}_i(q_i',e_i)\le\bar{u}_i(q_i',\sqrt{\frac{q_i}{q_i'}}e_i').$$

Authorized licensed use limited to: The Ohio State University. Downloaded on May 05,2023 at 23:56:56 UTC from IEEE Xplore. Restrictions apply.

Then it follows from Lemma 3 that

$$\bar{u}_i(q_i', \sqrt{\frac{q_i}{q_i'}}e_i') \le \bar{u}_i(q_i, \sqrt{\frac{q_i}{q_i}}e_i') = \bar{u}_i(q_i, e_i').$$

Hence we conclude that

$$\bar{u}_i(q_i', e_i) \le \bar{u}_i(q_i, e_i').$$

Proof of Theorem 2

We first express the estimation loss l as a function of the quality $\boldsymbol{q}$ and efforts $\boldsymbol{e}$ only, and then accordingly find the optimal effort assignment $\boldsymbol{e}^*$ for the requester's expected payoff as a function of the quality $\boldsymbol{q}$, and lastly show that the optimal effort assignment $\boldsymbol{e}^*(\boldsymbol{q})$ satisfies condition (9) when $c \ge 1/(9\underline{q})$.

The requester's posterior distribution of the interested event X conditioned on all the data $\boldsymbol{d}$ collected from the users is given by

$$X|\boldsymbol{d}(\boldsymbol{q},\boldsymbol{e}) \sim \mathcal{N}\left(\frac{\sum_{i\in\mathcal{N}} d_i e_i/q_i}{1+\sum_{i\in\mathcal{N}} e_i/q_i}, \frac{1}{1+\sum_{i\in\mathcal{N}} e_i/q_i}.\right)$$

Hence the MMSE estimator defined in (4) is given by [21]

$$x^* = \frac{\sum_{i\in\mathcal{N}} d_i e_i/q_i}{1+\sum_{i\in\mathcal{N}} e_i/q_i},$$

and the corresponding estimation loss defined in (5) is given by

$$l(\boldsymbol{d},\boldsymbol{q},\boldsymbol{e}) = \frac{1}{1+\sum_{i\in\mathcal{N}} e_i/q_i} \tag{26}$$

which can be written in short as $l(\boldsymbol{q},\boldsymbol{e})$.

Using (10) and (11), the expected reward paid to user i is given by

$$\begin{aligned} &E_{X,D_i(q_i,e_i)}[r_i(X,D_i,\boldsymbol{q},e_i)] \\ &\quad = c\int_{q_i}^{\bar{q}} \frac{e_i(q,\boldsymbol{q}_{-i})}{q}dq + 2ce_i(\boldsymbol{q}) - \frac{ce_i^2(\boldsymbol{q})}{q_i}\frac{q_i}{e_i(\boldsymbol{q})} \\ &\quad = c\int_{q_i}^{\bar{q}} \frac{e_i(q,\boldsymbol{q}_{-i})}{q}dq + ce_i(\boldsymbol{q}). \end{aligned} \tag{27}$$

It follows from (26) and (27) that the expected requester's payoff defined in (8) is given by

$$\begin{aligned} &E_{X,\boldsymbol{D}(\boldsymbol{q},\boldsymbol{e})}[u_0(X,\boldsymbol{D},\boldsymbol{q},\boldsymbol{e}(\boldsymbol{q}))] \\ &\quad = -\frac{1}{1+\sum_{i\in\mathcal{N}} e_i(\boldsymbol{q})/q_i} \\ &\qquad -\sum_{i\in\mathcal{N}}\left(c\int_{q_i}^{\bar{q}} \frac{e_i(q,\boldsymbol{q}_{-i})}{q}dq + ce_i(\boldsymbol{q})\right). \end{aligned} \tag{28}$$

For brevity, define

$$\bar{u}_0(\boldsymbol{e}(\boldsymbol{q})) \triangleq E_{X,\boldsymbol{D}(\boldsymbol{q},\boldsymbol{e})}[u_0(X,\boldsymbol{D},\boldsymbol{q},\boldsymbol{e}(\boldsymbol{q}))].$$

Hence the optimal effort assignment defined in (15) is given by

$$\{\boldsymbol{e}^*(\boldsymbol{q}),\forall\boldsymbol{q}\} = \arg\max_{\{\boldsymbol{e}(\boldsymbol{q}),\forall\boldsymbol{q}\}} E_{\boldsymbol{Q}}[\bar{u}_0(\boldsymbol{e}(\boldsymbol{Q}))]. \tag{29}$$

Since we observe that

$$\begin{aligned} &E_{Q_i}\left[c\int_{Q_i}^{\bar{q}} \frac{e_i(q,\boldsymbol{q}_{-i})}{q}dq\right] \\ &= \int_{\underline{q}}^{\bar{q}} f(q')c\int_{q'}^{\bar{q}} \frac{e_i(q,\boldsymbol{q}_{-i})}{q}dqdq' \\ &= \left[F(q')c\int_{q'}^{\bar{q}} \frac{e_i(q,\boldsymbol{q}_{-i})}{q}dq\right]_{q'=\underline{q}}^{q'=\bar{q}} + \int_{\underline{q}}^{\bar{q}} F(q')c\frac{e_i(q',\boldsymbol{q}_{-i})}{q'}dq' \\ &= \int_{\underline{q}}^{\bar{q}} f(q')\frac{F(q')}{f(q')}c\frac{e_i(q',\boldsymbol{q}_{-i})}{q'}dq' = E_{Q_i}\left[c\frac{F(Q_i)}{f(Q_i)}\frac{e_i(Q_i,\boldsymbol{q}_{-i})}{Q_i}\right] \end{aligned} \tag{30}$$

where the second equality follows from integration by parts, then using (28) we have

$$\begin{aligned} &E_{\boldsymbol{Q}}[\bar{u}_0(\boldsymbol{e}(\boldsymbol{Q}))] \\ &= -E_{\boldsymbol{Q}}\left[\frac{1}{1+\sum_{i\in\mathcal{N}} e_i(\boldsymbol{Q})/Q_i}\right] \\ &\quad -\sum_{i\in\mathcal{N}} E_{\boldsymbol{Q}}\left[c\int_{Q_i}^{\bar{q}} \frac{e_i(q,\boldsymbol{Q}_{-i})}{q}dq + ce_i(\boldsymbol{Q})\right] \\ &= E_{\boldsymbol{Q}}\left[-\frac{1}{1+\sum_{i\in\mathcal{N}} e_i(\boldsymbol{Q})/Q_i} - \sum_{i\in\mathcal{N}} ce_i(\boldsymbol{Q})\left(\frac{F(Q_i)}{f(Q_i)Q_i}+1\right)\right] \end{aligned} \tag{31}$$

where the second equality follows from (30). Hence finding $\{\boldsymbol{e}^*(\boldsymbol{q}),\forall\boldsymbol{q}\}$ in (29) is equivalent to solving the following problem for each $\boldsymbol{q}$ independently:

$$\max_{\boldsymbol{e}(\boldsymbol{q})} -\frac{1}{1+\sum_{i\in\mathcal{N}} e_i(\boldsymbol{q})/q_i} - \sum_{i\in\mathcal{N}} ce_i(\boldsymbol{q})\left(\frac{F(q_i)}{f(q_i)q_i}+1\right). \tag{32}$$

Substituting

$$\bar{e}_i(\boldsymbol{q}) \triangleq \frac{e_i(\boldsymbol{q})}{q_i}, \forall i$$

into (32), we have

$$\max_{\bar{\boldsymbol{e}}(\boldsymbol{q})} -\frac{1}{1+\sum_{i\in\mathcal{N}} \bar{e}_i(\boldsymbol{q})} - \sum_{i\in\mathcal{N}} \alpha(q_i)\bar{e}_i(\boldsymbol{q}). \tag{33}$$

We can see that the optimal solution of (33) must satisfy $\bar{e}_k(\boldsymbol{q}) > 0$ for $k \triangleq \arg\min_i \alpha(q_i)$ and $\bar{e}_i(\boldsymbol{q}) = 0$, $\forall i \ne k$. Hence we can find $\bar{e}_k(\boldsymbol{q})$ by solving

$$\max_{\bar{e}_k(\boldsymbol{q})} -\frac{1}{1+\bar{e}_k(\boldsymbol{q})} - \alpha(q_k)\bar{e}_k(\boldsymbol{q}),$$

which yields

$$\bar{e}_k(\boldsymbol{q}) = \max\left\{\frac{1}{\sqrt{\alpha(q_k)}} - 1, 0\right\}.$$

Hence we have (16).

To show that $\boldsymbol{e}^*(\boldsymbol{q})$ satisfies condition (9) when $c \ge 1/(9\underline{q})$, it suffices to show that $e_1^*(q_1^*, \boldsymbol{q}_{-1})$ is decreasing in q_1^* when $c \ge 1/(9\underline{q})$. Observe that we have

$$\frac{\partial\alpha(q_i)}{\partial q_i} = \frac{\partial c\left(\frac{F(q_i)}{f(q_i)} + q_i\right)}{\partial q_i} = \frac{\partial c(2q_i - \underline{q})}{\partial q_i} = 2c.$$

Authorized licensed use limited to: The Ohio State University. Downloaded on May 05,2023 at 23:56:56 UTC from IEEE Xplore. Restrictions apply.

Suppose $e_1^*(\boldsymbol{q}) = q_1^*(\frac{1}{\sqrt{\alpha(q_1^*)}} - 1) \geq 0$. Then we have

$$\frac{\partial e_1^*(\boldsymbol{q})}{\partial q_1^*} = \frac{1}{\sqrt{\alpha(q_1^*)}} - 1 - \frac{cq_1^*}{\alpha(q_1^*)\sqrt{\alpha(q_1^*)}} = \frac{\alpha(q_1^*)(1-\sqrt{\alpha(q_1^*)}) - cq_1^*}{\alpha(q_1^*)\sqrt{\alpha(q_1^*)}} \leq 0$$

where the inequality follows from the following lemma.

Lemma 6: For $c \geq 1/(9\underline{q})$, $g(q) \triangleq c(2q-\underline{q})(1-\sqrt{c(2q-\underline{q})}) - cq \leq 0$, $\forall q \geq \underline{q}$.

Proof: For any $q \geq \underline{q}$, we have

$$\frac{\partial g(q)}{\partial q} = 2c - 2c\sqrt{c(2q-\underline{q})} - \frac{c^2(2q-\underline{q})}{\sqrt{c(2q-\underline{q})}} - c = c(1-3\sqrt{c(2q-\underline{q})}) \leq c(1-3\sqrt{c\underline{q}}) \leq 0$$

where the inequality follows from the condition $c \geq 1/(9\underline{q})$. Then since we observe that

$$g(\underline{q}) = c\underline{q}(1-\sqrt{c\underline{q}}) - c\underline{q} \leq c\underline{q} - c\underline{q} = 0,$$

we have $g(q) \leq 0$, $\forall q \geq \underline{q}$. □

Proof of Proposition 1

We first show the claims for $E_{\boldsymbol{Q}}[u_0(\boldsymbol{e}^*(\boldsymbol{Q}))]$. Suppose $e_1^*(\boldsymbol{q}) = q_1^*(\frac{1}{\sqrt{\alpha(q_1^*)}} - 1) \geq 0$. Substituting (16) into (31), we have

$$\begin{aligned} &E_{\boldsymbol{Q}}[u_0(\boldsymbol{e}^*(\boldsymbol{Q}))] \\ &= E_{\boldsymbol{Q}}\left[-\frac{1}{1+e_1^*(\boldsymbol{Q})/Q_1^*} - ce_1^*(\boldsymbol{Q})\left(\frac{F(Q_1^*)}{f(Q_1^*)Q_1^*}+1\right)\right] \\ &= E_{Q_1^*}\left[-\frac{1}{1+e_1^*(Q_1^*)/Q_1^*} - \frac{\alpha(Q_1^*)e_1^*(Q_1^*)}{Q_1^*}\right] \\ &= E_{Q_1^*}\left[-\sqrt{\alpha(Q_1^*)} - \alpha(Q_1^*)\left(\frac{1}{\sqrt{\alpha(Q_1^*)}}-1\right)\right] \\ &= E_{Q_1^*}\left[-2\sqrt{\alpha(Q_1^*)} + \alpha(Q_1^*)\right] \end{aligned}$$

where the second equality follows from that $e_1^*(\boldsymbol{q}) = e_1^*(q_1^*)$. Since

$$\frac{\partial(-2\sqrt{\alpha(q_1^*)}+\alpha(q_1^*))}{\partial q_1^*} = \frac{\partial(-2\sqrt{\alpha(q_1^*)}+\alpha(q_1^*))}{\partial \alpha(q_1^*)}\frac{\partial \alpha(q_1^*)}{\partial q_1^*} = 2c\left(1-\frac{1}{\sqrt{\alpha(q_1^*)}}\right) \leq 0,$$

and that the best quality $Q_1^*(N)$ for N users stochastically dominates the best quality $Q_1^*(N')$ for N' users for any $N' > N$, i.e.,

$$Q_1^*(N) \geq_{st} Q_1^*(N'),\ \forall N' > N, \tag{34}$$

it follows that $E_{\boldsymbol{Q}}[u_0(e_1^*(\boldsymbol{Q}))]$ is increasing in N. Since

$$\frac{\partial(-2\sqrt{\alpha(q_1^*)}+\alpha(q_1^*))}{\partial c} = \frac{\partial(-2\sqrt{\alpha(q_1^*)}+\alpha(q_1^*))}{\partial \alpha(q_1^*)}\frac{\partial \alpha(q_1^*)}{\partial c} = 2q\left(1-\frac{1}{\sqrt{\alpha(q_1^*)}}\right) \leq 0,$$

it follows that $E_{\boldsymbol{Q}}[u_0(e_1^*(\boldsymbol{Q}))]$ is decreasing in c.

Proof of Proposition 2

It follows from (26) that finding the SO effort assignment $\boldsymbol{e}^{so}(\boldsymbol{q})$ is equivalent to solving the following problem for each $\boldsymbol{q}$ independently:

$$\max_{\boldsymbol{e}(\boldsymbol{q})} -\frac{1}{1+\sum_{i\in\mathcal{N}} e_i(\boldsymbol{q})/q_i} - \sum_{i\in\mathcal{N}} ce_i(\boldsymbol{q}).$$

The above problem can be solved using a similar argument as that of solving problem (32) in the proof of Theorem 2, which yields (20).

Proof of Proposition 3

Suppose $e_1^{so}(\boldsymbol{q}) = q_1^{so}(\frac{1}{\sqrt{cq_1^{so}}} - 1) \geq 0$. Using (26), substituting (20) into (18), we have

$$\begin{aligned} v(\boldsymbol{e}^{so}(\boldsymbol{q})) &= -\frac{1}{1+e_1^{so}(\boldsymbol{q})/q_1^{so}} - ce_1^{so}(\boldsymbol{q}) \\ &= -\sqrt{cq_1^{so}} - cq_1^{so}\left(\frac{1}{\sqrt{cq_1^{so}}}-1\right) \\ &= -2\sqrt{cq_1^{so}} + cq_1^{so}. \end{aligned} \tag{35}$$

Since

$$\frac{\partial(-2\sqrt{cq_1^{so}}+cq_1^{so})}{\partial q_1^{so}} = c\left(1-\frac{1}{\sqrt{cq_1^{so}}}\right) \leq 0,$$

it follows from (34) and $q_1^* = q_1^{so}$ that $E_{\boldsymbol{Q}}[v(\boldsymbol{e}^{so}(\boldsymbol{Q}))] = E_{Q_1^{so}}[v(\boldsymbol{e}^{so}(\boldsymbol{Q}))]$ is increasing in N. Similarly, we have

$$\frac{\partial(-2\sqrt{cq_1^{so}}+cq_1^{so})}{\partial c} = q\left(1-\frac{1}{\sqrt{cq_1^{so}}}\right) \leq 0,$$

and it follows that $E_{\boldsymbol{Q}}[v(\boldsymbol{e}^{so}(\boldsymbol{Q}))]$ is decreasing in c.

Suppose $e_1^*(\boldsymbol{q}) = q_1^*\left(\frac{1}{\sqrt{\alpha(q_1^*)}} - 1\right) \geq 0$. Using (26), substituting (16) into (18), we have

$$\begin{aligned} v(\boldsymbol{e}^*(\boldsymbol{q})) &= -\frac{1}{1+e_1^*(\boldsymbol{q})/q_1^*} - ce_1^*(\boldsymbol{q}) \\ &= -\sqrt{\alpha(q_i)} - \frac{cq_1^*}{\sqrt{\alpha(q_i)}} + cq_1^*. \end{aligned} \tag{36}$$

Since

$$\begin{aligned} &\frac{\partial(-\sqrt{\alpha(q_i)} - \frac{cq_1^*}{\sqrt{\alpha(q_i)}} + cq_1^*)}{\partial q_1^*} \\ &= -\frac{2c}{\sqrt{\alpha(q_1^*)}} + \frac{c^2q_1^*}{\alpha(q_1^*)\sqrt{\alpha(q_1^*)}} + c \\ &= -c\left[\left(\frac{1}{\sqrt{\alpha(q_1^*)}}-1\right) + \frac{1}{\sqrt{\alpha(q_1^*)}}\left(1-\frac{cq_1^*}{\alpha(q_1^*)}\right)\right] \leq 0, \end{aligned}$$

where the inequality follows from that $\alpha(q_1^*) = c(2q_1^* - \underline{q}) \geq cq_1^*$, it follows from (34) that $E_{\boldsymbol{Q}}[v(\boldsymbol{e}^*(\boldsymbol{Q}))] = E_{Q_1^*}[v(\boldsymbol{e}^*(\boldsymbol{Q}))]$ is increasing in N.

Authorized licensed use limited to: The Ohio State University. Downloaded on May 05,2023 at 23:56:56 UTC from IEEE Xplore. Restrictions apply.

Similarly, we have

$$\frac{\partial(-\sqrt{\alpha(q_i)}-\frac{cq_1^*}{\sqrt{\alpha(q_i)}}+cq_1^*)}{\partial c}$$
$$=-\frac{2q_1^*}{\sqrt{\alpha(q_1^*)}}+\frac{{q_1^*}^2c}{\alpha(q_1^*)\sqrt{\alpha(q_1^*)}}+q_1^*$$
$$=-q_1^*\left[\left(\frac{1}{\sqrt{\alpha(q_1^*)}}-1\right)+\frac{1}{\sqrt{\alpha(q_1^*)}}\left(1-\frac{cq_1^*}{\alpha(q_1^*)}\right)\right]\le 0,$$

and it follows that $E_{\boldsymbol{Q}}[v(\boldsymbol{e}^*(\boldsymbol{Q}))]$ is decreasing in c.

Proof of Proposition 4

Using (35) and (36), we have

$$v(\boldsymbol{e}^{so}(\boldsymbol{Q}))-v(\boldsymbol{e}^*(\boldsymbol{Q}))=\sqrt{\alpha(q_1^*)}+\frac{cq_1^*}{\sqrt{\alpha(q_1^*)}}-2\sqrt{cq_1^*} \quad (37)$$

where we use the fact that $q_1^*=q_1^{so}$. Then we have

$$\frac{\partial\left(\sqrt{\alpha(q_1^*)}+\frac{cq_1^*}{\sqrt{\alpha(q_1^*)}}-2\sqrt{cq_1^*}\right)}{\partial q_1^*}=\frac{\partial\frac{\left(\sqrt{\alpha(q_1^*)}-\sqrt{cq_1^*}\right)^2}{\sqrt{\alpha(q_1^*)}}}{\partial q_1^*}$$
$$=-\frac{c}{\alpha(q_1^*)\sqrt{\alpha(q_1^*)}}\left(\sqrt{\alpha(q_1^*)}-\sqrt{cq_1^*}\right)^2+\frac{1}{\sqrt{\alpha(q_1^*)}}$$
$$2\left(\sqrt{\alpha(q_1^*)}-\sqrt{cq_1^*}\right)\left(\frac{c}{\sqrt{\alpha(q_1^*)}}-\frac{c}{2\sqrt{cq_1^*}}\right)$$
$$=\frac{c\left(\sqrt{\alpha(q_1^*)}-\sqrt{cq_1^*}\right)\left(\sqrt{\alpha(q_1^*)cq_1^*}+cq_1^*-\alpha(q_1^*)\right)}{\alpha(q_1^*)\sqrt{\alpha(q_1^*)cq_1^*}}\ge 0$$

where the inequality follows from

$$\sqrt{\alpha(q_1^*)}-\sqrt{cq_1^*}=\sqrt{c(2q_1^*-\underline{q})}-\sqrt{cq_1^*}\ge 0$$

and

$$\sqrt{\alpha(q_1^*)cq_1^*}+cq_1^*-\alpha(q_1^*)=\sqrt{c(2q_1^*-\underline{q})cq_1^*}+cq_1^*-c(2q_1^*-\underline{q})$$
$$\ge c\sqrt{(q_1^*-\underline{q})(q_1^*-\underline{q})}-c(q_1^*-\underline{q})=0.$$

Then it follows from (34) that $E_{\boldsymbol{Q}}[v(\boldsymbol{e}^{so}(\boldsymbol{Q}))]-E_{\boldsymbol{Q}}[v(\boldsymbol{e}^*(\boldsymbol{Q}))]=E_{\boldsymbol{Q}}[v(\boldsymbol{e}^{so}(\boldsymbol{Q}))-v(\boldsymbol{e}^*(\boldsymbol{Q}))]$ is decreasing in N. Furthermore, using (37) we have

$$\lim_{N\to\infty}\left(E_{\boldsymbol{Q}}[v(\boldsymbol{e}^{so}(\boldsymbol{Q}))]-E_{\boldsymbol{Q}}[v(\boldsymbol{e}^*(\boldsymbol{Q}))]\right)$$
$$=\lim_{N\to\infty}E_{\boldsymbol{Q}}[v(\boldsymbol{e}^{so}(\boldsymbol{Q}))-v(\boldsymbol{e}^*(\boldsymbol{Q}))]$$
$$=\lim_{N\to\infty}E_{Q_1^*}\left[\sqrt{\alpha(Q_1^*)}+\frac{cQ_1^*}{\sqrt{\alpha(Q_1^*)}}-2\sqrt{cQ_1^*}\right]$$
$$=\lim_{q_1^*\to\underline{q}}\left(\sqrt{\alpha(q_1^*)}+\frac{cq_1^*}{\sqrt{\alpha(q_1^*)}}-2\sqrt{cq_1^*}\right)$$
$$=\sqrt{\alpha(\underline{q})}+\frac{c\underline{q}}{\sqrt{\alpha(\underline{q})}}-2\sqrt{c\underline{q}}$$
$$=\sqrt{c\underline{q}}+\frac{c\underline{q}}{\sqrt{c\underline{q}}}-2\sqrt{c\underline{q}}=0,$$

where the third equality follows from that

$$\lim_{N\to\infty}f_{Q_1^*(N)}(\underline{q})=\infty,$$

and

$$\lim_{N\to\infty}f_{Q_1^*(N)}(q)=0,\ \forall q\ne\underline{q}.$$

Proof of Theorem 3

The proof is similar to the proofs of Lemmas 1, 2, and 3 and Theorem 1, and we present the differences as follows. We first show that the DIC property is achieved. Similar to the proof of Lemma 2, given that any user i reports any quality q_i' and truthfully report its data d_i, its optimal effort to make is

$$e_i=\sqrt[3]{\frac{q_i}{q_i'}}e_i'(\boldsymbol{q}').$$

Similar to the proof of Lemma 3, for $q_i'=a>q_i$, we have

$$\hat{u}_i(q_i,\boldsymbol{q}'_{-i},q_i,e_i')-\hat{u}_i(a,\boldsymbol{q}'_{-i},q_i,e_i')$$
$$=c\int_{q_i}^{a}\frac{e_i'(q,\boldsymbol{q}'_{-i})}{q}dq-\frac{3}{2}ce_i'(a,\boldsymbol{q}'_{-i})\left(1-\left(\frac{q_i}{a}\right)^{\frac{2}{3}}\right)$$
$$\ge ce_i'(a,\boldsymbol{q}'_{-i})\left[\ln\frac{a}{q_i}-\frac{3}{2}\left(1-\left(\frac{q_i}{a}\right)^{\frac{2}{3}}\right)\right]\ge 0$$

where the second inequality follows from the following fact:

$$g(y)\triangleq\ln y-\frac{3}{2}\left(1-\left(\frac{1}{y}\right)^{\frac{2}{3}}\right)\ge 0,\ \forall y\ge 1,$$

which can be proved in a similar way as the proof of Lemma 4. Also, for $q_i'=b<q_i$, we have

$$\hat{u}_i(b,\boldsymbol{q}'_{-i},q_i,e_i')-\hat{u}_i(q_i,\boldsymbol{q}'_{-i},q_i,e_i')$$
$$=c\int_{b}^{q_i}\frac{e_i'(q,\boldsymbol{q}'_{-i})}{q}dq+2ce_i'(b,\boldsymbol{q}'_{-i})\left(1-\sqrt{\frac{q_i}{b}}\right)$$
$$\le ce_i'(b,\boldsymbol{q}'_{-i})\left[\ln\frac{q_i}{b}+\frac{3}{2}\left(1-\left(\frac{q_i}{b}\right)^{\frac{2}{3}}\right)\right]\le 0$$

where the second inequality follows from the following fact:

$$g(y)\triangleq\ln y+\frac{3}{2}\left(1-y^{\frac{2}{3}}\right)\le 0,\ \forall y\ge 1,$$

which can be proved in a similar way as the proof of Lemma 5.

Then we show that the IR property is also achieved. Similar to the proof of Theorem 1, we have

$$\tilde{u}_i(d_i,q_i,e_i')=c\int_{q_i}^{\bar{q}}\frac{e_i'^2(q,\boldsymbol{q}'_{-i})}{q}dq+\frac{3}{2}ce_i'^2(q_i,\boldsymbol{q}'_{-i})$$
$$-\frac{3}{2}c\left(\frac{q_i}{q_i}\right)^{\frac{2}{3}}e_i'^2(q_i,\boldsymbol{q}'_{-i})$$
$$=c\int_{q_i}^{\bar{q}}\frac{e_i'^2(q,\boldsymbol{q}'_{-i})}{q}dq\ge 0.$$

REFERENCES

[1] X. Gong and N. B. Shroff, "Truthful mobile crowdsensing for strategic users with private qualities," in *Proc. 15th Int. Symp. Modeling Optim. Mobile, Ad Hoc, Wireless Netw. (WiOpt)*, May 2017, pp. 1–8.

[2] G. Ding *et al.*, "Robust spectrum sensing with crowd sensors," *IEEE Trans. Commun.*, vol. 62, no. 9, pp. 3129–3143, Sep. 2014.

[3] M. Mun *et al.*, "PEIR, the personal environmental impact report, as a platform for participatory sensing systems research," in *Proc. 7th Int. Conf. Mobile Syst., Appl., Services (MobiSys)*, Jun. 2009, pp. 55–68.

Authorized licensed use limited to: The Ohio State University. Downloaded on May 05,2023 at 23:56:56 UTC from IEEE Xplore. Restrictions apply.

[4] R. K. Rana, C. T. Chou, S. S. Kanhere, N. Bulusu, and W. Hu, "Ear-phone: An end-to-end participatory urban noise mapping system," in *Proc. 9th ACM/IEEE Int. Conf. Inf. Process. Sensor Netw. (IPSN)*, Apr. 2010, pp. 105–116.
[5] *Atmos: Crowd Sensing Weather Platform*. Accessed: 2017. [Online]. Available: http://beja.m-iti.org/web/?q=node/1
[6] L. Duan, T. Kubo, K. Sugiyama, J. Huang, T. Hasegawa, and J. Walrand, "Incentive mechanisms for smartphone collaboration in data acquisition and distributed computing," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Mar. 2012, pp. 1701–1709.
[7] N. B. Shah and D. Zhou, "Double or nothing: Multiplicative incentive mechanisms for crowdsourcing," in *Proc. Conf. Neural Inf. Process. Syst. (NIPS)*, 2015, pp. 1–9.
[8] W. Wang, L. Ying, and J. Zhang, "The value of privacy: Strategic data subjects, incentive mechanisms and fundamental limits," in *Proc. ACM SIGMETRICS Int. Conf. Meas. Modeling Comput. Sci.*, Jun. 2016, pp. 249–260.
[9] L. Pu, X. Chen, J. Xu, and X. Fu, "Crowdlet: Optimal worker recruitment for self-organized mobile crowdsourcing," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Apr. 2016, pp. 1–9.
[10] 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. (MobiCom)*, Aug. 2012, pp. 173–184.
[11] I. Koutsopoulos, "Optimal incentive-driven design of participatory sensing systems," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Apr. 2013, pp. 1402–1410.
[12] A. Tarable, A. Nordio, E. Leonardi, and M. A. Marsan, "The importance of being earnest in crowdsourcing systems," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Apr./May 2015, pp. 2821–2829.
[13] Y. Luo, N. B. Shah, J. Huang, and J. Walrand, "Parametric prediction from parametric agents," in *Proc. 10th Workshop Economics Netw., Syst. Comput. (NetEcon)*, 2015.
[14] N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani, *Algorithmic Game Theory*, vol. 1. Cambridge, U.K.: Cambridge Univ. Press, 2007.
[15] P. Bolton and M. Dewatripont, *Contract Theory*. Cambridge, MA, USA: MIT press, 2005.
[16] Y. Cai, C. Daskalakis, and C. H. Papadimitriou, "Optimum statistical estimation with strategic data sources," in *Proc. Conf. Learn. Theory (COLT)*, Jun. 2015, pp. 280–296.
[17] D. R. Karger, S. Oh, and D. Shah, "Budget-optimal crowdsourcing using low-rank matrix approximations," in *Proc. IEEE Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Sep. 2011, pp. 284–291.
[18] D. Lee, J. Kim, H. Lee, and K. Jung, "Reliable multiple-choice iterative algorithm for crowdsourcing systems," in *Proc. ACM SIGMETRICS Int. Conf. Meas. Modeling Comput. Sci. (SIGMETRICS)*, Jun. 2015, pp. 205–216.
[19] Y. Liu and M. Liu, "An online learning approach to improving the quality of crowd-sourcing," in *Proc. ACM Int. Conf. Meas. Modeling Comput. Sci. (SIGMETRICS)*, Jun. 2015, pp. 217–230.
[20] 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)*, Jun. 2015, pp. 167–176.
[21] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, vol. 1. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.
[22] V. Krishna, *Auction Theory*. New York, NY, USA: Academic, 2009.
[23] P. K. Varshney, *Distributed Detection and Data Fusion*. New York, NY, USA: Springer-Verlag, 1997.

Xiaowen Gong (S'12–M'15) received the B.Eng. degree in electronics and information engineering from the Huazhong University of Science and Technology in 2008, the M.Sc. degree in communications from the University of Alberta in 2010, and the Ph.D. degree in electrical engineering from the Arizona State University in 2015. From 2015 to 2016, he was a Post-Doctoral Researcher with the Department of Electrical and Computer Engineering, The Ohio State University. He is currently an Assistant Professor with the Department of Electrical and Computer Engineering, Auburn University. His current research interests are edge computing and data crowdsourcing. He received the Runner-up Best Paper Award from the IEEE INFOCOM 2014.

Ness B. Shroff (S'91–M'93–SM'01–F'07) received the Ph.D. degree in electrical engineering from Columbia University in 1994. Immediately, he joined the School of Electrical and Computer Engineering, Purdue University, as an Assistant Professor. At Purdue, he became a Full Professor of ECE and the Director of the university-wide Center on Wireless Systems and Applications in 2004. In 2007, he joined The Ohio State University, where he holds the Ohio Eminent Scholar Endowed Chair in networking and communications with the departments of ECE and CSE. He holds or has held Visiting (chaired) Professor positions at Tsinghua University, Beijing, China, Shanghai Jiao Tong University, Shanghai, China, and IIT Bombay, Mumbai, India. He has received numerous best paper awards for his research and is listed in Thomson Reuters' "The World's Most Influential Scientific Minds," and was noted as a Highly Cited Researcher by Thomson Reuters. In 2014, he also received the IEEE INFOCOM Achievement Award. He currently serves as the Steering Committee Chair for the ACM Mobihoc and as an Editor at Large of the IEEE/ACM TRANSACTIONS ON NETWORKING.

Authorized licensed use limited to: The Ohio State University. Downloaded on May 05,2023 at 23:56:56 UTC from IEEE Xplore. Restrictions apply.