`A Quality Constraint Mechanism Based on Two Consensuses of Blockchain in Crowdsensing

Abstract *—*One important problem in crowdsensing network is to improve data quality of tasks. Several methods have been proposed to solve this problem, however, all these methods are based on a trusted third party. While in practice, all networks work without a fully trusted third party. In addition, the real network has the character of dynamic, dispersibility and isolation. This leads to more problems that reduce data quality, e.g. negative work, unfair payment, cooperative cheating, etc. Considering the decentralization of blockchain, in this work we build a crowdsensing network model based on the blockchain for data quality control without a trusted third party. The blockchain can also prevent transaction data from being illegal modification, so a data quality constraint method based on two consensuses is proposed. According to the actual calculate ability of nodes, this article put forward a lightweight consensus mechanism DPOR based on reputation rather than mining. For further constraint on task data quality, this article introduces truth discovery to evaluate the performance of workers. Experiments on two open datasets show that our new method outperforms existing methods in improving quality of task data.

Index Terms — Crowdsensing, Decentration, Quality constraint, Blockchain, Two consensuses.

# Introduction

W

ith the continuous innovation of mobile Internet services, participatory sensing[1] becomes one of the research hotspots. By recruiting several mobile participants with terminals like cellphones, it can solve many complex tasks such as data collection[2], image identification[3] and solution-finding[4]. Based on the participation of a large number of mobile workers, participatory sensing can collect data effectively by crowdsourcing and incentives. The participants can execute crowdsensing tasks in exchange of monetary reward. To gather data of high quality is the ultimate purpose of participatory sensing. However, due to the isolation and dynamic nature of crowdsensing network, constraint on data is necessary as well as challenging in a dynamic network. [5]

Currently, there are two main factors that lead to the risk of data quality decrease. First, for egoism of nodes, the task participants may have negative work and forge data by cooperative cheating. Therefore, to select participants of high quality data becomes a method to constrain data quality[6]. Pre-judgement idea is most used in existing methods. It selects participants according to the past performance of candidate nodes[7][8]. However, pre-judgement only increases the probability of collecting higher quality data and it can’t prevent participants from negative work. The selected workers may use fake data to maximize their interests. The second factor is that the task distributors have the risk of pay unfairly to maximize their own benefits. This will damage the working enthusiasm of participants and decrease the quality of task data. Several incentive mechanisms[9][10] have been proposed and performed well in data quality constraint. Truth discovery is the most common method based on incentive mechanisms. And it is especially used when most tasks in reality have no standard answers. However, only when data of all participants has been collected can quality control happen, so there exists a waste of resources and computing power.

All constraint methods above can improve data quality in different extent. Nevertheless, due to the centralized feature of traditional crowdsensing models[11], they are based on trusted third parties. They all ignore the lack of trusted third parties in real networks. With an untrusted central platform, the task data can be modified illegally. In addition, the nodes can deny their behaviors

By using technologies including data encryption algorithms, distributed consensus and smart contracts, the blockchain maintains a distributed chained data structure. Decentralized trust is the core advantage of blockchain. Among existing consensus mechanisms, the mining method which depends on the calculation power of nodes is the most common. In addition, time series data can’t be tampered with and smart contract can be executed automatically in this data structure. These enables blockchain to not only become the underlying support for Bitcoin, Ethereum and BitShares, but also play a role in identity management, mobile device services[12], privacy protection[13] and data quality constraints[14] and other aspects. So far, the blockchain framework applicable to the field of encrypted digital currency has been mature, but the application of blockchain for participatory sensing is still in exploration.

In view of the two above factors, this paper proposes a crowdsensing working model using two data quality constraint processes. The first process is based on pre-judgement and the second is based on truth discovery. By these two data quality constraint processes, the probability of gathering high quality data is increased while the whole cost decreases. Furthermore, blockchain is introduced to the crowdsensing model to deal with the lack of trusted third parties. Considering the actual calculation power of nodes in real networks, this paper proposes a lightweight consensus mechanism DPOR based on reputation instead of mining. In addition, two consensuses are added to prevent task data from illegal modification. Therefore, this paper

## Summary of key contributions

The focus of this work is to deal with the aforementioned problems by designing a data quality constrain mechanism based on two consensuses of the blockchain.

The main contributions of the paper are as follows.

* This article constructs a crowdsensing task model based on the blockchain. Decentralized trust can be implemented in this new model. Compared with the traditional crowdsourcing model, the data quality problems caused by the lack of trusted third parties can be solved. Any one in the blockchain can check the transaction history while there exists no single user that can change it.
* Considering the limited computing power in reality, a lightweight consensus mechanism DPOR based on reputation is proposed instead of the conventional consensus mechanism Mining. In addition, a method based on two consensuses is proposed to constrain the task data quality. The first consensus can keep the hash data of each task in case that the task distributor pays unfairly or refuses to pay once it gets all task data. The second consensus saves the complete transaction information which ensures the irreversibility and undeniableness of the blockchain.
* Combining the advantages of both pre-judgment and incentive methods, this paper proposes a data quality constraint method based on two quality constraint processes. By two quality constrain processes, the waste of computing resources can be avoided and the problem of negative work of workers can be solved. In addition, the use of truth discovery improves the cost of collaborative cheating and reduces the possibility of collaborative cheating.

## Paper Organization

The remainder of this paper is organized as follows. Section II reviews and summarizes existing related studies. Section III introduces the crowdsensing working model based on the blockchain and describes the whole process. Section IV provides a detailed discussion of the proposed mechanism and how it works to constrain the task data quality, and Section V evaluates its performances through simulations and experiments. Finally, conclusions are drawn in Section VI.

# RELATED WORK

Pre-judgement and incentive are two categories of data quality constraint strategies. By selecting participants that are more likely to provide high quality data, the first strategy improves the data quality before the task execution. The second strategy is used to evaluate the data quality of participants after the execution of task.

To select efficient and high-quality worker sets has become a general method to constrain the data quality in crowdsensing networks. Pre-judgement strategy focuses on how to select the participants of high quality. Karma[15] constructs a Bayesian model to predict the behaviors of workers. The task data quality is evaluated by being compared with standard answers in database. In this method, the expected benefit and the cost of hiring workers can achieve balance. Ren[16] proposes an effective matrix completion technique to select a minimum set of participants. During the sensing task execution, this technique can maintain the QoS and reduce the cost of system efficiently. However, the methods above can’t make predictions about individual workers. Roman[17] presents a crowdsourcing platform and predicts the available workforce by availability of nodes. In addition, an indicator suitability is defined as the standard to select workers. Jurca[18] proposed an average reputation model, introducing the concept of reputation to evaluate the performance of workers. The mean value of all historical service quality of a worker will be taken as its reputation which becomes the basis for worker selection. However, the method ignores the influence of the time factor. Workers should take more reputation from more recent tasks, rather than taking the mean value indifferently. However, the pre-judgement quality constraint works before the task execution so the selected participants may work negatively and submit fake task data.

Incentive mechanism works after the execution of tasks. By giving relevant reward to workers according to their performance, the incentive mechanism improves the enthusiasm of workers and encourages workers of high quality to participant in the future tasks. Yaron[19] presents constant-competitive incentive compatible mechanisms for maximizing the number of tasks under a budget, and for minimizing payments given a ﬁxed number of tasks to complete. Duan[20] designs a distributed auction framework and proposes two distributed auction schemes, CPAS and TPAS to achieve budget-balance and computational-efficiency. Considering the reliability of the data source, Li proposed the CRH algorithm to solve the truth discovery problem of heterogeneous data, and the iterative process included in this method has a higher convergence rate and accuracy. For long tail phenomenon, a CATD method which can automatically discovery the truth is proposed. In this method, different participation degrees are used to reflect the reliability of data sources. There are also studies that introduce third parties for evaluation to improve the quality of task data. Baba design another evaluation process after the task execution, using two crowdsourcing processes to improve data quality. However, this method is based on the hypothesis that the introduced evaluation party is credible while there exists no fully trusted third parties in reality.

Truth discovery is widely used in data quality constraint. According to the number of truth value, truth discovery is divided into multiple truth discovery and single truth discovery. This paper studies the single truth discovery problem. Yin[21] et al. first defined the truth discovery problem and proposed the TruthFinder algorithm to iteratively calculate the reputation of the fact and the data source.

In view of the methods above, they are all based on a trusted central platform. With the decentralization consensus and irreversibility of data in blockchain, this paper introduces blockchain into the crowdsensing model. In addition, considering the disadvantages of the two strategies above, two data quality constraint processes are necessary in the model to improve the data quality while decrease the task cost.

# Crowdsensing model based on two consensuses

Table 1 lists frequently used notations.

This paper proposed a decentralized crowdsensing model to constrain the quality of task data. For lack of trusted third parties in network, blockchain is introduced to the proposed



**Fig.1. The crowdsensing working model based on the blockchain**

model. In case of the illegal modification of data and the unfair payment, a data quality constraint mechanism with two consensuses is proposed. Considering the actual calculation ability of candidate nodes in real network, a lightweight consensus mechanism DPOR is proposed instead of the common consensus mechanism based on mining. In DPOR, verifiers will be selected according to reputation rather than computing power. To constrain the data quality effectively and prevent participants from working negatively, a method based on two data quality constraint processes is put forward.

There are three characters of task distributors, verifiers and task participants in the proposed crowdsensing model. There are five stages in the crowdsensing process, task distribution, node selection, the first consensus, data analyze and the second consensus. The data quality constraint mechanism is illustrated in Fig 1. First, the task distributors will broadcast the task information in the network. The task information includes type of the task, determined number of verifiers and participants and payments. Second, the node selection stage includes two parts: verifiers selection and participants selection. Verifiers of this task will be selected by DPOR. In DPOR, this paper designs a function V() to select verifiers and assign weights according to the reputation of candidate nodes. Then according to the task suitability of candidate nodes, required number of participants will be selected by the first quality constraint process. With the first quality constraint process, the selected participants are more likely to provide high quality data. After the execution of task, the distributor should gather task data from all participants. Then happens the first consensus in case of the denial of the task distributor. All task data should be hashed and sent to the verifiers, and the hash data will be kept in the most recent block. Once the first consensus is finished, all participants can send their data to the distributor. The task distributor will carry on the analyze of task data to get the function F(\*) of data quality rating and write F(\*) in the smart contract. The second data quality constraint process happens during the data analyze and truth discovery is used to evaluate data quality of each participant. In addition, the verifiers will calculate the rating result and payment of participants by the function F(\*). Last, the second consensus happens when the verifiers keep the complete transaction record in the block.

# The blockchain-based quality assessment mechanism

In the blockchain-based quality constraint model, DPOR is proposed as the lightweight consensus mechanism. In addition, TCM and TDQC are proposed to constrain the task data quality.

## Task distribution

Because of complexity and large amount of workload, some tasks are difficult to be implemented by the task distributor itself. Therefore, the task distributor u will broadcast the task information in the crowdsensing network.

The task information includes five parts, data requirement, total reward, the number of verifiers, the number of participants. According to the task information, several nodes in the network can volunteer to implement the task and join the candidate node set. Once all nodes in the crowdsensing network receive the information, the smart contract will be triggered and the function V() will be executed automatically to select the suitable verifiers set for the current task.

## Node selection

Node selection includes two parts: verifiers selection and participants selection. In the first part, DPOR is proposed to select the verifier instead of the mining consensus mechanism. The first quality constraint process implements in the participants selection. Pre-judgement is used to select suitable participants from the candidate node set.

* **Verifiers selection**

Verifiers are selected by the function V() written to the smart contract. V() is based on the reputation of task participants. In typical blockchain applications, the consensus mechanism mainly refers to POW(Proof of Work), which selects the verifier according to the calculation power. Under the circumstance of crowdsensing, participants generally lack large-scale calculation power and power resources. However, the cost requested by POW is too much beyond the actual ability of nodes. DPOS(Delegated Proof of Stack) is another mechanism used in consensus. Nodes in the network vote to select the verifier according to the stack of nodes. Considering the real source condition in the crowdsensing network, reputation is introduced as the standard to evaluate ballot weights.

Therefore, a consensus mechanism DPOR based on reputation is proposed to select verifiers. Nodes in the network are given different reputation and reputation determines the ballot weight that nodes have. By voting, the first node will be selected as the verifier of this task. Lightweight consensuses are implemented in network and the expenditure is decreased by DPOR.

**Definition 1.** Reputation of a node is introduced as the standard to select verifiers. It is defined as

(1)

where denotes the reputation, denotes sensitive parameter the denotes the past performance of the node, denotes the verification correct rate.

* **Participants selection**