Data-Oriented Mobile Crowdsensing: A Comprehensive Survey

Yutong Liu, *Student Member, IEEE*, Linghe Kong, *Senior Member, IEEE*, and Guihai Chen

***Abstract*—Mobile devices equipped with rich sensors, such as smartphones, watches, or vehicles, have been pervasively used all around the world. Their high penetration and powerful sensing ability enable them to carry out heavy sensing projects by splitting tasks into small pieces. Since ordinary participants can simply employ their mobile devices to sense and upload the required data, the mobile crowdsensing (MCS) technology is gaining great popularity. However, there are still some challenges in building a complete and sustainable MCS system. Researchers these years have proposed plenty of strategies to solve these challenges in order to improve the MCS technology. In this survey, we aim to provide a comprehensive literature review on recent advances in MCS. Oriented to the data flowing in MCS projects, we survey researches from five popular aspects in three stages: 1) incentive mechanism; 2) security protection; and 3) privacy preserving, together with resource optimization in the data collection stage; the data analysis stage; and the data application stage. To provide the convenience to interested researchers, some available testbeds, simulators, and commercial service platforms are also summarized in this survey. As the MCS technology still needs further development, we discuss some lessons learned from introduced researches as well as future research directions at last.**

***Index Terms*—Mobile crowdsensing, incentive mechanism, privacy preserving, resource optimization, multimodal data mining.**

I. Introduction

Mobile devices nowadays have been equipped with various sensors including accelerometers, gyroscopes, contact image sensors, and so on. According to the statistics of International Telecommunications Union (ITU) in 2018 [1], global mobile-cellular telephone subscriptions have grown more than 30% in the last five years and are expected to reach 8 billion by the end of 2018. Furthermore, there are even more mobile-cellular subscriptions than people on the planet, where every 100 inhabitants have around 110 subscriptions [1]. Inspired by such popularity of mobile devices, MCS is proposed [2] to realize more flexible and efficient data acquisition, analysis, and application than fixed wireless sensor network (WSN). The formal definition of MCS can be expressed as [3]:

MCS is a new sensing paradigm that empowers ordinary citizens to contribute data sensed or generated from their mobile devices, aggregates and fuses the data in the cloud for crowd intelligence extraction and human-centric service delivery.

Manuscript received August 4, 2018; revised November 2, 2018, January 24, 2019, and March 11, 2019; accepted April 5, 2019. Date of publication April 12, 2019; date of current version August 20, 2019. This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFB1004703, and in part by China NSF under Grant 61672349 and Grant 61672353. *(Corresponding author: Linghe Kong.)*

The authors are with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: isabelleliu@sjtu.edu.cn; linghe.kong@sjtu.edu.cn; chen-gh@sjtu.edu.cn).

Digital Object Identifier 10.1109/COMST.2019.2910855

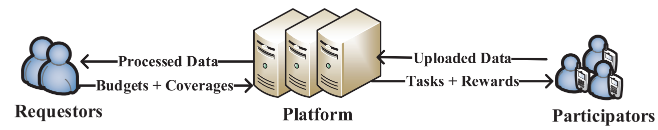

Fig. 1. The illustration of MCS structure.

As its name implies, MCS has three main characteristics:

- *"Mobile":* The participators and devices are in moving status subjectively, so the coverage of participators and the quality of sensing data are required to be controlled.
- *"Crowd":* The large group of individuals having mobile devices are capable of participating in the sensing projects. But different trajectories or preferences of participators will lead to the data heterogeneity, such as the uneven distribution of sensing data.
- *"Sensing":* Participators are mostly required to execute some simple sensing tasks without massive computations.

Take city noise monitoring as an example. It is intuitive for relevant departments to build special infrastructures on noise detection, but it will cost a large amount of money and time for construction, usage, and maintenance. Benefit from MCS, users who are willing to participate in this monitoring project will employ their mobile devices (*e.g.*, smartphones, wearable devices, smart vehicles, etc.) to record the audio clips in their locations, for further noise analysis. These records together with their GPS information will be uploaded to the platform and finally sent to requestors (*e.g.*, agencies responsible for noise monitoring). From the requestors' perspective, the costs in MCS are lower than traditional facility constructions and maintenances, which include the rewards for participators, charges for platform services, transmission consumptions, and so on, deriving MCS as a meaningful research and application direction.

From the above-mentioned example, the typical MCS systems are consisted of three main components: participators, platform, and requestors. While in some scenarios the participators and requestors are the same entity (e.g., the requestors can also go to collect data), in this paper we discuss them separately since their related service types and problems are different. Their relationships are illustrated in Figure 1. Requestors will initiate sensing requests from either mobile devices or computers, and provide corresponding constraints

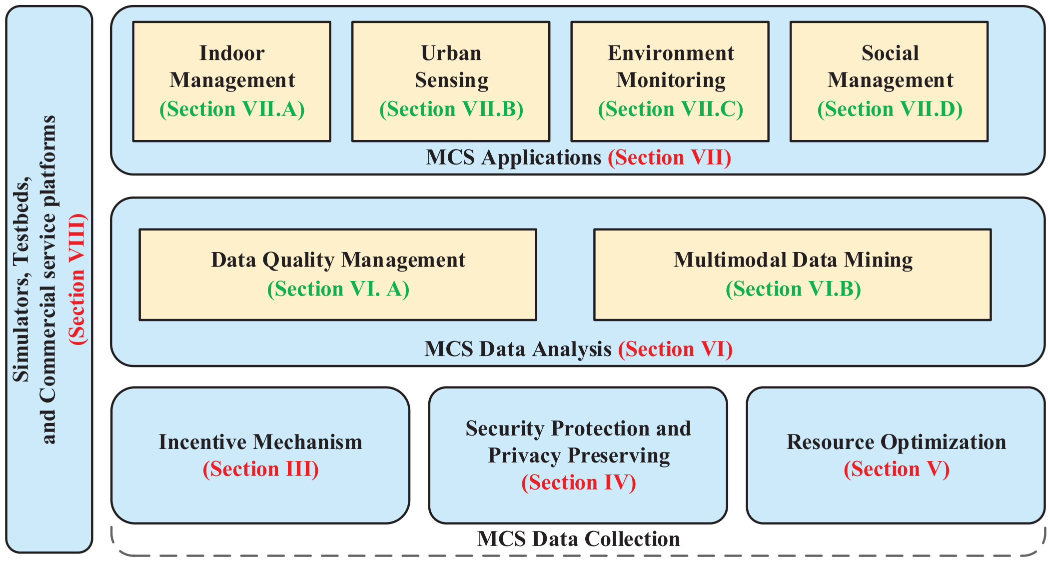

Fig. 2. The framework of data-oriented MCS survey.

including budgets and coverage requirements forwarded to the sensing platform. The platform side is responsible for assigning tasks and publishing rewards to a suitable subset of mobile device users, with the consideration of constraints. Participators will upload the sensed data after accepting their sensing tasks, which will be collected, stored, processed by the sensing platform and finally returned to requestors. This collaborative scheme builds the connection between industries, governments, organizations and ordinary people. The united environment it establishes plays an important role in supporting sustainable development goals, including economic, social and environmental requirements [4].

However, to build a sustainable and complete MCS system, several problems oriented to data in MCS waiting to be solved. *In the collection of data*, how to recruit enough participators? How to preserve their privacy and the security of their sensed data? And how to achieve best distribution and data quality under budget and coverage limits, together with several sustainable issues [5]? *In the analysis of collected data*, how to detect and correct missing or faulty data? And how to dig out the inner connection in multi-modal data? *In the data applications*, how to apply observations from data to improve our life quality? Correspondingly, in this survey, we discuss five main research aspects belonging to three data-oriented stages for MCS: incentive mechanism, security protection and privacy preserving, and resource optimization in the MCS data collection stage; the MCS data analysis stage; and the MCS application stage. Some existing simulation methods, testbeds, and commercial service platforms are summarized next. The organization of this survey is shown in Figure 2.

Since MCS has been researched in the past few years, there have been several survey papers discussing this area. One kind of surveys concentrates on the single aspect of MCS in detail: designs of incentive mechanisms [6], [7], strategies for privacy preserving and security [8]–[10], plans in resource optimization [11], or discussion on quality of information (QoI) [12]. Another kind of surveys comprehensively discusses MCS in different aspects [2], [3]. Ganti *et al.* [2] summarize the current state and future direction of MCS from a global view, while it only gives the brief introduction on the challenges and solutions where technology details are missed in this paper. Guo *et al.* [3] survey on the most related topic with us. On the one hand, we survey on different views of MCS. This survey is driven by the view of human intelligence, where the motivations of MCS projects are human participation, fusion and extraction of human-machine intelligence, and human-powered applications. But our paper is oriented to the data flowing in MCS projects, introduced from data collection, analysis, and application stages. On the other hand, plentiful advances have been made during the recent three years. The further survey on these related work is meaningful in our paper. The summarization on advances and differences of these survey papers are shown in Table I.

Specifically, the main contributions of our paper can be summarized as:

- This survey comprehensively considers the key techniques oriented to the data in MCS, which can give the global view on the main research aspects in MCS to readers.
- This survey selects plenty of related papers to present the recent advances in MCS research. As we systematically introduce and compare them, readers can clearly have a preliminary understanding of each aspect.
- This survey summarizes some existed simulation methods or testbeds, together with learned lessons, future research directions and potential solutions, which can inspire new or skillful researchers in MCS area to design and realize their own research topics.

The rest of this paper is organized as follows. Section II clears the relationships between MCS with other similar research areas, and summarizes key design considerations

TABLE I
THE COMPARISONS OF EXISTED SURVEY PAPERS

Type	Paper	Advances	Differences
Concentrated	Zhang et al. [6]	Survey on different **incentive** mechanisms	Category mechanisms by the kinds of incentives
	Jaimes et al. [7]	Discuss **incentive** with design constraints	Consider more about requirements and evaluations
	Pournajaf et al. [8]	Review **privacy** issues in task management	Include both task assignment and privacy preserving
	Christin et al. [9]	Consider **privacy** in applications	Discuss under real-world conditions
	Vergara et al. [10]	Summarize **privacy** protection in MCS system	Category protections to tasking and reporting sides
	Liu et al. [11]	Survey on **cost** deduction in MCS	Discuss more on future challenges
	Restuccia et al. [12]	Define and enforce the **QoI** in MCS	Include more mathematical background
Comprehensive	Ganti et al. [2]	Summarize the current state and future directions	Less technical details
	Guo et al. [3]	Introduce participation, extraction, and applications	Discuss MCS from human-machine intelligence view

in MCS. Sections III–VII respectively introduce the existed researches on incentive mechanism, security protection and privacy preserving, resource optimization, analysis techniques, and applications in MCS. Section VIII further lists state-of-the-art simulators, testbeds, and commercial service platforms served for MCS projects. Finally, Section IX concludes the paper and discusses the lessons learned from the mentioned researches and some possible future research directions.

II. PRELIMINARIES

In this section, we first clear the interactions between MCS with other similar research areas, and then summarize several factors in designing MCS system. This section can be regarded as a complementary background for MCS, which helps readers to understand the following sections easily.

A. Interactions With Similar Areas

MCS has a lot of similarities with another five research areas: the Internet of Things (IoT), cloud computing, edge computing, WSN, and crowdsourcing. The investigations between their interactions and differences will be presented in this subsection.

Firstly, the IoT is the network of physical devices, vehicles, home appliances, and other items embedded with electronics, softwares, sensors, and actuators. The connections between them enables the integration of the physical world into cyber systems [13], [14]. As the mobile devices with sensors in crowdsensing also connect with each other and exchange data, MCS is envisioned to co-exist with the IoT paradigm. It is proved to be a win-win strategy [15], where the potential synergy between the embedded IoT devices and the combined network support the MCS techniques, and reversely, the deep fusion of these embedded segments in MCS brings along the complementary benefits for IoT field without additional cost [16]. For example, the analysis of road sensory information collected from MCS through vehicle networks is beneficial for providing safer and more reliable vehicular transportation [17].

To overcome the computation, memory, and energy limitations, mobile cloud computing provides the possibility for MCS to consume various cloud resources via wireless networks. That is, the data analytic, storage and real-time processing are performed in the cloud, which is regarded as a cloud-centric IoT solution [18]. Additionally, since mobile devices are embedded with various sensors and associated with human users, they can collectively form a mobile cloud to provide pervasive services [19]. Thus, MCS is considered as a cloud-inspired model to solve large-scale problems, which follows "sensing as a service ($\mathrm{S^2aaS}$)" [20] and "sensing instruments as a service (SIaaS)" [21] models. One of the MCS schemes introduced in Section I is a typical $\mathrm{S^2aaS}$ model, where a mobile device user can be not only a cloud (service) user who is a requestor of sensing tasks but also a participator who fulfills sensing tasks [20]. Besides, as described in SIaaS model, the cloud provider (platform) is not the owner of shared resources but can share and manage participators' sensing instruments using cloud infrastructure through a virtualized system [21].

Although it is proved that assigning all computational tasks to the cloud is efficient for data analyzing according to its outstanding computing capability, the limitations of speed and bandwidth for data transmission have become the bottleneck for this cloud-centric scheme [22]. To deal with it, mobile edge computing (MEC) tends to transfer network functions, contents and resources closer to mobile users [23]. As one of the edge computing scheme, fog computing in wireless networks enables IoT applications running at the edge side. A multi-tiered architecture for fog computing is composed of three layers: the device layer accumulating all IoT devices, the fog layer containing all intermediate servers, and the cloud layer with cloud servers. Benefit from this hierarchical structure, MCS tasks can be assigned according to specific application requirements (*i.e.*, latency and cost constraints) and conditions of each layer (*i.e.*, bandwidth, battery, or computational capability). In general, the data collection is mainly assigned to edge devices, while the data analysis can be pre-processed on sub-servers, where redundancy can be filtered and near-optimal data can be extracted before transmitting to the final cloud servers. The transmission resources can be saved and the computational burden of cloud servers can also be relieved. In addition to this provisioning benefits, human-enabled mobile MEC ($\mathrm{M^2EC}$) shows the possibility of the integration between MEC and MCS, where the efficient resource usage and perceived service quality can be realized by their joint exploitation [24].

As one of the most important elements in the IoT paradigm, WSN has been used to carry out sensing tasks. It collects data from the environment and transmits them to the sink nodes [25]. But it requires specifically designed sensor nodes and network architectures for certain sensing tasks, where the limited node coverage, high installation and maintenance fee, and the lack of scalability make it difficult to be extended [26]. Different from traditional WSNs deployed with dedicated stationary sensors, MCS is powered by sensor-embedded mobile devices. As a result, MCS is capable to execute flexible sensing tasks by enlarging the sensing scale and coverage granularity [27].

Coined in 2005 by Howe [28], the term "crowdsourcing" is a portmanteau of crowd and outsourcing. The core idea of crowdsourcing is to divide work between participators to achieve a cumulative result. The crowd wisdom from large, relatively open and rapidly-evolving groups of Internet users can be applied to find a creative idea or satisfactory solution to solve the problem, or achieve some financial or research targets [29]. Inspired by crowdsourcing, crowdsensing specifically aims at the sensing targets, where the highly-distributed mobile devices from ordinary users can be served as sensors to collect and upload data. In other words, the mobile crowdsourcing used in sensing project [30]–[33] is equal to MCS.

B. Matrix

This subsection introduces several main matrices needed to be considered in MCS design. Note that the selections on these matrices are according to the targeting problems, where five technical aspects in the following sections have different concerns. Both their general meanings as well as varied explanations in different technical aspects will be introduced in this subsection.

1) QoI: The QoI represents the quality (or trustworthiness) of sensed data, which will be affected by two factors: (i) outright device malfunction, transmission error, or low sensory accuracy caused by environmental issues (*e.g.*, mobile phone kept in a pocket while sampling the street-level noise); and (ii) wrong or invalid readings stemming from malicious users. The former factor can be represented as the hard reputation, which can be predicted by statistical methods [34]. And the latter one can be represented as the soft reputation, but it is unpredictable [35].

Given that most modern mobile devices can achieve above 97% sensory accuracy [36], several incentive mechanisms (Section III) specially design to avoid malicious activities when encouraging more participation. The soft reputations of participators are evaluated in these reputation-based incentive mechanisms. For erroneous data caused by the former factor, outlier detection or correctness strategies are proposed in quality management (Section VI-A). As for privacy preserving and security protection aspect (Section IV), most strategies are applied on data itself. The QoI in this section should be evaluated after data recovery (or decryption), which is represented by information loss. While in resource optimization (Section V), the QoI performs as a driven for the design of strategies. That is, the selection of participators, task assignment, even transmission optimization should enhance or at least guarantee the QoI requirement.

2) Budget Limits: The budgets provided by requestors indicate the sum of expenditures used in the whole MCS system. It includes the reward (incentives) provided to participators, service charges for the platform, communication consumptions, etc. It performs as a constraint for the design of mechanisms, or the boundary of optimization algorithms. This factor is mainly considered in Sections III and V.

3) Coverage & Distribution: The coverage of MCS projects is the largest spatial scope covered by sensed data, while distribution shows the density of data at each sensed area. Caused by subjective trajectories and preferences of participators, the distribution is unbalanced. For example, less sensed data is acquired from countryside areas, compared with the center of the city. Together with the budget limits, the coverage of the sensed area also becomes a constraint condition. The goal of strategies to improve MCS is to keep a more balanced distribution until reaching the coverage constraint. This factor is emphasized in Sections III, V, and VI.

4) Energy: Mobile devices are severely constrained by their energy matters, including battery, memory, temperature, even computational and communication consumptions. As one of the concerned factors for participators, energy-cost MCS projects cannot attract or maintain the enthusiasm of users [10]. This factor is mainly concerned in Sections IV and V.

III. INCENTIVE MECHANISM

As one of the bases of MCS systems, large enough participation is needed for plenty of sensed data. However, there are two main challenges [37]: (i) Resource consumptions to execute MCS tasks are nonnegligible for participators, including battery consumption, data subscription plan, time, and effort; and (ii) Long-term commitment requirement distracts participator's daily plan. So a proper incentive mechanism is critical in MCS systems to motivate the active and persistent participation of mobile users. To decide an optimal reward strategy, besides the price it sets, the QoI, budget limits, utilities and fairness should also be considered. Utility matrices can be defined both from the perspective of the platform and users to quantify the cost-reward balance between them [35]. And the fairness is to ensure that rewards should be earned by fair competition among participators. However, these factors are not necessary for each incentive mechanism. For instance, in some remote area, where the QoI of collected data is difficult to be guaranteed, participators should still be rewarded as long as they contribute.

In this section, we classify the incentive mechanisms into three categories according to their pricing strategies: classical economy based strategies, game theory based strategies, and mathematical optimization based strategies. A taxonomy in Figure 3 and a summarized table in Table II are provided, illustrating the relationships between related papers and the comparisons on their concerned factors, promotions, and solutions.

A. Classical Economy Based Strategies

In classical economy based pricing strategies, the pricing strategies of selling the sensing data are divided into two

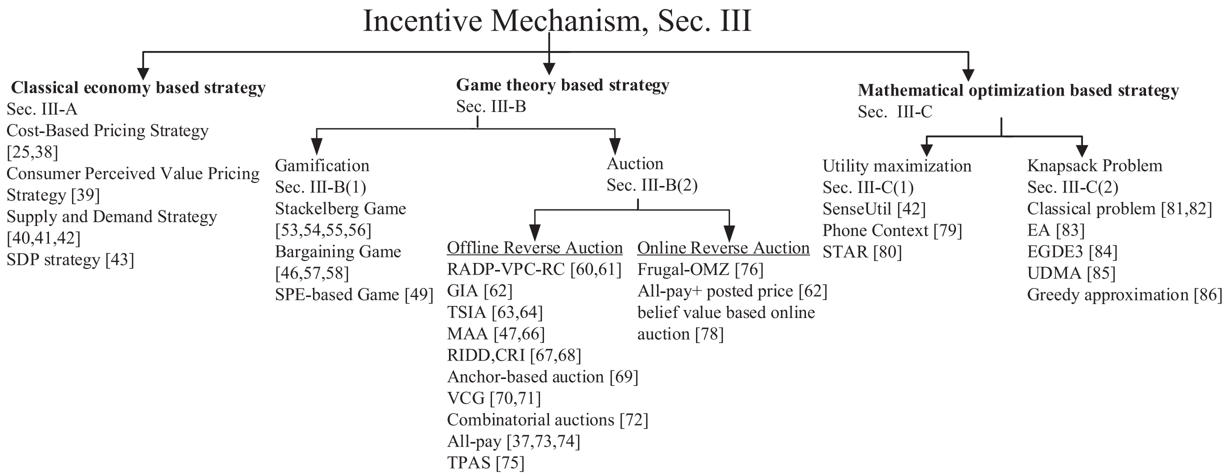

Fig. 3. A taxonomy of the incentive mechanism in MCS.

parts: the fixed costs depending on the physical consumptions in sensing process, *e.g.*, data transmission cost or battery consumption, and the variable costs depending on different flexible situations, *e.g.*, market conditions.

Cost-Based Pricing Strategy: The cost-based pricing strategy simply sets the selling price according to the consumption. The price p is mathematically denoted as $p = C \times (1 + m)$, where C is the total cost and m is the desired profit percentage [25]. In another papers, the pre-defined percentage m can also be represented by the sampling rate of each sensor node [38].

Consumer Perceived Value Pricing Strategy: Although the cost-based pricing model is easy to implement, it is also easily copied by competitors. Harmon *et al.* [39] present a consumer perceived value pricing strategy, which externally considers five perception factors both on the buyer side and the seller side: the buyer's perception on cost, data utility, seller's reputation, situation context, and buying motivations. The comprehensive consideration on these five factors derives the final decision of selling price for sensed data.

Supply and Demand Strategy: Although the consumer perceived value pricing strategy has considered on market factors, it still ignores the supply and demand relationship. Correspondingly, Pindyck and Rubinfeld [40] present the supply and demand model. As the basis of this model, the law of demand and supply is related to the price. The higher the price is, the lower the quantity demanded, and vice versa. When these two factors become equal, it reaches to a special point called market equilibrium. The profit of the strategy is the highest and a stable QoI can be achieved [41] at this point, because the amount of goods being supplied is exactly the same as that being demanded. The pricing adjustment can be easily understood that any excess supply or demand will lead to a movement in price towards this equilibrium point. Based on this design, if no one reports sensed data, the reward price will be automatically increased to promote participation [42].

Smart Data Pricing (SDP) Strategy: In the aforementioned strategies, they do not consider the congestion situation. This situation happens when a large number of users leverage the lacking resources simultaneously and reach to the peak demand. The SDP Strategy [43] can avoid this situation by varying the price over different time and different usage of resources. For instance, a customer has to pay higher electricity unit price, if its own usage level of electricity is higher than the average level in the wide of community.

B. Game Theory Based Strategies

Game theory researches on theories and mathematical models to solve competitive problems about the interactions (*i.e.*, conflicts or cooperations) between the rational decision-makers [44]. Typically, we can recognize a lot of game theory terms widely used in the design of incentive mechanisms, including Nash Equilibrium (NE), Stackelberg Equilibrium (SE), Subgame Perfect Equilibrium (SPE), etc. As the monetary reward is a powerful motivator, it is popular to apply game theory to incentive mechanisms for guaranteeing the extension and sustainability of participation in recent researches. In this survey, we would like to show some specific models by dividing them into two groups according to their outer forms: gamification and auction.

1) Gamification: Gamification is one of the forms in MCS incentive design. It introduces the game mechanism which is applied to non-game contexts to motivate the active participation of mobile device users [45]. A typical gamification method is composed of three functions: tasks, rewards, and communications [46]. Task means the sensing tasks in MCS incentive mechanisms. Rewards can be either monetary rewards, like money as a common one, coupons [47], and services [48], or non-monetary rewards, like badges [46], [49], scores [50], and credits [51], [52]. Communications like rankings can encourage the competition among participators. Next, we will introduce three main game mechanisms employed in gamification incentive designs, including the Stackelberg Game, the Bargaining Game, and the SPE-based Game.

Stackelberg Game: In this game, there are two kinds of players, the leader and the follower. The assumption of this game is that the follower will give a response on the leader's strategy, and the leader will consider this for its next strategy decision, which indicates the interplay relationship with each

TABLE II
SUMMARY OF INCENTIVE MECHANISMS

Type of incentive				Limited budget	Utility	QoI	Fair	Promotion	Solution
Classical economy based strategy			Cost-based Pricing					Easy computation	Simple cost function
Classical economy based strategy			Consumer Perceived Value Pricing		X			Situation context	Buyers' perception
Classical economy based strategy			Supply and Demand Pricing			X		Market situation	Market equilibrium
Classical economy based strategy			Smart Data Pricing			X		Market situation+congestion	Prices vary over time and usage
Game theory based strategy	Game		Stackelberg Game		X	X		Market situation	Nash Equilibrium
Game theory based strategy	Game		Bargaining Game				X	Distribution property	Nash Equilibrium
Game theory based strategy	Game		SPE-based Game			X		Partly join the game	Nash Equilibrium
Game theory based strategy	Auction	Offline	RADP-VPC-RC		X		X	Dynamic price and dropout protection	Virtual credit
Game theory based strategy	Auction	Offline	GIA	X	X		X	Dynamic price, dropout protection and coverage	Location information
Game theory based strategy	Auction	Offline	TSIA	X	X		X	Dynamic price, dropout, coverage and distribution	Timeslot separation
Game theory based strategy	Auction	Offline	MAA		X	X	X	Multi-attribute and situation context	Utility maximize
Game theory based strategy	Auction	Offline	RIDD,CRI			X	X	Reputation consideration	Nash Equilibrium
Game theory based strategy	Auction	Offline	VCG	X		X	X	Incentive compatible	Bayesian NE
Game theory based strategy	Auction	Offline	Anchor-assisted and vote-based auctions			X	X	Trustworthiness	Vote-based calculation
Game theory based strategy	Auction	Offline	Combinatorial auction			X	X	Combination of tasks	Optimal solution
Game theory based strategy	Auction	Offline	All-pay	X		X	X	Situation context	Quality maximize
Game theory based strategy	Auction	Offline	TPAS	X	X	X	X	Partial fulfillment, attributive diversity and price diversity	First-come-first-serve +greedy algorithm
Game theory based strategy	Auction	online	Frugal-OMZ, all-pay+posted, belief value based	X		X		Dynamic learning	Sequential computation and greedy algorithm
Mathematical optimization based strategy			Utility Maximization		X			Coverage and situation context	Graph-based method
Mathematical optimization based strategy			Knapsack problem	X	X	X		Coverage and distribution	Optimal solution

other. A typical example is the MSensing game [53]. Both the platform and the users are players, where the platform is the leader and the users are followers. Its sensing process can be described as:

- The platform announces its reward R as its strategy and each user arranges its sensing time accordingly to maximize its own utility at the next stage.
- The strategy of user i is its expected sensing time t_i. Let $t = (t_1, t_2, \ldots, t_n)$ denote the strategy profile consisting of all users strategies, and t_{-i} denotes the strategy profile excluding t_i. This stage is a non-cooperative game, which is called the Sensing Time Determination (STD) game.

This sensing process is mathematically proved to achieve the NE [54]. It implies that there is a stable set of strategies gaining nothing by unilaterally changing its current strategy for a given reward R, and this stable strategy set is unique. Hence, the platform can maximize its utility by choosing the optimal R. This optimal R together with the NE of the STD game are called SE, which implies the best response for each player.

Zeng *et al.* [55] propose a Stackelberg Game based incentive mechanism in the Word of Mouth Mode (WoM). In this mode, contributors can invite more participators through their social networks. The decisions made by contributors about the quantity of sensing data can directly influence the decisions of invitees. As each player in WoM has its only best response, this game exists the unique SE.

Furthermore, the Trustworthy Sensing for Crowd Management (TSCM) model [56] enhances the Stackelberg Game by introducing reputation-awareness and trustworthiness of participators. An outlier detection algorithm [34] proposed accordingly is used to detect possibly altered data which directly affects the trustworthiness of the corresponding participator.

Bargaining Game: Other than the non-cooperative Stackelberg Game, the Bargaining Game [57] is a cooperative scheme, which considers the joint balance on distributing objects among multiple players. So the final price accepted by everyone reaches to the bargaining success.

In MCS problems, a Nash model [58] is commonly used as a bargaining model. This model assumes that each node is selfish, expecting its own reward to be increased. As introduced below, the bargaining process between two selfish nodes is realized by their message exchange, which is considered as a win-win solution:

- Two nodes create their own message lists $\widehat{L_a}, \widehat{L_b}$ including their data types, sequence numbers, and appraisal information;
- Node *a* and node *b* exchange their message lists and give new candidate lists as:

$$\widetilde{L_a} = \widehat{L_b} - \left(\widehat{L_a} \cap \widehat{L_b}\right) \tag{1}$$

and

$$\widetilde{L_b} = \widehat{L_a} - \left(\widehat{L_a} \cap \widehat{L_b}\right); \tag{2}$$

- Based on the NE, the optimal Nash solution finally gets two lists L_a and L_b, indicating a successful bargain.

A possible solution to increase the sustainability in Bargaining Game is to increase the bargaining power of loyal and failed participators [46]. The higher weights can be assigned to the lists $\widehat{L_a}, \widehat{L_b}$ of these participators in the last bargain round, to maintain their continuous participation.

SPE-based Game: In the SPE-based Game, participators in the game will join a subset of this game whose strategies follow the NE. Participators can optionally decide to vote (accept) or not vote (refuse) for a sensing task [49]. To keep a high QoI of contributed data, the reputation of each participator will be evaluated by their voting capacity. This capacity will be increased only when the dissimilarity between uploading and uploaded data is above a certain threshold, efficiently avoiding redundant data collection. Correspondingly, the rewards to voters will increase when their data dissimilarity is larger, or decrease vice versa. As the reward system is tightly bounded with the reputation of voters, this game can efficiently promote the trustworthiness of participators.

2) Auction: To stimulate users' participation, it is popular to use reverse auction in incentive mechanisms. In an auction, an auctioneer requests bids and provides corresponding goods in exchange for payments from their selected winners [59]. The "reverse" means that the service provider is the buyer and the users are the sellers who sell their sensing data with claimed bid prices. At the first beginning, some researches propose offline reverse auctions, which can assure the minimizing and stabilizing incentive costs while maintaining an adequate number of participants. But these incentive mechanisms require deep participation, even the participators sometimes are forced to change their daily plan for fulfilling MCS sensing tasks. To avoid this shortcoming, the online reverse auction is proposed to recruit participators depending on their real-time spatial-temporal information. Next, we will discuss these two kinds of auctions in detail.

Offline Reverse Auction: As the data quality differs from the professionalism of participators, sensing data types, or spatial-temporal situations, it is impractical to set a fixed optimal incentive price. So a combined Reverse Auction Dynamic Price (RADP) and Virtual Participant Credit (VPC) incentive mechanism is proposed [60]. Briefly speaking, a service provider selects a pre-defined number of lower bid price users who will get these prices as their rewards. But it will cause the "drop-out" on the next round for those unsatisfied users, who will dramatically decrease the price competition and increase the provided reward. To maintain continuous participation, a virtual credit is taking into account. The loser in the former round will get the higher credit and a much lower bidding price, ensuring its probability to win in the next round.

Although this incentive mechanism can achieve its goals with 60% simulated accuracy, it does not consider the location of the users, the coverage, and the budget constraints. Jaimes *et al.* [61] combine the Reverse Auction Dynamic Price with Recruitment (RADP-VPC-RC) algorithm and the Greedy Budgeted Maximum Coverage (GBMC) algorithm [62] to create the Greedy Incentive Algorithm (GIA). GIA is a reverse auction-based incentive mechanism concerning more about the participators' distributions to cover the areas of interest within a given budget. Similarly, the Time Slots Incentive Algorithm (TSIA) [63] leverages the popular model SPEAD [64] to acquire a better sensing distribution through the target area at the minimum price, and to assure a minimum number of participants at each time slot.

In addition to the location coverage and distribution, the QoI is also considered, that is, new mechanisms should pay as how well the participants do [65]. Multi-attributive Auction (MAA) mechanisms [47], [66] combine multiple attributes to evaluate the QoI. They select the sensing data of the highest QoI and give users corresponding incentives. Not only the multi-attributive data but also the multi-attributive users should be considered to get high-quality results, like Reputation-based Incentives for Data Dissemination (RIDD) [67] and Cheating-Resilient Incentive (CRI) [68]. As the trustworthiness of data is directly related to users' reputations, Pouryazdan *et al.* [69] propose an anchor-assisted and vote-based incentive mechanism. Before recruitment, the controller selects some anchor points who have 100% trustworthiness during a pre-defined time period. The trustworthiness of every node will be voted by any other node where the anchor node has the full capacity on voting.

Instead of the multi-attributive character, Koutsopoulos [70] propose an optimal reverse auction with multiple winners based on Vickery-Clark-Groves (VCG) auction [71]. This method utilizes the incentive compatible mechanism. The strategies where each user reports its true cost follow the Bayesian NE. This auction scheme is a type of the second price auction for multiple items. Any user tends to give a high bidding to win, but it can not win at all. Actually, the winning cost of this user depends on the second user's bid, so such a high bidding price will also increase others' social costs. Only when the user's bidding price is equal to the true value of an object, this user can finally win.

Similarly, Jin *et al.* [72] incorporate the QoI to design an incentive mechanisms based on the reverse combinatorial auctions. Here, the users can bid on the combination of different kinds of commodities. This paper not only studies the single-minded scenario where every user is willing to execute one subset of tasks, but also investigates the multi-minded cases in which any user might be interested in executing multiple subsets of tasks. Since the winner determination in this auction is an NP-hard problem, this paper designs a computationally efficient mechanism with the close-to-optimal social welfare.

Further enhanced, Sun and Ma [73] propose a behavior-based incentive mechanism with budget constraints by applying sequential all-pay auctions. All participators have to pay their irrevocable bids and fulfill the bids regardless of who the winner is [74]. This mechanism incorporates a winning probability into the utility function and thereby makes the all-pay equivalent to winner-pay auctions [37]. In this auction scheme, all competitors will try their best to win the bid, which can improve the QoI and promote continuous participation.

However, all of the auctions mentioned above depend on the full participation of sensing tasks. But in practice, the sensing time of users is limited depending on their daily schedules.

Motivated by this concern, Duan *et al.* [75] propose a Time schedule-Preferred Auction Scheme (TPAS), considering the partial fulfillment, attributive diversity, and price diversity. For instance, the partial fulfillment means the sensing tasks require the sensing time from 9:00 pm to 11:00 pm, but the user can only sense from 10:00 pm to 10:30 pm. The attributive diversity represents different sensing abilities for users like the quality of their sensors and their initial locations. And the price diversity is the different requirement on rewards varied by users. The TPAS follows the first-come-first-serve principle and greedily chooses the potential winner until no task or candidate can be selected. The final evaluations prove its computationally efficient, individually rational, budget balanced and truthful performances.

Online Reverse Auction: In practice, the users can freely arrive or depart according to their daily plans, deriving the design of online auctions. This type of auction aims to minimize the total payment while completing a certain number of tasks. The Frugal-OMZ mechanism [76] is a typical online reverse auction, using a multiple-stage sampling-accepting strategy. It dynamically enlarges the selected user set and learns the bid threshold for future decisions. As long as the user claims a bid lower than this threshold and there are remaining tasks, it can be chose to join the sensing project and get corresponding rewards.

Similarly, Sun and Ma [62] combine the all-pay auction and the posted price mechanism [77] for such dynamic decision. And a belief values based incentive scheme [78] is also designed for joint social states and the real-time throughput.

C. Mathematical Optimization Based Strategies

Two common mathematical optimization based strategies: utility maximization and knapsack problem, will be discussed in this part. Different from the utility considered in game theory based strategies discussed before, utility maximization incentive in this subsection is to apply mathematical optimization methods to solve max-utility functions, where the results are the final reward decisions. It can be applied both in single-objective sensor tasks (*i.e.*, either to make full coverage or under the limited budget), and multi-objective scenarios (*i.e.*, make satisfied coverage together with high QoI and limited budget).

1) Utility Maximization: The utility function is a mathematical form to represent the level of preference in microeconomics. In different research problems, the definitions of utility functions are different. In SenseUtil [42], the utility function represents the sensing task execution situation in a certain sensing area. If the sensing tasks are not done, the utility increases with the time, implying the high demand on participation. For another example, Han *et al.* [79] care more about the phone contexts (*e.g.*, indoor or outdoor, in the pocket or out of it). The utility is calculated by the sum of contexts and the corresponding status (*i.e.*, performing the sensing task or not) of mobile devices.

Actually, the goal of the utility maximization problem is to get the maximum amount of rewards as incentives. As defined in STAR [80], the problem can be written as:

$$\underset{f_{ij}^S, f_{ij}^R}{\text{maximize}} \sum_{i,j: e_{ij} \in E^R} U_{ij} f_{ij}^R, \tag{3}$$

where U_{ij} is the utility function and f_{ij}^R is the amount of services provided by user *i* to user *j*. They solve this problem using the graph-based algorithm inspired by the cycle-canceling algorithm.

2) Knapsack Problem: The knapsack problem is popular in combinatorial optimization [81]. Given a set of items with their corresponding weights and utilities, the problem is to determine the number of items included in a bag, where the total weight should be within a given limit and get to the largest total value. This classical problem is discussed in the sample selection for crowdsensing tasks [82].

In MCS, this problem can be described as [83]:

The participant *i* sells data with the bid price bp_i (weight) and the quality q_i (utility). Only w_{cons} bidders of *n* participants ($w_{cons} < n$) are selected to collect the data, which is also known as the constraint in the knapsack problem. The objective is to select the smaller bid prices and the higher quality within the w_{cons} constraint in order to get the largest profit.

To effectively solve this NP-Complete problem in pseudo-polynomial time, Pham *et al.* [83] use Evolutionary Algorithm (EA) to discover the near-optimal solutions. Firstly, the bp_i and q_i are used to calculate the domination and Crowding Distance (CD) of every participant. By doing pair-wise tournament selection, which will select more participants of the maximum CD or highest non-domination degree, the population diversity can be increased. Then, using the selection step of the Third Evolution step in Generalized Differential Evolution (EGDE3) [84], the solutions are proposed as the parents for the next generation. Furthermore, the Univariate Model Distribution Algorithm (UDMA) [85] is used to build the next generation, which can convert individuals from infeasible to feasible, and finally solve this Knapsack problem. Such a near-optimal solution of the Knapsack problem is always made by Greedy approximation algorithm and fully polynomial time approximation scheme [86].

IV. Security Protection and Privacy Preserving

Security issues are severely threatening MCS systems, because of the large data flow and the lack of qualified security mechanism [87]. For the former perspective, mobile device users frequently contribute data to the platform via WiFi access points or the cellular infrastructure, involving lots of sensitive data like contact lists, device IDs, and location information as well. These sensitive data can simply disclose real identities, locations, or trajectories of participants. That is also the reason for decreasing their participation willingness [8], [9], [88]. For the latter perspective, it is common to see that many MCS applications on smartphones simply inform users of which sensitive data is used rather than telling them the exact reasons for using it. That makes users feel uncomfortable with the concern of privacy leakage [89]. To guarantee enough participation for MCS, it is necessary to protect the data security and solve the

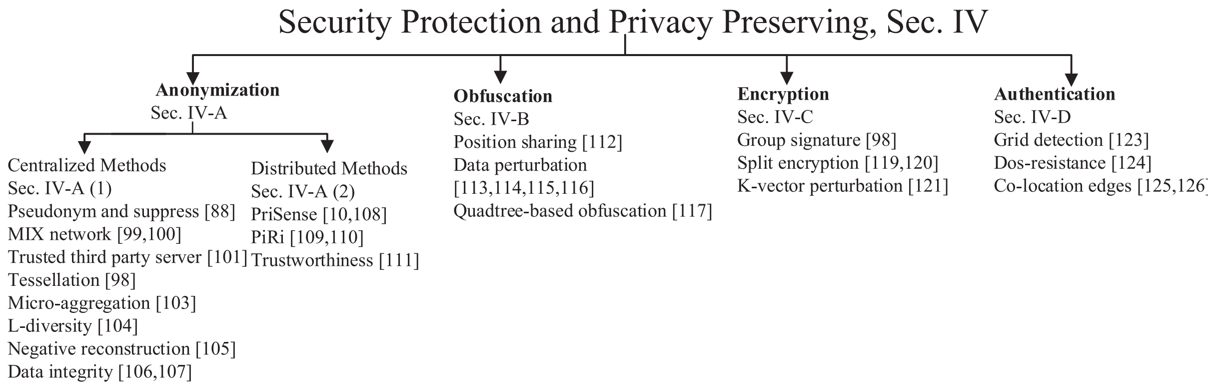

Fig. 4. A taxonomy of the privacy preserving.

privacy-preserving concern, so that people can feel at ease to hand over their sensed data. To sum up, there are three main security issues in the design of MCS systems [90]:

- *Privacy:* As most sensors will be carried by people, privacy-preserving methods should necessarily keep the user participation anonymous, unobservable and unlinkable, which contains three levels of protection: (i) Both user and device identities should not be revealed; (ii) Both external (*e.g.*, network providers) or internal (*e.g.*, MCS requestors) observers cannot infer the link between participators and certain uploaded data; (iii) No entity have access to infer any two or more reports contributed by a same participator [91].
- *Integrity:* Since most privacy-preserving methods intend to hide the real information, the integrity of the original data is difficult to be guaranteed. Especially for the data anonymization, it is inversely proportional to the accuracy of data analytics for the service provider. Malicious participators can easily disguise themselves behind the darkness of privacy-preserving mechanism without scruple. So the balance between privacy and data integrity is also one of the security issues.
- *Availability:* The adversaries to MCS systems may arise different kinds of attacks. As one of the examples, adversaries can gain the control of sensing devices and flood the network to dramatically affect the network capability by the denial of service (DoS) attack [92]. The authentication on the identity of participators can effectively prevent this attack and ensure the availability of MCS systems.

Actually, these three issues should be considered in each MCS stage. In data analysis, some validation strategies [93] are used for availability, and some other retrieval technologies like compressive sensing can be applied for integrity. For data application, three issues have different priorities in specific applications. For instance, in location-dependent applications, the availability is at the top priority to be protected, because of the tag attacks including illegal remove, tag falsify, or misplacement [94]. In this section, we mainly introduce the protection mechanisms used in the data collection stage: anonymization, obfuscation, encryption, and authentication. An efficient protection strategy is required where computation complexity, communication overhead, energy consumption

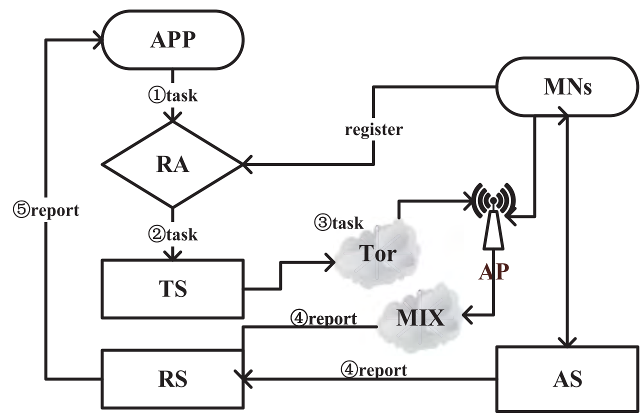

Fig. 5. The architecture of Anonysense.

and information loss are considered. Specifically, computation complexity can be represented by the running time in the sensing system, and the communication overhead is caused by extra message transmission for protection.

In the following subsections, we will discuss the related techniques for each mechanism. Figure 4 presents a taxonomy of this section to classify security and privacy preserving mechanisms. Their comparisons between different examples are summarized in Table III.

A. Anonymization

The main idea of anonymization is to generalize the usersąŕ private data into the same kind of group, which will confuse adversaries [95]. Before introducing different anonymous methods, a classical anonymization system will be described firstly, which is called "Anonysense" [96].

As shown in Figure 5, there are nine important components in this system:

- *Mobile nodes (MNs):* The mobile nodes (*e.g.*, smartphones) are devices with sensing, computation, memory, and wireless communication capabilities. In the crowdsensing process, each MN is recognized as the processor of sensing tasks and the carrier of sensing data.
- *Access points (APs):* The access points are the intermediate points between MNs and different services (*i.e.*, AS, RS, TS).

TABLE III
SUMMARY OF SECURITY PROTECTION AND PRIVACY-PRESERVING MECHANISMS (L-LOW, M-MEDIUM, H-HIGH, C-CENTRALIZED, D-DISTRIBUTED)

Type of preserving			Computation complexity	Communication overhead	Energy consumption	Information loss	Type of system
Anonymization	Centralized	Pseudonym and suppress	L	L	L	L	C
		MIX network	L	M	M	M	C
		Trusted third party server	L	L	M	M	C
		Tessellation	H	M	H	H	C
		Micro-aggragation	L	M	M	L	C
		L-diversity	L	M	M	L	C
		Negative reconstruction	L	L	L	H	C
		Data integrity	H	L	L	L	C
	Distributed	PriSense	H	M	H	L	D
		PiRi	H	H	H	L	D
		Trustworthiness	L	L	L	L	D
Obfuscation	Position sharing		H	M	H	L	D
	Data perturbation		M	L	L	L	D
	Quadtree-based obfuscation		M	L	L	L	D
Encryption	Group signature		H	H	H	L	D
	Split encryption		H	H	H	L	D
	K-vector perturbation		M	H	H	L	D
Authentication	Grid detection		M	L	L	L	C
	Dos-resistance		M	H	H	L	D
	Co-location edges		M	L	L	L	D

- *Anonymization service (AS):* A trusted anonymization service will apply anonymization techniques to ensure that each report is mixed with other similar reports. It will apply blurring techniques by adding uncertainty to the location in reports to ensure that the MN cannot be uniquely identified [97].
- *Report service (RS):* A RS will receive reports from AP through MIX network or AS and send them to the final application.
- *MIX network:* An asynchronous MIX network serves as an anonymous channel linking MNs and RS, routing reports through multiple servers, inserting delays, and mixing reports with other messages.
- *Task service (TS):* It is a task transfer station between RA and the registered MNs.
- *Anonymizing network (Tor):* It serves to protect the network identity and the location of MN by anonymization when it connects to TS to download new tasks.
- *Registration authority (RA):* It is a core responsible for the whole connection in this system. It should certify system components, MNs, and privacy-safe tasks so that it can release tasks securely.
- *Application (APP):* The application is the task publisher and the final report receiver, which serves as the terminal of this system.

Followed by the processing sequence labelled in Figure 5, the data flow in this system is:

The App submits a task (shown as 1) and the RA verifies whether the task respects carrier privacy before releasing them to the TS (shown as 2). At random intervals, the MN registers on RA, downloads new tasks from TS, and chooses the taken tasks (shown as 3). After collecting the required data, the MN sends these reports either to the RS or to the AS depending on the sensitivity of them. If the report contains sensitive information, the reports are sent to the AS directly or to the RS via a MIX network (shown as 4). Finally, an App can retrieve reports from the RS, and verify the integrity of them (shown as 5). Since a lot of anonymization methods can be used in this system, we will next introduce some typical ones.

1) Centralized Methods: If the data are collected or analyzed only by service providers, the mode of anonymization methods is centralization. Several typical centralized anonymization methods are described below:

Pseudonym and suppress: This is the simplest method in anonymization, which only give a pseudonym or suppression on user's identities to hide the real ones. However, it is too easy to be de-anonymized by adversaries. For example, they can analyze your device usage habits to disclose your identity [88].

MIX network: In Anonysense described before [96], [98], the MIX network is a developed anonymization technology. It is a statistical-based anonymizing infrastructure achieving the k-anonymity property, where the information for each person cannot be distinguished from at least $k-1$ individuals [99]. This concept is used in Hot-Potato Privacy-Protection Algorithm (HP^3) [100]. The data is sent to a random friend and this friend will choose another one to transfer this data in the next hop.

Considering that this approach does not require any additional computation in mobile devices, the calculation complexity is at a low level. But transferring data to intermediate nodes instead of the platform directly is considered as extra communication costs. According to evaluations [100], the average number of hops reaches to 7 as the hop threshold is set to 10^{-5}, which leads the energy consumption in transmission to a medium level. Besides, the information loss is linearly increasing with the number of malicious users. HP^3 gets 5% information loss when there are 7 malicious users, which is considered as a medium level.

Trusted third party server: To solve the limitation of confidentiality, Trajectory Privacy-Preserving Framework (TrPF) [101] adds a trusted third party server between an application service (ApS) and RS, which stores user-related information such as certificates and pseudonyms. The certificates are used for authentication to defense malicious attack, while the pseudonyms are used to break the potential linkage between spatial-temporal information and identities.

As this method anonymizes the sensitive locations on or nearby participator's trajectories by trajectory mix-zones graph model, the information loss of suppressed locations is lower than trajectory k-anonymity [102], which is considered as a medium level. From the evaluation results, the information loss reaches to 1×10^8 while trajectory k-anonymity gets 3×10^8 as k equals to 4. Additionally, pseudonyms are used in different trajectory segments, so the storage of them is the extra energy consumption, which requires $O(k)$ but lower than $O(n \times k)$ of trajectory k-anonymity, considering as a medium level.

Tessellation: Since the above methods are not suitable for some applications which need detailed location information, the tessellation method can handle this problem. Assume that there is no trusted intermediate entity, the location points should be processed directly [98]. Each location point is enlarged to a region called a **tile**, which should contain k points of participators' locations. Then the location of these k points are replaced by the center of this range to get anonymization.

However, since every tile has to meet the k-points requirements to ensure k-anonymity property, the information loss can be very large if these k points are far away from each other. The original location of every point should be delivered first to get its anonymized location. The total number of bytes exchanged in evaluation is 33,025 bytes (32.3 Kbytes), which can be decided as a medium communication overhead. But both retrieving the list of open access points and computing group signature for reports take a lot of computation time, which occupy 46.6% and 49.1% energy in the whole sensing process respectively. So the energy consumption is at a high level.

Microaggregation: To solve the large information loss problem in tessellation, Domingo-Ferrer and Torra [103] propose a variant of the Maximum Distance to Average Vector (MDAV). Here, the tile is classified based on the average vector of records, which is also an attribute-wised first-step. Then, the largest and the second distance record of that average vector are selected, which are denoted as d_r and d_s respectively. Finally, two clusters around d_r and d_s are formed. One cluster contains d_r and the $k - 1$ records closest to d_r. The other cluster contains d_s and the $k - 1$ records closest to d_s. Similar records form a more compact cluster leading to a more appropriate representative of cluster centroid to reduce information loss.

According to the evaluations on real-world datasets, this method gets 16.9% information loss and less than 1 second running time, indicating lower information loss and computation complexity. However, to select the similar records, all distances between records should be transmitted and stored resulting in a middle level of communication overhead and energy consumption.

L-diversity: Although we assume that an adversary does not know the true values of the times and locations in reports, the adversary in practice can still find out the spatial-temporal properties and identities. It bases on the prior knowledge of a user which is called background knowledge attack. To stop the adversary from knowing this knowledge and preserve user's privacy, Huang *et al.* [104] plan to give the user multiple values for its location attribute, which is also known as *l-diversity* method enhanced from MDAV.

The first step of *l-diversity* is to define the size $k \times l$ of each group, where k means k-anonymity property and l means the level of diversity. It leverages the MDAV over the temporal dimension. And the second step further applies MDAV over the spatial dimension.

In evaluations, as the diversity level is 2, information loss reaches to 15% which is less than MDAV and considered at a low level. Additionally, the other performances are similar to MDAV, which has low computational complexity, medium communication overhead, and medium energy consumption.

Negative reconstruction: Different from all above methods, Horey *et al.* [105] propose a negative reconstruction anonymization method, where sensor nodes transmit a negative sample of the data to a base station instead of transmitting their actual data. The base station then uses these negative samples to reconstruct a histogram of the original sensor readings, which are the data samples that are not collected.

As there is no more extra reporting messages, the communication overhead and energy consumption are low. The computation is also easy. Proved by their evaluations, the information loss revealing from the construction accuracy is high, which increases with the number of categories in a near-linear fashion.

Data integrity: With the k-anonymity, it will damage the data integrity to some extent. However, achieving data integrity while proposing privacy is an indispensable requirement for keeping the trustworthy and user-friendly service in crowdsensing. Murshed *et al.* [106] find that the specific selection on the $k - 1$ PoIs in k-anonymity can realize the compensation to the actual price of PoIs. This compensation can provide the maximum match with deduced prices, which keeps the data integrity at the application platform side. Certainly, the k-anonymity applied here preserves the privacy. It is obvious that the picking process will increase the computation complexity to a medium level. But no more extra message means the low communication overhead and energy consumption. For 5 PoIs and 2 anonymity requirements, it achieves 93.51% integrity implying a low information loss.

Another direction to keep a balance between privacy and data integrity is to design a privacy-preserving incentive mechanism. Alsheikh *et al.* [107] decide the level of protection into the reward allocation to encourage participators to upload true data. This paper adapts the k-anonymity into privacy-preserving design, while participators can freely select their privacy levels without knowing the preferences of others. The contribution of each participator will be calculated on the accuracy of their uploaded data. Compared with the accuracy acquired from the whole collected dataset, if there is no accuracy enhancement, this participator will get zero or negative payoff. The only performance difference between this design with k-anonymity is this accuracy calculation, which increases the computation complexity while other factors are the same.

2) Distributed Methods: If each mobile node maintains its own local database and no central entity needs to know these locations, this mode can be considered as a distributed one [10]. Some typical security strategies are introduced as follows.

PriSense: This method is combined with three steps: slicing, mixing, and aggregating. Generally speaking, each mobile node has to slice its data into $n + 1$ slices and randomly choose n other nodes to send a unique data slice. The aggregation means that every node sends their remaining slices together with other received node slices to the platform [108].

Since every node has to make its own slices and send them, the computation complexity and energy consumption to store and transmit these slices are high. For communication overhead, after calculating the sum of communication cost incurring from count query, the result can be 10^4 Bytes (around 10 KB), which is set as a medium level. But the information loss is low because data themselves are not modified [10].

PiRi: The term "PiRi" represents Partial-inclusiveness and Range independence. Considering that the range of queries sent by users have significant overlaps, only a group of representative participators are chosen to be protected [109].

This method firstly assumes that participators trust each other, and do not reveal any sensitive information about their peers, so the main approach is to communicate with the peers instead of a centralized platform. Each participator can determine its privacy level: K and A. K determines the k-anonymity, and A specifies the minimum resolution of the cloaked region. Different from the MDAV, the user communicates with its $k - 1$ closest peers to define its own cloaked region rather than letting platform do this. Then, each user computes its Voronoi cell [110] and defines the smallest circle containing this Voronoi cell. This circle is used as the influence radius r_u for calculating the score of each user. Finally, the platform only receives the data from one user in the cloaked region who has the highest score.

Obviously, the Voronoi computation process increases the level of computation complexity. The communication process with peers increases the level of communication overhead and energy consumption. In this paper, the authors use privacy leak matric to evaluate the information loss and show a low-level result which is only 3% difference.

Trustworthiness: To make sure the trustworthiness in distributed methods, the Anonymous Authentication of Visitors (AAV) is proposed [111] with two phases: certified pseudonym issuing phase and subsequent interaction phase.

In the certified pseudonym issuing phase, the user generates its pseudonym P and utilizes the partially blind signature scheme to hide P in a blinded message B. The mobile app then sends the ticket ID along with B to the app platform. The app platform verifies the validity of ticket ID and inputs an expiry date while digitally signing B. As a result, the app platform has no clue about user's pseudonym and possible linkages.

In the subsequent interaction phase, the mobile app uses P and S to send user's sensitive information. Since the signature on B from the app platform has been unblinded to the bare pseudonym P, the app platform can easily verify whether the pseudonym P sent by the mobile app matches with the pseudonym signed in the signature S and also whether it is within the expiry date to finish this authentication process.

As there is no more required processing or transmission in the AAV, these four evaluation elements (*i.e.*, computation complexity, communication overhead, energy consumption, and information loss) are all at a low level. According to their evaluations, the average response time of servers is below 6ms even with the increasing requests per second.

B. Obfuscation

Different from anonymization, obfuscation tends to modify the original data of users independently, without mixing with other users' data. Depending on this character, there are more distributed methods than centralized methods introduced below.

1) Position Sharing: Dürr *et al.* [112] propose a novel position sharing approach for the secure management of position information in partially trusted systems of location servers and location-based services. Users split up their precise position into position shares of limited precision, which are also known as the user-defined trust level. Then, these obfuscated shares will be distributed among a set of location servers of different operators. After that, even a revealing happens, the leaked information will only within the strictly limited precision range.

It is a more flexible method as the trust level can be defined depending on different Location-Based Service (LBS) precision requirements and different user requirements. Since no entity can get the whole location information of nodes, this method is a distributed one. But the generation os shares has to be performed frequently for every position, so the computation complexity and energy consumption are high. The average share generation time can easily be over 1 second for 128 positions. However, each share is in a small size (40 bytes) containing 32 bytes' user id and 8 bytes' translation vector, and the transmission in UDP and binary format will cost 550 bytes for 8 shares. So the frequent transmission only leads to a medium level of communication overhead. Besides, no modification to actual data causes low information loss.

2) Data Perturbation: Based on the lack of hierarchical trust structure, it is necessary to protect original data from its source using data perturbation methods [113]. The general idea is to add random noises with a known distribution to the user's data, after which a reconstruction algorithm is used to estimate the distribution of the original data. This technique allows users to perturb private measurements before sharing locally, so it can be considered as another distributed method leading to low communication overhead and energy consumption. However, this local computation also adds the computation complexity for every mobile node, and it also brings noise into data, which makes the information loss is at a medium level.

Additionally, the sensing node in MCS task can dynamically join or leave. To meet this situation and decrease computation

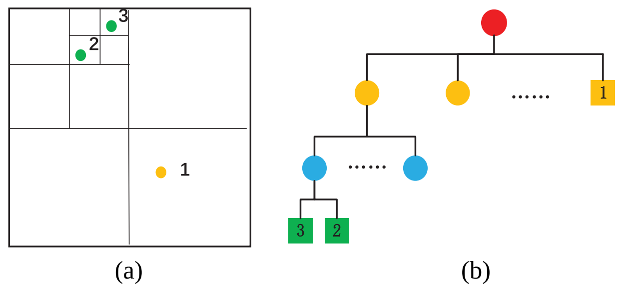

Fig. 6. An example of Quadtree-based obfuscation. (a) A region and its decomposition quadrants. (b) Quadtree representation of the region in (a).

complexity, Li and Cao [114] leverage a novel ring-based interleaved grouping technique. In this method, only a small number of nodes need to update their cryptographic keys when the situation of these nodes are changed. They add $O(1)$ noise to the sum to ensure differential privacy. Here, the definition of differential privacy is [115].

Definition 1 (ε-Differential Privacy): Consider the data D as input and data f as output, then the algorithm is ε-differential privacy when

$$\frac{P(f(D) \in S)}{P(f(D') \in S)} \leq e^{\varepsilon}, \tag{4}$$

where all measurable $S \subset T$ of the input range. For all data sets, D and D' differ in a single item.

The noise continues to be enhanced. The noise is computed by loss function (e.g., element-wise independent *Laplace* noise) and then added to the average gradient to gain ε-differential privacy property. As only a small set of nodes needs to upload their encrypted data, this ε-differential privacy is welcomed in sparse crowdsensing [116]. The introduction of sparse crowdsensing can be found in Section V-C.

Although the running time is only 4.6 ms for both encryption and decryption when the node number n is 10^3, it exponentially increases to 15.4 ms for $n = 10^4$ and 123.4 ms for $n = 10^5$. Compared with other methods, we decide its computation complexity at a medium level. As for communication cost, it is affected by the maximum fraction of nodes compromised. Set this fraction as 0.1, the number of updated nodes in the system is only 120 for both join and leave stages, which is at a low cost. And the aggregation error is only 11.4%, indicating also a low information loss.

3) Quadtree-Based Obfuscation: Instead of the above solution of obfuscating location by defining a circle or square around the current location of the participants, Krontiris and Dimitriou [117] propose a **quadtree-based obfuscation**. It is a flexible method as the users can define their own radius of the blurring region around their true locations. These locations can be represented by quadtrees. The concerned space is first partitioned into two dimensions, then decomposed into four equal quadrants, which are followed by subquadrants, and so on, until a predefined limit (the radius) is reached. Secondly, they build the correlated quadtree. The root node is the region they focused, and the child node is the quadrant of the region. The deeper they go, the more accurate the location will be, until reaching limitation. Finally, based on this quadtree, the location information can be obfuscated by declaring their position on these fixed quadrants.

As shown in Figure 6 (a), the biggest rectangular represents the focused region and the points inside are Mobile Objects (MO). Each object obfuscates its location at different level. MO_1 has higher obfuscation level than MO_2 and MO_3. Its corresponding quadtree is shown in Figure 6 (b). The MOs with the same granularity will be linked in the same layer. Since the maximal allowed location granularity is predefined as $f_{max} = 3$, the height of this quadtree is also 3.

In this method, the users can obfuscate the location information through their own agents, so no entity knows the whole data points, implying the distributed property of this method. Proved in their experiments, most (51.39%) traversed trees are created at the first level of query, which costs the medium level of computational delay in the system. Other properties like energy consumption, communication overhead, and information loss are all low.

C. Encryption

The main idea of encryption is using cryptographic methods or building secure channels during the reporting process [98], [118], which also bring high communication overhead, energy consumption, and computation complexity. But rather than modifying the actual data, it seems more like a lock on data, so the information loss is low. Besides, in this distributed mode, no entity knows the real locations of users. Next, we will describe several encryption methods used in MCS.

1) Group Signature: In Anonysense [98], the group signature can not only protect the users' identities without being disclosed, but also guarantee the integrity of the reported data. The registration authority in this system gives registered users a group of unique certifications which are used to encrypt their own data and identities. For integrity, the RS only receives the data from the appropriate certification, so the information loss in this system is low. But the storage space for all certifications and the cost of corresponding computations are nonnegligible.

2) Split Encryption: Another method to ensure privacy and integrity simultaneously is the split encryption [119]. In its probabilistic privacy architecture, it splits the actions of authentication and data processing into two different entities in the platform: an ID proxy server and an application server. The application server uses its own public key to encrypt sensed data and uses the public key from ID proxy to encrypt identity information. So the ID proxy can only guarantee the integrity of the data with its authentication system rather than knowing the real data measurement. Considering that it is difficult for adversaries to control both these two entities at the same time, the entire message is hard to be disclosed.

To deal with the costly property of this method, Vergara-Laurens and Labrador [120] introduce an energy-efficient and accurate privacy-preserving scheme, which combines this split encryption with anonymized methods. It divides the sensed data into two sets: the first one uses this split encryption while the other uses an anonymization scheme to report the data. According to the evaluation in this paper, the similar information loss will be caused when the thresholds are set as 0.7 and 0.5, where the half encryption on messages (threshold of 0.5) can save more energy. It means that this mode can not only achieve the desired privacy and accuracy of location information, but can also become energy-efficient.

3) K-Vector Perturbation: In Privacy-Preserving Compressive Sensing (PPCS) [121], the author uses the *K*-Vector Perturbation (KVP) to obfuscate the incomplete location data at the first step. The inverse KVP is then leveraged to restore the original trajectory. The main idea of KVP is to use *K* other trajectories to perturb the target one while maintaining the homomorphic obfuscation property for compressive sensing. Specifically, the private user *i* randomly downloads *K* public vectors $D_{(1)}, D_{(2)}, \ldots, D_{(k)}$ from the platform. This random process can bring more uncertainty for privacy preserving. Then the user *i* generates a $(K+1)$ random vector as its private key without being known by anyone else, so the encrypted vector $S_i{}'$ can be calculated by the fusion of original trajectory with public vectors. Finally, the encrypted vector $S'_{(i)}$ is then transmitted to the platform.

Proved by simulated results, over 90% recovery errors are lower than 250 meters, which can also be eliminated by map matching [122]. So the information loss of it is at a low level. Additionally, the storage of random vectors and the encrypted vector calculation both increase the computation complexity and energy consumption, while the communication overhead is more because of the downloading and uploading stages.

D. Authentication

Malicious false attack in MCS is easily ignored but devastating. On the one hand, some mischievous or malicious users seek to fool the system by falsifying data reporting. On the other hand, adversaries will pretend to be the normal user and report a flood of copies to overload crowdsensing applications (*i.e.*, DoS attack), who are called Sybil devices. So the proper authentication is needed to against this attack. In the next subsections, some authentications in practice are described. As the original data is not modified in all of these methods, the information loss is all at a low level.

1) Grid Detection: Fatemieh *et al.* [123] regard the area of interest as a grid of square cells. This proposed mechanism is based on identifying outlier measurements inside of these cells, as well as corroboration among neighboring cells in a hierarchical structure to find out malicious nodes.

Consider a cell C_j containing *m* nodes and a dispute threshold for this cell d_0, which is the maximum acceptable difference between the measurements of two nodes in that cell. Assuming that each pairwise comparison nodes are N_i and N_j, if the difference is greater than d_0, the dispute counts c_i and c_j for N_i and N_j respectively, are increased by one. After all pairwise comparisons, if $\frac{c_i}{m}$ is greater than or equal to the outlier threshold, the node is flagged as an outlier.

As no more extra messages need to be sent, the communication overhead and energy consumption are low. But because of the detection algorithm introduced above, its computation complexity is at a medium level.

2) DoS-Resistant Authentication: Due to the openness of the MCS system, the malicious attack is likely to generate task abortion by giving Denial of Service (DoS) attack. As the improvement of Multi-level Timed Efficient Stream Loss-tolerant Authentication (μTESLA), Ruan *et al.* [124] formulate the attack-defense model as an evolutionary game, and then presents an optimal solution, which achieves security assurance along with minimum resource cost.

This method firstly sets multiple buffers for nodes and randomly selects packages stored in node buffers. Secondly, Message Authentication Codes (MACs) are broadcasted, and then μMACs calculated with a hash function, are stored in nodes. Finally, after the key is disclosed, the message will be sent and the receiver can use this disclosed key to compute the theoretical MAC of the received packet. After comparing it with its attached MAC, the received packets are authenticated.

The 56Kb storage needed for each packet and the bandwidth required for MACs is 0.6G which shows a high level of communication overhead and energy consumption. But only the authentication comparison should be calculated, so the computation complexity is medium.

3) Co-Location Edges Authentication: Wang *et al.* [125] intend to defend against Sybil devices based on co-location edges. These edges are the authenticated records that attesting to the one-time physical co-location of a pair of devices. As Sybil devices cannot physically interact with real devices, the edges between them and the rest of the network cannot be formed. Based on this, the problem of detecting ghost riders is simplified as a detection problem on the proximity graph. As a detail, after creating the co-location edges graph, SybilRank [126] algorithm firstly computes the landing probability for short random walks from trusted nodes to land on all other nodes. Then normalized by the nodes' degrees, it calculates their landing probabilities as the trust scores for ranking. As short random walks from trusted nodes are very unlikely to traverse the few attack edges to reach Sybil nodes, the ranking scores of Sybil devices will be lower. So a cutoff threshold can be set on the trust score, and the tail of the ranked list is labeled as Sybil devices.

Intuitively, the ranking algorithm increases the computation complexity. But no more extra messages are sent, so the communication overhead and energy consumption are low. Reversely, according to its simulation results, the cost for Sybil attacks to break its defense is tremendous, which requires 60k attack edges to maintain 3k Sybil devices, illustrating the feasibility of this method.

V. Resource Optimization

One key factor considered by MCS requestors is the total cost required, including the incentives paid to MCS participants and/or the total energy consumption of all mobile devices (*e.g.*, battery and bandwidth consumption, subscription contract, or transportation fees) [127]. Generally speaking, requestors should consider not only the budget limits, but also the coverage requirement and data quality for a satisfactory sensing result. The trade-off between these considerations leads to the proposal of some resource optimization strategies summarized in this section. We review the related research papers from four aspects: participant selection, task

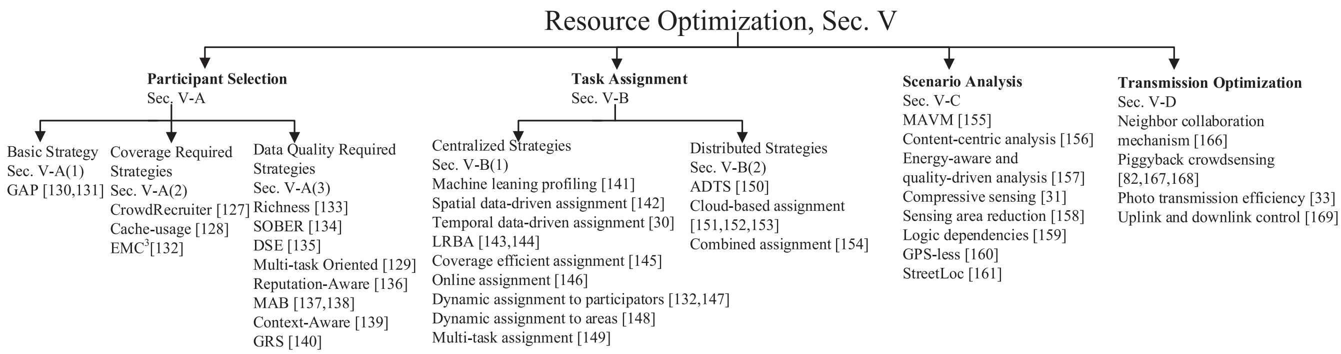

Fig. 7. A taxonomy of the resource optimization.

assignment, scenario analysis and transmission optimization, to avoid redundancy and improve efficiency in sensing projects. All related work will be summarized in taxonomy Figure 7 and Table IV.

A. Participant Selection

Different from the incentives rewarding for participators, participant selection strategies are targeting on the cost optimization while ensuring coverage and QoI from the requestor side. While selecting the large number (even the whole group) of participants can bring better sensing coverage and more accurate analysis, the costs mentioned before can be tremendous [128]. Besides, participators vary from each other on different incentive requirements, sensing capabilities and mobility patterns [129]. In order to optimize the **required sensing coverage** and **data quality** within budget limits, how to select the most appropriate participators in a large amount of candidates is discussed in this subsection.

1) Basic Strategy: The Participant Selection Problem (PSP) is a special case of the General Assignment Problem (GAP) [130]. The definition of this problem is to arrange objects in bins under the constraint of the weight limit and make sure the maximization of the total value. In crowdsensing PSP, participants are objects and the constraint is set by strategy makers, about task budgets or spatial coverage. This GAP problem is considered as the NP-hard problem where the optimizing method should be applied to solve it, and the straightforward solution is to use the greedy algorithm. Although there are numbers of variations introduced next, they are all enhanced from the greedy algorithm. That is, continuous data reporting until reaching a certain amount of contribution is an effective way to harvest data from the crowd [131].

Chu *et al.* [130] describe the greedy algorithm used in crowdsourcing tasks, which is called PSP-G. The iterative processes of this method are as follows:

- Firstly, every participator's Benefit-To-Cost (B2C) factor needs to be calculated by dividing its value to its cost.
- Secondly, all B2C factors are sorted in descending number.
- Thirdly, the highest B2C factor is selected and the corresponding person is dispatched to the region where he can contribute the most under constraints.
- Finally, the former steps are repeated until all participators are dispatched or the total value reaches to the peak.

2) Coverage Required Strategies: The goal of PSP in this part is to select the smallest number of participators to satisfy the coverage constraint. CrowdRecruiter [127] shows a two-phase participant selection framework: the first phase is to map participants' traces and predict their mobility, and the second phase is the calculation of the joint coverage probability for multiple users and iterative selection on near-optimal sets. Different from GAP [130], the selection criteria in CrowdRecruiter is the future locations of participators, which are predicted by the mobility model constructed on historical records. The threshold of this iteration is a pre-defined spatial coverage.

In addition to the location information, Li *et al.* [128] pay attention to the participators' sensing contributions and proposes a dynamic participant selection strategy. It uses cache to dynamically record the changing of sensing tasks and sensing data, where the new sensing data and new tasks can enter the cache at any time. For participant record, when new data are uploaded, the record of this participator's contribution value will be increased. And any selected participators who have "0" contribution values will be removed before the next sensing cycle. For task record, when the sensing data hits a certain sensing task, the corresponding frequency of this task will be reduced by 1. Similarly, the "0" frequency task will also be removed. These updating processes on task set and selection set represent the meaning of "dynamic". Through the design of cache, not only the historical call and location traces of mobile users can be learned, but also the distribution of possible future tasks can be predicted, where the suitable participators will be selected by the evaluation on their abilities to a certain task set.

Another dynamic PSP solution is inspired from piggyback transmission which is named as EMC^3 [132]. By collecting the historical call records of individuals, it is possible to construct the call/mobility pattern of a user and predict her next move. If this user is most likely to place two calls in the target area or will arrive at a low-density area, she will be selected as the candidate.

3) QoI Driven Strategies: The other group of methods are selecting the most qualified participants under budget constraint, where the QoI is the selection criteria.

TABLE IV

SUMMARY OF RESOURCE OPTIMIZATION (TYPE: S=STATIC, D=DYNAMIC; SCH.=SCHEME: C=CENTRALIZED, D=DISTRIBUTED; SC= SPATIAL COVERAGE; TE=TIME EFFICIENCY; QOI=QUALITY OF INFORMATION; EC=ENERGY CONSUMPTION; CON.=CONTEXT)

	Name	Type	Sch.	SC	TE	QoI	EC	Con.	Pros	Cons
Participant Selection	GAP	S	C			X	X		Easy to implement	Not practical
	CrowdRecruiter	S	C	X			X		Increasing the speed of selection in coverage requirement	No QoI guarantee
	Cache-usage	D	C	X	X				Dynamic selection	Low robustness
	EMC^3	D	C	X			X		Dynamic energy consideration	Complex comparison Low robustness
	Richness	D	C			X	X		Rich considerations	
	SOBER	S	C		X	X	X		Both subjective and objective considerations	
	DSE	S	C	X		X	X		Relatively rich considerations	
	Multi-task Oriented	D	C	X		X			Consider both data granularity and quantity	Ignore energy consumption
	Reputation-aware	D	C			X			Consider sociability character	Too simple willingness representation
	MAB	S	C			X			More accurate leaning	High selection delay
	Context-Aware	S	C					X	No need for ground truth	Coarse evaluation
	GRS	S	C	X	X	X	X	X	Rich Consideration	High computation overhead
Task Assignment	ML Profiling	S	C	X			X	X	User preference consideration	Less QoI monitoring
	Spatial data-driven assignment	D	C	X					Accurate location analysis	Less QoI monitoring
	Temporal data-driven assignment	D	C		X				Efficient time allocation	Less QoI monitoring
	LRBA	S	C+D	X	X	X			Rich consideration	Less QoI monitoring
	Coverage efficient assignment	D	C	X	X				Efficient coverage insurance and sensing time reduction	Less QoI monitoring
	Online assignment	D	C	X	X	X			Online concerning and fairness	High computational burden
	Dynamic assignment to participators	D	C	X		X			Practical consideration	High computational burden
	Dynamic assignment to areas	D	C	X					Avoid data redundancy	Less QoI monitoring
	Multi-task assignment	S	C	X		X			Comprehensive consideration	Computational overhead
	ADTS	S	D	X	X		X		Fairness and privacy preserving	Low robustness
	Cloud-based assignment	D	D	X		X	X	X	Rich consideration	High transmission overhead
	Combined assignment	S	C+D	X			X		More flexible	The tradeoff between global assessment and individual fine-tuning is difficult
Scenario Analysis	MAVM	S	D					X	Reduce bandwidth occupation	Computational burden for participators
	Content-centric analysis	S	D					X	Redundancy reduction	Transmission overhead
	Energy-aware and quality-driven analysis	S	C	X	X	X	X	X	Rich consideration	Computational overhead
	Compressive sensing	S	C	X	X				Data reduction	Reconstructed loss
	Sensing area reduction	S	C	X	X	X			Sensor reduction and quality guarantee	Reconstructed loss
	Logic dependencies	S	C	X	X			X	Energy reduction and quality guarantee	Specific application target
	GPS-less	S	C	X			X		Energy efficiency	Coverage loss
	StreetLoc	S	C	X	X		X	X	Observed based less GPS	Coarse estimation
Transmission Optimization	Neighbor collaboration mechanism	S	C	X		X			Efficient communication, avoid frequent neighbor discovery, low frequency of link refresh, decrease packet collision	Less QoI monitoring
	Piggyback Crowdsensing	S	C	X		X			Reduce user involvement, balance the current and future opportunity, efficient coverage	Privacy ignorance
	Photo transmission efficiency	S	C	X	X			X	Low communication overhead, low battery consumption, low CPU occupation, bandwidth occupation reduction	Accuracy loss
	Uplink and downlink control	S	C				X		Low communication overhead, bandwidth occupation reduction	Computational burden

Intuitively, the evaluation on equipped devices for participators can indicate their qualifications, including the battery, network bandwidth, sensor state, number of sensors, and rated sensor powers [133]. Similar to the Sociability-Oriented and Battery-Efficient Recruitment (SOBER) [134], Fiandrino *et al.* [135] design a participant recruitment strategy named DSE. "D" inside is *Distance*, representing the distance between the candidate and sensing task location; "S" is *Sociability*, implying the willingness of a candidate the participate in sensing tasks; and "E" is *Energy*, concerning about the remaining battery of the mobile device. To evaluate the quality of candidate i, her recruitment factor R_i is calculated by the weighted sum of D, S, and E, where the weight of each factor is application-dependent. As the sum of these three parameters equals unity, each higher value implies the preference of certain recruitment strategy. Finally, only the recruitment factor

R which is above the threshold will be considered and the highest several *R*s will be selected until reaching the budget constraint.

Song *et al.* [129] dynamically select a minimum subset of participants to provide the best QoI satisfaction metrics for all tasks. Although it still uses the greedy algorithm, the selection criteria of this paper are comprehensive, including the expected amount of collected data, the required QoI and users' reward expectations. In this method, the QoI metrics is calculated by data granularity and quantity. For example, the sensing tasks require the amount of data as $(3, 3, 2, 1)$ in four areas respectively, and there are three participants (*a, b, c*). If we choose (*a, b*), they have abilities to assembly collect $(3, 2, 2, 1)$ in these areas, while if we choose (*b, c*), the collect result is $(3, 3, 1, 0)$. Although (*b, c*) finish the former two tasks, they collect less data in the latter two tasks. Compared with them, (*a, b*) finish tasks more balanced and efficient. Mathematically, the authors calculate the ratio between the required metrics and the collected metrics to define this QoI. They declare that the QoI of (*a, b*) is higher than (*b, c*).

What's more, Wang *et al.* [136] take reputation values of participants into consideration. Two attributes are needed to define this reputation value: participation willingness and data quality. Inspired by the social principle, willingness is calculated by the average time gap between two collected behaviors. The shorter the time gap is, the more enthusiastic the participator will be. So, the feedback value for each participator can be represented by the combination of this calculated willingness, data quality, and rewards. After participators contribute sensing data to the platform, their reputation values will be monitored by watchdog and updated dynamically according to this feedback.

Although the PSP problem is considered similar to the knapsack problem, the uncertainty of sensing values and quality make it more challenging than the knapsack problem. To minimize the difference between the achieved total sensing revenue and the optimal one, Han *et al.* [137] give an online learning algorithm based on the Multi-Armed Bandit (MAB) paradigm [138] to acquire the statistical information on sensing values and qualities during the participation selection process. However, the lack of ground truth data leads to the failure of data quality estimation. To solve this problem, a context-aware data quality estimation scheme is proposed [139]. Depending on the historical sensing data, a context-quality classifier indicates the relationship between the context information of participators (i.e., keeping still, walking, or running) and their sensing data quality. Some volunteers like the running ones (have low sensing quality in noise sensing) will be filtered out by this context recognition rather than data quality estimation.

As the above-mentioned strategies are all considering individual-based selection, Azzam *et al.* [140] design a Group-based Recruitment System (GRS) to assess the QoI in a group of participators collectively. The QoI of each group will be represented by their own fitness value, including the coverage of sensing tasks, members' distributions, device availability, the reputation of participators, sampling frequency, residual energy, and group cost. A mutate algorithm is applied to select the most fit groups until occurring the convergence, where maximal QoI remains unchanged after several iterations.

B. Task Assignment

In this subsection, we will describe two main kinds of task assignment strategies: centralized strategies and distributed strategies, that is, platform-centric decision-making strategies and user-centric strategies.

1) Centralized Strategies: To globally monitor the processing of MCS tasks, requestors would like to regard the platform as the controller for task assignment. But it does not mean that the platform can make exclusive decisions without considering participants. Some researchers [132], [141] pay attention to user preferences to make a more practical arrangement for platform-centric mode. According to them, user preferences include five factors: energy consumption, the payment of contributing data, distance away from the place of interest, task context, and privacy. Karaliopoulos *et al.* [141] take the logistic-regression machine learning technique to profile users from the past data on user preferences. So the assignment problem converts to a sigmoid optimization problem, which determines the task offered to the most suitable users.

Different from the user-centric model, it is difficult for the platform to know the spatial and temporal information of participants in real time. So, except the preferences, the locations and the execution time of users can also be meaningful. Spatially, Pournajaf *et al.* [142] collect the historical trajectories of participators to predict their next locations, which can be the substitution of the real-time location report. Temporarily, Boutsis and Kalogeraki [30] decide to analyze the historical execution time to estimate the finishing probability before the deadline, which is supported by the Power Law Distribution Theory proposed by Ipeirotis.

Both spatially and temporarily, He *et al.* [143] combine both spatial and temporal information and observe the location-dependent character of crowdsensing, that is, sensing task at specific place requires a certain amount of traveling time for different mobile users. Considering the time budget of users, the authors design an approximated Local Ratio Based Algorithm (LRBA) to iteratively solve the allocation problem, which can efficiently promote users' sensing ability. Further proved, LRBA can also be performed in a distributed way, which is operated on the device side, and achieves the same result with the centralized scheme [144]. Similarly, Obinikpo *et al.* [145] target on efficient coverage on sensing areas. Firstly, they model the entire coverage system as a birth-and-death mechanism, where any new participator arriving implies a birth process and her exit after finishing sensing tasks signifies a death process. As for the spatial analysis, based on time constraints, the mean covered number of target areas can be calculated to reduce redundancy. The task in a certain sensing area will be assigned to the closest participator. And as for the temporal analysis, the average waiting time for a target area to be sensed will then be calculated. The sensing target prepared for any busy participator will be automatically passed to the next available participator to save this waiting time. Moreover, Wang *et al.* [146] emphasize on the online

MCS, which can capture location diversity of mobile participators and dynamic task arrival time. The model is further extended to maximize the proportional fairness for a fair task allocation among participators.

Reversely, another type of tasks is called time-delay tasks. For example, the temperature analysis needs the temperature data which are allowed to collect from a place during the whole day and be uploaded before the next day. Without time bounding, researchers can consider dynamic solutions.

Xiong *et al.* [132], [147] propose a dynamic task assignment strategy to ensure the full coverage of the target area. The current states of candidates will be evaluated and assigned to corresponding tasks. Three states are considered: (i) If a participator has not called the first time, she will receive her task information; (ii) If a participator has finished her first call, she is required to upload his data at her next call; (iii) If a participator has finished two calls in a sensing cycle, she will not be assigned in this cycle. During the iteration of this assignment, the performances of candidates are assessed according to their contribution and the task will be dynamically assigned to the most suitable participator in the future round. And this iteration will stop until the received data reaches the pre-defined amount or the sensing areas is fully covered.

Besides dynamic task assignment, researchers also consider about dynamic task area assignment. Depending on the observation that in adjacent time and areas, sensing data is similar in one sensing cycle. So the minimum cells can be selected and assigned suitable tasks to avoid data redundancy [148].

From the above discussion, for the time-sensitive tasks, the instant locations of users are analyzed. While for the time-delay tasks, the historical trajectory and mobility profiling are considered to make assignment. But the afore-mentioned strategies only talk about the single-task assignment, Guo *et al.* [149] discuss the multi-task assignment for both of these two kinds of tasks and a greedy-enhanced genetic algorithm is proposed to solve them.

2) Distributed Strategies: As the central strategies considering the real mobility data of participators, the privacy involved is sensitive to participators [150]. To solve this problem, researchers tend to put users as the centric of the task control system and let them choose any task they like without reporting their own situations. Asynchronous and Distributed Task Selection (ADTS) algorithm [150] represents people to make decisions using their information offline. By inputting the cost on movement, speed, location and the number of participants, this algorithm will find the largest payoff route and report to the platform as participator's selection. This algorithm considers both participators' interests and fairness, and it will not leak private information because of its local operation.

Another distributed framework [151] lets each device compute their data collection utility, sensing potential, and environmental context to comprehensively determine whether to take the sensing tasks. The data collection utility is depending on the amount of already-collected data in certain sensing area which is feedback by cloud collectors [152]. The higher requirement for further uploading leads to a larger value of this parameter. The sensing potential is represented by the local energy consumption for mobile devices to sense and upload sensing data. And the environmental context circumstances the status of the mobile device, such as the location or the mobility pattern [153].

Moreover, there are some methods combining both centric and distributed models. Pournajaf *et al.* [154] introduce a two-stage optimization approach: the first stage is a global assignment according to the cloaked locations of users, and the second stage is a client-end fine-tuning stage where the users can slightly adjust their interested tasks without destroying the balance of the whole assignment picture. Here, the first stage is centric and the second is distributed, which combines the advantages of the two modes and make it more flexible.

C. Scenario Analysis

In MCS tasks, the data-providers could continuously upload their sensed data to the platform in the moving situation. But sometimes, their mobility may lead to redundant coverage on certain areas. On the task side, different application interests on different cells. For example, the traffic monitoring applications focus on the highways and roads while the parking applications pay more attention to the parking lots. For both of these two reasons, it is necessary to give sensor management according to the specific scenario analysis.

In the relevant papers, words like "context", "semantics", "location dependent" are often linked with scenario analysis. It reveals that the goal of scenario analysis strategies is to find the context of sensing situation which can be represented by the specific interests on locations or semantics. To solve the large-scale mobile crowdsensing task, Xiao *et al.* [155] divide the target areas into lots of "cloudlets". These cloudlets can be either a business, community or even a data center which means a place for a special group. This decentralized method is operated by a master application called Master Application Virtual Machine (MAVM), which is beneficial to reduce bandwidth occupation. Another explanation of semantics is the concept of "content-centric", where the semantics of sensed data are represented by its expressed contents. Corresponding to different kinds of mobile devices (*e.g.*, smartphones, wearable devices, smart vehicles, etc.), the sources can be used from the crowd are various, deriving complex semantics discussed here. For instance, the semantics of sensed data in vehicular clouds can be road construction, gas station information, parking availability, and so on [156]. While the semantics of sensed data in smartwatch can be jogging steps, heartbeats, or gym locations.

Besides the context needed to be considered, the inner structure of crowd data can also be mined to reduce sampling burdens, including sensor and data reduction. Marjanović *et al.* [157] intend to give an energy-aware and quality-driven sensor management algorithm. The "energy-aware" means fewer sensor selection and "quality-driven" means the insurance of enough sensing coverage. Firstly, the algorithm gives an application-specific valuation function for sensor selection, including the different weight for each factor varied on different applications, battery level, trustworthiness level, sensor location goodness, data speed, and the current state of sensors (active or inactive). By evaluating different values of each sensor, the algorithm chooses the top *k* sensors which can satisfy the needed coverage to finish crowdsensing tasks. This *k* index is also application dependent. For instance, to monitor air quality, the number of needed sensors in the city should be bigger than in the park.

Additionally, Xu *et al.* [31] care more about data reduction. Assuming that there is the spatial-temporal relationship among data, which is called the underlying sparse structure, this paper proposes compressive sensing to reduce sampling rate depending on this relationship. This method can get a lower sampling rate than the Nyquist sampling rate and can still guarantee an accurately reconstructed signal.

To better understand, they give a function:

$$Y = \Psi X, \tag{5}$$

where *Y* denotes their interested signal, Ψ is the base of converts and *X* is a *k*-sparse coefficient vector who only has *k* non-zero entries. This function represents the compressibility of signal *Y*. If *Y* can be projected into a Discrete Cosine Transform (DCT) base and only get several sparse representations, then the number of these representations is *k*.

In $Z = \Phi Y = \Phi\Psi X$, Φ is a random partial identity matrix which is known before. It represents the sampling process which randomly picking on the rows of ΨX. After acquiring sampled data, reconstruction should be made to recover the original signal. It combines two process: 1-norm minimization to find *X* and base training to get Φ. As the evaluation in this paper, the reconstructed rate can reach over 90%.

Similarly, Wang *et al.* [158] also make use of compressive sensing to reduce the amount of sensing data, by narrowing down the sensing areas instead. Based on the spatial-temporal relationships between adjacent urban areas, this paper only allocates individual task in a small subset areas and calculates the data on the other sets of areas by the missing data inference algorithm in compressive sensing. Through this sparse sensing, the satisfied data quality can be achieved and the required participant sensors are reduced at the same time.

Together with the underlying sparse structure of data, the specific application knowledge and logic dependencies among data items can be exploited to reduce the underlying network bandwidth consumption [159]. Some severe scenarios, like disaster response or humanitarian assistance missions, have well-defined protocols for carrying out the sensing tasks and specific concerned areas or objects. For example, the earthquake rescuing task pays attention to the destroyed degree of buildings and some closest cabins which can become temporary shelters. Only the pictures on architectures and the aerial images are needed to analyze the maintenance or occupation conditions in surroundings. Further enhanced, the population distribution data, can imply some possible shelters which is unlikely to be occupied, then aerial images covering the candidate paths do not need to be gathered.

Another innovation on energy-efficient strategy is the GPS-less sensing scheduling [160]. As a matter of the high consumption on GPS location, the authors tend to weaken the function of GPS and use a probabilistic model for sensing coverage without accurate location information. Actually, this paper uses Google location service to get a rough location of users and leverages the relationship between location disk and the target disk to calculate the coverage probability. The disk mentioned here is a circular coverage model defined in the paper. As the known of reported location and target location, this calculation is feasible. Similarly, another coarse location prediction is proposed by StreetLoc [161]. As the target sensing area of this paper is streets in urban areas. Its instant location scheme is built depending on three observations: (i) Pedestrians generally follow the linear path of the walking street; (ii) The walking speeds of participators are considered uniform in a street segment; (iii) These walking speeds are spatially and temporarily consistent. According to these three observations, only the entrance and exit of a street needed to be detected by GPS. Together with the walking speed estimated from the historical mobility pattern of each participator, the instant location can be calculated and GPS is also deactivated along the street.

D. Transmission Optimization

To process unprecedented amounts of data in MCS, traditional methods tend to give a better WiFi offloading or a small heterogenous cellular network supporting their transmission. As the wireless data has become a "tsunami" [162], continuous reporting is proved as energy consuming [163]. The leverage of point-to-point technologies (*e.g.*, WiFi Direct, LTE Direct) gives rise to opportunistic transmission.

There are two main characters for opportunistic transmission [164]: Firstly, only when two participatory points are in the range of direct radio communication can they contact sporadically. That means, opportunistic transmission is an evolution of the ad hoc network. Secondly, an opportunistic relay is needed to enhance transmission opportunity, instead of the single hop. Making a share among the whole network picture can be of great help [162]. Depending on the large portion of participators, exploiting the natural gregarious attributes among humans and making them cooperate with each other is the foothold to achieve this character [165].

Targeting to this cooperative factor, a neighbor collaboration mechanism is proposed [166]. This paper firstly plans to let each sensing node save its sensed data locally and upload until it shows up around the hotspots of WiFi. Although it can reduce consumption to a large extent, the delay is too long and unstable. Then, this paper tends to make short-distance radio communication with relevant neighboring nodes and opportunistically upload to the platform. However, the frequent neighbor discovery and link establishment still consume a lot of energy. So this paper finds out three problems to solve: trade-off between message delivery delay and energy consumption, robustness against node density and overhead for link establishment.

Given an observation that in crowded areas like shopping malls or parks, people always play in droves. The analysis on their radio connected history can help on finding groups. Each group is regarded as a cluster who has its own local network. The frequent neighbor discovery can be avoided and the link establishment does not need to be often refreshed. By the way, the authors take Bluetooth as the point-to-point technology to solve the packet collision problem in the ad hoc network.

Another improved version of opportunistic sensing is Piggyback Crowdsensing (PCS) [82], [167]. It mainly collects and uploads data on the special situation when sensors have activated a WiFi/LTE like a phone call, an e-mail checking, a website browsing. The collection and uploading processes are running in the background, which greatly reduce user involvement. Besides this channel, some users without the data plan can transfer data to a relay device via "zero-cost" network (*i.e.*, Bluetooth), which can also avoid extra transmission cost [168]. However, the frequent applications on smart devices bring more sensing opportunities, which makes it difficult to choose the best one [82]. Simply using the greedy strategy to allocate opportunities may miss a better opportunity, so a predictive model to balance current and future opportunities is needed. Capturing each user's special app usage pattern, *e.g.*, when the user often calls or surfs on the Internet, the authors can predict and decide when to sample, sense and upload data to the platform. The coverage constraint here should also be considered [132], which we have described in Section V-B.

However, taking photo quality estimation as an example, if we run computer vision algorithm on phones, no matter how the design of algorithm is adaptive or light, the basic consumption on battery charge or CPU will still be large [33]. To make comparison between collected data and the ground truth, those true photos should be downloaded firstly, which will increase the burden of limited bandwidth. So, most researchers decide to run quality management on the platform side. Restricted by bandwidth resource, the authors then transmit metadata of collected photos instead of photos themselves to the platform in order to identify redundant or irrelevant photos. This quality measuring assisted transmission optimization is also applied for data collection for machine learning. Both the uploading and querying cost on sensing data and labels respectively should be considered under the budget limits. Xu and Zheng [169] decide to use both the local and global servers to control it. The distributed local server controls the upload upper limit, while the global server decides the querying limit on annotations based on active learning. Only the most informative samples are floating in their MCS system, which greatly saves the transmission consumption from both uplink and down-link aspects.

VI. Data Analysis

An important task in the MCS system is the aggregation of user-contributed sensing data. However, the large amount of raw data collected from mobile devices needs to be preprocessed before being applied in applications. There are two techniques proposed in this section: quality management and multimodal data analysis. The quality management technique is to evaluate the quality of collected data, delete invalid data, and supplement missing data. Moreover, the multimodal data analysis technique is to mine the potential relationship between the managed data and provide guidance for the next application. In this section, we will present existed analysis techniques on these two parts for a glance. The summarized taxonomy and table will be respectively shown in Figure 8 and Table V.

A. Quality Management

The strategies designed in Sections III and V for QoI improvement are realized by improving the soft reputations of participators, while the hard reputation of sensing devices or other environmental issues should be further managed. In this subsection, the data quality management in MCS is systematically introduced in three stages: before sensing, after sensing but before uploading data, and after uploading stages.

Before sensing, the goal of quality management is to avoid "fake" participators (*i.e.*, programs or virtual machines). As the first firewall in data quality management, various authentication methods have been applied in current MCS systems, including slider dragging, verification code inputting, categorized photo selecting, etc. Similar with the Amazon Turk's login system, some direct questions like "what is the name of this project?" can also work well for simple identity authentication [32]. A real user can quickly give the right answer, while a program will fail without specific coding.

After filtering out virtual participators, the data quality evaluation system is further designed *before uploading stage*. This system aims to improve the quality of uploaded data by pre-processing on the sensed data, where only satisfied data can be selected for transmission. To be efficiently applied in mobile devices, this system should be light-weight, where few computational resources are required in its processing. A light-weight comprehensive validation system optimizes and integrates several evaluation operations in one algorithm [93]: clustering, classification, change detection and frequent patterns analysis. This algorithm is also robust for different types of devices, where parameters related to battery level, CPU usage level, data stream rate, and other systematical settings can be customized.

After uploading, both the detection of faulty data (also called as truth discovery [170] or outlier detection [171]), and the correction on them can help to improve the quality of collected data. As for the detection, the comparison between the collected data and the ground truth can detect redundant or missing data. Take photo quality estimation as an example, some high-quality photos on certain areas can be downloaded online and the comparison is made to filter redundant and irrelevant photos. A straightforward comparison method is to apply computer vision techniques where the target will be recognized in photos (*i.e.*, to evaluate relevance) and the similarity between two photos should be lower than the threshold (*i.e.*, to detect redundancy) [172]–[174]. However, the large cost to run these machine learning models including time consumption on model training, the CPU and GPU operation consumption, and the bandwidth occupation to transmit full pictures or download ground truth cannot be neglected. To solve this problem, Wu *et al.* [33] develop a resource-friendly photo coverage model to quantify the value of photos. Only the metadata of photos are analyzed to infer their coverage to target areas and corresponding quality values. These metadata include locations, orientations, and views of the camera, which will be

Data Analysis, Sec. VI

Quality Management
Sec. VI-A
Question authentication [32]
Combined validation System [93]
Computer vision-based comparison [171-174]
Resource-friendly photo coverage model [33]
Trace graph comparison [175]
Further photo analysis [176]
Correlation regularization model [170]
DETECT-and-CORRECT [171]

Multimodal Data Analysis
Sec. VI-B
Multimodal data process [179]
Lasagna [183]
Unified analysis [175]
OMM [184]
Fuzzy situation inference [180]
CAP [181]

Fig. 8. A taxonomy of the data analysis.

caught by GPS, accelerometers and magnetic field sensors in mobile devices. Consider them as just a series of floating numbers, then the transmission, computation, and storage process can all be light-weight and cost-efficient.

Another observation used in comparison is the similarity of collected data between neighboring areas or during a short term spatially and temporarily. For example, the weather among sub-areas in a district will be similar, and the prices for goods in one supermarket will also be similar during a short period. Taking this observation into consideration, Zhang *et al.* [175] leverage graph comparison on historical human traces to correct the obvious numerical errors and detect missing records. After comparing the data collected in a short time, the authors can filter out the duplicated records. Advanced in technology, Talasila *et al.* [176] require both collected photos and correlated Bluetooth scan results to reveal the device's belonging Bluetooth communication area. Firstly, the framework manually or automatically validates some photos' trustworthiness. This step can be operated by human eyes or graphics recognition algorithm, which takes the ground truth photos collected by trusted experimenters as a baseline. The locations and time of these validated photos will be the referenced data to extend verification with nearby collected point results in the same time. Here, the Bluetooth scan results are used to justify "nearby" data. After analyzing the location of the photo, if it is in the referenced area, it is considered to be true. In Figure 9, if task 12 claims to belong to area A. But after graphic recognition and Bluetooth verification, it is not in A's circle. It will be regarded as a malicious point.

This observation is further applied in an optimization-based truth discovery problem where the ground truth for comparison is unknown [170]. The correlation between entities divides the whole group of data into different independent sets, where the correlated data gather in the same cluster. An objective function is applied to each cluster to measure the difference between the collected data and its unknown truth, adding the reliability of participators as its unknown weights. The ground truth will be estimated and updated depending on the optimization of this function and the correlation regularization terms will punish the deviations in the truths between correlated entities until satisfying the coverage criteria.

For correction, Wang *et al.* propose a DETECT-and-CORRECT framework [171] based on compressive sensing [177]. In the DETECT stage, a time-series based outlier detection algorithm [178] is applied to detect suspicious data. And in the CORRECT stage, these suspicious data will be marked as missing data, and the whole data set will be put into compressive sensing for reconstruction. The difference between this reconstructed matrix and the raw data matrix is iteratively evaluated to re-check the detection result at the first stage. If the difference is lower than the threshold, the detection is considered right and vice versa. The final reconstruction value until the coverage fulfilled is the final correction result. These two stages make each other possible: (i) Compressive sensing can effectively reconstruct missing values effectively; (ii) The biggest difficulty for compressive sensing is the existence of faulty data and it can be eliminated by the DETECT phase.

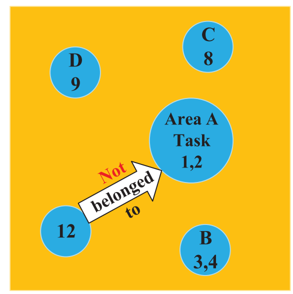

Fig. 9. Real distribution of tasks in areas detected by Bluetooth verification mechanism.

B. Multimodal Data Analysis

The big data collected by MCS have diversed or multimodal characteristics. Different papers have different meanings on "multimodal": some claim that it means different types of data including videos, pictures, sounds, or letters [179], [180]; some consider that it means different resources of data such as cellphone data, transport data, social media data, and so on [175], [181], [182]; and there are still some researchers think that it means different levels of data (*e.g.*, from activity to walk to slow pace), which specifically represent heterogenous motion status [183]. No matter which meaning is referred in the related work, data analysis methods in MCS can be divided into two kinds: separated methods and unified methods.

TABLE V
SUMMARY OF ANALYSIS TECHNIQUES

Model		Method	Pros	Cons
Quality Management	Question authentication	Asking questions	Easy to implement	Simple filter on fake user cannot guarantee the sensing quality
	Combined validation system	Combine clustering, classification, change detection and frequent patterns analysis	1. Integration to be light-weight. 2. Adapt to different kinds of devices.	Cannot avoid the loss in sensing process
	Computer vision-based comparison	Photo similarity machine learning algorithm	Accurate analysis	High computation burden
	Resource-friendly photo coverage model	Meta data transmission	light-weight and cost-efficient transmission, computation and storage process	Accuracy loss
	Trace graph comparison	Node and edges comparison	Missing and duplicated data detection	Unavoidable transmission burden
	Further photo analysis	Graphic recognition algorithm and Bluetooth scan validation	Comprehensive validation	Bluetooth scan consumption
	Correlation regularization model	Divide entities into independent sets and update optimization function	Solve the unknown ground truth problem	Accuracy loss
	DETECT-and-CORRECT	Time-series based outlier detection and data reconstruction	Realize both detection and correction	Computational overhead
Multimodal Data Analysis	Multimodal data process	Separate different type of data into different classifier	Increase analysis accuracy	Computational overhead
	Lasagna	Heterogeneous representation for all activities at multiple resolutions without prior knowledge	Increase analysis accuracy	Cannot represent combined situation
	Unified analysis	1. graphical method 2. Mapping coverage method	Avoid bias	Computational overhead
	OMM	Clustering method	High speed real-time data process	Accuracy loss
	Fuzzy situation inference	LWC + fuzzy inference system	Fine-grained representation	Computational overhead
	Community representation	CAP method, combining SVD, clustering and tensors analysis methods	Accurate analysis	Storage overhead

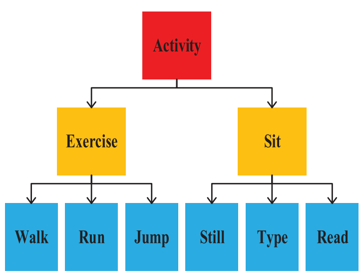

Fig. 10. The illustration of hierarchical structure in Lasagna.

The *separated methods* focus on the processing on each piece of data. The machine learning method is commonly used here to train a large amount of data and classify them one by one. As different type of data cannot be trained in the same model, they should be partitioned into different classifier first. To analyze place-centric contexts of collected data, Chon *et al.* [179] propose four sensor-data classifiers: Optical Character Recognition (OCR), object recognition, speech recognition, and sound classification. Depending on the type of uploaded data, the algorithm allocates the data into the corresponding classifier and gets results on different contexts. For example, if a picture of an attraction is coming, this picture will be input to object recognition model to find how many spots are in this picture so that the correlated location can be recognized.

The machine learning method can be further used to understand human moving patterns by distinguishing different behaviors of them, like walking or smoking. To train such a physical model, pre-knowledge or labeled data are needed. However, some comprehensive exercises are difficult to label, like smoking while walking. Even there are solutions to label those exercises, the fining problem that whether to label them roughly or specifically is still a barrier. Lasagna [183] designs a universal representation for all activities at multiple resolutions without prior knowledge. By applying unsupervised learning, the labeled problem can be avoided and a hierarchical structure to represent activities can deal with fining concern. The illustration of this hierarchical structure is shown in Figure 10, where the finer representations on exercise and sit are given by the next layer.

Nevertheless, it is difficult to process every piece of message when the scale of data becomes extremely large. It derives the *unified methods*, which fuse data and process on them together. Graph theory is widely used here, where collected data can be represented by nodes on a map and the relationship of data can be denoted by the links between nodes. Some graphical methods like mapping coverage, clustering, and Singular Value Decomposition (SVD) can be used to analyze the data.

Zhang *et al.* [175] point out the importance of comprehensively analyzing multimodal data from different sources: avoiding bias. "Bias" means partially seek the patterns only from one side. For example, if researchers want to analyze the transportation patterns of citizens, the leverage of the data from bus station can only reveal the bus pattern, where the subway and car patterns are ignored. Dealing with this matter, the paper gives a unified process on the whole transportation data together with cellphone data which can also reveal the tendency of mobility. By labeling the location of the collected data as a node into the map, the authors find out that cellphone data are highly distributed on larger coverage than transit data, which is more regular and limited. So this paper selects $G^c + G^{\overline{c}}$ (cellphone data and non-cellphone data predicted by historical data) to denote the whole mobility picture and requires this adding result must cover the transit graph.

Other than this mapping coverage method, the clustering method is also considered [180], [182]. They apply the generic toolkit for mobile data mining namely Open Mobile Miner (OMM) [184] where the Light Weight Cluster (LWC) algorithm uses data cluster technique to match high-speed data streams. The paper [182] also applies the OMM on data mining while the data includes both collected sensing data and social media data to specifically support real-time queries pertaining to locations of interest. Moreover, Zaslavsky *et al.* [180] further define a fuzzy situation inference in mining. These fuzzy situations often occur on the transition of two situations, like a situation between hot and cold, which are called "a little cold" or "quite hot". This definition is indispensable to accurately represent data, especially when monitoring the health status of the human body.

Furthermore, the Community Activity Prediction (CAP) method [181] combines SVD, clustering and tensors analysis methods together. Firstly, the authors merge the collected data into an individual-community map where the row represents the community and the column represents the individual. The node (*i, j*) denotes the relationship between individual and community. If individual *i* belongs to community *j*, then the value of this node is 1, otherwise, the value is 0. Secondly, apply SVD to this map and select the last two dimensions of the left singular vectors to project. Thirdly, make the clustering process to limit the type of community. Finally, use a three-fold tensor <*Time*, *Community*, *Activity*> to indicate the collected data, and make Tucker decomposition to get three projected matrices, which are useful for representative approximate tensors. By doing all of these, the final representation results can be acquired.

VII. Applications

As the last step in data-oriented MCS projects, the collected data from the crowd and the processing on them give solutions to various applications for constructing a smart city. The MCS-applied smart city tries to better support and simplify our daily life [185], generally speaking, from three main aspects: environmental, infrastructural, and social sensing [2]. According to the specific application scenarios, in this section, we deeply divide these three aspects into four types: (i) *indoor localization*, where techniques are specially designed for the indoor environment; (ii) *urban sensing*, which involves public infrastructural phenomena like traffic issues and urban planning; (iii) *environment monitoring*, which breaks the limits on monitoring severe environments, (iv) *social management*, which concerns more about human and their relationships. Related researches are introduced in each subsection. To better illustrate their relationships, the taxonomy of this section is drawn in Figure 11. Combining with the above-mentioned sections, we will analyze the related MCS techniques applied in each application (*i.e.*, incentive, safety and privacy preserving, resource optimization, and data analysis methods) and summarize its adoption ratio in Table VI.

A. Indoor Management

GPS technique is widely used in locating or mapping. But it is inefficient in the indoor scenario, where too many wall barriers and architecture shielding weakening its functions. To deal with this problem, researchers tend to apply MCS systems to build the indoor floor plans.

At first, Radio Frequency (RF) fingerprint based on WiFi or cellular signal is proposed [186], [187]. But it needs some prior knowledge and user-specific information including users' initial locations, stride-lengths and their device placements (*e.g.*, in/out the pocket), which makes this difficult to be applied into practice. So Rai *et al.* [188] propose Zee, a ZEro-Effort MCS system, which only needs an indoor map accompanied with WiFi and inertial sensor measurements to infer location over time. There are two key technologies innovated in Zee: Placement Independent Motion Estimator (PIME) and Augmented Particle Filter (APF). PIME uses mobile sensors (*i.e.*, accelerometer, compass, and gyroscope) to estimate the motion of participators. And the APF combines this estimated motion result with the floor map as inputs to track the location of participators. Zee also periodically scans for beacons from proximate WiFi Access Points (AP) and records the Received Signal Strength Index (RSSI) labeled by its timestamp, to further increase the location accuracy on the floor. On the one hand, this application does not consider the incentive or privacy preserving strategy in real-life MCS appliance, leading to a relatively low adoption ratio. On the other hand, the dynamic location estimation can be achieved without any prior knowledge, but it relies too much on the WiFi network which means interference in a certain range. So a single network data collection is not useful.

Instead, a hybrid indoor mobile phone localization mechanism is needed. Pazl [189] aims to combine PDR with WiFi fingerprinting to get accurate locations. Pedestrian Dead Reckoning (PDR) is a local estimation algorithm utilizing the sensors to measure numbers of steps, stride-lengths, and directions of pedestrians to infer locations and traces. WiFi fingerprinting is to make an offline location database with AP information and give an online comparison to locate an object. Although the PDR has the error accumulation problem and WiFi fingerprint has less efficiency on places where AP is not covered, the combination of them can make up for their own shortcomings. As declared in this paper, the WiFi fingerprint can correct the error of PDR and increase localization accuracy. However, a continuous localization application needs to monitor the transition from outdoors to indoors for participators to start sensing. Pazl fails to provide this related solution, which indicates low real-life adoption ratio.

However, the indoor map needed in Zee is sometimes not available. Then Gao *et al.* propose Jigsaw [190]. It uses pictures taken by participators to extract positions, sizes, and orientations of objects on each floor. Combined with user mobility traces, the Maximum Likelihood Estimation (MLE) algorithm [190] can obtain the spatial relationship between

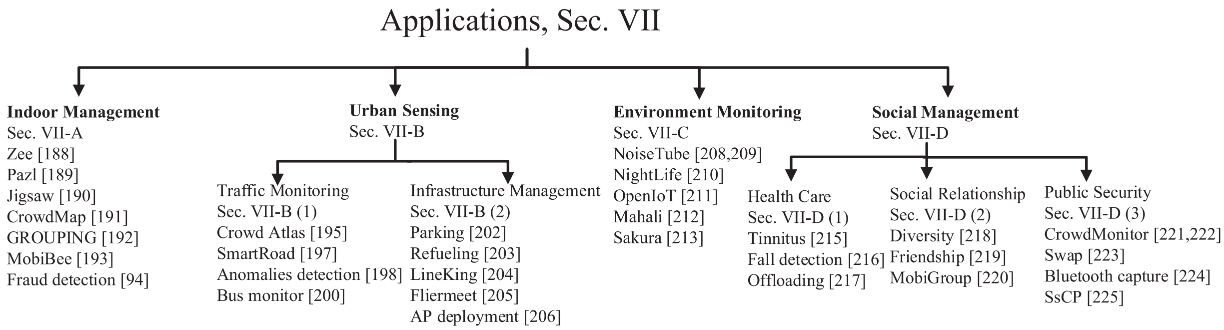

Fig. 11. A taxonomy of applications.

adjacent landmarks for drawing the whole floor plan. By applying computer vision and MCS techniques, Jigsaw can stably acquire landmarks with rich information and low overhead. Jigsaw defaults that the incentive is supported by the service providers. And the privacy preserving in collected photos is to blur the contained customers' faces. So the real-life adoption ratio is in middle level as these two factors are considered in implementation.

Nevertheless, only a series of photos is not enough for some restricted areas where people cannot arrive. CrowdMap [191] uses mobile phone video instead of pictures to get the sequential relationship from the consecutive frames of videos. This system is divided into three stages: data collection including spatial information, videos, and inertial data collection; indoor path modeling including trajectory aggregation and layout reconstruction; and final floor plan reconstruction. Similar to Jigsaw, they assume that participators actively get involved in sensing tasks, so several incentive mechanisms will be further developed before deploying in reality. But this paper stops at the research stage, which leads to a low real-life adoption ratio.

There are also some people proposing geomagnetic fingerprints rather than WiFi fingerprints [192], which is more stable, more lightweight, totally independent to any wireless infrastructure. This proposed system is called GROPING, which comprehensively integrates three implemented functions: mapping, localization, and navigation. This system is composed of the client side and the server side. In the client side, each client provides collected data, as well as visualizing the constructed map, the current (estimated) location, and the navigation routes. And the server side is built on cloud, which consists of floor map building, location estimation, and real-time navigation. The revised Monte Carlo Localization (MCL) algorithm is applied to solve the mapping problem. The incentive mechanisms applied range from monetary to valuable services. And they recruit extra participators from AMT, resulting in high adoption ratio.

Both WiFi or geomagnetic fingerprints mentioned above belong to location-dependent fingerprints. MobiBee [193] provides a mobile participatory game to collect these fingerprints. While scanning the QR codes posted on walls or pillars, the corresponding WiFi RSS and magnetic field strength will be uploaded. But in real world, to quickly get the reward from requestors, some malicious participators will design some tricks on these QR tags. Three attacks on tags used in MCS project should be detected: tag forgery, misplacement, and removal. Accordingly, Xu *et al.* [94] designs a fraud detection mechanism. For the first attack, a truth discovery algorithm is proposed to detect falsified data. And for the latter two attacks, visiting patterns of participators can be utilized here. To better support an accurate localization application in the future, these detections are necessary to be applied as a protection mechanism, leading to a high adoption ratio.

B. Urban Sensing

The data sensed from urban areas can reversely serve for urban management, especially when a large amount of data can reveal the tendency of urban development. On the one hand, the analysis of these sensed data can contribute to traffic or transportation management [194]. On the other hand, they are also beneficial for infrastructure management, including urban planning, local business management, and district management. Typical applications are discussed in this subsection following these two aspects.

1) Traffic Monitoring: To generally monitor the traffic situation for all kinds of transportation, the GPS traces are always chosen to be sensed. CrowdAtlas [195] uses an automated road inference algorithm with GPS probes to rectify existing maps. The inference algorithm includes Hidden Markov Model (HMM) map matching algorithm [122] to find the difference between the uploaded traces and the existing routes on maps, and a clustering-based inferencing algorithm called polygonal principal curve (PPC) algorithm [196] to update these routes. This system can automatically solve the drawbacks of existing digital road maps and update them in real time. The participators in CrowdAtlas can choose to send only unmatched segments to the server, reducing privacy exposure. This system is attracted to users without Internet connections, and users who want to customize their own travelling maps, so the adoption ratio is high.

Another GPS-based application is traffic regulator identification. To find out all of the traffic regulators, such as traffic lights and stop signs, the field survey on entire road can provide detailed and accurate results, but the cost on manpower or gasoline consumption cannot be neglected. In SmartRoad [197], researchers tend to use statistical classification techniques to process closely related GPS trace data including trajectories and time and extract five features: final

TABLE VI
SUMMARY OF APPLICATIONS (INC.=INCENTIVE MECHANISMS, S&P=SECURITY PROTECTION AND PRIVACY PRESERVING MECHANISMS, AR=ADOPTION RATIO)

Type	Name	Inc.	S&P	Resource Optimal	Data Analysis	AR
Indoor Management	Zee	-	-	-	PIME+APF	L
	Pazl	-	-	-	PDR	L
	Jigsaw	Provided by services	Face blurring	-	MLE+Computer vision	M
	CrowdMap	-	-	-	Trajectory Aggregation + floor plan reconstruction	L
	GROUPING	Monetary + valuable services	-	-	MCL	H
	MobiBee	Monetary + entertainment + competition	-	-	QR code validation + fraud detection	M
	Fraud detection	Monetary + entertainment + competition	-	-	Ground truth comparison + truth discovery	H
Urban Sensing	Crowd Atlas	Unmatched segments option	-	-	Clustering-based PPC + HMM map matching	H
	SmartRoad	Green routing	-	Information aggregation + feature selection	Statistical classification	H
	Anomalies detection	-	-	-	Map matching	L
	Bus monitor	-	-	Cellular signatures + audio signals instead of GPS	MLE+clustering	L
	Parking	-	Anonymous	-	Trivial analysis	M
	Refueling	-	-	-	Context-aware collaborative filtering+queue system	H
	LineKing	Advertising	-	-	MAP scan	M
	Fliermeet	Monentary	-	-	Clustering	H
	AP deployment	Better connection + cheaper service	-	-	Map matching	M
Environment Monitoring	NoiseTube	-	Cryptographic	-	Map presentation	H
	NightLife	Monetary	Local process	Transmit when connecting WiFi	Automatic ambiance +feature extraction	H
	OpenIoT	-	-	-	Data filtering	L
	Mahali	-	-	-	Tomographic analysis	L
	Sakura	-	-	k-stage sensing	Histogram-based color analysis +region-based fractal dimension analysis	L
Social Management	Tinnitus	-	-	-	Social combined analysis	L
	Fall detection	-	-	-	Activity recognition	H
	Wearable devices	-	-	-	computing and communication offloading model	H
	Diversity	-	-	-	Correlation analysis	M
	Friendship	-	Anonymous	-	Location trail prediction	H
	MobiGroup	-	Length limitation +local extraction	-	heuristic rule-based strategy +context-based group computing +generic group activity model +cross-community mechanism	H
	CrowdMonitor	-	Anonymous	-	Clustering	M
	Swap	-	Without camera	Touch screen swap direction transmission	Clustering	H
	Crowd density	-	-	-	Collaborative estimation	H
	SsCP	-	-	-	MoST	M

stop duration, minimum crossing speed, number of decelerations, number of stops and distance from intersections. For example, the stop durations between red light and stop signs are different and have obvious patterns to be distinguished. Combining the information aggregation and feature selection schemes, they decrease the classification scenarios and reduce transmission consumption. Moreover, the SmartRoad is piggybacked on a navigation app, which requires zero user effort for the setup and running. The incentive provided to participators is the free use of system services, like green routing. So the adoption ratio can be high as they have claimed in their paper.

After finding out these regulators, some traffic anomalies can also be detected based on human mobilities [198]. Instead of analyzing traffic volume and velocity on roads, the differences between the real-time routing behavior and typical patterns can suggest the happening of some traffic anomalies, like accidents, traffic control, protests, sports events, and disasters etc. As the routing behaviors are denoted by the sub-graph, the anomalies recognition can be processed by the map matching method. In this paper, the social media data on the related spot are also added to better explain the semantic meaning of the events, which can increase the reliability of the final result. But the incentive, security and privacy, and resource optimization are all ignored in practical design, the adoption ratio of this system is low.

However, for some GPS-less applications, researchers may make use of special transportation, for example, bus. As the traffic map of bus is fixed, if their travel time and average speed are changing, the situation of traffic can be easily detected. And only cellular signatures and audio signals recording accelerated information are needed for this calculation [199], which greatly save energy consumption caused by GPS. To estimate the bus arrival time, both drivers and passengers can report their nearby cell tower IDs to analyze their routes on real-time and give estimated arrival time feedbacks [200]. The MLE algorithm is applied to do road matching and the bus stop is detected by the clustering algorithm. Considering there is no practical incentive mechanism or even company supported, the access to public transportation is difficult to get. So the adoption ratio of this system is regarded at a low level.

2) Infrastructure Management: With the proliferation of private cars, the needs for parking spaces and gas stations are constantly increasing. Even when you want to get off the car and go to the nearby coffee shop, there is still another long line waiting for you. To get rid of this frustration, the crowd data collected in urban areas can be used for infrastructure management. Accordingly, we will introduce several typical applications related to infrastructure management, like parking and refueling management, coffee shop waiting time management, advertisement posting, and WiFi AP deployment.

In our daily life, people often fail to know where the best parking locations are and whether a parking place will be available when they arrive. Some researches tend to apply MCS methods into parking management, and try to make the method agility, large-scale and low-cost [201]. Coric and Gruteser [202] use vehicles' pre-installed parking sensors and GPS locations to classify on-street areas into legal or illegal parking spaces, then map them. It concerns the street parking areas rather than professional parking lots which are strictly controlled by certain corporations. Luckily, MCS methods can provide more measurements for an accurate construction. Lots of parking sensors can detect the distance between the target objects on the parking spaces, which used to distinguish the cars on the parking spaces with other objects. Then the trivial analysis on this uploaded big data finally gives feedback on real-time parking information. As the GPS trace data will reveal real locations, the anonymous privacy preserving method is applied. But there is no incentive mechanism design which makes the adoption ratio only at a medium level.

To reduce the waiting time in refueling stations, the status of nearby gas station should be notified to drives. However, as the energy use data on the gas station is difficult to acquire and estimate, few researches have done on this refueling problem. But MCS methods make it possible. The GPS trajectories of cars can detect gas station visiting rates and measure the time spent on each station, through which the overall demand can be estimated. If the fuel events data is sparse, Zhang *et al.* [203] apply the context-aware collaborative filtering approach and queue system to calculate the final rate. Depending on this, the spatial-temporal fueling behavior can be analyzed properly. According to their real-life experiments, they evaluate on the road network of Beijing, China, which contains 106,579 road nodes, 141,380 road segments, and 369,668 POIs. Together with 30,000 taxicabs targets, we consider its adoption ratio as high level.

Specifically, the LineKing [204] can also address the wait-time detection in coffee shop scenario. The calculation of wait time starts/ends from the proximity for users to enter/exit the shop, which is periodically scanned by the Wireless Access Points (WAP) around the users. The recruitment of participators is only advertising on social media without monetary incentives, so the adoption ratio can only be middle level.

Traditional methods for advertisement posting on campus is via flyers pasted on bulletin boards. But this way has limited spatial-temporal coverage, lacking order and low search speed problems. It is intuitive if we take these flyer form the physical space to the cyber space, for collectively control and broadcast. Fliermeet [205] applies MCS to collect flyers in different times and places, then cluster duplicated fliers to choose the best picture as a representation, which will be reposted to the Internet and broadcast all over the world. Considering that this application can also be extended to any flyers in urban areas, its adoption ratio is high.

Another concerning scenario is WiFi AP deployment management. Suitable distribution and configuration of open APs at a city level are needed to build a better urban communication. Farshad *et al.* [206] recruit participators with mobile phones traveling on public buses and collect WiFi interface data. A cloud-based WiFi spectrum management service is proposed for better interference management in urban area. As the GPS data is also required in data collection, the lack of privacy-preserving mechanism leads to a middle adaption ratio of this work.

C. Environmental Monitoring

The recent discussion [207] has investigated some big data derived applications in terms of environmental issues. It points out the lack of mature solutions on the sensing cost, durability, scalability, and interoperability of specially designed WSNs. Although these WSNs have dense coverage on urban areas in the daylight, some unattended areas like the air quality or special scenery in suburbs, the noise at night, even the space weather at the earth level are lacking in monitoring. To solve these special cases, some researches [208]–[213] prove the flexibility, scalability, and cost-efficiency of MCS techniques.

For noise monitoring, NoiseTube [208], [209] is widely used in MCS projects. It measures noise, localizes it, tags it, and shows it on Google Maps in real time. Accompanied by the provided questionnaires, the feedbacks on users about noise can be added to get more information, like the ranking questions can reveal the prioritization for the information shown on the noise map. Towards large-scale adoption, the privacy-preserving mechanism [214] leverages cryptographic techniques and distributed computations in the cloud.

Further developed, researchers tend to survey on nightlife patterns in urban areas. It can not only monitor sound and light pollution at night, but also reveal economic opportunities and potential safety risks. Santani *et al.* [210] consider both the self-reported data including images or videos labeled manually and the data scratched from social media. Extracted from all of these data, the places they usually hang out, the corresponding social context and night life activities can be extracted by the use of automatic ambiance features. Through the practical report, the authors decide 100 CHF monetary reward for each participator, and the uploaded data can only be shared within the research group to protect data privacy and safety. To optimize battery life, data transmission from phones to the back-end server will only be performed when connecting WiFi. As they are supported by the ethical review board of Vaud and Zurich cantons in Switzerland, the adoption ratio can be guaranteed.

For air quality monitoring, the mobile phones cannot do this work independently as no equipped sensors have such a function. So an auxiliary equipment is designed [211], which can upload data firstly to a mobile phone and get orders from it through Bluetooth communication. Then an OpenIoT crowdsensing platform can collect data from mobile phones and filter them to satisfy the suitable coverage rate. Because of this auxiliary device overhead, the adoption ratio of this method is low.

And for a high-level space weather monitoring, Pankratius *et al.* [212] propose the Mahali Project. They aimed to analyze the electron density variations in the ionosphere to get conditions of space environment. The tomographic analysis based on the collection of GPS data, which can acquire the total integrated plasma density between a ground receiver and a transiting satellite. Compared with the sparse coverage of weather monitors, this GPS acquisition can be achieved by individual mobile phones, utilizing the "division" idea in crowdsensing method. As the accuracy based on this method is not high enough, we consider the adoption ratio of this method is low.

If a special scenery in a city appeared (*e.g.*, a beautiful view or a geographical change), immediate detection can bring great economic benefits for the city manager. Morishita *et al.* [213] detect a beautiful "sakura" scenery through video sensing by cars. This paper proposes a *k*-stage sensing strategy to dynamically shorten the sensing intervals of these cars: once the flowering cherries are detected, the nearby passing cars have to sense at shorter distance intervals and narrow step by step until the *k* requirement is reaching. After collecting sensing pictures, the authors use histogram-based color analysis and region-based fractal dimension analysis for accurately detecting flowering cherries and their degrees. Nevertheless, as the face information containing in pictures are not protected, the adoption ratio of this method is low.

D. Social Management

As mobile devices have become an indispensable part of our life, they can easily record the social behaviors of human beings, including both motion and interactive status. As one of the big portion in these records, enormous data from social media and social networks can reveal individual or crowd characters. If the target is on single person, her social data can reveal her daily habits, preferences, or health conditions, etc. While if the research targets a group of people, the social relationships even security issues can be revealed or predicted. Correspondingly in this subsection, we category the social management into three aspects: health care, social relationship, and security monitoring, for the benefits of our daily life.

1) Health Care: As we all know, the treatments for different patients have individual variability, especially for some epidemic diseases. The big data is declared to give great help on considering correct treatments depending on the analysis of plenty of historical cases [215]. Although traditional clinical trials can also find out the corresponding treatment, they consume too much time and heavy efforts. So the collection of information from numerous patients is the new solution on health care. Taking tinnitus as an example, questionnaires are provided to analyze the symptoms of patients and collected the related environmental sound level to prove their thoughts [215]. By doing this in a large range, these big data can lead to more accurate treatments to patients. In the health care domain, the privacy of patients is an important research aspect, but this paper fails to give preserving strategy which makes a low adoption ratio.

In addition to the professional clinical data gathered from patients, health monitoring can be achieved by the commonly-used wearable devices for large-scale ordinary users. These devices like smart watches, glasses, rings, gloves, and helmets are popular in our daily life. They are incorporated in clothing or worn on the body to track or monitor the physiological characters and make records or analysis. He *et al.* [216] develop a fall detection system based on this kind of data collected from a waist-mounted smartphone. Correspondingly, an automatic Multimedia Messaging Service (MMS) will be sent as an emergency treatment backup for pre-selected people. Time, GPS coordination, and fall location are packed in MMS. As the smartphone is readily available to most people, this application is highly adopted.

Although these collected data can be of great help to monitor our health status, the storage, computing and communication limitations are still challenging for its appliance. Ragona *et al.* [217] introduce a computing and communication offloading model to make a tradeoff between energy and execution delay. The energy savings brought by this model makes its adoption ratio to a high level.

2) Social Relationship: Compared with traditional questionnaires and auxiliary equipment carried ways, MCS applied in social relationship analysis is more accurate, objective and humanized.

The interconnected nature of people and places is shown by MCS data collection. The social diversity of urban locations can be captured through social networks. It also provides opportunities to dig out the mobility patterns of participators [218]. The social diversity here indicates their social role (whether it attracts diverse individuals or regulars). Correlating with the wellbeing indicators of neighbourhoods, the relationship between the prosperity of people and places can be analyzed, providing the suggestions on future urban polity or socially-aware applications.

As one of the social relationship, friendship can be predicted by similar moving patterns between two users. Location trails dug out from a location sharing social network can be used to indicate friendship [219]. Assisted by Facebook, privacy preferences can be guaranteed, which makes a high adoption ratio in real life.

This relationship between people and place can also be beneficial for travel planning. Guo *et al.* [220] propose a three-step planning:

- Activity preparation: For public activities, a heuristic rule-based strategy is designed to initiate potential attendees according to the group preferences. And for private activities, a context-based group computing is applied for group recommendation;
- Group activity mining: A generic group activity model is trained to classify the characters and stages of group activities;
- Activity suggestion: For both online and offline communities, a cross-community mechanism is proposed to extract the interaction features among them and decide the final suggestion list.

For privacy preserving, the MobiGroup architecture in this paper limits the length of the uploaded audio clips and the feature extraction is processed locally. As the target of this paper is a wide range of groups, the adoption ratio is high.

3) Public Security: The monitor on public security includes a lot of aspects: emerging crimes, traffic accidents, emergencies, and hazards. The related departments or sparsely covered monitors ignore a lot of security problems in inaccessible places. Typically, the related information for public-safety and volunteer-initiated activities implied in social media data can indicate security issues [221], [222]. Combined with these online witnesses into traditional monitoring can decrease the overall uncertainty and take protective measures on real-time. To protect the privacy contained in sensing data, the anonymization is considered and the adoption ratio is at a medium level.

For sorely make use of these eye witnesses, Ouyang *et al.* [223] design an event localization application on smartphones. If the citizens have seen something, they just need a swipe from their locations towards the happening places. It seems like the steps are so easy to perform, but the reported data can have significant meanings. Clustering the big reported data, they can estimate the exact place the events happened. Without the utilization of cameras, the authors achieve a low-cost and near-continuous monitoring of outdoor events, leading to a high adoption ratio.

Considering that the big event held in a stadium can easily suffer from potential safety risks caused by the crowd congestion. Taking crowd stadium as an example, Weppner and Lukowicz [224] fuse different Bluetooth-captured crowd data and finds six relative features to collaboratively make estimation: the average sum of distinct devices discovered by all sensors in scan window, Bluetooth link structure, crowd movement, team-wise diversity, semi-continuous unique devices, and average duration. By analyzing all of these six features, the whole populated picture can be drawn for this big stadium. As they evaluate the method on a big data set during 3 days at the European soccer championship public viewing event in Kaiserslautern containing thousands of visitors, the adoption ratio is high. To further help event organizers manage the venue with minimized risk, the Smart Stadium Crowd Planner (SsCP) [225] can detect activity, position and entrance/exit of spectators to/from a stadium fence. It designs a Mobile Sensing Technology (MoST) for activity detection, geolocalization, and geofencing (*i.e.*, the capability of detecting user proximity to a geographical location). As the SsCP prototype is widely available for the Android system and supporting 98% off-the-shelf devices. The adoption ratio of this method can reach to the middle level.

VIII. Testbeds and Simulators

For the convenience of researching or applications, MCS systems should be implemented on a comprehensive architecture across mobile devices and cloud computing platforms with extensive supports to application developers and end users. It will integrate the service-oriented architecture with all necessary functions discussed in this survey paper: incentive mechanism design, security protection and privacy preserving, participant recruitment, task assignment, data collection and processing. On the one hand, this architecture should dynamically adjust the environment parameters depending on situation changes to ensure enough sensing coverage and quality. On the other hand, it should provide public interfaces to upper applications which allow researchers to customize their own sensing tasks. Even a visualized and generalized platform is designed to relieve the heavy pressure on freshmen or human-subject researchers. There are three kinds of architectures for implementation: testbeds, simulators, and commercial service platforms. As a testbed, it requires real-life collected inputs and its outputs are working for practical application. While for a simulator, simulated inputs are used to evaluate large scale systems that cannot be realized in a testbed or prototype settings. And a commercial service platform is publicly provided for MCS projects by an organization with its generality, profitability, and reliability. In this section, we will present several typical testbeds, simulators and commercial service platforms for the need of MCS researchers. They are all summarized in Table VII, which compared by their present usage status together with the available download links, supported functions and advantages. As we introduce these related

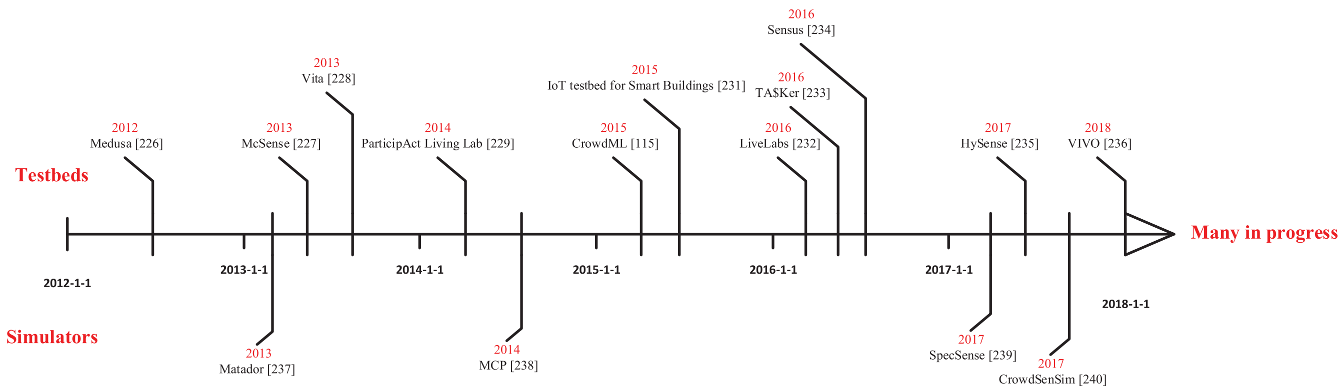

Fig. 12. A taxonomy of Testbeds and Simulators.

work chronologically, the development timeline is shown in Figure 12.

A. Testbeds

The earlier accepted testbed is Medusa [226]. It finds out the requirements for crowdsensing and builds a whole processing system from incentive design to automatedly data collection. The recruitment of real workers is coming from Amazon Mechanical Turk[1] for further realize other function parts. For simplifying the task assignment process, it designs its own high-level programming language, called MedScript. After that, testbeds proposed in the following years realize different functions and have unique characters.

In 2013, McSense [227] and Vita [228] are introduced. For McSense [227], the functions of it are more comprehensive. Instead of the basic steps described in Medusa, it can also evaluate participators' performance including the achievement ratio and completion time to give a suggestion on the next task assignment step. So its assignment policy is depending on the default parameter values or the previous results of every sensing cycle. The implemented McSense mobile app which has a simple graphical user interface (GUI) interaction is installed by 44 participators who often visit the New Jersey Institute of Technology (NJIT) campus in Newark and decided to participate as potential workers. Additionally, the customized platform of Vita enables intelligent deployments of tasks between humans in the physical world, and dynamic collaborations of services between mobile devices and cloud computing platform during run-time of mobile devices with service failure handling support. the outstanding point of Vita is its developed ability where the APIs of this testbed are freely provided to application developers and the third party service providers to design their own crowdsensing environment [228].

All of these presented testbeds are two-side architectures: the client side and the platform side. Here, the client side is used to perform task receiving, data collection and uploading tasks, while the platform side decides participant recruitment, task assignment and collect sensing data. In 2014, a new architecture named ParticipAct Living Lab is proposed [229]. It started within the city of Bologna and connected to the students of the University of Bologna. As an on-going testbed, it realizes three goals:

- It designs and tests for the genetic MCS system which realizes the whole processing functions inside.
- It evaluates the machine-learning methods which are useful for MCS systems on a large scale.
- It assesses the challenges to manage human resources in MCS systems, and designs a less burden platform which moves the recruitment and task monitoring process to a third party, called MCS manager.

This testbed also gets extension on Cambridge, and MIT university [230].

As the proliferation of big data mining, machine learning methods have acquired wide attention. So in 2015, a machine learning based testbed has come out. This CrowdML architecture [115] integrates sensing, learning and privacy preserving together, where the optimal parameters are automatically calculated by the collected data using the Stochastic Gradient Descent (SGD) function rather than customizing them manually. Also, a user-centric design style has been used to build an Internet of Things (IoT) testbed for smart buildings this year [231]. It equips a lot of incentive mechanisms to guarantee participators profits. The users can declare their preferences on different situations, actively receive acceptable incentive offers and execute sensing tasks.

But as pointed out before, GUI interfaces are beneficial for fresh or social-study researchers. So, in 2016, LiveLabs [232], TA$Ker [233] and Sensurs [234] all have their well-designed GUI interactive interfaces. While the LiveLabs can give a detailed and visualized experiment feedback, Sensus designs a GUI-based sensing plans instead of the programming process. They both are easy for new researchers to perfect their crowdsensing experiments. Especially, Sensus is a cross-platform, general-purpose design for MCS-based human-subject researches. A controversial design in this testbed is in worker recruitment. Users in Sensus typically post advertisements in conspicuous places offline (e.g., outside of clinics or schools) and online (e.g., Web forums). Participants who are interested in the topic will connect with the researchers via mail, telephone, or e-mail. Although it can select the high quality of participators from the face-to-face meeting between volunteers and requestors, it is time-consuming for a large

[1] https://www.mturk.com/

requirement on participant population. For another testbed, the task bundling, differential pricing and cheating analytic of TA$Ker can be more helpful in practical applications which is designed, developed and experimented with a real-world mobile crowd-tasking platform on Singapore Management University (SMU) campus.

In 2017, more specific problems are solved in new testbeds. These new testbeds are built based on previous ones so the basic functions and visualized characteristic are inherited. To deal with the imbalance distribution of sensing data, a hybrid testbed called HySense is proposed [235]. It not only integrates the mobile sensing data with static sensing data to solve the data-sparse problem, but also migrate the redundant users from densely populated places to sparse ones.

In 2018, enhanced testbeds are invented. Take VIVO [236] as an example, it is an enrolled crowd-sensing model, which allows the simultaneous deployment of multiple experiments. The privacy and data security are preserved before the data leaving the devices. The collected data can be then processed either offline after gathering together, or online in real-time. This testbed is implemented by distributing an experiment to 40 volunteers scattered over the whole Switzerland.

B. Simulators

The lacking in large-scale real-life sensing data sometimes limits the construction of testbeds. Simulators are derived to prove the feasibility or efficiency of designed MCS frameworks by taking the input of simulated data.

Matador [237] proposed in the year of 2013 designs for the task assignment and data collection steps. Its assignment language is XML with the construction of "Context + Action". And it presents relevant tasks to users and commits to preserve the battery life of mobile phones. This energy-efficient context sampling algorithm is validated with a small scale field study where only one user carries the mobile phone and drives on the road, showcasing the potential of the proposed solution.

Hu *et al.* [238] claim that the consideration on context can bring a better understanding of users' situations, leading to a better allocation and energy-efficient execution on sensing tasks. So they propose a Mobile Context-aware Platform (MCP) depending on Vita [228], which can provide environment and services for users' participation and suitable task assignments depending on specific contexts. Low computational and communication overhead makes it efficient as a simulator.

In SpecSense [239], considering the cost on RF sensors which are used to monitor spectrum occupation is high, this system gives a sensor selection algorithm to choose few suitable sensors to solve the overhead problem with the limit of sensing coverage. Although the incentive mechanisms are used, inadequate sensing opportunities sometimes come out, especially for some sparsely populated areas. Furthermore, the CrowdSenSim [240] is specially designed for realistic urban environment simulations. It concludes the participant selection, data collection and process. Finally, it can give the visual result on the real urban mapping, which will be a great help for urban planners and decision makers.

C. Commercial Service Platforms

Except for these testbeds and simulators in the research area, there are also some available platforms provided by commercial organizers. We also summarize them in Table VII. The typical examples are Gigwalk and Streetspotr.

Gigwalk [241] was founded in 2010 with the goal of reinventing work in a mobile world. This Gigwalk infrastructure supports the MCS task publish, monetary reward payment, real-time data collection from requestor side; and task acceptance and execution, sensing data upload, privacy preserving and reward earning from participator side. It is founded by several investors including Nokia Growth Partners (NGP), August Capital, Harrison Metal etc. As it has solid investors, the service it provides is more trustable and easy to be implemented.

Streetspotr [242] has provided services in qualifying and engaging MCS for over 5 years. It basically provides a platform for task publish, execution, incentive and security protection and privacy preserving. The spotlight is its best-in-class data analytics dashboard. They create and continuously improve this analytics dashboard which illustrates all main insights from the data collected through the crowd in real-time. After applying their immense filter capabilities (postcodes, Nielsen areas, retail channels, store sizes, distribution center), requestors can easily screen the individual KPIs at a glance, where only simple downloading can realize plenty of formats they need.

IX. Conclusion, Lessons Learnt, and Future Research Directions

A. Conclusion

Our survey filters out the related work in the five aspects of crowdsensing. We follow the processing framework of MCS tasks to introduce these five aspects, including incentive mechanism, security protection and privacy preserving, and resource optimization strategies in MCS data collection; MCS data analysis; and MCS applications. Then we show the available testbeds and simulators for MCS tasks. After all relevant papers are introduced, the taxonomies and the comparison tables attached on each section can help readers understand them from a global view.

B. Lessons Learned

Although we discuss MCS by separated sections, they actually are combined as an entire entity. Only when all of these parts work together can realize their own values. In the era of data explosion and the growing of AI, MCS provides a systematical structure for data collection, analysis, to application. The MCS makes it possible for a large quantity of data applied to create a better life. Separately, some lessons learned from above-mentioned sections are as follows:

- As the involvers of MCS systems are most human-beings, their willingness should be aroused by incentives and their privacy and data security should be preserved. It implies that for any human participant experiment, human rights should be well protected with a specially designed mechanism.

TABLE VII
SUMMARY OF TESTBEDS (INC.: INCENTIVE MECHANISM; PRI.: SAFETY AND PRIVACY PRESERVING; PS: PARTICIPANT SELECTION; TA: TASK ASSIGNMENT; DC&DP: DATA COLLECTION AND PROCESSING)

	Name	Status	Link	Inc.	Pri.	PS	TA	DC &DP	Sportlights
Testbeds	Medusa	Using	https://code.google.com /archive/p/medusa-crowd -sensing/downloads	X	X			X	MedScript programming language for TA.
	McSense	Using	https://web.njit.edu/ mt57 /mcsense/	X	X	X	X	X	Random, attendance, recency policies for TA.
	Vita	Using	http://mobilesoa.appspot.com/				X	X	Low computation and communication overhead
	ParticipAct Living Lab	Using	http://http://participact.unibo.it/	X	X	X	X	X	Freely mix all kinds of passive and active sensing actions in one task.
	CrowdML	Using	https://github.com/ jihunhamm/Crowd-ML		X			X	Calculate optimal parameters from collected big data using SGD function.
	IoT testbed for Smart Buildings	Maintained	-	X			X	X	User-centric design combining a lot of incentive mechanisms.
	LiveLabs	Using	http://is.gd/livelabs		X		X	X	1.Well-established interfaces. 2.Detailed experiment feedback.
	TA$Ker	Maintained	-	X	X		X	X	The former three functions added on the LiveLabs testbed.
	Sensus	Using	https://play.google.com /store/apps/details? id=edu.virginia.sie .ptl.sensus			X	X	X	GUI-based sensing plans.
	HySense	Maintained	-	X	X	X	X	X	Migrate redundant users from densely populated areas to sparse ones.
	VIVO	Using	https://crowd.unige.ch /noiseMapSwiss		X			X	Simultaneously support multiple applications in one system.
Simulators	Matador	Maintained	-				X	X	1.Energy-efficient context sampling algorithms. 2.XML task assignment language.
	MCP	Maintained	-			X	X	X	Consider the context of sensing tasks based on Vita.
	SpecSense	Maintained	-			X	X	X	RF sensor selection.
	CrowdSenSim	Using	https://crowdsensim. gforge.uni.lu/			X		X	Specially designed for urban environment simulation.
Commercial	Gigwalk	Using	http://www.gigwalk.com/	X	X	X	X	X	1. Comprehensive system. 2. Solid investors and large amount of users. 3. Trustable. 4. Easy to be implemented.
	Streetspotr	Using	https://streetspotr.com/	X	X	X	X	X	1. Expert data analytics. 2. Easy operation.

- To build a more "real" MCS system, that is, a high adoption ratio in our daily life, the computational capability, transmission resource, and budget limits cannot be ignored. The balance between low cost and high quality should be emphasized in strategy design.
- Dealing with large quantity data, some falsified, redundant, and missing data should be filtered or fixed first. Both the quantity and the quality of data should all be guaranteed for further application.

Besides the existing work introduced in this survey, there are still some challenges as well as future research directions discussed in the following subsection, which we hope to provide some possible selections for new or experienced researchers.

C. Future Research Directions

We suggest three future directions enhanced from MCS, targeting on data collection, analysis and application aspects respectively. The first direction is the customized design for data collection. As for safety and privacy protection, the "user-centric" concept can be applied to design more customize-available preserving systems. On the one hand, users can freely decide the running time of preserving systems and the preserving degrees. On the other hand, heterogeneous organizations have different levels of requirements. For example, the large corporations have higher requirements on reliable data protection mechanism than small corporations, as they have more clients' expectations. Actually, as another popular technology these days, blockchain [243] can also be incorporated into crowdsensing. As the basic structure of Bitcoin [244], blockchain have the mutual system for token money. If we record the contributions of different participator on blockchains, it can support a token money rewarding system for these users. These contributions should be recorded as an unalterable and irrevocable way, which is the superiority of blockchain.

The second direction is the light-weight MCS data analysis, like applying machine learning into micro-devices. In the Internet of Anything (IoA) [29], the light-weight devices are employed in MCS systems. However, they have a lot of operation limitations: low ROM and RAM storage, low CPU frequency, low network bandwidth support etc. The preprocessing algorithms applied on these devices (e.g., deep learning model) should be designed as light-weight. For instance, as declared in DeepMon [245], the convolutional layer consumes most processing time for image modeling. A cache-based background filter algorithm is applied to optimize the pre-trained model to a lighter level. But in design, the tradeoff between shrink program and corresponding accuracy loss should also be considered carefully. For another perspective, we recommend the combination of crowdsensing and edge computing. The edge computing will decompose the large processing service handled by the central platform to smaller and easily managed parts. These parts will be scattered to the edge node, which is called cloudlet. Different from the cloudlet in [155], what we mentioned here is more like a sub-platform. It is closer to the user terminal device, so it can speed up the data processing and transmission speed, reduce the delay and release the processing burden of main servers. User terminals, cloudlets and main servers can construct a heterogeneous system to process big data step by step, where suitable cloud resources are assigned according to their specific offload situations [246].

Moreover, mobile crowd sensing and computing (MCSC) can be also introduced to data analysis enhancement. This definition is proposed by Guo *et al.* [247], presenting the fusion of human and computer intelligence. The flexibility and scalability supported by human side can lead to extending coverage, while the computational ability provided by computers can solve the big data process limitations in human-side devices. We trust this win-win strategy can be the main scheme of MCS design in the future. Besides, the human-enabled mobile MEC mentioned in Section II-A can also be further researched for better resource provisioning and optimization.

The third direction is to extend MCS applications to IoA. According to the discussion in Section VII, most of the applications are benefit from data collected from mobile phone, including sensor data or social media data. As the prosperous of IoA, data from anything can be accumulated and make some change to every aspect in our daily life. The mobile phone plays an important role as an intermediate node to collect the data from crowd sensors, which is also implied as the life extension of these sensors [248]. Every social environment like offices, restaurants, hospitals, cinemas, and so on, can become the targets of MCS projects. For example, an intelligent coffee breaking plan can be made by MCS in offices. The coffee machine can record the timestamp of every new cup making, where the rush hour of coffee making can be calculated. Assisted by the workers' mobile phones, the waiting time to make a coffee and the location of the coffee room can be uploaded. Combining these two collected data, valuable suggestions can be given for the most efficient time to make a coffee, even for coffee machine maintenance.

MCS technology can help to execute heavy sensing tasks in the big data era. There are still a lot of blanks on this research field needed to be filled. As it can also be integrated with other techniques, we believe that the further dig out on the potential of MCS can make it as a very powerful sensing technology in the future.

REFERENCES

[1] *ICT Facts and Figures 2017*. Accessed: Oct. 2017. [Online]. Available: https://www.itu.int/en/ITU-D/Statistics/Pages/facts/default.aspx

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

[3] B. Guo *et al.*, "Mobile crowd sensing and computing: The review of an emerging human-powered sensing paradigm," *ACM Comput. Surveys*, vol. 48, no. 1, pp. 1–7, 2015.

[4] J. Wu, S. Guo, H. Huang, W. Liu, and Y. Xiang, "Information and communications technologies for sustainable development goals: State-of-the-art needs and perspectives," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 2389–2406, 3rd Quart., 2018.

[5] J. Wu, S. Guo, J. Li, and D. Zeng, "Big data meet green challenges: Greening big data," *IEEE Syst. J.*, vol. 10, no. 3, pp. 873–887, Sep. 2016.

[6] X. Zhang *et al.*, "Incentives for mobile crowd sensing: A survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 1, pp. 54–67, 1st Quart., 2016.

[7] L. G. Jaimes, I. J. Vergara-Laurens, and A. Raij, "A survey of incentive techniques for mobile crowd sensing," *IEEE Internet Things J.*, vol. 2, no. 5, pp. 370–380, Oct. 2015.

[8] L. Pournajaf, L. Xiong, D. A. Garcia-Ulloa, and V. Sunderam, "A survey on privacy in mobile crowd sensing task management," Dept. Math. Comput. Sci., Emory Univ., Atlanta, GA, USA, Rep. TR-2014-002, 2014.

[9] D. Christin, A. Reinhardt, S. S. Kanhere, and M. Hollick, "A survey on privacy in mobile participatory sensing applications," *J. Syst. Softw.*, vol. 84, no. 11, pp. 1928–1946, 2011.

[10] I. J. Vergara-Laurens, L. G. Jaimes, and M. A. Labrador, "Privacy-preserving mechanisms for crowdsensing: Survey and research challenges," *IEEE Internet Things J.*, vol. 4, no. 4, pp. 855–869, Aug. 2017.

[11] J. Liu, H. Shen, and X. Zhang, "A survey of mobile crowdsensing techniques: A critical component for the Internet of Things," in *Proc. IEEE ICCCN*, Waikoloa, HI, USA, 2016, pp. 1–6.

[12] F. Restuccia, N. Ghosh, S. Bhattacharjee, S. Das, and T. Melodia, "Quality of information in mobile crowdsensing: Survey and research challenges," *ACM Trans. Sensor Netw.*, vol. 13, no. 4, pp. 1–41, 2017.

[13] C. Alcaraz, P. Najera, J. Lopez, and R. Roman, "Wireless sensor networks and the Internet of Things: Do we need a complete integration?" in *Proc. 1st Int. Workshop Security Internet Things (SecIoT)*, vol. 139, no. 3, 2010, pp. 170–174.

[14] R. Atat *et al.*, "Big data meet cyber-physical systems: A panoramic survey," *IEEE Access*, vol. 6, pp. 73603–73636, 2018.

[15] J. G. P. Rodrigues, A. Aguiar, and J. Barros, "SenseMyCity: Crowdsourcing an urban sensor," *arXiv preprint arXiv:1412.2070*, pp. 1–10, 2014.

[16] A. Capponi *et al.*, "Assessing performance of Internet of Things-based mobile crowdsensing systems for sensing as a service applications in smart cities," in *Proc. IEEE CloudCom*, Hong Kong, 2017, pp. 456–459.

[17] V. Petrov *et al.*, "Vehicle-based relay assistance for opportunistic crowdsensing over narrowband IoT (NB-IoT)," *IEEE Internet Things J.*, vol. 5, no. 4, pp. 3710–3723, Oct. 2018.

[18] A. Antonić, M. Marjanović, K. Pripužić, and I. P. Žarko, "A mobile crowd sensing ecosystem enabled by CUPUS: Cloud-based publish/subscribe middleware for the Internet of Things," *Future Gener. Comput. Syst.*, vol. 56, pp. 607–622, Mar. 2016.

[19] J. Ren, Y. Zhang, K. Zhang, and X. Shen, "Exploiting mobile crowdsourcing for pervasive cloud services: Challenges and solutions," *IEEE Commun. Mag.*, vol. 53, no. 3, pp. 98–105, Mar. 2015.

[20] X. Sheng, J. Tang, X. Xiao, and G. Xue, "Sensing as a service: Challenges, solutions and future directions," *IEEE Sensors J.*, vol. 13, no. 10, pp. 3733–3741, Oct. 2015.

[21] R. D. Lauro, F. Lucarelli, and R. Montella, "SiaaS—Sensing instrument as a service using cloud computing to turn physical instrument into ubiquitous service," in *Proc. IEEE ISPA*, Leganes, Spain, 2012, pp. 861–862.

[22] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: Vision and challenges," *IEEE Internet Things J.*, vol. 3, no. 5, pp. 637–646, Oct. 2016.

[23] S. Wang *et al.*, "A survey on mobile edge networks: Convergence of computing, caching and communications," *IEEE Access*, vol. 5, pp. 6757–6779, 2017.

[24] P. Bellavista, S. Chessa, L. Foschini, L. Gioia, and M. Girolami, "Human-enabled edge computing: Exploiting the crowd as a dynamic extension of mobile edge computing," *IEEE Commun. Mag.*, vol. 56, no. 1, pp. 145–155, Jan. 2018.

[25] N. C. Luong *et al.*, "Data collection and wireless communication in Internet of Things (IoT) using economic analysis and pricing models: A survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 4, pp. 2546–2590, 4th Quart., 2016.

[26] B. Guo, C. Chen, Z. Yu, D. Zhang, and X. Zhou, "Building human-machine intelligence in mobile crowd sensing," *IT Prof.*, vol. 17, no. 3, pp. 46–52, May/Jun. 2015.

[27] H. Yao, M. Xiong, C. Liu, and Q. Liang, "Encounter probability aware task assignment in mobile crowdsensing," *Mobile Netw. Appl.*, vol. 22, no. 2, pp. 275–286, 2017.

[28] J. Howe, "The rise of crowdsourcing," *Wired*, vol. 14, no. 14, pp. 1–5, 2011.

[29] E. Estellés-Arolas and F. González-Ladrón-de-Guevara, "Towards an integrated crowdsourcing definition," *J. Inf. Sci.*, vol. 38, no. 2, pp. 189–200, 2012.

[30] I. Boutsis and V. Kalogeraki, "On task assignment for real-time reliable crowdsourcing," in *Proc. IEEE ICDCS*, Madrid, Spain, 2014, pp. 1–10.

[31] L. Xu, X. Hao, N. D. Lane, X. Liu, and T. Moscibroda, "More with less: Lowering user burden in mobile crowdsourcing through compressive sensing," in *Proc. ACM UbiComp*, Osaka, Japan, 2015, pp. 659–670.

[32] J. Goncalves *et al.*, "Crowdsourcing on the spot: Altruistic use of public displays, feasibility, performance, and behaviours," in *Proc. ACM UbiComp*, Zürich, Switzerland, 2013, pp. 753–762.

[33] Y. Wu, Y. Wang, W. Hu, X. Zhang, and G. Cao, "Resource-aware photo crowdsourcing through disruption tolerant networks," in *Proc. IEEE ICDCS*, Nara, Japan, 2016, pp. 374–383.

[34] Y. Zhang, N. Meratnia, and P. J. M. Havinga, "Outlier detection techniques for wireless sensor networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 12, no. 2, pp. 159–170, 2nd Quart., 2010.

[35] M. Pouryazdan, B. Kantarci, T. Soyata, L. Foschini, and H. Song, "Quantifying user reputation scores data, trustworthiness, and user incentives in mobile crowd-sensing," *IEEE Access*, vol. 5, pp. 1382–1397, 2017.

[36] Y. He and Y. Li, "Physical activity recognition utilizing the built-in kinematic sensors of a smartphone," *Int. J. Distrib. Sensor Netw.*, vol. 9, no. 4, pp. 481–580, 2013.

[37] T. Luo, S. S. Kanhere, J. Huang, S. K. Das, and F. Wu, "Sustainable incentives for mobile crowdsensing: Auctions lotteries, and trust and reputation systems," *IEEE Commun. Mag.*, vol. 55, no. 3, pp. 68–74, Mar. 2017.

[38] S. Yang, U. Adeel, and J. A. McCann, "Selfish mules: Social profit maximization in sparse sensornets using rationally-selfish human relays," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 6, pp. 1124–1134, Jun. 2013.

[39] R. Harmon, H. Demirkan, B. Hefley, and N. Auseklis, "Pricing strategies for information technology services: A value-based approach," in *Proc. IEEE HICSS*, 2009, pp. 1–10.

[40] R. S. Pindyck and D. Rubinfeld, *Microeconomics*, 6th ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2005.

[41] C.-K. Tham and T. Luo, "Quality of contributed service and market equilibrium for participatory sensing," *IEEE Trans. Mobile Comput.*, vol. 14, no. 4, pp. 829–842, Apr. 2015.

[42] T. Tsujimori, N. Thepvilojanapong, Y. Ohta, Y. Zhao, and Y. Tobe, "History-based incentive for crowd sensing," in *Proc. ACM IWWISS*, Saint-Étienne, France, 2014, pp. 1–6.

[43] J. K. MacKie-Mason and H. R. Varian, "Pricing congestible network resources," *IEEE J. Sel. Areas Commun.*, vol. 13, no. 7, pp. 1141–1149, Sep. 1995.

[44] R. J. Aumann and S. Hart, *Handbook of Game Theory With Economic Applications*, vol. 2, Elsevier, 1992.

[45] S. Deterding, D. Dan, R. Khaled, and L. Nacke, "From game design elements to gamefulness: Defining 'gamification,"' in *Proc. ACM MindTrek*, Tampere, Finland, 2011, pp. 9–15.

[46] Y. Arakawa and Y. Matsuda, "Gamification mechanism for enhancing a participatory urban sensing: Survey and practical results," *J. Inf. Process.*, vol. 24, no. 1, pp. 31–38, 2016.

[47] A. Albers, I. Krontiris, N. Sonehara, and I. Echizen, "Coupons as monetary incentives in participatory sensing," in *Proc. IFIP Adv. Inf. Commun. Technol.*, vol. 399, 2017, 226–237.

[48] T. Luo and C.-K. Tham, "Fairness and social welfare in incentivizing participatory sensing," in *Proc. IEEE SECON*, Seoul, South Korea, 2012, pp. 425–433.

[49] M. Pouryazdan *et al.*, "Intelligent gaming for mobile crowd-sensing participants to acquire trustworthy big data in the Internet of Things," *IEEE Access*, vol. 5, pp. 22209–22223, 2017.

[50] R. Kawajiri, M. Shimosaka, and H. Kashima, "Steered crowdsensing: Incentive design towards quality-oriented place-centric crowdsensing," in *Proc. ACM UbiComp*, Seattle, WA, USA, 2014, pp. 691–701.

[51] C.-M. Chou, K.-C. Lan, and C.-F. Yang, "Using virtual credits to provide incentives for vehicle communication," in *Proc. IEEE ITST*, Taipei, Taiwan, 2012, pp. 579–583.

[52] K. Lan and H. Wang, "On providing incentives to collect road traffic information," in *Proc. IEEE IWCMC*, Cagliari, Italy, 2013, pp. 1–5.

[53] D. Yang, G. Xue, X. Fang, and J. Tang, "Crowdsourcing to smartphones: Incentive mechanism design for mobile phone sensing," in *Proc. ACM MOBICOM*, Istanbul, Turkey, 2012, pp. 173–184.

[54] S. W. Salant, S. Switzer, and R. J. Reynolds, "Losses from horizontal merger: The effects of an exogenous change in industry structure on Cournot–Nash equilibrium," *Quart. J. Econ.*, vol. 98, no. 2, pp. 185–199, 1989.

[55] F. Zeng, R. Wang, and J. Wu, "How mobile contributors will interact with each other in mobile crowdsourcing with word of mouth mode," *IEEE Access*, vol. 7, pp. 14523–14536, 2019.

[56] B. Kantarci and H. T. Mouftah, "Trustworthy sensing for public safety in cloud-centric Internet of Things," *IEEE Internet Things J.*, vol. 1, no. 4, pp. 360–368, Aug. 2014.

[57] X. Xie, H. Chen, and H. Wu, "Bargain-based stimulation mechanism for selfish mobile nodes in participatory sensing network," in *Proc. IEEE SECON*, San Diego, CA, USA, 2009, pp. 1–9.

[58] A. Muthoo, "Bargaining theory with applications," *Econ. J.*, vol. 112, no. 483, pp. F596–F597, 2002.

[59] H. Amintoosi, S. S. Kanhere, and M. N. Torshiz, "A socially-aware incentive scheme for social participatory sensing," in *Proc. IEEE ISSNIP*, Singapore, 2015, pp. 1–6.

[60] J.-S. Lee and B. Hoh, "Sell your experiences: A market mechanism based incentive for participatory sensing," in *Proc. IEEE PerCom*, Mannheim, Germany, 2010, pp. 60–68.

[61] L. G. Jaimes, I. Vergara-Laurens, and M. A. Labrador, "A location-based incentive mechanism for participatory sensing systems with budget constraints," in *Proc. IEEE PerCom*, Lugano, Switzerland, 2012, pp. 103–108.

[62] J. Sun and H. Ma, "Collection-behavior based multi-parameter posted pricing mechanism for crowd sensing," in *Proc. IEEE ICC*, Sydney, NSW, Australia, 2014, pp. 227–232.

[63] L. G. Jaimes, I. Vergara-Laurens, and A. Raij, "A crowd sensing incentive algorithm for data collection for consecutive time slot problems," in *Proc. IEEE LATINCOM*, Cartagena, Colombia, 2014, pp. 1–5.

[64] L. G. Jaimes, I. Vergara-Laurens, and A. Chakeri, "SPREAD, a crowd sensing incentive mechanism to acquire better representative samples," in *Proc. IEEE PERCOM Workshops*, Budapest, Hungary, 2014, pp. 92–97.

[65] D. Peng, F. Wu, and G. Chen, "Pay as how well you do: A quality based incentive mechanism for crowdsensing," in *Proc. IEEE SECON*, Seattle, WA, USA, 2015, pp. 177–186.

[66] I. Krontiris and A. Albers, "Monetary incentives in participatory sensing using multi-attributive auctions," *Int. J. Parallel Emergent Distrib. Syst.*, vol. 27, no. 4, pp. 317–336, 2012.

[67] J. Li, X. Wang, R. Yu, and R. Liu, "Reputation-based incentives for data dissemination in mobile participatory sensing networks," *Int. J. Distrib. Sensor Netw.*, vol. 11, no. 12, pp. 1–13, 2015.

[68] C. Zhao *et al.*, "Cheating-resilient incentive scheme for mobile crowdsensing systems," in *Proc. IEEE CCNC*, Las Vegas, NV, USA, 2017, pp. 377–382.

[69] M. Pouryazdan, B. Kantarci, T. Soyata, and H. Song, "Anchor-assisted and vote-based trustworthiness assurance in smart city crowdsensing," *IEEE Access*, vol. 4, pp. 529–541, 2017.

[70] I. Koutsopoulos, "Optimal incentive-driven design of participatory sensing systems," in *Proc. IEEE INFOCOM*, Turin, Italy, 2013, pp. 1402–1410.

[71] L. M. Ausubel and P. Milgrom, "The lovely but lonely Vickrey auction," in *Combinatorial Auctions*, vol. 17. Cambridge, MA, USA: MIT Press, 2006, pp. 22–26.

[72] 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 MobiHoc*, Hangzhou, China, 2015, pp. 167–176.

[73] J. Sun and H. Ma, "A behavior-based incentive mechanism for crowd sensing with budget constraints," in *Proc. IEEE ICC*, Sydney, NSW, Australia, 2014, pp. 1314–1319.

[74] T. Luo, H.-P. Tan, and L. Xia, "Profit-maximizing incentive for participatory sensing," in *Proc. IEEE INFOCOM*, Toronto, ON, Canada, 2014, pp. 127–135.

[75] Z. Duan, W. Li, and Z. Cai, "Distributed auctions for task assignment and scheduling in mobile crowdsensing systems," in *Proc. IEEE ICDCS*, Atlanta, GA, USA, 2017, pp. 635–644.

[76] Z. Dong, H. Ma, and L. Liu, "Frugal online incentive mechanisms for mobile crowd sensing," *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3319–3330, Apr. 2017.

[77] K. Han, H. Huang, and J. Luo, "Posted pricing for robust crowdsensing," in *Proc. IEEE SECON*, London, U.K., 2016, pp. 261–270.

[78] J. Sun, "An incentive scheme based on heterogeneous belief values for crowd sensing in mobile social networks," in *Proc. IEEE GLOBECOM*, Atlanta, GA, USA, 2013, pp. 1717–1722.

[79] Y. Han, Y. Zhu, and J. Yu, "Utility-maximizing data collection in crowd sensing: An optimal scheduling approach," in *Proc. IEEE SECON*, Seattle, WA, USA, 2015, pp. 345–353.

[80] X. Gong, X. Chen, J. Zhang, and H. V. Poor, "Exploiting social trust assisted reciprocity (STAR) toward utility-optimal socially-aware crowdsensing," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 1, no. 3, pp. 195–208, Sep. 2015.

[81] S. Martello and P. Toth, *Knapsack Problems: Algorithms and Computer Implementations*. Chichester, U.K.: Wiley, 1990.

[82] N. D. Lane *et al.*, "Piggyback crowdsensing (PCS): Energy efficient crowdsourcing of mobile sensor data by exploiting smartphone app opportunities," in *Proc. ACM SenSys*, Roma, Italy, 2013, pp. 1–14.

[83] H. N. Pham, B. S. Sim, and H. Y. Youn, "A novel approach for selecting the participants to collect data in participatory sensing," in *Proc. IEEE SAINT*, Munich, Germany, 2011, pp. 50–55.

[84] S. Kukkonen and J. Lampinen, "GDE3: The third evolution step of generalized differential evolution," in *Proc. IEEE CEC*, Edinburgh, U.K., 2005, pp. 443–450.

[85] H. Mühlenbein, "The equation for response to selection and its use for prediction," *Evol. Comput.*, vol. 5, no. 3, pp. 303–346, 1997.

[86] V. V. Vazirani, *Approximation Algorithms*. Heidelberg, Germany: Springer, 2013.

[87] R. Atat *et al.*, "Enabling cyber-physical communication in 5G cellular networks: Challenges, spatial spectrum sensing, and cyber-security," *IET Cyber Phys. Syst. Theory Appl.*, vol. 2, no. 1, pp. 49–54, Apr. 2017.

[88] D. He, S. Chan, and M. Guizani, "User privacy and data trustworthiness in mobile crowd sensing," *IEEE Wireless Commun.*, vol. 22, no. 1, pp. 28–34, Feb. 2015.

[89] J. Lin *et al.*, "Expectation and purpose: Understanding users' mental models of mobile app privacy through crowdsourcing," in *Proc. ACM UbiComp*, Pittsburgh, PA, USA, 2012, pp. 501–510.

[90] A. Kapadia, D. Kotz, and N. Triandopoulos, "Opportunistic sensing: Security challenges for the new paradigm," in *Proc. IEEE COMSNETS*, Bengaluru, India, 2009, pp. 1–10.

[91] S. Gisdakis, T. Giannetsos, and P. Papadimitratos, "Security, privacy, and incentive provision for mobile crowd sensing systems," *IEEE Internet Things J.*, vol. 3, no. 5, pp. 839–853, Oct. 2016.

[92] A. D. Wood and J. A. Stankovic, "Denial of service in sensor networks," *IEEE Comput.*, vol. 35, no. 10, pp. 54–62, Oct. 2002.

[93] S. Krishnaswamy, J. Gama, and M. M. Gaber, "Mobile data stream mining: From algorithms to applications," in *Proc. IEEE MDM*, Bengaluru, India, 2012, pp. 360–363.

[94] Q. Xu, R. Zheng, and E. Tahoun, "Detecting location fraud in indoor mobile crowdsensing," in *Proc. ACM CrowdSenSys*, Delft, The Netherlands, 2017, pp. 44–49.

[95] I. J. Vergara-Laurens, D. Mendez-Chaves, and M. A. Labrador, "On the interactions between privacy-preserving, incentive, and inference mechanisms in participatory sensing systems," in *Proc. NSS*, Madrid, Spain, 2013, pp. 614–620.

[96] M. Shin *et al.*, "Anonysense: A system for anonymous opportunistic sensing," *Pervasive Mobile Comput.*, vol. 7, no. 1, pp. 16–30, 2011.

[97] B. Hoh and M. Gruteser, "Protecting location privacy through path confusion," in *Proc. IEEE SECURECOMM*, Athens, Greece, 2005, pp. 194–205.

[98] C. Cornelius *et al.*, "Anonysense: Privacy-aware people-centric sensing," in *Proc. ACM MobiSys*, Breckenridge, CO, USA, 2008, pp. 211–224.

[99] L. Sweeney, "k-anonymity: A model for protecting privacy," *Int. J. Uncertainty Fuzziness Knowl. Based Syst.*, vol. 10, no. 5, pp. 557–570, 2002.

[100] L. Hu and C. Shahabi, "Privacy assurance in mobile sensing networks: Go beyond trusted servers," in *Proc. IEEE PERCOM Workshops*, Mannheim, Germany, 2010, pp. 613–619.

[101] S. Gao, J. Ma, W. Shi, G. Zhan, and C. Sun, "TrPF: A trajectory privacy-preserving framework for participatory sensing," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 6, pp. 874–887, Jun. 2013.

[102] O. Abul, F. Bonchi, and M. Nanni, "Never walk alone: Uncertainty for anonymity in moving objects databases," in *Proc. IEEE ICDE*, Cancun, Mexico, 2008, pp. 376–385.

[103] J. Domingo-Ferrer and V. Torra, "Ordinal, continuous and heterogeneous k-anonymity through microaggregation," *Data Min. Knowl. Disc.*, vol. 11, no. 2, pp. 195–212, 2005.

[104] K. L. Huang, S. S. Kanhere, and W. Hu, "Preserving privacy in participatory sensing systems," *Comput. Commun.*, vol. 33, no. 11, pp. 1266–1280, 2010.

[105] J. Horey, M. M. Groat, S. Forrest, and F. Esponda, "Anonymous data collection in sensor networks," in *Proc. IEEE MobiQuitous*, Philadelphia, PA, USA, 2007, pp. 1–8.

[106] M. Murshed, T. Sabrina, A. Iqbal, and K. M. Alam, "A novel anonymization technique to trade-off location privacy and data integrity in participatory sensing systems," in *Proc. IEEE NSS*, Melbourne, VIC, Australia, 2010, pp. 345–350.

[107] M. A. Alsheikh *et al.*, "The accuracy-privacy trade-off of mobile crowdsensing," *IEEE Commun. Mag.*, vol. 55, no. 6, pp. 132–139, Jun. 2017.

[108] J. Shi, R. Zhang, Y. Liu, and Y. Zhang, "PriSense: Privacy-preserving data aggregation in people-centric urban sensing systems," in *Proc. IEEE INFOCOM*, San Diego, CA, USA, 2010, pp. 1–9.

[109] L. Kazemi, and C. Shahabi, "A privacy-aware framework for participatory sensing," *ACM SIGKDD Explor. Newslett.*, vol. 13, no. 1, pp. 43–51, 2011.

[110] W. Alsalih, K. Islam, Y. N. Rodríguez, and H. Xiao, "Distributed Voronoi diagram computation in wireless sensor networks," in *Proc. ACM SPAA*, Munich, Germany, 2008, p. 364.

[111] D. M. Konidala, R. H. Deng, Y. Li, H. C. Lau, and S. E. Fienberg, "Anonymous authentication of visitors for mobile crowd sensing at amusement parks," in *Proc. ISPEC*, Lanzhou, China, 2013, pp. 174–188.

[112] F. Dürr, P. Skvortsov, and K. Rothermel, "Position sharing for location privacy in non-trusted systems," in *Proc. IEEE PerCom*, Seattle, WA, USA, 2011, pp. 189–196.

[113] R. K. Ganti, N. Pham, Y.-E. Tsai, and T. F. Abdelzaher, "PoolView: Stream privacy for grassroots participatory sensing," in *Proc. ACM SenSys*, Raleigh, NC, USA, 2008, pp. 281–294.

[114] Q. Li and G. Cao, "Efficient privacy-preserving stream aggregation in mobile sensing with low aggregation error," in *Proc. PETS*, Bloomington, IN, USA, 2013, pp. 60–81.

[115] J. Hamm, A. C. Champion, G. Chen, M. Belkin, and D. Xuan, "Crowd-ML: A privacy-preserving learning framework for a crowd of smart devices," in *Proc. IEEE ICDCS*, Columbus, OH, USA, 2015, pp. 11–20.

[116] L. Wang, D. Zhang, D. Yang, B. Y. Lim, and X. Ma, "Differential location privacy for sparse mobile crowdsensing," in *Proc. IEEE ICDM*, New York, NY, USA, 2017, pp. 1257–1262.

[117] I. Krontiris and T. Dimitriou, "Privacy-respecting discovery of data providers in crowd-sensing applications," in *Proc. IEEE ICDCS*, Philadelphia, PA, USA, 2013, pp. 249–257.

[118] Y. Liu *et al.*, "Stop unauthorized access to your smart devices," in *Proc. IEEE ICPADS*, Singapore, 2018, pp. 228–235.

[119] B. Hoh *et al.*, "Virtual trip lines for distributed privacy-preserving traffic monitoring," in *Proc. ACM MobiSys*, Breckenridge, CO, USA, 2008, pp. 15–28.

[120] I. J. Vergara-Laurens and M. A. Labrador, "Preserving privacy while reducing power consumption and information loss in LBS and participatory sensing applications," in *Proc. IEEE GLOBECOM Workshops*, Houston, TX, USA, 2011, pp. 1247–1252.

[121] L. Kong *et al.*, "Privacy-preserving compressive sensing for crowdsensing based trajectory recovery," in *Proc. IEEE ICDCS*, Columbus, OH, USA, 2015, pp. 31–40.

[122] P. Newson and J. Krumm, "Hidden Markov map matching through noise and sparseness," in *Proc. ACM SIGSPATIAL GIS*, Seattle, WA, USA, 2009, pp. 336–343.

[123] O. Fatemieh, R. Chandra, and C. A. Gunter, "Secure collaborative sensing for crowd sourcing spectrum data in white space networks," in *Proc. IEEE DySPAN*, Singapore, 2010, pp. 1–12.

[124] N. Ruan *et al.*, "Toward optimal DoS-resistant authentication in crowdsensing networks via evolutionary game," in *Proc. IEEE ICDCS*, Nara, Japan, 2016, pp. 364–373.

[125] G. Wang *et al.*, "Defending against sybil devices in crowdsourced mapping services," in *Proc. ACM MobiSys*, Singapore, 2016, pp. 179–191.

[126] Q. Cao, M. Sirivianos, X. Yang, and T. Pregueiro, "Aiding the detection of fake accounts in large scale social online services," in *Proc. USENIX NSDI*, San Jose, CA, USA, 2012, p. 15.

[127] D. Zhang, H. Xiong, L. Wang, and G. Chen, "CrowdRecruiter: Selecting participants for piggyback crowdsensing under probabilistic coverage constraint," in *Proc. ACM UbiComp*, Seattle, WA, USA, 2014, pp. 703–714.

[128] H. Li, T. Li, F. Li, W. Wang, and Y. Wang, "Enhancing participant selection through caching in mobile crowd sensing," in *Proc. IEEE/ACM IWQoS*, Beijing, China, 2016, pp. 1–10.

[129] Z. Song, C. H. Liu, J. Wu, J. Ma, and W. Wang, "QoI-aware multitask-oriented dynamic participant selection with budget constraints," *IEEE Trans. Veh. Technol.*, vol. 63, no. 9, pp. 4618–4632, Nov. 2014.

[130] E.-H. Chu, C.-Y. Lin, P. Tsai, and J. Liu, "Participant selection for crowdsourcing disaster information," *WIT Trans. Built Environ.*, vol. 133, pp. 231–240, Feb. 2013.

[131] M. Tomasoni *et al.*, "Profiling energy efficiency of mobile crowdsensing data collection frameworks for smart city applications," in *Proc. IEEE MobileCloud*, Bamberg, Germany, 2018, pp. 1–8.

[132] H. Xiong, D. Zhang, L. Wang, and H. Chaouchi, "EMC^3: Energy-efficient data transfer in mobile crowdsensing under full coverage constraint," *IEEE Trans. Mobile Comput.*, vol. 14, no. 7, pp. 1355–1368, Jul. 2015.

[133] G. Bajaj and P. Singh, "Load-balanced task allocation for improved system lifetime in mobile crowdsensing," in *Proc. IEEE MDM*, Aalborg, Denmark, 2018, pp. 227–232.

[134] F. Anjomshoa and B. Kantarci, "SOBER-MCS: Sociability-oriented and battery efficient recruitment for mobile crowd-sensing," *Sensors*, vol. 18, no. 5, p. 1593, 2018.

[135] C. Fiandrino *et al.*, "Sociability-driven framework for data acquisition in mobile crowdsensing over fog computing platforms for smart cities," *IEEE Trans. Sustain. Comput.*, vol. 2, no. 4, pp. 345–358, Oct./Dec. 2017.

[136] W. Wang, H. Gao, C. H. Liu, and K. K. Leung, "Credible and energy-aware participant selection with limited task budget for mobile crowd sensing," *Ad Hoc Netw.*, vol. 43, pp. 56–70, Jun. 2016.

[137] K. Han, C. Zhang, and J. Luo, "Taming the uncertainty: Budget limited robust crowdsensing through online learning," *IEEE/ACM Trans. Netw.*, vol. 24, no. 3, pp. 1462–1475, Jun. 2016.

[138] S. Bubeck and N. Cesa-Bianchi, "Regret analysis of stochastic and nonstochastic multi-armed bandit problems," *Found. Trends Mach. Learn.*, vol. 5, no. 1, pp. 101–112, 2012.

[139] S. Liu, Z. Zheng, F. Wu, S. Tang, and G. Chen, "Context-aware data quality estimation in mobile crowdsensing," in *Proc. IEEE INFOCOM*, Atlanta, GA, USA, 2017, pp. 1–9.

[140] R. Azzam, R. Mizouni, H. Otrok, A. Ouali, and S. Singh, "GRS: A group-based recruitment system for mobile crowd sensing," *J. Netw. Comput. Appl.*, vol. 72, pp. 38–50, Sep. 2016.

[141] M. Karaliopoulos, I. Koutsopoulos, and M. Titsias, "First learn then earn: Optimizing mobile crowdsensing campaigns through data-driven user profiling," in *Proc. ACM MobiHoc*, Paderborn, Germany, 2016, pp. 271–280.

[142] L. Pournajaf, L. Xiong, and V. Sunderam, "Dynamic data driven crowd sensing task assignment," *Procedia Comput. Sci.*, vol. 29, pp. 1314–1323, Jun. 2014.

[143] S. He, D.-H. Shin, J. Zhang, and J. Chen, "Toward optimal allocation of location dependent tasks in crowdsensing," in *Proc. IEEE INFOCOM*, Toronto, ON, Canada, 2014, pp. 745–753.

[144] S. He, D. Shin, J. Zhang, and J. Chen, "Near-optimal allocation algorithms for location-dependent tasks in crowdsensing," *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3392–3405, Apr. 2017.

[145] A. A. Obinikpo, Y. Zhang, H. Song, T. H. Luan, and B. Kantarci, "Queuing algorithm for effective target coverage in mobile crowd sensing," *IEEE Internet Things J.*, vol. 4, no. 4, pp. 1046–1055, Aug. 2017.

[146] X. Wang, R. Jia, X. Tian, and X. Gan, "Dynamic task assignment in crowdsensing with location awareness and location diversity," in *Proc. IEEE INFOCOM*, Honolulu, HI, USA, 2018, pp. 2420–2428.

[147] H. Xiong, D. Zhang, L. Wang, J. P. Gibson, and J. Zhu, "EEMC: Enabling energy-efficient mobile crowdsensing with anonymous participants," *ACM Trans. Intell. Syst. Technol.* vol. 6, no. 3, 2015, Art. no. 39.

[148] L. Wang *et al.*, "CCS-TA: Quality-guaranteed online task allocation in compressive crowdsensing," in *Proc. ACM UbiComp*, Osaka, Japan, 2015, pp. 683–694.

[149] B. Guo, Y. Liu, W. Wu, Z. Yu, and Q. Han, "ActiveCrowd: A framework for optimized multitask allocation in mobile crowdsensing systems," *IEEE Trans. Human–Mach. Syst.*, vol. 47, no. 3, pp. 392–403, Jun. 2017.

[150] M. H. Cheung, R. Southwell, F. Hou, and J. Huang, "Distributed time-sensitive task selection in mobile crowdsensing," in *Proc. ACM MobiHoc*, Hangzhou, China, 2015, pp. 157–166.

[151] A. Capponi, C. Fiandrino, D. Kliazovich, and P. Bouvry, "Energy efficient data collection in opportunistic mobile crowdsensing architectures for smart cities," in *Proc. IEEE INFOCOM WKSHPS*, Atlanta, GA, USA, 2017, pp. 307–312.

[152] A. Capponi, C. Fiandrino, D. Kliazovich, P. Bouvry, and S. Giordano, "A cost-effective distributed framework for data collection in cloud-based mobile crowd sensing architectures," *IEEE Trans. Sustain. Comput.*, vol. 2, no. 1, pp. 3–16, Jan./Mar. 2017.

[153] Ö. Yürür *et al.*, "Context-awareness for mobile sensing: A survey and future directions," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 1, pp. 68–93, 1st Quart., 2016.

[154] L. Pournajaf, L. Xiong, V. Sunderam, and S. Goryczka, "Spatial task assignment for crowd sensing with cloaked locations," in *Proc. IEEE MDM*, Brisbane, QLD, Australia, 2014, pp. 73–82.

[155] Y. Xiao, P. Simoens, P. Pillai, K. Ha, and M. Satyanarayanan, "Lowering the barriers to large-scale mobile crowdsensing," in *Proc. IEEE HotMobile*, 2013, Art. no. 9.

[156] P. Talebifard and V. C. M. Leung, "Towards a content-centric approach to crowd-sensing in vehicular clouds," *J. Syst. Archit.*, vol. 59, no. 10, pp. 976–984, 2013.

[157] M. Marjanović, L. Skorin-Kapov, K. Pripužić, A. Antonić, and I. P. Žarko, "Energy-aware and quality-driven sensor management for green mobile crowd sensing," *J. Netw. Comput. Appl.*, vol. 59, pp. 95–108, Jan. 2016.

[158] L. Wang *et al.*, "Sparse mobile crowdsensing: Challenges and opportunities," *IEEE Commun. Mag.*, vol. 54, no. 7, pp. 161–167, Jul. 2016.

[159] S. Hu *et al.*, "On exploiting logical dependencies for minimizing additive cost metrics in resource-limited crowdsensing," in *Proc. IEEE DCOSS*, Fortaleza, Brazil, 2015, pp. 189–198.

[160] X. Sheng, J. Tang, X. Xiao, and G. Xue, "Leveraging GPS-less sensing scheduling for green mobile crowd sensing," *IEEE Internet Things J.*, vol. 1, no. 4, pp. 328–336, Aug. 2014.

[161] A. Khan, S. K. A. Imon, and S. K. Das, "Ensuring energy efficient coverage for participatory sensing in urban streets," in *Proc. IEEE SMARTCOMP*, Hong Kong, 2014, pp. 167–174.

[162] S. Liu and A. D. Striegel, "Exploring the potential in practice for opportunistic networks amongst smart mobile devices," in *Proc. ACM/IEEE MobiCom*, 2013, pp. 315–326.

[163] M. Tomasoni *et al.*, "Why energy matters? Profiling energy consumption of mobile crowdsensing data collection frameworks," *Pervasive Mobile Comput.*, vol. 51, pp. 193–208, Dec. 2018.

[164] M. Karaliopoulos, O. Telelis, and I. Koutsopoulos, "User recruitment for mobile crowdsensing over opportunistic networks," in *Proc. IEEE INFOCOM*, Hong Kong, 2015, pp. 2254–2262.

[165] M. Conti and M. Kumar, "Opportunities in opportunistic computing," *Computer*, vol. 43, no. 1, pp. 42–50, Jan. 2010.

[166] T. Higuchi, H. Yamaguchi, T. Higashino, and M. Takai, "A neighbor collaboration mechanism for mobile crowd sensing in opportunistic networks," in *Proc. IEEE ICC*, Sydney, NSW, Australia, 2014, pp. 42–47.

[167] Y. Chon, N. D. Lane, F. Li, H. Cha, and F. Zhao, "Automatically characterizing places with opportunistic crowdsensing using smartphones," in *Proc. ACM UbiComp*, Pittsburgh, PA, USA, 2012, pp. 481–490.

[168] L. Wang, D. Zhang, and H. Xiong, "EffSense: Energy-efficient and cost-effective data uploading in mobile crowdsensing," in *Proc. ACM UbiComp*, Zürich, Switzerland, 2013, pp. 1075–1086.

[169] Q. Xu and R. Zheng, "When data acquisition meets data analytics: A distributed active learning framework for optimal budgeted mobile crowdsensing," in *Proc. IEEE INFOCOM*, Atlanta, GA, USA, 2017, pp. 1–9.

[170] C. Meng *et al.*, "Truth discovery on crowd sensing of correlated entities," in *Proc. ACM Sensys*, Seoul, South Korea, 2015, pp. 1–14.

[171] B. Wang *et al.*, "I(TS, CS): Detecting faulty location data in mobile crowdsensing," in *Proc. ICDCS*, Vienna, Austria, 2018, pp. 808–817.

[172] U. Weinsberg *et al.*, "CARE: Content aware redundancy elimination for challenged networks," in *Proc. ACM HotNets Workshop*, Redmond, WA, USA, 2012, pp. 1–6.

[173] S. Zhang, H. Sheng, C. Li, J. Zhang, and Z. Xiong, "Robust depth estimation for light field via spinning parallelogram operator," *Comput. Vis. Image Understand.*, vol. 145, pp. 148–159, Apr. 2016.

[174] H. Sheng, S. Zhang, X. Cao, Y. Fang, and Z. Xiong, "Geometric occlusion analysis in depth estimation using integral guided filter for light-field image," *IEEE Trans. Image Process.*, vol. 26, no. 12, pp. 5758–5771, Dec. 2017.

[175] D. Zhang *et al.*, "Exploring human mobility with multi-source data at extremely large metropolitan scales," in *Proc. ACM/IEEE MobiCom*, 2014, pp. 1–12.

[176] M. Talasila, R. Curtmola, and C. Borcea, "Improving location reliability in crowd sensed data with minimal efforts," in *Proc. IEEE WMNC*, Dubai, UAE, 2013, pp. 65–85.
[177] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
[178] S. Basu and M. Meckesheimer, "Automatic outlier detection for time series: An application to sensor data," *Knowl. Inf. Syst.*, vol. 11, no. 2, pp. 137–154, 2007.
[179] Y. Chon, N. D. Lane, Y. Kim, F. Zhao, and H. Cha, "Understanding the coverage and scalability of place-centric crowdsensing," in *Proc. ACM UbiComp*, Zürich, Switzerland, 2013, pp. 3–12.
[180] A. Zaslavsky, P. P. Jayaraman, and S. Krishnaswamy, "ShareLikesCrowd: Mobile analytics for participatory sensing and crowd-sourcing applications," in *Proc. IEEE ICDEW*, Brisbane, QLD, Australia, 2013, pp. 128–135.
[181] Y. Zhang, M. Chen, S. Mao, L. Hu, and V. Leung, "CAP: Community activity prediction based on big data analysis," *IEEE Netw.*, vol. 28, no. 4, pp. 52–57, Jul./Aug. 2014.
[182] W. Sherchan *et al.*, "Using on-the-move mining for mobile crowdsensing," in *Proc. IEEE MDM*, Bengaluru, India, 2012, pp. 115–124.
[183] C. Liu *et al.*, "Lasagna: Towards deep hierarchical understanding and searching over mobile sensing data," in *Proc. ACM/IEEE MobiCom*, New York, NY, USA, 2016, pp. 334–347.
[184] S. Krishnaswamy *et al.*, "Open mobile miner: A toolkit for mobile data stream mining," in *Proc. ACM KDD Demo*, Paris, France, 2009, pp. 1–4.
[185] M. de Almeida Buosi *et al.*, "A crowdsensing campaign and data analytics for assisting urban mobility pattern determination," in *Proc. IEEE ISCC*, Natal, Brazil, 2018, pp. 224–229.
[186] K. Chintalapudi, A. P. Iyer, and V. N. Padmanabhan, "Indoor localization without the pain," in *Proc. IEEE Mobicom*, Chicago, IL, USA, 2010, pp. 173–184.
[187] A. Goswami, L. E. Ortiz, and S. R. Das, "WIGEM: A learning-based approach for indoor localization," in *Proc. Conext*, 2011, pp. 1–12.
[188] A. Rai, K. K. Chintalapudi, V. N. Padmanabhan, and R. Sen, "Zee: Zero-effort crowdsourcing for indoor localization," in *Proc. ACM/IEEE MobiCom*, Istanbul, Turkey, 2012, pp. 293–304.
[189] V. Radu, L. Kriara, and M. K. Marina, "Pazl: A mobile crowdsensing based indoor WiFi monitoring system," in *Proc. IEEE CNSM*, Zürich, Switzerland, 2013, pp. 75–83.
[190] R. Gao *et al.*, "Jigsaw: Indoor floor plan reconstruction via mobile crowdsensing," in *Proc. ACM/IEEE MobiCom*, 2014, pp. 249–260.
[191] S. Chen, M. Li, K. Ren, and C. Qiao, "Crowd map: Accurate reconstruction of indoor floor plans from crowdsourced sensor-rich videos," in *Proc. IEEE ICDCS*, Philadelphia, PA, USA, 2015, pp. 1–10.
[192] C. Zhang, K. P. Subbu, J. Luo, and J. Wu, "GROPING: Geomagnetism and crowdsensing powered indoor navigation," *IEEE Trans. Mobile Comput.*, vol. 14, no. 2, pp. 387–400, Feb. 2015.
[193] Q. Xu and R. Zheng, "MobiBee: A mobile treasure hunt game for location-dependent fingerprint collection," in *Proc. ACM UbiComp*, Heidelberg, Germany, 2016, pp. 1472–1477.
[194] H. Yu, H. Pang, and M. Zhang, "Value-added effects of transit-oriented development: The impact of urban rail on commercial property values with consideration of spatial heterogeneity," *Papers Reg. Sci.*, vol. 97, no. 4, pp. 1375–1396, 2017.
[195] Y. Wang *et al.*, "CrowdAtlas: Self-updating maps for cloud and personal use," in *Proc. ACM Mobisys*, Taipei, Taiwan, 2013, pp. 1–14.
[196] B. Kegl, A. Krzyzak, T. Linder, and K. Zeger, "Learning and design of principal curves," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 3, pp. 281–297, Mar. 2000.
[197] S. Hu, L. Su, H. Liu, H. Wang, and T. F. Abdelzaher, "SmartRoad: Smartphone-based crowd sensing for traffic regulator detection and identification," *ACM Trans. Sensor Netw.*, vol. 11, no. 4, 2015, Art. no. 55.
[198] B. Pan, Y. Zheng, D. Wilkie, and C. Shahabi, "Crowd sensing of traffic anomalies based on human mobility and social media," in *Proc. ACM SIGSPATIAL GIS*, Orlando, FL, USA, 2013, pp. 344–353.
[199] P. Zhou, S. Jiang, and M. Li, "Urban traffic monitoring with the help of bus riders," in *Proc. IEEE ICDCS*, Columbus, OH, USA, 2015, pp. 21–30.
[200] P. Zhou, Y. Zheng, and M. Li, "How long to wait: Predicting bus arrival time with mobile phone based participatory sensing," in *Proc. ACM Mobisys*, Ambleside, U.K., 2012, pp. 11093–11096.
[201] B. Hoh *et al.*, "TruCentive: A game-theoretic incentive platform for trustworthy mobile crowdsourcing parking services," in *Proc. IEEE ITSC*, Anchorage, AK, USA, 2012, pp. 160–166.
[202] V. Coric and M. Gruteser, "Crowdsensing maps of on-street parking spaces," in *Proc. IEEE DCOSS*, Cambridge, MA, USA, 2013, pp. 115–122.
[203] F. Zhang, D. Wilkie, Y. Zheng, and X. Xie, "Sensing the pulse of urban refueling behavior," in *Proc. ACM UbiComp*, Zürich, Switzerland, 2013, Art. no. 37.
[204] M. F. Bulut, M. Demirbas, and H. Ferhatosmanoglu, "LineKing: Coffee shop wait-time monitoring using smartphones," *IEEE Trans. Mobile Comput.*, vol. 14, no. 10, pp. 2045–2058, Oct. 2015.
[205] B. Guo *et al.*, "FlierMeet: Cross-space public information reposting with mobile crowd sensing," in *Proc. ACM UbiComp*, Seattle, WA, USA, 2014, pp. 371–372.
[206] A. Farshad, M. K. Marina, and F. J. Garcia, "Urban WiFi characterization via mobile crowdsensing," in *Proc. IEEE NOMS*, Kraków, Poland, 2014, pp. 1–9.
[207] J. Wu, S. Guo, J. Li, and D. Zeng, "Big data meet green challenges: Big data toward green applications," *IEEE Syst. J.*, vol. 10, no. 3, pp. 888–900, Sep. 2016.
[208] E. D'Hondt, M. Stevens, and A. Jacobs, "Participatory noise mapping works! An evaluation of participatory sensing as an alternative to standard techniques for environmental monitoring," *Pervasive Mobile Comput.*, vol. 9, no. 5, pp. 681–694, 2013.
[209] N. Maisonneuve, M. Stevens, M. E. Niessen, and L. Steels, "NoiseTube: Measuring and mapping noise pollution with mobile phones," *Environ. Sci. Eng.*, vol. 2, no. 6, pp. 215–228, 2009.
[210] D. Santani *et al.*, "The night is young: Urban crowdsourcing of nightlife patterns," in *Proc. ACM UbiComp*, Heidelberg, Germany, 2016, pp. 1–12.
[211] A. Antonić *et al.*, "Urban crowd sensing demonstrator: Sense the Zagreb air," in *Proc. IEEE SoftCom*, Split, Croatia, 2014, pp. 423–424.
[212] V. Pankratius, F. Lind, A. Coster, P. Erickson, and J. Semeter, "Mobile crowd sensing in space weather monitoring: The Mahali project," *IEEE Commun. Mag.*, vol. 52, no. 8, pp. 22–28, Aug. 2014.
[213] S. Morishita *et al.*, "SakuraSensor: Quasi-realtime cherry-lined roads detection through participatory video sensing by cars," in *Proc. ACM UbiComp*, Osaka, Japan, 2015, pp. 695–705.
[214] G. Drosatos *et al.*, "A privacy-preserving cloud computing system for creating participatory noise maps," in *Proc. IEEE CompSac*, Izmir, Turkey, 2012, pp. 581–586.
[215] R. Pryss, M. Reichert, J. Herrmann, B. Langguth, and W. Schlee, "Mobile crowd sensing in clinical and psychological trials—A case study," in *Proc. IEEE CBMS*, São Carlos, Brazil, 2015, pp. 23–24.
[216] Y. He, Y. Li, and S.-D. Bao, "Fall detection by built-in tri-accelerometer of smartphone," in *Proc. IEEE EMBS*, Hong Kong, 2012, pp. 184–187.
[217] C. Ragona, F. Granelli, C. Fiandrino, D. Kliazovich, and P. Bouvry, "Energy-efficient computation offloading for wearable devices and smartphones in mobile cloud computing," in *Proc. IEEE GLOBECOM*, San Diego, CA, USA, 2015, pp. 1–6.
[218] D. Hristova, M. J. Williams, M. Musolesi, P. Panzarasa, and C. Mascolo, "Measuring urban social diversity using interconnected geo-social networks," in *Proc. WWW*, Montreal, QC, Canada, 2016, pp. 21–30.
[219] J. Cranshaw, E. Toch, J. I. Hong, A. Kittur, and N. M. Sadeh, "Bridging the gap between physical location and online social networks," in *Proc. ACM UbiComp*, Copenhagen, Denmark, 2010, pp. 119–128.
[220] B. Guo, Z. Yu, L. Chen, X. Zhou, and X. Ma, "MobiGroup: Enabling lifecycle support to social activity organization and suggestion with mobile crowd sensing," *IEEE Trans. Human–Mach. Syst.*, vol. 46, no. 3, pp. 390–402, Jun. 2016.
[221] T. Ludwig, C. Reuter, T. Siebigteroth, and V. Pipek, "CrowdMonitor: Mobile crowd sensing for assessing physical and digital activities of citizens during emergencies," in *Proc. ACM CHI*, Seoul, South Korea, 2015, pp. 4083–4092.
[222] H. Roitman, J. Mamou, S. Mehta, A. Satt, and L. V. Subramaniam, "Harnessing the crowds for smart city sensing," in *Proc. ACM CrowdSens*, Maui, HI, USA, 2012, pp. 17–18.
[223] R. W. Ouyang *et al.*, "If you see something, swipe towards it: Crowdsourced event localization using smartphones," in *Proc. ACM UbiComp*, Zürich, Switzerland, 2013, pp. 23–32.
[224] J. Weppner and P. Lukowicz, "Bluetooth based collaborative crowd density estimation with mobile phones," in *Proc. IEEE PerCom*, San Diego, CA, USA, 2013, pp. 193–200.
[225] G. Cardone *et al.*, "Crowdsensing in urban areas for city-scale mass gathering management: Geofencing and activity recognition," *IEEE Sensors J.*, vol. 14, no. 12, pp. 4185–4195, Dec. 2014.

[226] M.-R. Ra, B. Liu, T. F. La Porta, and R. Govindan, "Medusa: A programming framework for crowd-sensing applications," in *Proc. ACM MobiSys*, Ambleside, U.K., 2012, pp. 337–350.
[227] G. Cardone *et al.*, "Fostering participaction in smart cities: A geo-social crowdsensing platform," *IEEE Commun. Mag.*, vol. 51, no. 6, pp. 112–119, Jun. 2013.
[228] X. Hu, T. H. S. Chu, H. C. B. Chan, and V. C. M. Leung, "Vita: A crowdsensing-oriented mobile cyber-physical system," *IEEE Trans. Emerg. Topics Comput.*, vol. 1, no. 1, pp. 148–165, Jun. 2013.
[229] G. Cardone, A. Cirri, A. Corradi, and L. Foschini, "The ParticipAct mobile crowd sensing living lab: The testbed for smart cities," *IEEE Commun. Mag.*, vol. 52, no. 10, pp. 78–85, Oct. 2014.
[230] S. Chessa *et al.*, "Mobile crowd sensing management with the ParticipAct living lab," *Pervasive Mobile Comput.*, vol. 38, pp. 200–214, Jul. 2017.
[231] C. M. Angelopoulos *et al.*, "A user-enabled testbed architecture with mobile crowdsensing support for smart, green buildings," in *Proc. IEEE ICC*, London, U.K., 2015, pp. 573–578.
[232] K. Jayarajah, R. K. Balan, M. Radhakrishnan, A. Misra, and Y. Lee, "LiveLabs: Building in-situ mobile sensing & behavioural experimentation testbeds," in *Proc. ACM MobiSys*, Singapore, 2016, pp. 1–15.
[233] T. Kandappu *et al.*, "TASKer: Behavioral insights via campus-based experimental mobile crowd-sourcing," in *Proc. ACM UbiComp*, Heidelberg, Germany, 2016, pp. 392–402.
[234] H. Xiong, Y. Huang, L. E. Barnes, and M. S. Gerber, "Sensus: A cross-platform, general-purpose system for mobile crowdsensing in human-subject studies," in *Proc. ACM UbiComp*, Heidelberg, Germany, 2016, pp. 415–426.
[235] G. Han, L. Liu, S. Chan, R. Yu, and Y. Yang, "HySense: A hybrid mobile crowdsensing framework for sensing opportunities compensation under dynamic coverage constraint," *IEEE Commun. Mag.*, vol. 55, no. 3, pp. 93–99, Mar. 2017.
[236] L. Luceri *et al.*, "VIVO: A secure, privacy-preserving, and real-time crowd-sensing framework for the Internet of Things," *Pervasive Mobile Comput.*, vol. 49, pp. 126–138, Sep. 2018.
[237] I. Carreras, D. Miorandi, A. Tamilin, E. R. Ssebaggala, and N. Conci, "Matador: Mobile task detector for context-aware crowd-sensing campaigns," in *Proc. IEEE PERCOM Workshops*, San Diego, CA, USA, 2013, pp. 212–217.
[238] X. Hu, X. Li, E. C.-H. Ngai, V. C. M. Leung, and P. Kruchten, "Multidimensional context-aware social network architecture for mobile crowdsensing," *IEEE Commun. Mag.*, vol. 52, no. 6, pp. 78–87, Jun. 2014.
[239] A. Chakraborty, M. S. Rahman, H. Gupta, and S. R. Das, "SpecSense: Crowdsensing for efficient querying of spectrum occupancy," in *Proc. IEEE INFOCOM*, Atlanta, GA, USA, 2017, pp. 1–9.
[240] C. Fiandrino *et al.*, "CrowdSenSim: A simulation platform for mobile crowdsensing in realistic urban environments," *IEEE Access*, vol. 5, pp. 3490–3503, 2017.
[241] (2010). *Gigwalk*. [Online]. Available: http://www.gigwalk.com/
[242] (2011). *Streetspotr*. [Online]. Available: https://streetspotr.com/
[243] J. Wang *et al.*, "A blockchain based privacy-preserving incentive mechanism in crowdsensing applications," *IEEE Access*, vol. 6, pp. 17545–17556, 2018.
[244] I. Miers, C. Garman, M. Green, and A. D. Rubin, "Zerocoin: Anonymous distributed e-cash from Bitcoin," in *Proc. IEEE Symp. Security Privacy*, Berkeley, CA, USA, 2013, pp. 397–411.
[245] H. N. Loc, Y. Lee, and R. K. Balan, "DeepMon: Mobile GPU-based deep learning framework for continuous vision applications," in *Proc. ACM MobiSys*, Niagara Falls, NY, USA, 2017, pp. 82–95.
[246] M. Hassan, H. Chen, and Y. Liu, "DEARS: A deep learning based elastic and automatic resource scheduling framework for cloud applications," in *Proc. IEEE ISPA*, Melbourne, VIC, Australia, 2018, pp. 541–548.
[247] B. Guo, D. Zhang, D. Zhang, Z. Yu, and A. Chin, "Mobile crowd sensing and computing: When participatory sensing meets participatory social media," *IEEE Commun. Mag.*, vol. 54, no. 2, pp. 131–137, Feb. 2016.
[248] G. Texier and V. Issarny, "Leveraging the power of the crowd and offloading urban IoT networks to extend their lifetime," in *Proc. IEEE LANMAN*, Washington, DC, USA, 2018, pp. 104–109.

Yutong Liu received the B.E. degree in computer science and technology from the Ocean University of China in 2017. She is currently pursuing the Ph.D. degree in computer science with Shanghai Jiao Tong University. Her research interests include wireless communications, artificial intelligence, and mobile computing.

Linghe Kong (S'09–M'13–SM'18) received the B.E. degree in automation from Xidian University in 2005, the master's degree in telecommunication from TELECOM SudParis in 2007, and the Ph.D. degree in computer science from Shanghai Jiao Tong University in 2012. He was a Post-Doctoral Fellow with Columbia University and McGill University. He is currently a Research Professor with Shanghai Jiao Tong University. His research interests include wireless communications, sensor networks, and mobile computing.

Guihai Chen received the B.S. degree from Nanjing University in 1984, the M.E. degree from Southeast University in 1987, and the Ph.D. degree from the University of Hong Kong in 1997. He is a Distinguished Professor with Shanghai Jiao Tong University, China. He had been invited as a Visiting Professor by many universities, including the Kyushu Institute of Technology, Japan, in 1998; the University of Queensland, Australia, in 2000; and Wayne State University, USA, from 2001 to 2003. He has a wide range of research interests with focus on sensor network, peer-to-peer computing, high performance computer architecture, and combinatorics.