A Survey on the Mobile Crowdsensing System life cycle: Task Allocation, Data Collection, and Data Aggregation

Xia Zhuoyue, Azween Abdullah, and S.H. Kok ,
xiazhuoyue@sd.taylors.edu.my azween.abdullah@taylors.edu.my mike.kok@taylors.edu.mycp.edu
School of Computer Science and Engineering, Taylor's University, Malaysia

Summary
The popularization of smart devices and subsequent optimization of their sensing capacity has resulted in a novel mobile crowdsensing (MCS) pattern, which employs smart devices as sensing nodes by recruiting users to develop a sensing network for multiple-task performance. This technique has garnered much scholarly interest in terms of sensing range, cost, and integration. The MCS is prevalent in various fields, including environmental monitoring, noise monitoring, and road monitoring. A complete MCS life cycle entails task allocation, data collection, and data aggregation. Regardless, specific drawbacks remain unresolved in this study despite extensive research on this life cycle. This article mainly summarizes single-task, multi-task allocation, and space-time multi-task allocation at the task allocation stage. Meanwhile, the quality, safety, and efficiency of data collection are discussed at the data collection stage. Edge computing, which provides a novel development idea to derive data from the MCS system, is also highlighted. Furthermore, data aggregation security and quality are summarized at the data aggregation stage. The novel development of multi-modal data aggregation is also outlined following the diversity of data obtained from MCS. Overall, this article summarizes the three aspects of the MCS life cycle, analyzes the issues underlying this study, and offers developmental directions for future scholars' reference.

Keywords:
MCS, Task Allocation, Data Collection, Data Aggregation

1. Introduction

The prominence and rapid development of portable gadgets, wireless networks, and communication technologies have led to the emergence of MCS, a novel sensing paradigm that integrates the crowdsourcing concept with smart device sensing ability under the Internet of Things (IoT) for large-scale and fine-grained sensing and computing tasks. Notably, the rapid growth of crowd-sensing with regards to deployment, maintenance, sensing range and granularity, and reusability proves advantageous in the current era of intelligence compared to traditional wireless sensor networks.

A global data analysis agency, IDC, reported smartphone shipments of approximately 1.31 billion units in 2022 with a projected rise of 5% towards the end of 2023. This gradual increase, which could surpass 1.3 billion annually, implies that each global citizen would own at least one smartphone. Parallelly, wearable smart devices (smart watches, glasses, and bracelets) and wireless communication technology reflect drastic growth. The rapid transition from 4G to 5G technology and high transmission rates also strengthen the smart device-data transmission interconnection.

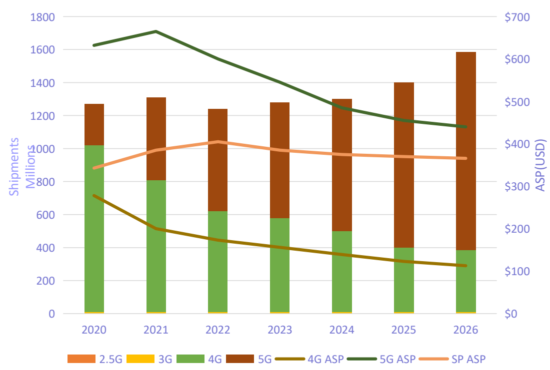

Fig.1: Worldwide Smartphone Forecast, 2022 Q1

The MCS could monitor urban road and environmental noise [61,62, 64] and detect air quality [63] through users' smart devices (as sensing nodes) to develop a sensing network and perform sensing tasks for fine-grained MCS. For example, urban dwellers could utilize MCS to monitor the 5G signal coverage in cities. Such tasks can be distributed to mobile users, who would analyze and ascertain the aforementioned coverage through their feedback data to improve poor-coverage areas and mitigate communication operators' testing costs. Recent works on crowd-sensing algorithms and applications have been published in top computer network conferences and academic journals. Specific MCS applications and platforms, including a Google-owned traffic navigation application that collects GPS information to analyze road traffic flow (Waze), have also been released locally and globally. The system offers drivers in over 190 countries improved driving routes through relevant computation and analyses.

Manuscript received March 5, 2023
Manuscript revised March 20, 2023
https://doi.org/10.22937/IJCSNS.2023.23.3.4

The MCS, which results from the integration of crowdsourcing ideas and the sensing ability of mobile intelligent devices, primarily capitalizes on the extensive usage of mobile devices and accelerated communication transmission speed to gather sensing data through user-owned intelligent devices for large-scale sensing and computing tasks. Compared to conventional sensing methods, this affordable and simplified system is beneficial based on sensing range and granularity in the present era of intelligent sensing.

2. Mobile Crowd-sensing System

Generally, an MCS system entails a sensing platform and platform users presented in Figure 2 [1]. The three-part MCS life cycle encompasses task allocation, data collection, and data aggregation, albeit with specific technical complexities. First, the sensing platform performs task release.

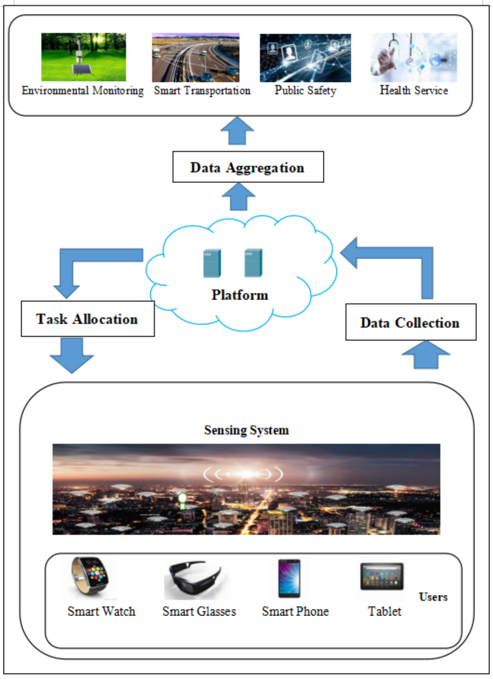

Fig.2: The MCS System

User-released sensing task was classified into several sensing sub-tasks during this process and subsequently released to mobile users. A specific incentive mechanism was concurrently adopted to encourage mobile users' involvement. Appropriate participants (mobile users) were selected to perform each sensing sub-task with sensing equipment for sub-task completion and report the sensing data to the sensing platform, thus compensating for their sensing costs incurred from the adopted incentive mechanism. The derived data were processed and analyzed with vital knowledge mined for environmental monitoring [2-3], intelligent transportation [4], public safety [5-6], and health services [7]. Notably, the overall system would operate reliably and efficiently with a seamless interconnection of all three MCS links. Recent empirical works were then summarized following three aspects: task allocation, data collection, and data aggregation.

3. Task Allocation in MCS

Task allocation is a key MCS issue impacting sensing task quality and costs. Figure 3 illustrates the task allocation process, which selects appropriate users to complete sensing tasks and fulfill relevant quality and budget requirements. Notably, it is vital to determine budget-saving strategies under the premise of meeting the same requirements.

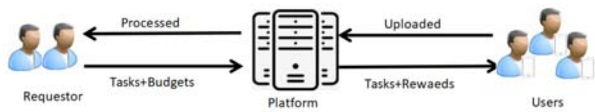

Fig.3: The Illustration of Task Allocation

Numerous studies on task allocation, which is relatively basic at this stage, have been performed through specific stages, with an emphasis on single-task allocation to seek users for individual tasks. Empirical MCS applications to multiple scenarios render single-task allocation insufficient in fulfilling user needs. Thus, multi-task allocation issues arise due to the presence of multiple-sensing tasks following a whole sensing demand. Multiple-task assignation strategies strive to assign tasks optimally using limited resources. Allocation issues with time and space constraints have gained much scholarly attention. As the core MCS component, users with time and space attributes could meaningfully examine multi-task allocation under the aforementioned limitations. Past task allocation studies would be duly summarized based on specific aspects.

3.1 Single Task Allocation

Given the issues underpinning sensing task allocation in MCS, this study aimed to prioritize task coverage based on task allocation, albeit with complexities on how to balance low cost and high coverage. Early task assignation works focused on single-task sensing following its relative novelty at the time.

Past scholars on single-task assignation primarily investigated participant selection methods. Zhang et al. [8] recommended a user selection framework, known as Crowdrecruiter, to select a minimum number of participants for low-cost task performance. The framework estimates and selects the minimum effective set following the prediction outcomes for goal attainment by analyzing historical information. In this vein, user recruitment is a key task allocation challenge in motivating motivate real and efficient user collaborations for optimal performance. From the participants' standpoint, Anjomshoa et al. [9] proposed a novel recruitment policy of social activity awareness (social orientation and high battery efficiency) for MCS based on the essentiality of mobile energy. The socialization and power of smart devices, two key Sober-MCS system indicators, were duly identified. Participants' task performance preparedness is judged by socialization, while battery power considers participant suitability. Both attributes were integrated for suitable worker recruitment. Resultantly, Sober-MCS could be both energy-saving and highly efficient.

Task coverage is extensively researched in single-task allocation. Coverage quality is primarily regarded from two aspects: (i) how to elevate the coverage rate and (ii) achieve the minimum cost when the coverage rate is guaranteed. Potential studies could examine the most optimal quality awareness coverage in the mobile crowd network. The appropriate subset would also be selected from mobile users to enhance traditional coverage concerns to complete the tasks under specific limitations. Experimental outcomes revealed the ability of the method to ensure maximum task coverage quality [10]. For example, Xiong et al.'s [11] CrowdTasker framework served to improve task allocation by first analyzing historical data to obtain users' mobility prediction. Suitable participants were selected in each task execution cycle for task performance parallel to the forecasted outcomes and incentives. The system could successfully elicit the most optimal coverage quality under the premise of satisfying the constraint conditions with this operation.

3.2 Multi-task Allocation

New sensing tasks in the sensing platform frequently require simultaneous performance following the increase in task allocation studies. As such, multi-task MCS allocations from past single-task allocation studies require further examination to adequately assign various tasks under constrained conditions and attain the minimum cost-maximum allocation effect. Participant selection, task coverage, and associated limitations characterize multi-task allocation factors, not unlike single-task allocation. Past single-task allocation scholars primarily selected a suitable participant subset for a single task, which is deemed simple. Regardless, participant selection in the multi-task allocation process proves pivotal for platform efficiency.

Wang et al. [12] recommended a multi-task allocation framework, which simplifies a task for participants who continue performing it, to reasonably assign various tasks within the multi-task allocation scenario for maximum task coverage under specific sharing constraints. For example, participants who are assigned a second task type upon performing the first would be rewarded to encourage more task performance. This framework, which incentivizes users who have performed multiple task types to compensate for their incurrence of additional costs, attains the goal through the greedy algorithm.

Xiao et al. [13] examined the task allocation issues underlying mobile crowdsensing in micro-decentralized neural networks compared to that of traditional task scheduling. Notably, all the tasks were classified uniformly upon arrival as conventional studies failed to categorize task allocation. Different tasks could be appropriately classified and assigned by reducing the average task allocation duration following this research. Task allocation was classified into two categories, online and offline, using the greedy algorithm to address this complexity with optimal outcomes.

The recent and extensively-studied emergence of mobile crowding is typically resolved with the greedy algorithm. This study proposed the Qos-sensitive task allocation of the mobile population with the greedy algorithm, where the number of user-performed tasks varies based on the simultaneous execution of multiple tasks. Correspondingly, users who execute more tasks would obtain more rewards. The same task could be assigned to multiple users for performance and optimal execution results [14].

Task execution efficiency would decrease in multi-task performance if the participants' situation is disregarded. In this study, their skills motivation, requirements, and sensing time were duly regarded for efficient task allocation under the multi-task correlation condition following past literature. Song et al. [15] recommended a multi-task-oriented participant selection strategy (DPS) to address the aforementioned challenges with three key design elements. The proposed strategy could achieve improved QoI satisfaction in all tasks through the QoI satisfaction measurement, optimization of participant selection through multi-task oriented QoI, and computation of the anticipated participant-gathered data. Overall, multi-task allocation under time and space constraints is also a core research component apart from the aforementioned empirical factors.

3.3 Space-time Constrained Task Allocation

Several researchers began focusing on multi-task allocation with limited resources upon realizing their essentiality. Particularly, Guo et al. [16] examined the issue underpinning multi-task allocation under time constraints and recommended ActiveCrowd, a participant selection framework. In terms of workers' choice, the tasks were first classified and then categorized into time-sensitive and time-insensitive tasks. It is deemed necessary to integrate workers' wishes in the task allocation process in ascertaining their willingness to move. Hence, these individuals must be selected based on their (i) preparedness to move and pursue the minimum moving distance and (ii) ability to rapidly complete time-sensitive tasks. Meanwhile, workers who need not move long distances for task completion should perform time-insensitive tasks. The study also examined Liu et al.'s [17] participant selection framework, TaskMe, based on two aspects: (i) FPMT (fewer participants and more tasks) problem with insufficient resources and (ii) MPFT (more participants and fewer tasks) problem with abundant resources. An optimal solution is attainable through the multi-TaskMe framework under time constraints, where participants with the highest number of completed tasks and smallest total travel distance or counterparts with the smallest total reward payment and travel distance.

Most scholars regarded the overall distribution effect to determine the implications of multitasking. The published individual tasks may not have been effectively executed for every task publisher following the distribution outcome despite minimizing or fulfilling the overall execution time or constrained condition of overall sharing, respectively. Wang et al. [18] proposed the MTASKER framework to define a threshold that measures effective task execution, in which only tasks that exceed the threshold is deemed effective. Specifically, the system would measure tasks with the time-space coverage. The task is considered a low-quality execution that does not elevate the overall task execution quality with a time-space coverage under the threshold. This framework could effectively identify suitable tasks for workers and ensure optimal task performance with the aforementioned methods.

Extensive research on time constraints highlighted space as an equally important limitation. This study proposed the generation of spatio-temporal coverage measure (based on coverage) to examine space-time constraints, not unlike Xiong et al.'s [19] study which first addressed the issues underpinning the measurement degree. Essentially, the recommended coverage must be thoroughly analyzed in relation to the (i) proportion of sub-areas that could be covered and (ii) sensor readings collected in each measured area, unlike its simplified counterpart. This study proposed a task allocation framework, known as iCrowd, to resolve the measurement issues and perform task allocation with a fixed budget to attain two optimization goals: (ii) minimum cost and (ii) maximum coverage. Notably, both objectives could be simultaneously attained.

A large number of vehicles is needed to collect multi-data types and complete the sensing task following the heterogeneity of vehicles and their multiple sensing interfaces for MCS application to vehicles under the premise of meeting the space-time coverage. Specifically, Liu et al. [20] recommended a heterogeneous sensor vehicle selection (HVS) method to address multivariate data collection issues. This study proposed using the utility function for computation purposes to determine the vehicle sensing ability. Factors involving vehicle division and sensing interface and coverage range were duly regarded in the estimation process. Meanwhile, a greedy algorithm was incorporated to resolve this knapsack issue post-comprehensive consideration.

An in-depth analysis of MCS led to the empirical focus on participants' positions, which significantly impact task assignation, and space-oriented studies. Cheng et al.'s [21] work on space-based task allocation under a similar scenario regarded both participants and space. In this vein, each participant possesses distinctive skills, with each space task being time-constrained. This study selected the most appropriate participants for task performance by considering the matching degree of their skills with the space task. The aforementioned method potentially enables these participants to obtain maximum profit upon fulfilling the constraint conditions. Furthermore, Deng et al. [22] investigated the worker choice task (WST) model, a task type where workers autonomously perform self-selections. The system provides a worker and group with multiple tasks. Each task is allotted its own position and time constraint. Empirically, two algorithms based on dynamic programming and branch-and-bound strategy were incorporated for task assignation and catalyzed each worker to perform multiple tasks and gain maximum profit.

Table 1: Recent Task Allocation Research in MCS

Ref.	***Problem/Purpose of the article***	***Method/Technique/Algorithm***	***Result/Features/Measurement***	***Future Direction/Limitation***
[8]	Minimize incentive payment by selecting a small number of participants while concurrently satisfying the probability coverage constraint.	Crowdrecruiter	Under the same coverage constraints, the proposed solution is significantly better than the three baseline algorithms by selecting 10.0%-73.5% fewer participants on average.	Redundant Cell Tower Coverage; Call/Mobility and Coverage Prediction; Sensing Coverage and Privacy; Leveraging Multiple Piggyback Sensing Opportunities; Different Incentive Payment Models; Using Real Sensing Datasets.
[9]	User recruitment using sociability and the residual power of participants' smartphones	Sober-MCS	Introduce battery savings of up to 18.5% while improving user and platform utilities by 12% and 20%, respectively.	Introducing mobility prediction and GPS-less sensing to SOBER-MCS.

[10]	Investigating the optimal quality-aware coverage for maximum coverage quality with a constrained budget.	Approximation Algorithm for Maximum Coverage Quality with Budget Constraint Problem	Resultantly, ours scheme outperformed the random selection scheme and one of the state-of-the-art in terms of total coverage quality by 2.4× and 1.5× (at most) and 1.4× and 1.3× (on average).	Strategies to leverage the temporal-spatial properties of both the mobile users and POIs to improve the coverage of MCS networks.
[11]	maximize the coverage quality of the sensing task while satisfying the incentive budget constraint	CrowdTasker	CrowdTasker significantly outperformed three baseline approaches by achieving 3%-60% higher coverage quality.	Using MaxUtils for Utility1 calculation; call/mobility prediction and privacy; coverage quality metrics and incentive models; leveraging multiple piggyback opportunities.
[12]	Allocating multiple tasks to participants and maximizing the overall data quality under the shared budget constraint	MTPS	This approach outperformed the baseline methods.	1) More complex constraints and models: 2) privacy-preserving mechanisms: 3) heterogeneity of multiple tasks.
[13]	Solving task allocation problems in MSNs	FTA/NTA	FTA is the optimal offline task allocation algorithm while NTA can achieve a smaller average makespan in the online decision case.	Nil
[14]	QoS-sensitive Task Assignment (QSTA) problem	QSTA Algorithm	Proposing a greedy approximation algorithm to solve this problem and proving the approximation ratio of the proposed QSTA algorithm.	This solution could support the Euclidean distance and be extended to support more complex Qos models.
[15]	Selecting the most appropriate participants to satisfy the quality-of-information (QoI) requirements of multiple concurrent tasks with different budget constraints.	DPS	Given the same budget, the proposed participant selection strategy can achieve higher QoI satisfaction for all tasks than selecting participants randomly or through reversed-auction-based approaches.	Extend the model to balance the incentive cost and the gain of QoI satisfaction; explore the field of selecting the most energy-efficient participants.
[16]	The problem of multi-task worker selection	ActiveCrowd 、GGA-I 、GGA-U	The proposed algorithms outperformed baseline methods under different experiment settings (scale of task sets, available workers, varied task distributions, etc.).	Other factors that may affect worker selection in multi-task MCS environments.
[17]	Based on a different number of participants and tasks to achieve different optimization goals.	TaskMe	The proposed algorithms of two problems outperformed the baseline approaches.	Diversity of participants; Various constraints; Incentive models; Methods for task allocation; Large-scale user study.
[18]	Maximizing the overall system utility by considering the completion quality of a single task.	MTasker	MTasker is superior to Naive-AG, Ru-AG, etc.	Various Factors and Constraints; Sensing Quality Metrics; Mobility Prediction; Task Dependency; Sub-areas; Privacy Concerns; Achievable Sensing Quality; Large-scale User Study.
[19]	Goal 1: near-maximal k-depth coverage without exceeding a given incentive budget or Goal 2: near-minimal incentive payment while meeting a pre-defined k-depth coverage goal	iCrowd	Goal 1 achieved 3%–60% higher k-depth coverage compared to baseline approaches under the same budget constraint. For Goal 2, iCrowd required 10.0% − 73.5% lesser in terms of overall incentive compared to baselines under the same k-depth coverage constraint.	Nil
[20]	Strategies to continuously collect comprehensive tempo-spatial sensing data with a limited number of heterogeneous sensing vehicles	HVS	The HVS algorithm could collect sensing data with a higher coverage ratio in a more uniform and continuous manner than current mobile crowd-sensing methods.	Nil

[21]	Multi-skill spatial crowdsourcing (MS-SC) to identify an optimal worker and-task assignment strategy	The MS-SC Greedy Algorithm	MS-SC algorithm is effective and efficient on both real and synthetic datasets.	MS-SC problem is also constrained by budget, time, and distance.
[22]	This study aims to identify a schedule for the worker that maximizes the number of performed tasks.	WST	Experimental evaluation on both real-world and synthetic data to verify the performance and accuracy of proposed approaches.	Developing algorithms to optimize dual criteria; considering other properties of spatial tasks; extending MTS to the SAT mode while addressing privacy issues.

Summarily, research on task allocation stemmed from single- to multi-task allocation and multi-task allocation to multi-task allocation under time and space constraints. Specific issues, such as the complexity of time constraints were also highlighted in these studies. Although users' distance and time are typically considered, time constraint does not merely imply users' time but include their task performance duration, idle time of users, and different user efficiency at different times. Thus, time constraint-oriented research could be further subdivided. In studies involving task allocation with time and space constraints, the order of both these constraints potentially affects the final multi-task allocation outcomes. Furthermore, the multi-task correlations remain in the preliminary stage. Two related tasks that are assigned to the same user would significantly save the task execution time and cost.

4. Data Collection

As the core component of MCS, data collection is vital to obtain accurate data sensing in the future. Past studies emphasized data collection quality and security. Regardless, pressure on the MCS system would continue mounting with the increase in MCS workload. Thus, data collection efficiency is also a key research point. This section elaborates on the data collection quality, security, and efficiency.

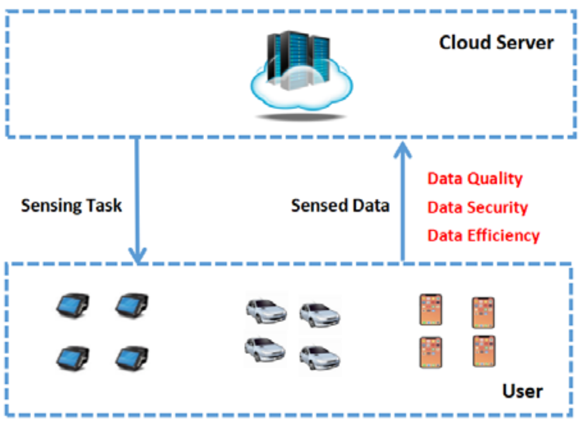

Fig.4: The Illustration of Data Collection

4.1 Data Collection Quality

Essentially, MCS employs a large number of user-owned mobile devices to gather sensing data with the following benefits: flexible topology, broad sensing-area coverage, and minimal deployment cost. Regardless, a large amount of redundant or low-quality data may exist in the gathered information as sensing users reflect personal knowledge, skills, behavior, habits, and degree of responsibility without professional training. Regarding the platform, problematic data could negatively impact subsequent data processing and analysis and lead to inaccurate data analysis outcomes or deviation from accurate results. In this vein, sensing data selection has garnered much scholarly interest. Meanwhile, data screening is classified into two sections: (i) screening users who gathered the data and (ii) screening the collected data.

In screening data-collecting users, suitable user selection for data-gathering purposes would guarantee the derived data quality. In other words, a user forms the basis of data collection. Past scholars employed an incentive mechanism to choose appropriate data-collecting users for optimal data quality. Peng et al. [23], who integrated users' sensing data quality with the incentive mechanism, made different payments based on their performance. Specifically, additional incentives were offered to high-performing participants as a source of motivation to gather high-quality data. Explicit evaluations of participant data quality would prove ineffective without any feedback. Consequently, Yang et al. [24] incorporated an incentive mechanism based on quality assessment for participants to guarantee their data quality and increase their user reputation. Problematic user data would be filtered and removed upon analyzing users' long-term data quality and reputation. Meanwhile, Gong et al. [25] integrated quality, effort, and data acquisition to design an actual crowdsourcing mechanism for participants to truthfully report their work quality and data to the task sender and complete data collection as requested. A linear programming and participant selection method was employed by Yu et al. [26] to avoid selecting low-performing participants who do not attain the required sensing task quality. Although both methods could effectively resolve small-scale issues, the greedy approach reflected optimal performance when the problem becomes more complex.

Sensing data requires screening for optimal data selection, processing, and analysis. Following past literature, majority voting is a common truth discovery algorithm [27] (albeit with some drawbacks) and highly competent in the presence of several low-quality users. Regardless its outcome may prove problematic when the number of low-quality users exceeds that of their high-quality counterparts. Overall, relevant scholars have significantly improved on this basis.

As a participant denotes the core MCS component, data screening based on truth value involves user screening. Yin et al. [28] proposed a truth discovery method following the expectation maximization strategy to address contradictory facts from multiple websites. Truth value and user reliability were employed as parameters, whereas the website credibility-factual evidence interdependence served to identify reliable websites. Li et al. [29] regarded the long tail phenomenon by modeling user reliability with the user probability model. A small number of experienced or highly-efficient users in data collection domains could characterize high and efficient data gathering. This small-scale database could contribute valid data, known as the long tail phenomenon. This study recommended a confidence-sensing truth discovery method, which could automatically discover truth values from conflicting data to resolve this intricacy. Specifically, this technique could distinguish the sources and evaluate reliable sources through analysis.

Data screening encompasses more MCS applications. For example, Uddin et al. [30] proposed a resource-constrained image transmission service, “PhotoNet”, which determines content priority by computing semantic similarity. In this vein, insufficient resources were first allocated to deliver the most “worthy” content and maximize the event coverage. Regarding mobile crowd photography (MCP), previously-delivered pictures were associated with pictures that were delivered later, which resulted in redundancy. Guo et al. [31] proposed Picpick, a general data collection framework, to actualize diversified MCP task specifications with the multidimensional task model. Meanwhile, the pyramid tree method functions to select the best picture set and effectively minimize redundancy. Meanwhile, Cheng et al. [32] recommended a DECO detection framework as MCS allows the participation of different user groups that possibly instigates poor data quality. As traditional error detection frameworks essentially require data integrity, it is deemed challenging to complete the detection amidst missing data. The DECO framework identifies the error value in the presence of missing data and rectifies the missing and erroneous values post-identification. Despite the extensive utilization of professional equipment or mobile professionals to monitor urban pollution, the incurred cost and available coverage are high and limited, respectively. Zappatore et al. [33] suggested data identification based on contextual content to gather reliable data despite the emergence of MCS, as non-professionals’ data collection would instigate inaccurate or low-quality data.

4.2 Data Collection Security Issues

Given the absence of a strict audit mechanism for participant selection in MCS, task distribution is typically accompanied by an incentive mechanism. Some participants fabricate or provide false data for the platform to obtain more rewards, while others may even feed malicious data to attack the platform following its large-scale collection of professional data. Relevant researchers have extensively discussed this shortcoming in multiple disciplines. Notably, this study aimed to design a reputation algorithm with high computational efficiency by (i) integrating the semi-matching concept with workers’ punishment based on label divergence and (ii) assigning a reputation score to each worker in detecting and filtering antagonistic counterparts. The current work paralleled Jagabathula et al.’s [34] examination of inferring the potential real labels of binary tasks by aggregating noisy labels. Meanwhile, Miao et al. [35] primarily examined the attack behavior in the MCS system with the truth discovery mechanism. This study examined two attack types, namely usability and target, following a high-quality attack framework, where attackers could optimize the attack utility. Intriguingly, this framework could enable attackers to disguise other counterparts as normal users to render the system challenging to identify. The system functions to safeguard data confidentiality, as MCS may encounter privacy issues during data collection, albeit with the risk of protecting malicious data as well. For example, Zhao et al.’s [36] examination of data quality recommended a set of data protection mechanisms by integrating the algorithm mechanism, game theory, and truth discovery. High-quality sensing data could be collected to protect users’ privacy information.

In line with pertinent studies, some malicious people may collaboratively submit false information: a conspiracy attack as an individual is easily detected by the system. Malicious users in a distributed system may possess multiple identities to launch attacks, such as Sybil. Consequently, Yu et al. [37] proposed a new system limit protocol that was optimized and upgraded based on the guard system [38]. The superiority of the algorithm was verified in a million-node system experiment. With regards to the Sybil attack, each IP could be limited to only one ID in resolving this problem. Feng et al. [39] proposed BMCA, a defense scheme, to suppress collaborative attacks with binary minimum clustering analysis for high resistance against collusion attacks. A new binary minimum clustering algorithm was used in this scheme to detect, evaluate, and mitigate collusion attacks, complete trust, and attackers′ trust values, respectively.

4.3 Data Collection Efficiency

Following past works, data collection efficiency is a key indicator of safety and reliability that brings novel MCS architecture through edge computing. As a network technology that positions the server near the device, edge computing reduces the system processing load and resolves data transmission delays. Edge computing could facilitate the system to significantly improve operating efficiency. By applying edge computing to the MCS system framework, some processing work could be completed in edge nodes to optimize the MCS system efficiency. Roy et al. [40] proposed a data collection framework based on fog computing, bioMCS, for effective energy and resource management. In this framework, fog nodes function as relay nodes to perform effective task sensing and forwarding. Data quality is also ensured by receiving the task data from reliable and intelligent devices. Although BioMCS demonstrates good performance in data delivery, delay, and energy efficiency, the need to collect large-scale data placed much pressure on base stations and servers following the MCS task expansion. Mobile device communication that is inspired by D2D technology could occur with the

extreme privacy protection potentially reduces data availability. Hence, the means to balance data quality and privacy protection is a challenging issue in receipt collection.

facilitation of nearby base stations. Wang et al. [41] recommended a reliable data collection (RDC) algorithm to authenticate the device-generated sensor data. The algorithm, which obtained reliable data through reliable user verification on the client's side, was uploaded, thus saving data collection costs.

Edge computing is widely acknowledged for enhancing data collection efficiency. Specifically, Xia et al. [42] suggested a novel data collection architecture with mobile edge computing (MEC) to significantly impact the sensing of frequent user movement. Data correlation in edge nodes minimized redundant data and selected suitable users. Meanwhile, the compressed sensing technology in the cloud served to recover high-quality data. Notably, the frame performance index was high post-verification.

The current research emphasized data privacy, security, and data in terms of data collection (see Table 2). Nevertheless, the complexity of sensing tasks inevitably increases sensing data. Data collection for large-scale data would significantly hamper the system efficiency. The advent of edge computing offers novel solutions to this problem. Although data quality and privacy protection are two key data collection issues,

Table 2: Recent Data Collection Research in MCS

Ref.	*Problem/Purpose of the article*	*Method/Technique/ Algorithm*	*Result/Features/Measurement*	*Future Direction/Limitation*
[23]	Adding data quality to the incentive mechanism to solve the problem of collecting low-quality data.	Quality-based incentive mechanism	This mechanism achieved superior performance compared to the uniform pricing scheme.	Nil
[24]	Combining quality estimation with financial incentives to ensure the high quality of collected data.	Quality Estimation Algorithm; Distance-Based Outlier Detection Algorithm; Shapley Value Approximation Algorithm	This approach achieved good performance in terms of quality estimation and surplus sharing.	The proposed system is a general framework for subsequent quality-aware crowdsensing designs.
[25]	Incentivizing strategic workers to truthfully report their private worker quality and data to the requester and maintain their integrity as requested.	QEDE	QEDE mechanisms resolved the lack of ground truth and coupling in the joint elicitation of worker quality, effort, and data.	Consider workers who do not know their quality
[26]	Aiming to select a minimum number of participants to achieve the quality required by a sensing task.	A "t-sweep k-coverage" data quality model; a participant selection algorithm based on linear programming; a participant selection algorithm based on greedy strategy.	For small problems, the results of both methods met the requirements. When the problem became more complex, the method based on greed revealed better results.	Choose participants based on the probability of their appearing in a certain location.
[27]	Accurate identification of a malaria patient from the given set of samples and	An ensemble method called voting of C4.5, Naive Bayesian, and KNN classifiers.	The voting ensemble method revealed a classification accuracy of 95.2% on imbalanced malaria	Nil

	classification when dealing with incongruent datasets.		disease data, whereas dealing with balanced malaria disease data voting ensemble denoted 92.1% of accuracy	
[28]	Ways to identify true facts from a large amount of conflicting information on the multiple subjects provided by different websites.	TruthFinder	TruthFinder successfully identified true facts among contradictory information and identified trustworthy websites better than popular search engines.	Nil
[29]	Solving the ubiquitous long-tail phenomenon in tasks.	CATD	The proposed method outperformed advanced truth discovery approaches by successfully discounting the effect of small sources	Nil
[30]	A significant overlap between their content could be identified when pictures are generated and serviced by a mission-driven network.	PhotoNet	Suppressed redundant content belonging to the same event and effectively improved resource rate usage.	Explore the limits of the approach usability and generalize it to heterogeneous content, more complex application goals, and multiple concurrent applications.
[31]	Eliminating the redundancy of the collected picture data and reducing network overhead.	PicPick	PTree could effectively reduce data redundancy while maintaining the coverage requests, with the overall framework remaining flexible.	Extend task model to address evolvable and dynamic constraint settings; improve PTree with branch pruning techniques and quality-oriented data selection strategies; investigate opportunistic collaboration of participants to achieve local data selection and optimized transmission.
[32]	False data detection and correction in crowd-sensing.	DECO	Detected false data and estimated both false and missing values for data correction	Nil
[33]	Exploiting context awareness to improve the reliability of MCS-collected data.	Machine learning; a computational solution working on the most usual and recurrent scenarios to monitor urban noise pollution; a transductive learning algorithm	The method, which was verified by some noise pollution, could improve the collected user data credibility.	Explore solutions where urban computing leverages intelligent systems to maximize the exploitation of crowdsourced data.
[34]	The problem of aggregating noisy labels from crowd workers to infer the underlying true labels of binary tasks.	Reputation algorithms	The reputation algorithm could significantly improve the accuracy of current label aggregation algorithms in real-world crowdsourcing datasets	Nil
[35]	How to effectively conduct two types of data poisoning attacks	Truth discovery framework; optimal attack framework	Compared to the native baseline schemes, the proposed attack framework could achieve higher attack utility and inadvertently allow malicious workers to gain higher reliability degrees that hamper easy detection.	To develop defense methods

[36]	Privacy protection in MCS and untrue reports lead to data quality problems.	Quality-driven sensing duty allocation algorithm; truth discovery & data quality evaluation algorithm	This method could ensure data quality without affecting user privacy.	Nil
[37]	Against Sybil attacks.	SybilLimit	The results on real-world social networks confirmed their rapid mix of property, thus validating the fundamental assumption underlying SybilLimit's (and SybilGuard's) approach	To implement SybilLimit within the context of specific real-world applications and demonstrate its utility
[38]	Against Sybil attacks.	SybilGuard	SybilGuard ensured that (i) the number and size of Sybil groups were properly bounded for 99.8% of the honest users and (ii) an honest node could accept and be accepted by 99.8% of other honest nodes.	Obtain real social network data to further validate SybilGuard.
[39]	Fighting against collusive CSDF attack.	BMCA	The BMCA scheme could enhance the accuracy of trust evaluation and successfully reduce the power of collusive CSDF attacks against data collection in MCS	Extensively study a defense scheme from the design idea of distinctive trust evaluation for different tasks to suppress such threats.
[40]	To achieve seamless communication and effective energy and resource management in MCS.	bioMCS 2.0	Experimented on the New York City map to demonstrate the efficacy of bioMCS 2.0 across different realistic mobility models for steady data delivery, latency, and energy efficiency compared to a standard shortest-path-based centralized data forwarding technique.	To explore effective thresholding methods that could distinguish between (i) harmless and malicious user behavior and human error factors and (ii) technical hardware and software glitches on smart devices that users employ to report task data.
[41]	High cost and low data quality occurred during large-scale data collection.	RDC algorithm	The RDC algorithm outperformed the compared benchmarks with at least a 24.6% improvement in estimating accuracy and 13.1% in saving data collection costs.	Nil
[42]	Data redundancy and poor data quality.	Quality-aware sparse data collection (QSDC) algorithm	The QSDC, which performed efficiently to reduce the amount of sampling data, could rapidly converge to cover the sampling matrix.	Nil

5. Data Aggregation in MCS

The inextricable link between data aggregation stage and data collection in the MCS life cycle implies that both stages would encounter the same problems. Consequently, the data aggregation stage was executed from two standpoints: security and quality.

Fig.5: The Illustration of Data Aggregation

5.1 Data Aggregation Security

User privacy protection was proposed in the data aggregation stage as early sensing data works typically encompass users' sensitive information sources. Data perturbation and encryption denote two of the most common privacy protection types.

5.1.1 Privacy Protection using Data Perturbation Method

The data perturbation method ensures data aggregation security by artificially incorporating random noise into the data pre-data submission to realize privacy protection. Jin et al. [43] proposed INCEPTION, a novel MCS system framework, with the incentive mechanism to skillfully incorporate the incentive mechanism (to compensate for the cost of users' privacy leakage), data aggregation, and data disturbance mechanism (to ensure workers′ satisfaction with the privacy protection). Such attributes ensure an optimal final data aggregation outcome. For example, Zhang et al. [44] developed an incentive mechanism (REAP) to compensate for users′ loss of privacy. The aforementioned study did not assume the data platform to be reliable, unlike the previous mechanism, and would allow users to incorporate corrected noise pre-data submission. This framework adopts the integration of differential privacy with an incentive mechanism to customize different contracts for users with different privacy levels. Thus, users are compelled to lower their privacy level with the incentive mechanism for a good data aggregation effect under the budget constraint.

Past works employed the incentive mechanism and differential privacy, which implies more application scenarios in secure aggregation, to perform secure data aggregation. Yang et al. [45] suggested an auction framework, in which the platform functions as an auctioneer for private data aggregation. The differential privacy data aggregation algorithm was used in this framework to clarify the relationship between workers' data privacy, their added noise, and the total noise in the summary outcomes. Appropriate workers were then chosen for task completion and optimal outcomes. In realizing the real incentive mechanism and effective privacy protection, this study enhanced the two-stage auction algorithm based on trust and privacy by integrating anonymous privacy protection with differential privacy protection and proposed k-ε- differential privacy protection for effective user location privacy protection. Following past research, Wang et al. [46] recommended a real incentive mechanism based on location privacy protection for MCS. Zhang et al. [47] subsequently developed an auction framework for privacy-preserving data aggregation in MCS, which functions as an auctioneer and recruits workers. The workers in this study were allowed to add noise when reporting data to ensure their privacy protection. As the platform control over aggregation potentially declines upon devolving privacy control rights from the platform to the workers, the system selected workers based on their privacy preferences for optimal privacy data aggregation and a better aggregation effect.

Protecting user location privacy is a key aspect of user data privacy in MCS. Most scholars failed to consider users' different protection needs and provide personalized protection despite much research on user location privacy protection. Consequently, Wang et al. [48] recommended a location privacy protection framework to effectively protect users' location privacy and prevent the exposure of their real location through fuzzy examples and personal privacy levels.

Aggregation algorithm-oriented studies, which previously depended on a centralized framework where users uploaded sensor data to a platform, instigated centralized user data leakage. Hence, Wang et al. [49] a decentralized data aggregation algorithm to fulfill different users′ privacy protection needs. This method could effectively protect user privacy through random noise and realize the balance between differential privacy protection and data accuracy with the distributed average algorithm.

5.1.2 Privacy Protection using Data Encryption Method

Research on the encryption algorithm to protect users' data aggregation privacy highlights homomorphic cryptography and function-based encryption as two common methods. This scheme, which depends on homomorphic cryptography, realizes users' sensitive data protection through data encryption. In many cases, the final aggregation outcome requires user-platform collaborations for computation purposes. Undeniably, the encryption method is more convenient than the privacy protection method with noise. Li et al. [50] proposed an effective summation and aggregation protocol that addressed how the untrusted aggregator in MCS could obtain user data without compromising their privacy.

Fog computing IoT proves beneficial in terms of low latency and location awareness with real-time MCS services. A lightweight privacy-preserving data aggregation scheme that incorporated data aggregation from the mixture network into the same device was proposed by Lu et al. [51]. Furthermore, homomorphic particle encryption, China remainder theorem, and a one-way hash chain technology filtered the false data inserted at the network edge in the early stage for data aggregation privacy. Despite being a novel solution to address the drawbacks identified by the traditional cloud computing architecture, mobile edge computing failed to meet the following requirements: low data latency, fast data access, and other services. Hence, Li et al. [52] incorporated edge computing into conventional privacy data aggregation methods to structure a new data aggregation model involving terminal equipment, edge server, and cloud service. The data aggregation process is completed through their cooperation. Terminal and cloud devices are accountable for data encryption and decryption, respectively. Meanwhile, the edge server functions as a relay node. Data privacy could be effectively protected through this framework, with the significant enhancement of the running model efficiency following the distributed processing mode.

Truth discovery functions, a common MCS method, identify meaningful information through data inferences and computations. The platform does not directly access user data to safeguard private information. Users' sensing information and user-completed task division prove necessary in sparse scenarios. A privacy-preserving truth discovery scheme with an additive homomorphic cryptographic system and additive secret-sharing was structured by Liu et al. [53] with two non-merged servers for high privacy protection requirements under low computation and communication.

The homomorphic encryption scheme is prevalent in data encryption domains to prevent servers from accessing data, particularly in outsourcing scenarios. Contrarily, function encryption enables the server to restore user-controlled information through secret key delegation. Kim et al. [54] developed a function to hide the application of inner product function encryption in biometric authentication and the nearest neighbor search of encrypted data following the lack of function encryption theory applications. Meanwhile, Ryffel et al. [55] realized the quadratic function encryption application to perform privacy-preserving machine learning.

5.2 Data Aggregation Quality

Data aggregation aims to obtain high-quality services, where users could upload unreliable or false data for privacy protection, their own ability limitation, or other reasons, which impact the data aggregation outcomes. Such complexities complicate the attainment of high-quality aggregation results. As such, relevant scholars emphasized the aggregated data quality by recommending a series of schemes.

The mechanism is bound to hide some real information and negatively affect the data quality despite the extensive utilization of the incentive mechanism for data privacy protection. Zhao et al. [56] proposed a PACE incentive scheme for improved data quality and privacy. Alternatively, this research aimed to measure data quality while ensuring privacy protection by comparing reliable data with the actual counterpart.

Several studies considered the cost incurred in ensuring data aggregation quality. A new LightLRFMS based on DRMF was proposed by Li et al. [57] for accurate and prompt false data detection in MCS. Based on the direct robust matrix decomposition, the current separation algorithm requires iteration with a large-scale data matrix, which results in extremely high cumulative computational costs. The current work proposed an intelligent false matrix separation algorithm with a lightweight and low ranking to re-use past matrix decomposition outcomes by observing the fast false data position. Thus, low computation costs significantly improve the algorithm efficiency.

Traditional data aggregation models that require data collection at one point instigate excessive load at the aggregation point and data delay. Unreliable servers and nodes in the fog-assisted data aggregation process would undermine the aggregation outcome accuracy despite the application of fog-assisted moving sensation, which optimally resolves this problem. Consequently, Yan et al. [58] suggested a fog-assisted data fusion scheme, which realizes double privacy protection that protects both user and aggregation outcome privacy and tolerates multiple issues without impacting the aggregation outcomes.

Mobile sensing devices could gather multiple data types by further upgrading mobile sensing devices. This collection necessitates higher data aggregation requirements as the aggregation of only one-dimensional data fails to meet the necessary prerequisites. Thus, multimodal data fusion was incorporated into multiple domains following current research.

Zhang et al. [59] proposed a multimodal learning method for facial expression recognition (FER) with face-to-face image texture and landmark mode to learn joint representation. Meanwhile, the structure regularization method improved the algorithm performance. Yang et al.'s [60] study on video recommendation in the video recommendation field proposed an online video recommendation system, where multi-channel fusion technology and relevant feedback achieved optimal recommendation effects. The system represents the video document as a multimodal combination of text, visual, and audio. This system utilizes the different relevance weights of three modes between two videos and automatically adjusts the weight of each mode with users' click feedback to make video recommendations. Notably, good video recommendation is attainable in place of user data. Individual moods could be gauged by analyzing user files on social media. Pang et al. [61] developed a joint density model in multi-modal input space with the Depth Bolz Mann machine (DBM). The model, which explicitly uses user-generated content data for training without any marking, reflects no strict regulations on data type inputs and outputs. As such, this model applies to the integration of multiple data types.

Table 3: Recent Data Aggregation Research in MCS

Ref.	***Problem/Purpose of the article***	***Method/Technique/Algorithm***	***Result/Features/Measurement***	***Future Direction/Limitation***
[43]	Selecting participants who are more likely to provide reliable data through an incentive mechanism to generate highly-accurate aggregated results.	INCEPTION	The algorithm could select workers who provide reliable data, compensate the cost of workers' privacy disclosure, and obtain accurate data aggregation results. User privacy was well-protected with the disturbance mechanism.	Nil
[44]	This paper aims to quantify privacy compensation for continuous data sensing while allowing FC to directly control PUs.	REAP	REAP realized accurate data aggregation and effective privacy protection.	Nil
[45]	Privacy-preserving data aggregation in MCS.	Differentially Private Data Aggregation Algorithm; Differentially Private Data Auction: Winner Determination Algorithm; Differentially Private Data Auction: Payment Determination Algorithm	The designed algorithm met the requirements of data aggregation accuracy and incurred the lowest sensing data purchase cost.	Worker privacy is still not fully guaranteed. Employees have more privacy protection rights.
[46]	To improve the efficiency and truthfulness of mobile crowdsourcing systems.	TATP; The k-ε-differential privacy-preserving	The proposed incentive mechanism with location privacy-preserving could inspire users to participate in sensing tasks, and protect users' location privacy effectively.	Strategies to improve the task assignment and auction algorithm.
[47]	Develop an auction framework for privacy-preserving data aggregation.	An auction framework for privacy-preserving data aggregation; DPDA algorithm; EDPDA algorithm	A subset of workers was selected to (nearly) minimize the cost of purchasing their private sensing data subject to the accuracy requirement of the aggregated result.	Nil
[48]	To the best of the author's knowledge, none of the existing privacy-preserving task allocation mechanisms could provide personalized location protection considering workers' different protection demands.	PWSM	The proposed framework provided personalized privacy protection and satisfied the truthfulness, profitability, and probabilistic individual rationality.	Nil
[49]	Implement a decentralized data aggregation framework, and realize user privacy protection in it.	Decentralized data aggregation algorithm	Extensive simulations and real-world tests demonstrated the algorithm effectiveness.	Nil
[50]	Strategies for an untrusted aggregator in mobile sensing to periodically obtain desired statistics over the data contributed by multiple mobile users	A new privacy-preserving protocol	This scheme demonstrated a much lower communication overhead than existing works.	Nil

	without compromising user privacy.			
[51]	Hybrid IoT devices' data aggregation problem.	LPDA	LPDA was deemed highly secure and privacy-enhanced with differential privacy techniques; LPDA proved to be lightweight in fog computing-enhanced IoT.	Evaluate the proposed scheme in some realistic IoT scenarios, consider a stronger adversarial model, and design new solutions under a novel model.
[52]	The cloud computing traditional architecture could not satisfy specific requirements, such as low latency and fast data access for IoT applications.	Privacy-preserving data aggregation for mobile edge computing assisted IoT applications.	This scheme guaranteed the data privacy of terminal devices with source authentication and integrity and save half of the communication cost compared to the traditional model.	Evaluate this scheme in the real IoT environment; research on security and privacy of mobile edge computing; lightweight authentication mechanism across trust domains.
[53]	Privacy protection truth discovery scheme in sparse data scenes.	A privacy-preserving truth discovery scheme targeted for privacy-preserving problems in sparse data scenarios; CATD,	The proposed scheme could satisfy strong privacy-preserving requirements with low computation and communication overhead.	Nil
[54]	Function-hiding inner product encryption	A fully secure, secret-key, function-hiding inner product encryption scheme	Using the construction, encryption, and decryption operations for vectors of length 50 complete in the 10^{th} of a second in a standard desktop environment.	The possibility of constructing function-hiding inner-product encryption with equally short secret keys and ciphertexts from a concrete assumption in bilinear groups.
[55]	Machine learning on encrypted data.	A new FE scheme; an adversarial training technique	Reflected little reduction in the model's accuracy and significantly improved data privacy.	The provision of privacy-preserving methods for all features, excluding public ones that are hard to identify in advance remains an open problem.
[56]	Resolve the privacy protection and sensing data quality problems of participants in the incentive mechanism.	PACE	PACE could evaluate data reliability while protecting privacy and preventing attacks initiated by dishonest task requesters.	Achieve privacy-preserving quality quantification via secure computation outsourcing.
[57]	Quick and accurate false data detection	LightLRFMS	LightLRFMS can realize fast and low-cost false data detection.	Nil
[58]	Obtain reliable data aggregation results in FA-MCS and ensure user and data privacy.	A verifiable, reliable, and privacy-preserving data aggregation scheme for FA-MCS	The proposed scheme preserved both user data and result aggregation privacy, thus enabling requesters to verify the aggregation result accuracy	To adjust this scheme to support privacy-aware and verifiable multi-function aggregation
[59]	Facial expression recognition using multi-modality	Multi-modal learning architecture for FER	The comprehensiveness of the facial expression was fully considered to manage subtle expressions and perform robustly to different facial image inputs	Nil
[60]	Online video push based on multi-mode	A novel online video recommendation system based on multi-modal fusion and relevance feedback	An extensive experiment using 20 source videos as users' current viewings, which were searched by 10 representative queries from an existing popular video site. The comparisons indicated the system effectiveness.	Utilization of high-level feature extraction to better describe video content, support more precise recommendations based on video shots instead of the whole video, and collect user profiles from click-through data for improved performance.
[61]	Research on sensing emotion based on multi-modal data	A joint density model over the space of multi-modal	The learned joint representation could be very compact and complementary to	Nil

		inputs, including visual, auditory, and textual modalities; a multi-pathway DBM architecture	hand-crafted features, thus leading to optimal performance in both emotion classification and cross-modal retrieval.	

Summarily, early works on data aggregation in MCS emphasized privacy protection. Developing privacy protection to a certain stage necessitates a balance between privacy and quality, as excessive privacy protection would significantly reduce data quality. Following the presence of multiple sensing data types and patterns with MCS development, multi-mode and multi-type data aggregation strategies must be thoroughly examined to attain optimal data.

6. Summary and Discussion

Recent MCS works in terms of task allocation, data collection, and data aggregation were summarized based on Tables 1, 2, and 3, which outlined the problems, methods, results, and future works, respectively.

Recent task allocation works summarized in Table 1Palpably, research on task allocation has transitioned from single-task to multi-task allocation, multi-task allocation under time and space constraints. More multi-task allocations that meet the constraint conditions were considered in the current multi-task allocation. Some studies have begun emphasizing the relationship between tasks. In this vein, task relevance would substantially influence the multi-task allocation outcomes, which is a future research direction.

Following Table 2, research on data collection ranged from (i) the initial data collection quality to data collection security and (ii) data collection security to data collection efficiency. The complexities underpinning sensing tasks inevitably increased the data amount despite much concern about data collection privacy, security, and quality. Currently, the increasing data collection pressure of the MCS system and ways to improve the system data collection efficiency have gained much scholarly interest. Existing works have also transitioned from a single data collection system to a composite counterpart to ensure data collection quality and security and consider the system data collection efficiency, which renders the system more comprehensive and complex to design.

Recent data aggregation works with an emphasis on data aggregation security and quality were summarized in Table 3. Early works on data aggregation in MCS emphasized privacy protection. A balance between privacy and quality proves necessary when privacy protection develops to a certain stage as extreme privacy protection potentially reduces data quality. With regard to the MCS system development and a multitude of sensor data types and patterns, multi-mode and multi-type data collection strategies must be reviewed to derive optimal data.

7. Conclusion

This paper mainly summarized and analyzed recent works on the MCS life cycle. following three life cycle links (task allocation, data collection, and data aggregation). Each link was summarized from several research directions. Recommendations for future research were also provided. Overall, this paper could be a useful reference for potential MCS scholars.

References

[1] J. Ji, Y. Guo, D. Gong, and X. Shen, "Evolutionary multi-task allocation for mobile crowdsensing with limited resource,"Swarm and Evolutionary Computation, vol. 63,2021, p. 100872, doi:10.1016/j.swevo.2021.100872.

[2] S. Kim, C. Robson, T. Zimmerman, J. Pierce, and E. M. Haber, "Creek Watch: Pairing usefulness and usability for successful citizen science,"in Proc. of the ACM Conference on Human Factors in Computing Systems, ser. CHI, 2011, pp. 2125–2134.

[3] M. Zappatore, A. Longo, and M. A. Bochicchio, "Crowd-sensing our Smart Cities: a Platform for Noise Monitoring and Acoustic Urban Planning," Journal of Communications Software and Systems, vol. 13, no. 2,2017, pp. 53–67.

[4]] J. Wan, J. Liu, Z. Shao, A. V. Vasilakos, M. Imran, and K. Zhou, "Mobile crowd sensing for traffic prediction in internet of vehicles," Sensors, vol. 16, no. 1, 2016,p. 88.

[5] P. Simoens et al., "Scalable crowd-sourcing of video from mobile devices," in Proc. 11th Annu. Int. Conf. Mobile Syst. Appl. Services,Taipei, Taiwan, 2013, pp.139–152.

[6] X Zhao, N Wanga, and R Han, "Urban infrastructure safety system based on mobile crowdsensing," International Journal of Disaster Risk Reduction, Vol. 27,2018, pp. 427-438.

[7] A. Clarke and R. Steele,"Smartphone-based public health information systems: Anonymity, privacy and intervention," Journal of the Association for Information Science and Technology, vol. 66, no. 12, 2015, pp. 2596–2608.

[8] D. Zhang, H. Xiong, L. Wang, and G. Chen, "Crowdrecruiter: selecting participants for piggyback crowdsensing under probabilistic coverage constraint," in Proc. of ACM UbiComp, 2014, pp.703 – 714, doi:10.1145/2632048.2632059.

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

[10] Zhang, M., Yang, P., Tian, C., Tang, S., Gao, X., Wang, B., & Xiao, F, "Quality-aware sensing coverage in budget-constrained mobile crowdsensing networks," IEEE Trans. Veh. Technol., vol. 65,no. 9, Sep. 2016, pp. 7698–7707.

[11] H. Xiong, D. Zhang, G. Chen, L. Wang, and V. Gauthier, “Crowdtasker: Maximizing coverage quality in piggyback crowdsensing under budget constraint,” in Pervasive Computing and Communications (PerCom), 2015 IEEE International Conference on, 2015, pp. 55 – 62, doi:10.1109/PERCOM.2015.7146509.
[12] J. Wang et al., “Fine-grained multitask allocation for participatory sensing with a shared budget,” IEEE Internet Things J., vol. 3, no. 6, 2016, pp. 1395 – 1405.
[13] M. Xiao, J. Wu, L. Huang, Y. Wang, and C. Liu, “Multi-task assignment for crowdsensing in mobile social networks,” in IEEE Conference on Computer Communications (INFOCOM), Apr 2015, pp. 2227 – 2235, doi: 10.1109/INFOCOM.2015.7218609.
[14] T. Hu, M. Xiao, C. Hu, G. Gao, and B. Wang, “A Qos-sensitive task assignment algorithm for mobile crowdsensing,” Pervasive and Mobile Computing, vol. 41, 2017, pp. 333 – 342.
[15] Z. Song, C. H. Liu, J. Wu, and W. Wang, “QoI-Aware Multitask-Oriented Dynamic Participant Selection With Budget Constraints,” IEEE Transactions on Vehicular Technology, vol. 63, no. 9, 2014, pp.4618-4632.
[16] B. Guo, Y. Liu, W. Wu, Z. Yu, and Q. Han, “Activecrowd: A framework for optimized multitask allocation in mobile crowdsensing systems,” IEEE Transactions on Human-Machine Systems, vol. 47, no. 3, 2017, pp.392 – 403.
[17] J. Wang, Y. Wang, D. Zhang, F. Wang, H. Xiong, C. Chen, Q. Lv, and Z. Qiu, “Multi-task allocation in mobile crowd sensing with individual task quality assurance,” IEEE Transactions on Mobile Computing, vol. 17, no. 9, 2018, pp. 2101 – 2113.
[18] H. Xiong, D. Zhang, G. Chen, L. Wang, V. Gauthier, and L. E. Barnes, “iCrowd: Near-Optimal Task Allocation for Piggyback Crowdsensing,” IEEE Transactions on Mobile Computing, vol. 15, no. 8, 2016, pp. 2010 – 2022.
[19] Y. Liu, J. Niu, and X. Liu, “Comprehensive tempo-spatial data collection in crowd sensing using a heterogeneous sensing vehicle selection method,” Personal and Ubiquitous Computing, vol. 20, no. 3, 2016, pp. 397 – 411.
[20] P. Cheng, X. Lian, L. Chen, J. Han, and J. Zhao, “Task assignment on multi-skill oriented spatial crowdsourcing,” IEEE Transactions on Knowledge and Data Engineering, vol.28, no.8, 2016, pp.2201-2215.
[21] D. Deng, C. Shahabi, and U. Demiryurek, “Maximizing the Number of Worker ’ s Self-Selected Tasks in Spatial Crowdsourcing,” Proc. 21st ACM SIGSPATIAL Int’l. Conf. Advances in Geographic Information Systems, 2013, pp. 324 – 33, doi:10.1145/2525314.2525370.
[22] D. Peng, F. Wu, and G. Chen, “Pay as how well you do: A quality based incentive mechanism for crowdsensing,” in Proc. 6th ACM Int. Symp. Mobile Ad Hoc Netw. Comput., 2015, pp. 177 – 186, doi:10.1145/2746285.2746306.
[23] S. Yang, F. Wu, S. Tang, X. Gao, B. Yang, and G. Chen, “On designing data quality-aware truth estimation and surplus sharing method for mobile crowdsensing,” IEEE Journal on Selected Areas in Communications, vol. 35, no. 4, 2017, pp. 832 – 847.
[24] X. Gong and N. Shroff, “Incentivizing truthful data quality for qualityaware mobile data crowdsourcing,” in Proceedings of the ACM Mobihoc ’ 18, 2018, pp. 161 – 170, doi:10.1145/3209582.3209599 .
[25] Z. Yu, J. Zhou, W. Guo, L. Guo, and Z. Yu, “Participant selection for t-sweep k-coverage crowd sensing tasks,” World Wide Web J., vol. 21, no. 3, 2018, pp. 741 – 758.
[26] Sajana, T., & Narasingarao, M. R. “Majority voting algorithm for diagnosing of imbalanced malaria disease,” In International conference on ISMAC in computational vision and bio-engineering . 2018, pp. 31-40, doi: 10.1007/978-3-030-00665-5_4.
[27] X. Yin, J. Han, and P. S. Yu, “Truth discovery with multiple conflicting information providers on the web,” Knowledge and Data Engineering, vol. 20, no. 6, 2008, pp. 796 – 808.
[28] Q. Li et al., “A confidence-aware approach for truth discovery on longtail data,” Proc. VLDB Endowment, vol. 8, no. 4, 2014, pp. 425 – 436.
[29] M. Y. S. Uddin, H. Wang, F. Saremi, G.-J. Qi, T. Abdelzaher, and T. Huang, “Photonet: a similarity-aware picture delivery service for situation awareness,” in Proc. of IEEE RTSS, 2011, pp. 317 – 326, doi: 10.1109/RTSS.2011.36.
[30] B. Guo, H. Chen, Z. Yu, X. Xie, and D. Zhang, “Picpick: a generic data selection framework for mobile crowd photography,” Personal and Ubiquitous Computing, vol. 20, no. 3, 2016, pp. 325 – 335, doi:10.1007/s00779-016-0924-x
[31] L. Cheng et al., “Compressive sensing based data quality improvement for crowd-sensing applications,” J. Netw. Comput. Appl., vol. 77, 2017, pp. 123 – 134.
[32] M. Zappatore, C. Loglisci, A. Longo, M. A. Bochicchio, L. Vaira, and D. Malerba, “Trustworthiness of context-aware urban pollution data in mobile crowd sensing,” IEEE Access, vol. 7, 2019, pp. 154141 – 154156.
[33] S. Jagabathula, L. Subramanian, and A. Venkataraman, “Reputation-based worker filtering in crowdsourcing,”In Proceedings of the Advances in Neural Information Processing Systems, 2014, pp. 2492 – 2500.
[34] C. Miao, Q. Li, H. Xiao, W. Jiang, M. Huai, and L. Su, “Towards data poisoning attacks in crowd sensing systems,” in Proceedings of the Eighteenth ACM International Symposium on Mobile Ad Hoc Networking and Computing, 2018, pp. 111 – 120, doi:10.1145/3209582.3209594.
[35] C. Zhao, S. Yang, and J. A. McCann, “On the data quality in privacypreserving mobile crowdsensing systems with untruthful reporting,” IEEE Transactions on Mobile Computing, 2019, pp. 1 – 1, doi: 10.1109/TMC.2019.2943468.
[36] H. Yu, P. B. Gibbons, M. Kaminsky, and F. Xiao, “Sybillimit: A nearoptimal social network defense against sybil attacks,” in Proc. IEEE Symp. SP, 2008, pp. 3 – 17,doi: 10.1109/SP.2008.13 .
[37] H. Yu, M. Kaminsky, P. Gibbons, and A. Flaxman, “Sybilguard: Defending against sybil attacks via social networks,” Proc. ACM SIGCOMM Comput. Commun. Rev., vol. 36, no. 4, 2006, pp. 267 – 278.
[38] J. Feng, T. Li, Y. Zhai, S. Lv, and F. Zhao, “Ensuring honest data collection against collusive CSDF attack with binary-minmaxs clustering analysis in mobile crowd sensing,” IEEE Access, vol. 7, 2019, pp. 124491 – 124501.
[39] Satyaki Roy, Nirnay Ghosh, Preetam Ghosh, and Sajal K Das, “biomcs 2.0: A distributed, energy-aware fog-based

framework for data forwarding in mobile crowdsensing," Pervasive and Mobile Computing, vol.73, 2021, pp. 101381.
[40] P. Wang, Z. Yu, C. Lin, L. Yang, Y. Hou, and Q. Zhang, "D2D-enabled reliable data collection for mobile crowd sensing," in 2020 IEEE 26th International Conference on Parallel and Distributed Systems (ICPADS), 2020, pp. 180 – 187, doi:10.1109/ICPADS51040.2020.00033.
[41] X. Xia, Y. Zhou, J. Li, and R. Yu, "Quality-aware sparse data collection in mec-enhanced mobile crowdsensing systems," IEEE Trans. Computat. Social Syst., vol. 6, no. 5, 2019, pp. 1051 – 1062.
[42] H. Jin, L. Su, H. Xiao, and K. Nahrstedt, "Incentive mechanism for privacy-aware data aggregation in mobile crowd sensing systems," IEEE/ACM Transactions on Networking (TON), vol. 26, no. 5, 2018, pp. 2019 – 2032.
[43] Z. Zhang, S. He, J. Chen, and J. Zhang, "REAP: An efficient incentive mechanism for reconciling aggregation accuracy and individual privacy in crowdsensing," IEEE Trans. Inf. Forensics Secur., vol. 13, no. 12, 2018, pp. 2995 – 3007.
[44] L. Yang, M. Zhang, S. He, M. Li, and J. Zhang, "Crowdempowered privacy-preserving data aggregation for mobile crowdsensing," in ACM MobiHoc, 2018, pp. 151-160, doi:10.1145/3209582.3209598.
[45] Y. Wang, Z. Cai, X. Tong, Y. Gao, and G. Yin, "Truthful incentive mechanism with location privacy-preserving for mobile crowdsourcing systems," Computer Networks, vol. 135, 2018, pp. 32-43.
[46] M. Zhang, L. Yang, S. He, M. Li, and J. Zhang, "Privacy-preserving data aggregation for mobile crowdsensing with
[53] S. Kim, K. Lewi, A. Mandal, H. W. Montgomery, A. Roy, and D. Wu, "Function-Hiding Inner Product Encryption is Practical," IACR Cryptology ePrint Archive, vol. 2016, 2016, pp. 440 – 456.
[54] T. Ryffel, D. Pointcheval, F. Bach, E. Dufour-Sans, and R. Gay, "Partially encrypted deep learning using functional encryption," in Proc. NeurIPS, Vancouver, BC, Canada, Dec. 2019, pp. 1 – 21.
[55] B. Zhao, S. Tang, X. Liu, and X. Zhang, "PACE: Privacy-preserving and quality-aware incentive mechanism for mobile crowdsensing," IEEE Transactions on Mobile Computing, vol. 20, no. 5, 2020, pp. 1924 – 1939.
[56] K. Xie, X. Li, X. Wang, G. Xie, D. Xie, Z. Li, J. Wen, and Z. Diao, "Quick and accurate false data detection in mobile crowd sensing," in Proc. IEEE INFOCOM, 2019, pp. 2215 – 2223, doi: 10.1109/TNET.2020.2982685.
[57] Yan, X., Ng, W. W., Zeng, B., Lin, C., Liu, Y., Lu, L., & Gao, Y, "Verifiable, reliable, and privacy-preserving data aggregation in fog-assisted mobile crowdsensing," IEEE Internet of Things Journal, vol. 8, no. 18, 2021, pp. 14127-14140.
[58] W. Zhang, Y. Zhang, L. Ma, J. Guan, and S. Gong, "Multimodal learning for facial expression recognition," Pattern Recognit., vol. 48, no. 10, 2015, pp. 3191 – 3202.
[59] B. Yang, T. Mei, X. Hua, L. Yang, S. Yang, and M. Li, "Online video recommendation based on multimodal fusion and relevance feedback," in CIVR, 2007, pp. 73 – 80, doi:10.1145/1282280.1282290.
[60] L. Pang, S. Zhu, and C.-W. Ngo, "Deep multimodal learning for affective analysis and retrieval," IEEE Transactions on Multimedia, vol. 17, no. 11, 2015, pp. 2008 – 2020.
externality: An auction approach," IEEE/ACM Trans. Netw., vol. 29, no. 3, 2021, pp. 1046 – 1059.
[47] Wang, Z., Hu, J., Lv, R., Wei, J., Wang, Q., Yang, D., & Qi, H, "Personalized privacy-preserving task allocation for mobile crowdsensing," IEEE Transactions on Mobile Computing, vol. 18, no. 6, 2018, pp. 1330-1341.
[48] Wang, X., Ying, C., & Luo, Y. , "Privacy-friendly decentralized data aggregation for mobile crowdsensing.," In GLOBECOM 2020-2020 IEEE Global Communications Conference ,2020, pp. 1-6, doi: 10.1109/GLOBECOM42002.2020.9322169.
[49] Q. Li, G. Cao, and T. F. L. Porta, "Efficient and privacy-aware data aggregation in mobile sensing," IEEE Transactions on Dependable and Secure Computing, vol. 11, no. 2, 2014, pp. 115 – 129.
[50] R. Lu, K. Heung, A. H. Lashkari, and A. A. Ghorbani, "A lightweight privacy-preserving data aggregation scheme for fog computing-enhanced IoT," IEEE Access, vol. 5, 2017, pp. 3302 – 3312.
[51] X. Li, S. Liu, F. Wu, S. Kumari, and J. J. P. C. Rodrigues, "Privacy preserving data aggregation scheme for mobile edge computing assisted IoT applications," IEEE Internet Things J., vol. 6, no. 3, 2019, pp. 4755 – 4763.
[52] F. Liu, B. Zhu, S. Yuan, J. Li, and K. Xue, "Privacy-preserving truth discovery for sparse data in mobile crowdsensing systems," in Proc. IEEE Glob. Commun. Conf., 2021, pp. 1 – 6, doi: 10.1109/GLOBECOM46510.2021.9685134.

[61] T. H. Silva, P. O. V. de Melo, A. C. Viana, J. M. Almeida, J. Salles, and A. A.Loureiro, "Traffic Condition Is More Than Colored Lines on a Map: Characterization of Waze Alerts," in Social Informatics, Springer, 2013, pp. 309 – 318, doi: 10.1007/978-3-319-03260-3_27.
[62] S. Mathur et al., "Parknet: Drive-by Sensing of Road-Side Parking Statistics," Proc. ACM MobiSys, 2010, pp. 123 – 36, doi:10.1145/1814433.1814448.
[63] P. Dutta, P. M. Aoki, N. Kumar, A. Mainwaring, C. Myers, W. Willett, and A. Woodruff, "Common sense: participatory urban sensing using a network of handheld air quality monitors," in Proc. of the ACM Conference on Embedded Networked Sensor Systems, 2009, pp. 349 – 350, doi:10.1145/1644038.1644095 .
[64] Alashaikh A S, Alhazemi F M, "Efficient Mobile Crowdsourcing for Environmental Noise Monitoring," IEEE Access, 2022, vol. 10, pp. 77251-77262.

Xia Zhuoyue received a Bachelor's degree in E-commerce from Hefei University of Technology in 2015 and a Master's degree in Software Engineering from Northwestern Polytechnical University in 2018. He is currently pursuing a Ph.D. in Computer Science at Taylor University and specializes in mobile swarm intelligence perception.

Azween Abdullah is a professional development alumnus of Stanford University and MIT with 30 years of experience as an academic in higher learning institutions and 15 years as a software engineer, systems analyst, computer software developer, and IT/MIS consultant and trainer in commercial companies. He was also the director of research and academic affairs at two higher learning institutions and the vice president of educational consultancy services.

SH Kok has 17 years of industrial experience before completing his Ph.D. in Computer Science from Taylor's University in 2021 and currently works as a senior lecturer.