Data Quality in Mobile Crowd Sensing Systems: Challenges and Perspectives

Konstantina Banti, Filomeni Katsimpoura, Malamati Louta
Department of Informatics and Telecommunications
Engineering School of Engineering,
University of Western Macedonia
Kozani, Greece
louta@uowm.gr

George Karetsos
Department of Computer Engineering
Technology Education Institute of Thessaly
Larissa, Greece
karetsos@teithessaly.gr

Abstract— Mobile Crowd Sensing (MCS) has become a new sensing paradigm, which leverages on the ubiquity of mobile devices with advanced multi-modal sensing features in conjunction with human intelligence so as to cost-efficiently monitor and analyze large-scale phenomena. MCS core characteristic is user involvement in data collection, processing, analysis and sharing; Due to the opt-in nature of MCS systems, a number of critical concerns is raised that should be efficiently addressed so as to enable MCS unimpeded advancement. Data reported by users could be unintentionally inaccurate and/or deliberately falsified. Thus, ensuring data quality and integrity in MCS constitutes one of its key challenges. This paper as a first step discusses on the data quality challenge and identifying its interdependencies with different underlying key issues. Subsequently, we comprehensively survey representative reputation-based trust establishment mechanisms proposed in related research literature in the context of MCS, as a potential solution to the data quality problem. Their distinct features are analyzed and their relative merits and weaknesses are identified and highlighted. Finally, we discuss on design aspects of reputation mechanisms and provide guidelines and future research directions.

Keywords—mobile crowd sensing; challenges; data quality; trustworthiness; reputation mechanisms; survey

I. INTRODUCTION

IN recent years, the Mobile Crowd Sensing (MCS) paradigm has emerged as a technological solution that fosters the collection of data sensed or generated from mobile devices and subsequently analyzed by employing proper data mining algorithms, so as to identify spatio-temporal patterns, generate models and/or make predictions on physical or social phenomena being observed in a cost-efficient manner. MCS leverages on the power and wisdom of crowd, exploiting human intelligence, ubiquity and mobility features. To this respect, MCS empowers people to contribute data, enabling efficient (in terms of cost and time) monitoring of large-scale phenomena that cannot easily be measured or would otherwise need costly investments (in terms of hardware and software) [1, 2]. MCS has recently attracted the attention of researchers with designed applications ranging from environmental monitoring (e.g., air quality [3]) to traffic planning [4], public safety [5] to smart parking [6-7].

However, this technology faces a number of critical issues and challenges, whose solution becomes imperative, so as MCS could reach its full potentials. Most importantly, the dynamic conditions and the limitations imposed to the mobile devices (related to the wireless medium, as well as to energy, bandwidth and computing resources availability) should be carefully considered. Additionally, task assignment and scheduling process, security related concerns, mainly in terms of users' privacy and data integrity, arising from the opt-in nature of MCS, should be efficiently addressed, in conjunction with the incentive mechanisms that should be applied in order to promote users' participation.

Maintaining data quality and integrity is an important and challenging issue that should be addressed in the context of MCS. MCS applications could suffer from inaccurate data provisioning, taking place unintentionally and/or intentionally due to their inherent open nature. Besides the cases of noisy and obsolete data, faults, low quality of the wireless medium, low expertise of the contributing users, the presence of selfish and/or malicious users deliberately providing low quality and/or falsified data should be taken into account.

Trust mechanisms evaluating users' trustworthiness and building trust relationships are generally considered to provide a solution to the data quality / integrity challenge in MCS. Trust is often described as the belief of an entity in the competence and benevolence of another entity to act honestly, reliably and dependably [8]. Reputation mechanisms establish trust by exploiting learning from experience concepts in order to obtain a trust related value in the form of rating based on past experiences, observations and other entities opinions. Even though trust and reputation constitutes a well investigated field in different settings with a wide variety of trust and reputation models with advanced features being developed in a number of different research areas (e.g., ecommerce [9], social networks and recommendation systems [10, 11], ad-hoc and opportunistic networks [12, 13], Web, cloud computing and pervasive systems [14, 15] they cannot be readily applied to MCS due to its specific characteristics. The aim of this paper is to survey reputation-based trust establishment in MCS setting, identify and discuss on their distinct characteristics, merits and weaknesses. Based on our findings, we will provide guidelines for future research.

The rest of the paper is structured as follows. Section II presents the general MCS architectural design. Section III presents, in detail, the data reliability issue focusing on the proposed solutions in related research literature. Section IV analyzes the critical aspects which should be taken into account when designing a reputation mechanism. Section V discusses on user reputation mechanisms which have be adopted to solve the data reliability challenge. Section VI discusses on our findings concerning the focal aspects of mobile crowd sensing system design identified, while section VII concludes the paper and

highlights our future plans.

II. MCS ARCHITECTURAL DESIGN

The general architectural design of an MCS system comprises the following main entities: a) the *Requestors*, b) the *Workers* and c) the *Crowdsensing Platform*. The Requestors submit sensing requests pertinent to their interests to the crowdsensing platform; see the answers provided by Workers and/or gain access to the knowledge acquired after the platform has analyzed the data collected. They may evaluate the Workers based on their responses and depending on the adopted incentive model provide the Workers' reward. The Workers are the main source of information and play major role in data collection. Depending on the assumed model, they may be assigned several tasks taking into account their owners' preferences and the requirements of both the Requestors and the crowdsensing platform or the Workers may select in which one task they want to participate and contribute. The Crowdsensing platform is the main communication link between Requestors and Workers. The platform stores, processes and analyzes data provided by Workers and the Requestors and depending on the adopted incentive model provide the Workers' reward.

As noted in [16], the different architectural frameworks proposed in recent related research literature lack coherence; the authors identified several issues where the architectural frameworks vary significantly and concluded that there is no consolidated and unifying set of well recognized principles for building MCS systems. Furthermore, they advance the incorporation of context awareness and self-adaptivity principles in conjunction with advanced cognitive capabilities to future MCS architectures, while they stress that each problem / challenge should not be addressed independently, but taking into account potential inter-dependencies of MCS key elements. Following, we summarize the key MCS architectural elements, inter-related to the data quality challenge.

In MCS, as already noted, data collection and sharing necessitates human involvement. On the one hand, a user may actively participate in sensing tasks, requiring his/her explicit actions so as to complete the sensing task (e.g., taking a picture), referred to *participatory sensing*. On the other hand, *opportunistic sensing* does not require explicit user's actions to perform a sensing task; the sensing task is executed in the background, without any explicit user involvement. Similarly, according to the *pull model*, users are required to retrieve the active tasks and select the ones they would like to contribute to, while according to the *push model*, user's control is diminished and the tasks are pushed to the mobile devices if specific requirements and criteria are met. Data may be collected from both physical and online worlds, so as to integrate and exploit their complementary features and merits [2]. Concerning data transmission, users may adopt a store-carry-forward behavior, waiting until a better transmission opportunity occurs (*opportunistic transmission*) or they may utilize a communication system (*infrastructure-based transmission*).

In the following section, focusing on the data integrity and quality challenge, we will further elaborate on the different key design aspects that are inter-related with the solutions proposed.

III. DATA QUALITY CHALLENGE IN MCS

In order to enable MCS unimpeded advancement, data integrity and quality should be maintained. This challenge is raised due to the open nature of MCS systems that rely on the contributions of users, which may be inaccurate, produced either intentionally and / or unintentionally. Unintentional inaccurate data provisioning or low quality data may be a result of faults, low quality of the wireless link, low resource levels of the mobile device, limited capabilities of the sensors, current context of mobile device's operation, less qualified Workers, environmental uncertainties and outdated data due to the time period elapsed prior to data uploading to the platform. Sensed data quality (in terms of accuracy and confidence as well as latency) varies significantly due to the devices' mobility, current resources' availability and context of operation. For example, Workers may unintentional keep their mobile devices in an undesirable position, when collecting sensor readings, decreasing thus the quality of the submitted data [2, 16]. Furthermore, limited availability of energy resources may lead to activation of different sensors, producing data for serving the same purpose, albeit with lower data quality, while necessitating lower energy consumption levels [17].

To this respect, task assignment and scheduling across multiple devices with diverse sensing capabilities, resource availabilities and limitations imposed may lead to higher or lower data quality. Selecting the right set of participants to execute the required tasks with the minimum cost, while guaranteeing maximum coverage and high sensing quality is a complex problem to be addressed [18]. While data sensed by few users can be considered not to lead to very accurate results, multiple sensing processes of the same task makes the data more valid. However, increasing the number of participants does not necessarily mean that the overall quality of the results will be improved [19]. In the light of the aforementioned, in related research literature, prediction models are proposed so as to estimate the number of the workers necessitated and provide a specific task assignment so as to achieve the required accuracy. Additionally, the quality of data may be affected by users who are not familiar with the target area or have to make a special effort because they don't know how to perform the task [20]. Many proposed solutions develop user profiles based on available social networking information [21], while others collect and process data in relation to the user (preferences, interests) and the surrounding environment so as to identify the suitable users to assign the required sensing tasks in order to improve on data quality. Low quality data can also be generated when a user accepts to perform multiple tasks at the same time [2], increasing the latency incurred in the provision of data. A maximum workload of the same user may be set so as to avoid the aforementioned situation. The solution provided by the task assignment problem should guarantee maximum coverage, i.e., sensing results should be provided for all the key points specified in the area of interest. Lack of sensing reports in some

sub-areas lowers the overall data quality, which cannot be compensated by potential report redundancy in several others and jeopardizes the knowledge extracted by MCS systems. Additionally, improving on the data quality level, while minimizing the consumption of resources necessitated (e.g., by deciding on the sensors to be used each time and their duty cycling, the transmission method adopted, the local processing of data on the mobile device on the basis of the available energy levels of the devices) is a second challenging issue to address, strongly inter-related with both the data quality challenge and the task assignment process.

Intentional inaccurate information provisioning covers the cases of selfish and/or malicious entities that provide in purpose low quality and / or erroneous information. Selfish parties may provide low quality data (e.g., non-fresh or random sensor readings) in purpose so as to minimize their own cost in terms of effort put and resources consumed or maximize their own utility, while being entitled to the reward specified for task execution. For example, in real-time traffic monitoring, selfish users can report false traffic congestion warnings to prevent traffic on their own routes [22]. Malicious parties in an attempt to harm the usefulness of the extracted information and the MCS applications, falsify the data provided. For example, a malicious participant can spoof a GPS location [23] and start providing falsified location data. The presence of selfish and malicious users leads to low quality contributions, which need to be identified and eliminated. Therefore, mechanisms for efficiently validating and ensuring the integrity of the collected data are necessitated. As a simple, first step solution, location validation mechanisms may be enforced. For example, current solutions proposed check if the users are in close proximity with the position they indicate in their answer. Additionally, the system can aggregate the responses of all users assigned with the same sensing task using a majority rule [24] or the system may check data quality based on potentially available gold answers [21].

Trust and reputation systems contribute significantly to systems where the lack of basic knowledge about participating users may lead to undesirable situations as to the reliability of the information they provide [25]. Avoiding erroneous data based on reputation values requires monitoring of participants' behavior (e.g., in terms of the quality of the contributions provided with respect to accuracy, confidence and latency) and attributing reputation scores [22, 26], which indicate their credibility in submitting sensing data. Users' trustworthiness evaluation may be exploited in order to identify the most trustworthy users to be involved at specific sensing tasks [27], taking also into account other design aspects and specified parameters of the task assignment process or the significance of data collected may be weighted according to the trustworthiness of the contributing participant [28]. Thus, selfish and malicious parties attributed with a low reputation value may be directly identified [26, 29] and potentially excluded from future task executions. In general, reputation mechanisms are considered to sustain cooperation and serve as an incentive for good behavior, rewarding good players and penalizing those ones behaving badly. However, trustworthiness evaluation may contradict users' security and privacy considerations. Mechanisms are required in order to link users to actions performed, without disclosing their identity [30]. In general, MCS systems raise privacy concerns. Users are reluctant in disclosing sensitive information, while in most cases, users would like to know and control the information they share, with whom and for how long. Additionally, since users are required to put personal effort or spend their time and consume own resources, they should be rewarded in order to retain their engagement to provide accurate contributions. Incentive mechanism should be carefully designed so as to reward participants according to their contributions (both quantitative and qualitative), taking into account and properly balancing its interplay with the task assignment process. Following, we present the basic design aspects of reputation mechanisms employed in related research literature.

IV. REPUTATION MECHANISMS FUNDAMENTALS

Hereafter, we summarize the critical aspects and issues considered when designing a reputation mechanism to promote cooperation amongst the involved entities and/or enhance the provisioned service quality. Reputation mechanisms highly depend on the underlying model and related mechanisms for trust-related information collection and analysis, reputation rating formation and actions taken when identifying a misbehaving entity. Reputation information collection mechanism may be based on direct experiences of the evaluator entity (requestor) concerning the behavior of the target entity under evaluation (worker), referred to as first-hand information, and on propagated reputation information on the target worker's behavior, referred to as second hand information. Without being exhaustive, issues related to reputation related information propagation involve the type of information that is propagated (e.g., alarms, positive and/or negative experiences, reputation ratings), the determination of the recipients of this information (e.g., entity's friends list, entities within a certain community, all entities in the network) and the time that this information is propagated. Considering propagated information in reputation rating formation process speeds up trust convergence; however, it entails the risk of handling falsified (intentionally and/or unintentionally) information, impacting negatively the accuracy of the reputation rating formed. Most related works introduce a trust level associated with each entity (witness) providing a recommendation on another entity. In line with [31], our view is that this trust level should reflect both the trustworthiness of witnesses in the eyes of the evaluator and confidence of the witness on the accuracy of the information that it possesses. Additionally, reputation rating formation should be based mostly on recent events, while the initial reputation value assumed for all unknown entities should be carefully considered. Finally, most systems proposed entail the gradual isolation of the misbehaving entities, punishing them for not complying with what promised. A

forgiving mechanism is also considered in order to enable misbehaving entities to re-enter in the system in case they exhibit good behavior.

V. REPUTATION MECHANISMS IN MCS

A. A Crowd-Sensing Framework for Allocation of Time-Constrained and Location-based Tasks

In [32], the authors propose a service computing framework for time constrained-task allocation in location based MCS systems. This framework relies on 1) a recruitment algorithm that implements task allocation, 2) queuing schemes to handle efficiently the sensing tasks, 3) a task delegation mechanism and 4) a reputation management component. The goal of this framework is to assign each task to the most appropriate set of users who will return high quality results within the required response time. In order to achieve the high quality results, the quality of information provided by a user is computed based on user's reputation, user's confidence to successfully perform the task (e.g., considering the battery level of the user's device) and the distance between the user and the location of the task. More specifically, regarding user's reputation, the system takes into account if the user can or can't finish their task, exploiting a delegation mechanism where if a user can't perform the assigned task recommends other users from his/her social network to finish the task. User's reputation is a parameter computed based on user's historical performance concerning the completion of the tasks assigned to him/her. If data reported by the user is close to the ground truth (the most common answer by users) of a task, then, it is assumed that the user completed the task successfully. On the whole, the user's ability to complete tasks affects their reputation, where the payment depends on the user reputation.

B. An Endorsement-based Reputation System for Trustworthy Crowdsourcing

In [33], the authors propose a reputation system to not only assess but also predict the trustworthiness of user contributions. In particular, they explore an inter-worker relationship called endorsement to improve trustworthiness prediction, while taking into account the heterogeneity of both users and tasks. Different users participate in various types of tasks, which may require different domain-specific knowledge or expertise. Users are connected to one another via endorsement links, which represent a trust or support relationship. Every endorsement link has a weight, indicating how much confidence has one user to another. The prediction of the trustworthiness of contributions takes into account each user's own reputation, user's expertise and the endorsement impact from his endorsers. User's reputation is calculated based on the requestors' ratings returned for the tasks which he/she has performed. Concerning user's expertise, the system checks his/her past performance in this task, as given by the requestors using machine learning methods because historical information may not be available. Time effects are taken into account and recent information is prioritized over the past. They motivate users to participate exploiting user reputation instead monetary rewards. Finally, for each task request, the system acquires the best candidate workers and returns them to the requestor, enabling him/her to select a subset of them to perform the requested task. After the task has been completed, the requestor rates them in order for the system to update their reputation.

C. DTRF

In [27] the authors propose a Dynamic-Trust-based Recruitment Framework (DTRF), in which real-time direct trust and feedback aggregation trust are combined to select the most suitable participants. Each task is characterized by a trust threshold value requirement, which denotes the minimum trust degree a user should possess in order to be able to perform a task. In order for a user to perform a task, his location must be in the range of the task and his/her aggregate trust degree shouldn't be lower than a pre-specified task's threshold. After the participant has sent the requested data, the requestor will evaluate the quality of the data and will send an evaluation report to the platform. Data quality is the degree of requestor's satisfaction. More specifically, trust represents the correctness and reliability of each user. The overall trust degree is computed according to direct and indirect trust. On the one hand, direct trust is the trust of one user to another based on their interactions in the recent past. On the other hand, indirect feedback trust is the overall feedbacks generated by the service requestors (positive/negative). Finally, the system adopts centralized architecture where the platform will evaluate the submitted data by the participants and the pertinent reports of requestors and give rewards or penalties to the relevant participants, updating also their trust degree.

D. CCIS

In [34], the authors propose a Crowd-based Credibility Improving Scheme (CCIS), which consists of a clustering algorithm used to place false and regular data into different groups. Also, the participants' reputation is introduced to identify and filter the incorrect data, improving the overall data reliability of collected data sent by the crowd. It is assumed that a given set of data is reliable, only if the location of task execution is valid (that is it lies within an acceptable distance from location of interest), while sensory value is assumed to be valid if it reflects the ground truth of physical phenomenon of the corresponding location. The framework of the proposed scheme is carried out sequentially in two phases: first, the initialization phase during which basic knowledge about the participants' reputation is obtained and the validity of the submitted data is checked, and second the filtering phase during which the data are categorized as untrue or normal, to be ignored or taken into account in the final result, respectively, according to the clustering algorithms and the available reputation information. Finally, valid data (as determined according to the aforementioned process) additive increases user's reputation, while invalid data multiplicative decreases it.

E. *SONATA with Anchors*

The authors in [35] adopt a vote-based trustworthiness and vote-based social trustworthiness assessment schemes with trusted entities, called trustworthy anchors of the system. An anchor user is fully trustworthy regardless of the accuracy of its sensor readings. Thus an anchor node can't be considered as malicious user. Also, an anchor node is fully capable of voting for the trustworthiness of other participants. Each node casts a vote for a newly node. The weight of the vote of a node is the product of its trustworthiness (is computed according to the sum of votes) and vote capacity (the number of votes each user can cast). Furthermore, each node casts a vote for each newly joining node. In this way, the reputations of new users are determined based on acquired votes. The authors consider that the success of vote-based trustworthiness is closely related to the i) ratio of malicious users in the crowd, ii) sensing task load on the MCS system, and iii) initial reputations of the users. Finally, they adopt reverse auction-based recruitment of users where any recruited node is guaranteed to be rewarded no less than its sensing cost where is scaled by its trustworthiness. For the winner selection, the system calculates a marginal value of node which is the difference between reputation-based values of the nodes set before and after recruiting node. The platform adds nodes until the difference between the marginal value of node and its sensing cost is non-positive.

F. *Are You Contributing Trustworthy Data? The Case for a Reputation System in Participatory Sensing*

The authors in [36] propose a novel reputation system for estimating devices' reputation score. Specifically, a high device reputation score is an indication that a particular device has been reporting reliable measurements in the past, while suggesting that the server should place a higher level of trust in the sensor readings from this device in the future, where most recent information is more relevant than the past. The proposed system consists of a watchdog and a reputation module, which associates a reputation score with each contributing device, reflecting in essence the level of trust about the data uploaded by that device over a period of time. The watchdog module produces ratings according to the sensors readings, which can be considered as the device's confidence level. These ratings act as input, denotes that reputation is the result of aggregating historical device information, to the reputation module which computes the device reputation score, exploiting the Gompertz function.

G. *Crowdsensing with social network-aided collaborative trust scores*

This paper [37] leverages upon a centralized reputation system by incorporating statistical and vote-based trust scores, using social network theory to evaluate the trustworthiness of crowdsensed data and of the involved mobile devices that provide sensing services, while taking into account attacks by malicious users. Specifically, the system considers each participant as a node in a social network, where each social network community is dynamically created based on common sensing tasks between participants. In the so formed social network communities, it is assumed that directly connected nodes can vote for their neighbors to contribute to their reputation assessment. Every node has a vote capacity which means the impact of node's vote. Vote-based trustworthiness is the impact of votes received by the neighboring nodes scaled by their vote capacities. The overall trustworthiness is computed based on statistical (ratio of positive values to the total) and social (vote-based) reputation.

H.*ARM*

In [38], the Accumulated Reputation Model (ARM) analyzes the data submitted by the participants and then evaluates the reliability of participants in successfully performing sensing tasks using an accumulated reputation rating that can minimize the impact of corrupted data and, ultimately, achieve a high precision result. Generally, the ARM which runs in the server processes all the data obtained from the participants to produce a sensing result, compute the contribution and reputation score. More specifically, the contribution score corresponds to the assessment of the quality of the sensed data provided by each participant and is computed in each act of participation. The reputation score of each participant is calculated based on the evaluated contribution score and his/her attendance frequency, which shows the history of the participations of each user. The reputation score of each participant is derived from the trimmed-mean method and is based on all participants' historical behaviors. The sensing data of each user is weighted according to his reputation score. Thus, the final sensing result equals to the sum of the weighted data of all users. Concluding, ARM treats the incorrect data, even in some cases where malicious users is a significant percentage among the participants.

I. *ARTSense*

In [39], the authors propose ARTSense, a framework to solve the problem of "trust without identity" in participatory sensing networks. Its goal is to achieve anonymity, reputation and trust. It exploits a privacy-preserving provenance model, a data trust assessment scheme and an anonymous reputation management protocol. Anonymous pseudonyms are used to achieve anonymity and a blind signature technique is used, converting experience reporting and reputation assessment into two distinct processes. The trust value (referred to as trust base) of the sensing report represents its reliability and correctness and is based on various factors such as sensing time, sensing location, sensor mode, user's traveling mode (i.e., standstill, walking, cycling, driving). A similarity factor is assigned to each report with respect to all collected reports for this task. The final trust value assigned to a report is proportional to the similarity factor and its trust base. Comparing the final trust and the reputation level of the participant, the server creates the reputation feedback level. If the final trust value of a report is greater/lower than participant's reputation level, then the feedback value is positive/negative, respectively. A negative feedback affects reputation more than a

positive one. After calculating the trust value of a report, the reputation feedback level for the participant is generated and encrypted within a reputation feedback coupon. The user receives it, removes the blinding factor and redeems the coupon from the server after the necessary security checks have been made. If all security checks are successful, the server gets the reputation feedback level and the user's password from the coupon and updates the corresponding record in the reputation table. Thus, trust and anonymity are succeeded at the same time.

J. PaySense

In [40], the authors propose PaySense, a general framework that encourages user participation and provides a mechanism for validating the quality of collected data based on user reputation. All these functions are performed, while ensuring privacy through the Bitcoin encryption system as a reputation and rewarding mechanism and maintaining anonymity of users, attributing different pseudonyms to the same user. The reputation of a user could be estimated as the sum of the reputation of all pseudonyms attributed to the user, as each of them has its own reputation score according to the user's behavior and the respective sensor capabilities. Once the sensed data has been validated or rejected, the user is entailed to a reward or punishment and his reputation will be accordingly updated. Since in PaySense, both values are associated with a Bitcoin payment, the reputation update will determine the reward / punishment value the user shall receive. PaySense provides a satisfactory response to the reputation problem in anonymity scenarios. In Paysense, the process of transferring reputation from an old to a new pseudonym causes the reduction of the reputation. However, because the user's reputation is reduced, the system gives to him an economical profit thanks to the fact that reputation is expressed directly in bitcoins.

K. A Reputation Framework for Social Participatory Sensing Systems

The proposed mechanism [28] takes into account both the quality of the contributed data (a group of parameters must be evaluated such as relevance to the campaign, ability in determining a particular feature, fulfillment of task requirements) and the trustworthiness level of participant (that is a combination of personal and social factors such as expertise, timeliness, locality, friendship duration, interaction time gap) within the social network. These two dimensions are combined through a fuzzy inference system so as to provide a final assessment of the trustworthiness of contributions (ToC). As soon as a task is launched, the participants begin to send a series of contributions. For each contribution, the requestor computes a value for the trustworthiness of the participant. Based on the ToC assigned to each contribution, the trust of requester upon the corresponding participant is updated. Subsequently, the participant's reputation score is updated centrally by the platform, adopting a reward / penalty policy, where the requestor's reputation score is used, outweighing accordingly the participant's reputation modification in case the requestor is unreliable. In essence, it is considered that an evaluation by a requester with high reputation score is more reliable than that originating from a requestor with low reputation score. In this way, the reputation score of each node depends on (i) the trust ratings that other nodes has assigned to him, and (ii) the reputation of those nodes. This process is repeated for all participants at the end of each task.

L. Credible and energy-aware participant selection with limited task budget for mobile crowd sensing

In [41], a crowd sensing system is proposed aiming to a) efficiently allocate tasks to the most appropriate participants and b) maximize participants' rewards so as to encourage them to contribute sensing data continuously, while considering the welfare of both platform and participants. In this context, the participant's reputation is defined and a mechanism for its evaluation / update is proposed, which takes into account the participant's willingness and quality of the data. A reputation score is used to select the most reliable users. This minimizes the damage and the threat of dishonest behavior and protects the system from possible abuse. Two metrics are introduced: a) Difficulty of Task that is used to weight the difficulty level of a task for participants and helps them to choose the right tasks to maximize their rewards, and b) Quality of Information that describes the quality level of the collected sensing data pertinent to the assigned task requirements, where sensing cost and participant's reputation value are used to predict the level of quality of the sensed data that the user can contribute. User's reputation consists of two parts: willingness, showing their enthusiasm to contribute to the sensing task (measured by the user's response time to submit data), and the quality of the data contributed by each user, contributing equally to the reputation score. When the platform receives a task from a publisher, it first selects a number of trusted participants based on their reputation scores. They contribute data, and based on their willingness and the quality of these data, positive or negative feedback is generated for each participant and their reputation is accordingly updated.

VI. DISCUSSION

After surveying the reputation mechanisms proposed and adopted in recent related research efforts, it is found that the different approached lack unity. Some mechanisms have adopted a centralized architecture (e.g., [32, 34, 36-40]), where users send their reports to the platform, which subsequently processes the reports and the quality of submitted data and accordingly update users' (participants and in some cases requestors) reputation score. On the other end, some systems have adopted a distributed architecture, where each user keeps separately a reputation score for each participant that was assigned and executed a task launched by the user in the past, based mostly on the quality of the contributed data. At this point, it should be noted that only a few of the presented mechanisms (e.g., [28,33,35,37]) consider the trustworthiness of the requestors' reports returned to the platform and take into account the reputation of the requestors so

	32	33	27	34	35	36	37	38	39	40	28	41
High Reputation = High Quality Data			x	x	x			x	x	x		
User Willingness (Time/Participant Attendance Frequency)	x							x			x	x
Location	x		x	x					x		x	x
User Reputation (quality of submitted data/ historical performances)	x	x	x	x	x	x	x	x	x	x	x	x
Initial Reputation for new users			x	x	x	x						
Endorsement/Vote-based between users		x	x		x	x	x				x	
User reputation of endorser		x			x						x	
Incentives according to user reputation (monetary/reputation)	x				x		x			x		x
Privacy/Security									x	x		
Distributed/Centralized/Semi-Centralized	C	C	S	C	D	C	C	C	C	C	S	C
Push/Pull model	x (pull)	x (pull)	x (push)				x (push)	x (push)				x (push)
Task Delegation	x											
Social Characteristics											x	
Task Difficulty												x
Sensor Capabilities									x			
User Expertise		x									x	
Malicious User Presence		x	x	x		x	x	x	x			x
Update of Reputation (Reward/Punishment)	x		x	x		x			x	x	x	x
Priority recent interaction over the past		x	x			x						

Table 1. Comparison of Reputation Mechanisms

as to outweigh the effect of a report originating from an untrustworthy requestor. The different systems take different aspects into account when estimating participants' reputation. Most reputation mechanisms take into account data quality, identifying obsolete and inconsistent users' contributions, accordingly rewarding / penalizing the different participants, while updating correspondingly their reputation. Only a few systems take into account also other aspects (user expertise, task difficulty, sensor capabilities, sensing cost, user willingness, social relations / characteristics and users' capability in executing the task by including delegation mechanisms or endorsements and vote based mechanisms). Most systems do not discriminate between inaccurate data submitted in purpose or unintentionally due for example to transmission errors or faulty sensor readings. The authors believe that this is a focal aspect that should be taken into account and properly addressed in MCS frameworks. Additionally, at this point it should be noted that in order to take into account the time effects, the latest users' interaction should be attributed more weight than the previous ones. This has been implemented to some of the proposed systems (e.g. [27,33,36]).

VII. CONCLUSION

In this paper, we discuss about data quality challenge in MCS, highlighting also its interrelation with several other open issues pertinent to overall MCS architectural design. Reputation mechanisms are often adopted as a potential solution to ensuring data quality and integrity despite the inherently open nature of MCS systems. In this context, a representative set of reputation mechanisms proposed in related research literature is critically surveyed, identifying and discussing on the different aspects proposed. We plan to continue our work towards that direction, proposing and implementing a reputation mechanism that incorporates a multitude of different factors that influence the reputation of the participants.

REFERENCES

[1] R. Ganti, F. Ye, H. Lei, "Mobile crowdsensing: current state and future challenges", IEEE Communications Magazine, vol. 49, issue.11, pp. 32-39, 2011.

[2] B. Guo, Z. Yu, D. Zhang, X. Zhou, "From Participatory Sensing to Mobile Crowd Sensing", in Proc. of the 12th IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshop), pp. 593 - 598, Budapest, Hungary, 2014.

[3] C. Leonardi, A. Cappellotto, M. Caraviello, B. Lepri, F. Antonelli, "SecondNose: an air quality mobile crowdsensing system", in Proc. of the 8th Nordic Conference on Human-Computer Interaction, pp. 1051-1054, Helsinki, Finland, 2014.

[4] B. Pan, Y. Zheng, D. Wilkie, C. Shahabi, "Crowd sensing of traffic anomalies based on human mobility and social media", in Proc. of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, pp. 344-353, Orlando, FL, USA, 2013.

[5] C-Y. Lin, Y-Y. Su, H-H. Chu, "BikeTrack: Tracking Stolen Bikes through Everyday Mobile Phones and Participatory Sensing", in Proc. of the 2nd International Workshop on Sensing Applications on Mobile Phones, Seattle, USA, 2011.

[6] V. Coric, M. Gruteser, "Crowdsensing Maps of On-street Parking Spaces", in Proc. of the 9th IEEE International Conference on Distributed Computing in Sensor Systems, pp. 115-122, Cambridge, MA, USA, 2013.

[7] R. Salpietro, L. Bedogni, M D. Felice, L. Bononi, "Park Here! a smart parking system based on smartphones' embedded sensors and short range Communication Technologies", in Proc. of the 2015 IEEE 2nd World Forum on Internet of Things, pp. 18-23, Milan, Italy, 2015.

[8] Gambetta, D., Ed.: 'Trust: making and breaking cooperative relations' (Basil Blackwell, New York, NY [UAS], 1988).

[9] J. Sun, H. Ma, "Incentive Mechanisms for Mobile Crowd Sensing: Current States and Challenges of Work", available online: arXiv preprint arXiv:1310.8364, 2013.

[10] M. Louta, D. Roussaki, L.Pechlivanos, "Reputation Based Intelligent Agent Negotiation Frameworks in the e-Marketplace", in ICE-B, pp. 5-12, 2006.

[11] I. Varlamis, M. Eirinaki, M. Louta, "A study on social network metrics and their application in trust networks" in IEEE International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp. 168-175, 2010.

[12] I. Varlamis, M. Eirinaki, M. Louta, "Application of social network metrics to a trust-aware collaborative model for generating personalized user recommendations", in the Influence of Technology on Social Network Analysis and Mining, Springer, pp. 49-74, Vienna, 2013.

[13] M. Louta, S. Kraounakis, A. Michalas, "A survey on reputation-based cooperation enforcement schemes in wireless ad hoc networks" in Proc. of the IEEE International Conference on Wireless Information Networks and Systems (WINSYS), pp. 1-4, 2010.

[14] N. Mantas, M. Louta, E. Karapistoli, G. Karetsos, S. Kraounakis, M. Obaidat, "Towards an incentive-compatible, reputation-based framework for stimulating cooperation in opportunistic networks: a survey", IET Networks, vol. 6, issue:6, pp. 169-178, 2017.

[15] Z. Malik, A. Bouguettaya, "Reputation bootstrapping for trust establishment among web services", IEEE Internet Computing, vol. 13, issue:1, pp. 40-47, 2009.

[16] M. Louta, K. Banti, G. Karetsos, T. Lagkas, "Mobile crowd sensing architectural frameworks: a comprehensive survey", in Proc. of the 7th IEEE International Conference on Information, Intelligence, Systems & Applications (IISA), pp. 1-7, 2016.

[17] M. Marjanović, L. Skorin-Kapov, K. Pripužić, A. Antonić, I. Žarko, "Energy-aware and quality-driven sensor management for green mobile crowd sensing" Journal of network and computer applications, vol. 59, pp.95-108, 2016.

[18] H. Li, T. Li, F. Li, W. Wang, Y. Wang, "Enhancing Participant Selection through Caching in Mobile Crowd Sensing", in 24th IEEE/ ACM International Symposium on Quality of Service (IWQoS), pp.1-10, Beijing, China, 2016.

[19] R. Azzam, R. Mizouni, H. Otrok, A. Ouali, S. Singh, "GRS: A Group-Based Recruitment System for Mobile Crowd Sensing", Journal of Network and Computer Applications, vol. 72, pp. 38-50, 2016.

[20] B. Guo, H. Chen, Z. Yu, W. Nan, X. Xie, D. Zhang, "TaskMe: Toward a Dynamic and Quality-Enhanced Incentive Mechanism for Mobile Crowd Sensing", International Journal of Human-Computer Studies, vol. 102, pp.14-26, 2017.

[21] D. E. Difallah, G. Demartini, P. Cudre-Mauroux, "Pick-a-crowd: tell me what you like, and i'll tell you what to do", in Proc. of the 22nd International Conference on World Wide Web, pp. 367-374, Rio de Janeiro, Brazil, 2013.

[22] L. Cheng, J. Niu, L. Kong, C. Luo, Y. Gu, W. He, "Compressive sensing based data quality improvement for crowd-sensing applications",Journal of Network and Computer Applications, vol. 77, pp. 123-134, 2017.

[23] Manoop Talasila, Curtmola, Reza, Cristian Borcea. "Mobile crowd sensing", Google Scholar, 2015.

[24] A. Slivkins and J. Wortman Vaughan. "Online decision making in crowdsourcing markets: Theoretical challenges", ACM SIGecom Exchanges, vol.12, issue:2, pp: 4-23, New York, USA, 2013.

[25] C. Tanas, J. Herrera-Joancomarti, "When users become sensors: can we trust their readings?", International Journal of Communications Systems, vol.28, issue:4, pp. 601-614, 2015.

[26] D. Christin, C. Roßkopf, M. Hollick, L. A. Martucci, S. S. Kanhere, "IncogniSense: An Anonymity-preserving Reputation Framework for Participatory Sensing Applications", in Proc. of the 10th IEEE International Conference on Pervasive Computing and Communications (PerCom), pp. 135–143, 2012.

[27] Y. Gao, X. Li, J. Li, Y. Gao, "DTRF: A Dynamic-Trust-based Recruitment Framework for Mobile Crowd Sensing System", in IFIP/IEEE 15th Symposium on Integrated Network and Service Management (IM), pp. 632-635, Lisbon, Portugal, 2017.

[28] H. Amintoosi, S. S. Kanhere, "A Reputation Framework for Social Participatory Sensing Systems", Mobile Networks and Applications, vol. 19, issue:1, pp.88-100, 2014.

[29] Y. L. Sun, Z. Han, W. Yu, K.J. Ray Liu, "Attacks on Trust Evaluation in Distributed Networks", in IEEE Proc. of the 40th annual conference on Information sciences and systems, pp. 1461-1466, USA, 2006.

[30] Thanassis Giannetsos, Stylianos Gisdakis, Panos Papadimitratos, "Trustworthy People-Centric Sensing: Privacy, Security and User Incentives Road-Map", in Proc. of the 13th Annual Mediterranean Workshop on Ad Hoc Networking, pp. 39-46, Piran, Slovenia, 2014.

[31] S. Kraounakis, I. Demetropoulos, A. Michalas, M. Obaidat, P. Sarigiannidis, M. Louta, "A robust reputation-based computational model for trust establishment in pervasive systems", IEEE Systems Journal, vol. 9 issue:3, pp. 878-891, 2015.

[32] R. Estrada, R. Mizouni, H. Otrok, A. Ouali, J. Bentahar, "A crowd-sensing framework for allocation of time-constrained and location-based tasks", IEEE Transactions on Services Computing, vol. PP, issue: 99, pp. 1-1, 2017.

[33] C. Wu, T. Luo, F. Wu, G. Chen, "An endorsement-based reputation-system for trustworthy crowdsourcing", in IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), pp. 89-90, 2015.

[34] T. Zhou, Z. Cai, M. Xu, Y. Chen, "Leveraging crowd to improve data credibility for mobile crowdsensing", in IEEE 21st Symposium on Computers and Communication (ISCC), pp.561-568, Messina, Italy, 2016.

[35] M. Pouryazdan, B. Kantarci, T. Soyata, H. Song, "Anchor-Assisted and Vote-Based Trustworthiness Assurance in Smart City Crowdsensing", IEEE Access, vol.4, pp. 529-541, 2016.

[36] K. Lun Huang, S. S. Kanhere, W. Hu, "Are you contributing trustworthy data? the case for a reputation system in participatory sensing" in Proc. of the 13th ACM international conference on Modeling, analysis and simulation of wireless and mobile systems, pp.14-22, Bordum, Turkey, 2010.

[37] B. Kantarci, P. M. Glasser, L. Foschini, "Crowdsensing with social network-aided collaborative trust scores", in IEEE Global Communications Conference (GLOBECOM), pp.1-6, San Diego, USA, 2015.

[38] R. Yu, R. Liu, X. Wang, J. Cao, "Improving Data Quality with an Accumulated Reputation Model in Participatory Sensing Systems", Sensors, vol.14, issue:3, pp.5573-5594, 2014.

[39] X. Wang, W. Cheng, P. Mohapatra, T. Abdelhazer, "ARTSense: Anonymous Reputation and Trust in Participatory Sensing", in 32nd IEEE International Conference on Computer Communications (INFOCOM), pp. 2517-2525, Turin, Italy, 2013.

[40] S. Delgado-Segura, C. Tanas, J. Henrera-Joancomartf, "Reputation and Reward: Two Sides of the Same Bitcoin", Sensors, vol.16, issue: 6, 2016.

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