INTEGRATED SECURITY AND PRIVACY PRESERVATION APPROACH IN MOBILE CROWD SENSING

OWOH NSIKAK PIUS

UNIVERSITI SAINS MALAYSIA

2020

INTEGRATED SECURITY AND PRIVACY PRESERVATION APPROACH IN MOBILE CROWD SENSING

by

OWOH NSIKAK PIUS

**Thesis submitted in fulfilment of the requirements
for the degree of
Doctor of Philosophy**

May 2020

ACKNOWLEDGEMENT

I am grateful to God Almighty who in His infinite mercy gave me the strength and wisdom to carry out this research work. The past years on my PhD research at Universiti of Sains Malaysia are sincerely remarkable, rewarding and precious time in my life. I would like to thank all the people who directly or indirectly supported my PhD study.

Firstly, I would like to express my sincere gratitude to my Supervisor Dr Manmeet Mahinderjit Singh for her patience, motivation and immense knowledge. Her guidance helped me throughout the writing of this thesis and the entire research. My sincere thanks also go to my Co-supervisor, Dr Zarul Fitri Zaaba, for his contributions and suggestions in the cause of my research.

Secondly, I thank the School of Computer Sciences and the Management of Universiti Sains Malaysia for granting me this opportunity to conduct my research at their institution. I also thank them for the financial assistance I enjoyed during my research.

I also place on record, my special thanks to my family, especially my dad, mum and sisters for standing by me throughout the period of conducting this PhD research. I truly appreciate all the prayers and motivation I enjoyed from you.

Last but never least, I thank my friends and colleagues for their encouragement and support during the entire journey of this research. Special thanks go to Dr Rogers for his technical input in this research work. I appreciate you all.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF SYMBOLS

|L|	Absolute value of L
|V|	Absolute value of V
|X|	Absolute value of X
|Y|	Absolute value of Y
|Z|	Absolute value of Z
$\triangle$	Additional state information
G	Additive Group
P	Arbitrary Generator of Group
$\in$	Belongs To
$\hat{e}$	Bilinear pairing
ℓ	Bit-length of plaintext
C_i	Ciphertext
(T_i, W_i)	Encryption parameters
$\bigoplus$	Exclusive OR
$\bigoplus$	Exclusive OR
F	Features (X, Y, Z, L, V)
Q, a, b	Group parameters
H	Hash function
h'_a	Hashed message
i	Identity
s	Master key
s	Master private key
P_{pub}	Master public key
m_i	Message

G_T	Multiplicative Group
$\{ID_i\}_i^n$	Number of users with identity
S	Observations
D_i	Partial private key
Y_i	Partial Public key
q	Prime order
d	Private (secret value)
$\cdot$	Product
r	Random number
ID_R	Receiver’s identity
k	Security parameter
ID_i	Sender’s identity
β_i	Signcryption parameter
K	Total number of clusters
n	Total number of samples

LIST OF ABBREVIATIONS

2D-DCT	Two Dimensional Discrete Cosine Transform
ADL	Activity of Daily Living
AES	Advanced Encryption Standard
API	Application Programming Interface
App	Application
BC	Blockchain
CIA	Confidentiality Integrity Authentication
CLASC	Certificateless Aggregate Signcryption
CLPKC	Certificateless Public Key Cryptography
DLP	Discrete Logarithm Problem
ECC	Elliptic Curve Cryptography
ECDH	Elliptic Curve Diffie-Hellman
ECDSA	Elliptic Curve Digital Signature
EHR	Electronic Health Record
ESP	Edge Service Provider
GPRS	General Radio Packet Services
GPS	Global Positioning System
GSM	Global System for Mobile Communication
HMM	Hidden Markov Model
HTTP	HyperText Transfer Protocol
HTTPS	HyperText Transfer Protocol Secure
IDC	International Data Corporation
IoT	Internet of Things
ISPPA	Integrated Security and Privacy-Preservation Framework

KGC	Key Generation Centre
KMC	Key Management Centre
KNN	K-Nearest Neighbour
LAN	Local Area Network
MCS	Mobile Crowd sensing
MFPCVI	Multiple Pair-Frequency Cluster Validation Index
mHealth	Mobile Health
MITM	Man-In-The-Middle
Mod	Modulo
MQTT	Message Query Telemetry Transport
MSF	Mobile Sensing Framework
NB	Naïve Bayes
NoSQL	Not Only Structured Query Language
PII	Personally Identifiable Information
PKC	Public Key Cryptography
PKI	Public Key Infrastructure
PLAKAP	Pairing-less Authenticated Key Agreement Protocol
RZS	Reduced Zigzag Scan
SCSD	Signcryption of Sensitive Data
SHA	Secure Hash Algorithm
SKA	Spatial k-Anonymity
SSK	Selective Seed Key
SSL	Secure Socket Layer
SVM	Support Vector Machine
TLS	Transport Layer Security
WSN	Wireless Sensor Network

LIST OF APPENDICES

KERANGKA PENCEGAHAN KESELAMATAN DAN PRIVASI BERSEPADU DALAM PENGESANAN KHALAYAK MUDAH ALIH

ABSTRAK

Proliferasi pelbagai peranti mudah alih seperti telefon pintar dan tablet dengan sensor terbenam dan ciri-ciri komunikasi memperkenalkan paradigma penderiaan novel yang dikenali sebagai penderiaan kumpulan mudah alih (MCS). Meskipun dengan pelbagai kelebihan dan peluang, penderiaan kumpulan mudah alih masih berhadapan isu-isu privasi dan keselamatan. Salah satu isu major dengan perlindungan data secara efektif (seperti maklumat lokasi) bagi pengguna-pengguna MCS ialah ketidakupayaan untuk mencatat bacaan mentah dari sensor GPS telefon pintar disebabkan keadaan “buka” dan “tutup” yang terwujud. Satu masalah lain ialah mengesan data hasad ketika peringkat penderiaan MCS. Kaedah yang diutamakan untuk menyelesaikan masalah yang dikenal pasti tanpa memberi kesan kepada pengalaman pengguna ialah dengan menulis data secara automatik “sensitif”, “tidak sensitif” dan “hasad”. Kaedah ini walau bagaimanapun memperkenalkan masalah yang terwujud dengan catatan data automatik yakni pengesahan berkelompok. Walau bagaimanapun, interaksi ini rentan dengan serangan-serangan seperti orang di tengah disebabkan tiadanya saling pengesahan antara pelayan dan pelombong. Berdasarkan masalah-masalah ini, penyelidikan ini mencadangkan kaedah pemeliharaan privasi dan keselamatan bersepadu dalam penderiaan kumpulan mudah alih. Kaedah yang dicadangkan melibatkan model yang mengesahkan kelompok-kelompok (sensitif dan tidak sensitif) yang dihasilkan dari algoritma *k*-means yang dirujuk sebagai Pengesahan Indeks Kelompok Kekerapan Pasangan Berganda. Set data sekunder tidak berlabel digunakan untuk melatih dan menguji model pengesahan dan catatan. Ia juga mengintegrasikan pemampatan data dan protokol Pengangkutan Telemetri Pertanyaan Mesej untuk meminimumkan overhed komunikasi dan pengkomputeran apabila melindungi data sensor menggunakan Tandatangan Penyulitan Agregat Tanpa Sijil. Tambahan pula, kaedah yang yang dicadangkan mengurangkan serangan orang di tengah antara pelayan penderiaan kumpulan mudah alih dan pelombong blok rantai dengan pelaksanaan protokol Kunci Perjanjian Pengesahan Tanpa Pasangan bagi menjamin saling pengesahan. Set-set data primer yang dikutip dari sensor-sensor telefon pintar juga serangan simulasi digunakan sebagai penanda aras dalam kajian ini. Keputusan dari model pengesahan yang dinilai menunjukkan kejituan 98.2 % untuk proses melabel kelompok dari algoritma *k*-means. Selain itu, data mampatan menggunakan Teknik Pengekodan Ruangan sebelum penyulitan tandatangan data dan penyepaduan protokol MQTT dalam protokol pemindahan hiperteks meminimumkan overhed komunikasi dan pengkomputeran secara luar biasa dengan 18.04 milisaat dan 48 bait masing-masing. Sementara itu, cadangan protokol Kunci Perjanjian Pengesahan Tanpa Pasangan untuk saling pengesahan menggunakan kos komunikasi dan pengkomputeran yang lebih baik dengan 18.04 milisaat dan 48 bait masing-masing. Kerja yang dicadangkan dalam penyelidikan ini adalah teguh terhadap serangan dalaman istimewa, serangan ulangan, serangan pemalsuan dan serangan identiti. Hasil boleh serah dari kaedah yang dicadangkan dalam penyelidikan ini ialah antaramuka pengaturcaraan aplikasi penderiaan selamat dan penghantaran data lokasi sensitif antara telefon pintar dan pelayan MCS di pelbagai domain MCS. Ini juga memastikan komunikasi selamat antara pelayan-pelayan MCS pelombong blok rantai apabila melaksanakan interaksi blok rantai IoT hibrid.

INTEGRATED SECURITY AND PRIVACY PRESERVATION APPROACH IN MOBILE CROWD SENSING

ABSTRACT

The proliferation of mobile devices such as smartphones and tablets with embedded sensors and communication features introduces a novel sensing paradigm called mobile crowd sensing (MCS). Despite its opportunities and advantages, mobile crowd sensing still faces security and privacy issues. One major issue with effectively protecting sensitive data (such as location information) of users in MCS is the inability to annotate raw readings from smartphone GPS sensor due to its inherent "on" and "off" state. Another problem is detecting malicious data at the sensing stage of MCS. A preferred method to solve the identified problem without affecting the user experience is to annotate "sensitive", "non-sensitive" and "malicious" data automatically. This technique, however, presents an inherent problem with automatic data annotation, which is cluster validation. Recently, Certificateless Aggregate Signcryption schemes have been employed in mobile crowd sensing for signing and encryption of data. Unfortunately, these schemes incur communication and computational overhead during implementation. Lastly, blockchain has been integrated with the Internet of Things to validate the integrity of stored data. However, this interaction is vulnerable to attacks such as man-in-the-middle due to the lack of mutual authentication between servers and miners. Based on these problems, this research work proposes an integrated security and privacy preservation approach in mobile crowd sensing. The proposed approach includes a model that validates clusters (sensitive and non-sensitive) generated from *k*-means algorithm, which is referred to as Multiple Pair Frequency Cluster Validation Index. Unlabelled secondary datasets are used to train and test the annotation and validation model. It also integrates data compression and Message Query Telemetry Transport protocol for the minimisation of computational and communication overhead when protecting sensor data using Certificateless Aggregate Signcryption schemes. Furthermore, the proposed approach mitigates man-in-the-middle attacks between mobile crowd sensing servers and blockchain miners by implementing a Pairing-less Authenticated Key Agreement protocol by ensuring mutual authentication. Primary datasets collected from smartphone sensors as well as simulated attacks are used to benchmark the proposed approach. Results from the evaluated validation model show an accuracy of 98.2% for the cluster labelling process from the *k*-means algorithm. Also, compressing data using the spatial coding technique before data signcryption and the integration of MQTT protocol in place of HyperText Transfer Protocol minimises computational and communication overhead remarkably to 18.04 milliseconds and 48 bytes respectively. Meanwhile, the proposed pairing-less authenticated key agreement protocol for mutual authentication uses even computational and communication cost of 4.03 milliseconds and 42 bytes respectively. The proposed approach in this research work is robust against privileged insider attack, replay attack, forgery attack, and identity attacks. The deliverable from the proposed approach in this research work serves as an application programming interface for the secure sensing and transmission of sensitive location data between smartphones and MCS servers in various MCS domains. It also ensures secure communication between MCS servers and blockchain miners when implementing a hybrid IoT-blockchain interaction.

CHAPTER 1

INTRODUCTION

1.1 Background

Internet of Things (IoT) is a dynamic and global network infrastructure for linking together the physical and virtual world, using standard and interoperable communication protocols (Distefano et al., 2013). In IoT, “things” can interact and communicate with each other and with the environment using service interfaces. Recent research published in Longo et al. (2018) shows that 85% of companies worldwide will implement IoT technology by the end of 2019. Before this, Gartner (Hung, 2017) predicted that a total of 26 billion IoT objects would be connected by 2020.

Many technologies that require the collection of ubiquitous information rely on the Internet of Things. Of recent interest is mobile crowd sensing (MCS) (Ganti et al., 2011). MCS refers to the numerous sensing platforms that enable carriers of sensing and computing devices such as smartphones, tablets and wearable devices to acquire and share essential data for various applications (Ganti et al., 2011). A mobile Sensing System (MSS) incorporates a user-level Application (APP) running on the phone that captures internal phone’s sensor(s). Captured sensor data from these sensors are transmitted to the server for further processing and storage (Macias et al., 2013). For this to happen, the phone’s operating system must provide an Application Programming Interface (API) to coordinate the data sensing and reporting processes. A typical MCS architecture consists of a requester (campaign administrator), who initiates a sensing task(s) via application servers or platforms; participants who contribute sensor data; and end-users who make use of aggregated data from sensing application servers. In recent years, MCS has revolutionised to become an important sensing mechanism.

Furthermore, the pervasive nature of mobile devices (smartphones and tablets) carried by users can be exploited to offer complex computation and sensing services (Yang et al., 2018). A report made available by the International Data Corporation (IDC) shows that as of 2018, about 1.9 billion smartphone units were shipped from different manufacturers (GSS, 2018). These figures justify a large number of smartphone users who are potential participants of sensing task and activities. The increasing number of mobile devices such as smartphones and tablets with embedded sensors and communication features is one reason why MCS is a widely accepted sensing platform (Wu et al., 2016; Miao et al., 2018).

Currently, smartphones do more than serving as computing and communication devices, as they can now perform sensing tasks using in-built sensors such as GPS, magnetometer, digital compass, accelerometer, camera and microphone (Lane et al., 2010). Together, these sensors aid the development of novel applications across several domains such as transportation (Ma et al., 2013), healthcare (Khan et al., 2013), social networks (Guo et al., 2015a), safety (Tsung-Te Lai et al., 2011), and environmental monitoring (Leonardi et al., 2014), thereby expanding the applicability of mobile crowd sensing. Based on the service-oriented architecture (SOA), mobile crowd sensing model comprises of three layers, which are, the sensing layer, the network layer and the application layer (Tiburski et al., 2015). At the sensing layer, raw data from humans and the environment are collected from smartphone sensors. Data pre-processing and annotation are performed at the sensing layer. Annotated data are aggregated at the network layer, then transmitted to the application layer, where they are visualised by end-users (Jing et al., 2014).

Different from traditional sensing approaches (such as wireless sensor networks) that depend on inert sensors or dedicated monitoring stations, MCS allows humans to carry out sensing tasks at different locations and times. Such a possibility is achieved with their mobile devices (Distefano et al., 2013). The advancement in mobile technology has been key to the advantages of MCS over traditional sensing technologies. Firstly, the availability of affordable smartphones with integrated sensors has enabled the development of several landmark applications. Furthermore, the programmability of smartphones supports novel sensing applications such as user's real-time activity shared with friends on social networks. Secondly, apart from sensing, mobile devices like smartphones have computing and communication features which allow programmers to deploy third-party applications. Thirdly, the availability of app stores by phone vendors allow sensing application developers to ship out novel applications at large-scale. Such large-scale sensing was not possible with previous sensing technologies like wireless sensing networks (WSNs). Fourthly, developers can offload mobile services to backend servers, thereby ensuring additional computing resources that aid advanced features in sensing applications (Lane et al., 2010). Examples are user feedback and persuasion apps.

Despite its benefits, MCS applications still face challenges that include quality and reliability of sensed data (data and user trustworthiness) (Talasila et al., 2015b); incentivisation of participants (Wen et al., 2015; Jin et al., 2016), energy consumption of mobile sensing devices (Ganti et al., 2011; Ma et al., 2014), sensor data annotation (Radu et al., 2016; Hammerla et al., 2016; Ronao and Cho, 2016; Ordóñez and Roggen, 2016), security and privacy (He et al., 2015; Zhang et al., 2014b). The quality and reliability of sensed data is a lingering issue in MCS applications, as participants could deliberately report low-quality or falsified data. Furthermore, the quality of sensed data reduces when data from faulty sensors are collected and recorded during sensing activities. To improve data quality in MCS, data selection, quality estimation and fault filtering techniques are necessary. However, user's participation determines the quality of collected data, which makes the incentivising of users important in achieving a successful MCS system (Liu et al., 2015).

Security and privacy is another pressing issue in MCS, raising concerns about the collection and usage of personal data. In MCS, sensitive information of users such as their location details is vulnerable to privacy attacks (Guo et al., 2014). An adversary can intercept MCS traffic and capture sensitive information of users contained in sensor data. For example, GPS sensor readings can be used by an adversary to infer personal information of individuals about their daily routes to work and their home locations (Ganti et al., 2011).

1.2 Motivation of the Study

Mobile crowd sensing applications use efficient data mining techniques to analyse data, detect Spatio-temporal patterns, generate models and perform predictions on observed physical phenomena (Ganti et al., 2011). Data in MCS are acquired both from the physical world (sensed data from smartphones) and from online communities (mobile social network services) (Guo et al., 2014). A typical scenario is with smart city applications that actualise the sustainable development goals (SDGs).

Smart city applications, for example, collect sensor data that contain sensitive information of participants, such as location traces, and time (Christin et al., 2011b; Huang et al., 2012). When participants contribute sensor data to sensing campaigns for the improvements of cities, their most visited and real-time locations can be revealed to an adversary, if not effectively protected. The potential security risk with sensor data might hinder the vision 2020 Sustainable Development Goals for which smart city is an essential element. More participants will take part in sensing tasks when they are assured that their privacy can be preserved (Huang et al., 2012). To achieve effective security of location information, validation of annotated sensor readings is necessary.

However, in the real-world, GPS signals are not constantly available during mobile sensing, especially when the users are indoors (Rawassizadeh et al., 2016). Also, users may deliberately turn off their GPS sensor to conserve their phone battery while using other sensors like accelerometer, gyroscope and magnetometer. In Miluzzo et al. (2008), the authors report that in a day, a typical smartphone user uses only 4.5% GPS signal. Securing location data is a difficult task due to the inconsistencies in the acquisition of GPS data streams (that is, “on” and “off” state of GPS sensor). For any MCS security framework to be holistic, there is need for the identification of the sensitive data that requires protection. Such identification can be achieved through data annotation

1.3 Problem Statement

In MCS, data is classified as sensitive when it contains personally identifiable information (PII) (Guo et al., 2015b; Xiao and Xiong, 2015). Such information includes users’ location or mobility traces (home or work addresses), as well as their identities. However, the issue of identity disclosure can be tackled with the use of pseudonyms (Ma et al., 2017). Nevertheless, privacy concerns of MCS applications participants remain due to the possible disclosure of their location traces to adversaries as a result of inadequate security mechanisms on such sensitive information (Pournajaf et al., 2014a; Li et al., 2018a).

Considering the three-layer MCS model (sensing, network and application layers), which is similar to the IoT architecture, sensitive location information of users needs to be protected at all layers. Security at the sensing layer requires protection of sensitive information during sensing. At the network layer, sensitive data in-transit must be protected. Meanwhile securing sensitive data at the application layer requires data protection during storage and visualisation. Presently, existing security and privacy frameworks such as PRISM (Das et al., 2010), AnonySense (Shin et al., 2011), PEPSI (De Cristofaro and Soriente, 2013) and InvisibleHand (Liu et al., 2017a) do not protect sensitive location attributes at all layers of the MCS model. Moreso, these frameworks do not label sensitive location information of users during sensing. This limitation is one reason for the successful attacks that lead to privacy leakage (Li et al., 2018a).

Secure sensing must be achieved at the perception (sensing) layer of MCS by ensuring that malicious data are detected while accurately annotating sensitive information (such as location attributes) of MCS users. Existing security and privacy framework mentioned above lack this feature. To this end, there is a need for a technique that models the peculiar “on” and “off” state of GPS signals on smartphones in order to automatically annotate sensitive location information. The lack of annotation of sensitive data in MCS before securing them makes security implementation a non-trivial process, bearing in mind the large amount of data e received from smartphone sensors at a given time. On the other hand, the non-detection of malicious data (bogus data) during sensing as exhibited by existing frameworks leads to aggregation and processing of corrupted data during sensing. Though automatic annotation can be used to detect malicious and sensitive data, one challenge that remains is the validation of generated clusters from the automatic annotation process.

At the second layer of the model (network layer), existing frameworks employ traditional cryptographic mechanism such as TLS/SSL to secure the channel used for the transmission of sensitive location information of MCS users. Using this security mechanism offers confidentiality, integrity and authentication. However, traditional PKC operations employed by TLS/SSL are computationally expensive (Basudan et al., 2017). Also, the use of certificate authorities (CAs) by existing frameworks introduces the key escrow problem (He et al., 2012). Lastly, since messages are not digitally signed when employing TLS/SSL, non-repudiation is not offered as a security service by existing frameworks (Eslami and Pakniat, 2014). A security scheme that addresses the identified challenges with the traditional PKC (that is, TLS/SSL) is the certificateless aggregate signcryption CLASC, which is an example of the certificate-less public key cryptography (CLPKC). This novel security scheme achieves signing and encryption in one logical step (Al-Riyami and Paterson, 2003). Apart from ensuring confidentiality, integrity and authentication, CLASC schemes also offer non-repudiation which is an important security service in MCS. It also eliminates the key escrow problem as KGC only generate partial keys to users. Proposed works in Lu and Xie (2011); Eslami and Pakniat (2014) and (Basudan et al., 2017) have employed CLASC to achieve CIA and non-repudiation by signing and encrypting sensitive information in MCS. Though less expensive than the traditional PKC (TLS/SSL), its implementation in MCS still requires minimisation of computational and communication overhead. To this end, techniques that will reduce these overheads are needed to enhance the performance of CLASC schemes proposed in Eslami and Pakniat (2014) and (Basudan et al., 2017).

Sensor data are stored and displayed to users in the third layer (application layer) of the MCS model. The integrity of sensor data needs to be validated at this layer to ensure that data has not been tampered with during storage. Integrity validation ensures that only unmodified data are displayed to users during visualization. Presently, existing security and privacy frameworks in MCS do not offer this functionality, which makes it possible for end-users to visualise falsified and inaccurate data. Recently, blockchain technology has been integrated with the IoT to offer security to stored data (Dorri et al., 2016). Its practicality is achieved when implementing a hybrid IoT-blockchain interaction (using smart contracts), where only the metadata of sensor data is stored in the blockchain while the sensor data are stored in off-chain storage. However, this interaction is susceptible to attacks such as MITM attack between the IoT server and the blockchain miner due to the lack of mutual authentication with smart contracts (Jesus et al., 2018; Fernández-Caramés and Fraga-Lamas, 2018; Kshetri, 2017). To thwart such attacks, there is a need for a key agreement protocol that will ensure mutual authentication between the server (MCS) and the blockchain miners.

The entire problem statement can be summarised as the lack of a framework or technique in MCS that ensures effective and efficient security to sensitive location information of users in all three layers of the MCS model. Specifically, the sensing layer of MCS lacks a model that can automatically annotate sensitive location data and detect malicious data. Existing frameworks in MCS also lack a model that can validate cluster outputs from the automatic annotation of unlabeled data. On the other hand, high computational and communication overheads are associated with the implementation of security mechanisms in the network layer. Lastly, existing security frameworks in MCS do not incorporate a mechanism that validates the integrity of stored sensor data. This problem can be addressed using the hybrid IoT-blockchain interaction which is achievable through smart contracts. However, the vulnerability of smart contracts to attacks due to lack of mutual authentication between the MCS server and the miners remains a challenge.

1.4 Research Questions

The research to be conducted is guided by the following detailed research questions:

1. Which technique can be used to validate annotated data in order to ensure secure sensing and preserve the privacy of users at the sensing layer of MCS?
2. Which efficient method(s) can be used to secure data-in-transit at the network layer of MCS?
3. Which secure approach can be applied to validate the integrity of stored sensor data at the application layer of MCS?

1.5 Research Objectives

The aim of the research is to develop a security and privacy-preserving approach in mobile crowd sensing that efficiently secures sensitive location information at the three layers of the MCS model. The specific objectives of this thesis can, therefore, be broken down into the following:

1. To propose a mathematical validation model for sensitive and non-sensitive clusters that are automatically annotated from *k*-means algorithm.
2. To integrate data compression and message query telemetry transport protocol to minimise computation and communication overhead when implementing certificateless aggregate signcryption schemes.
3. To propose a pairing-less authenticated key agreement protocol that ensures mutual authentication between MCS servers and blockchain miners in a hybrid IoT-blockchain interaction.

1.6 Scope

Sensor data that will be considered in the study are those from motion sensors (accelerometer, gyroscope, and magnetometer) and location (GPS) sensor in mobile devices (smartphones). The selection of these sensors is based on their availability in almost all smartphones. Moreover, these sensors collect readings that are used in several mobile crowd sensing domains, making them suitable for evaluating a generic approach such as the one proposed in this study. The proposed approach will use the Android operating system running version 4.4 (KitKat) and above as the test environment. The selected Android sensing apps used for security analysis in this research are used rather than the reviewed apps (presented in Section 2.3), because of the unavailability of the reviewed apps on the Google Play store.

The approach developed in this work focuses on sensitive location data of MCS participants. The proposed approach takes into consideration the fact that both participants and the MCS server are semi-honest users, which means that they may at any point in time backup users' information in order to infer sensitive information from sensor data. Potential adversaries envisioned in the research study are malicious participants and the MCS server (internal attackers) and other end users (external attackers) of the MCS system. The validation model in this research work only focuses on generated cluster labels from the *k*-means algorithm and not any other clustering algorithm.

When integrating a private blockchain in the proposed framework, only non-malicious (verified) nodes will be authenticated as miners for the validation and appending of metadata into the blockchain. This research work limits the number of miners to five. Meanwhile, integrity validation of off-chain data during retrieval and visualisation is outside the scope of this research. Lastly, SHA-256 is the hashing algorithm used in the smart contract.

1.7 Contributions of the Research

This research work revealed that sensitive location data of MCS users are not secured in all three layers of the MCS model. From the conducted analysis, most existing security and privacy frameworks only secure data at the network layer of the MCS model. More so, the security mechanisms employed in the network layer experiences performance issues such as computational and communication overheads. The main contribution of this research is to develop an approach that effectively and efficiently secures sensitive location data in MCS in order to preserve the privacy of users at all three layers of the MCS model. The detailed contributions of this research work are as follows.

1. Establish a mathematical model for the validation of cluster labels that are automatically annotated from *k*-means algorithm.

Recently in MCS, automatic annotation has been applied in some works for labelling activities (such as walking, running, sitting and driving). However, applying this technique to label sensitive data such as location attributes remains unexplored in MCS. Furthermore, automatic annotation faces a significant challenge which is validating cluster labels generated from clustering algorithms such as *k*-means. Few efforts have been made in tackling the identified problem. To this end, this research work presents a Multiple Pair-Frequency Cluster Validation Index (MPFCVI) for the validation of clusters generated from *k*-means algorithm. The proposed model validates sensor data grouped as “sensitive” and “non-sensitive”.

2. Minimise computational and communication overhead in certificateless aggregate signcryption scheme by integrating data compression technique and a message query telemetry transport protocol.

Certificateless aggregate signcryption (CLASC) is used to ensure confidentiality, integrity, authentication and non-repudiation of messages transmitted in client-server communication. However, existing CLASC schemes proposed in MCS experience high computational and communication overhead due to the colossal amount of sensor data as well as the application protocol used. A solution to the mentioned problem is proposed in this research work. In this research work, spatial coding is used as the data compression technique before signcryption is employed. This approach reduces remarkably the computational cost associated with CLASC. Also, implementing MQTT in place of HTTP for the delivery of signcrypted sensitive data minimises network traffic, hence reduces the communication overhead experienced when implementing CLASC schemes.

3. Introduce mutual authentication between MCS servers and blockchain miners in a hybrid IoT-blockchain interaction.

The integrity of stored sensitive data can be validated by incorporating blockchain technology with mobile crowd sensing. The numerous blockchain security functionalities can be harnessed in MCS using the Ethereum smart contracts. However, smart contracts are vulnerable to attacks such as MITMA. Existing works that integrate blockchain with MCS have failed to solve this problem with smart contracts. In light of the above, the proposed approach offers a solution to the lingering problem by implementing a pairing-less authenticated key agreement protocol that ensures mutual authentication between MCS servers and blockchain miners (that is, when using smart contracts).

1.8 Thesis Organization

The thesis is organised as follows. The first Chapter presents the background of the study relating to mobile crowd sensing (MCS) and identifies the problems associated with this emerging sensing paradigm. The Chapter also defined the aim and objectives of this research. The second Chapter discusses more on mobile crowd sensing, presenting its components, frameworks, architecture, and applications. Challenges in MCS are discussed further with an emphasis on security and privacy. Proposed solutions to curb the rising security and privacy threats on sensitive location information are presented as well in this Chapter. In Chapter three, the methodology used in developing and evaluating the proposed approach is discussed. Chapter four presents the detailed implementation of the first objective, which is the validation of cluster labels from the automatic annotation of unlabelled sensor data. Chapter five presents the second and third objectives which are, the integration of data compression and MQTT protocol to enhance the performance of CLASC and the integrity validation of stored sensor data using a hybrid IoT-blockchain interaction technique. In Chapter six, the evaluation of the integrated approach (involving unit and integrated testing) including the experiments, the results and the discussion on the results are presented. Chapter seven concludes the study by highlighting the challenges, limitations, and recommendation of the research.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

The literature review discusses more on the problem area, as highlighted in Section 1.3. The review starts by explaining what mobile crowd sensing is, then presents its characteristics and advantages over traditional sensing technology (wireless sensor network). The different application areas of mobile crowd sensing are discussed. The emerging technology (mobile crowd sensing) is without its challenges, as will be shown in this Chapter. The significant issues with MCS are discussed with proposed solutions to tackle such issues. However, emphasis is on data annotation, security, and privacy for data in motion as well as secure storage for data at rest. With respect to data annotation, three machine learning techniques (supervised, semi-supervised and unsupervised) are employed for sensor data labelling and will be discussed and proposed works that use them are presented. When dealing with security and privacy in MCS, anonymity-based approaches and cryptographic techniques are mostly used to tackle this persistent threat. Proposed schemes and frameworks that utilise each of these techniques will be discussed. Lastly, a novel method of achieving secure storage of data in IoT using blockchain technology is discussed and few implementations are presented. Figure 2.1 summaries the flow of the literature review in this research work.

2.2 Mobile Crowd Sensing (MCS)

Over the years, MCS has grown to become an ideal technology for acquiring sensor data from mobile devices, to detect Spatio-temporal patterns, and predict physical and social phenomena (Banti et al., 2018).

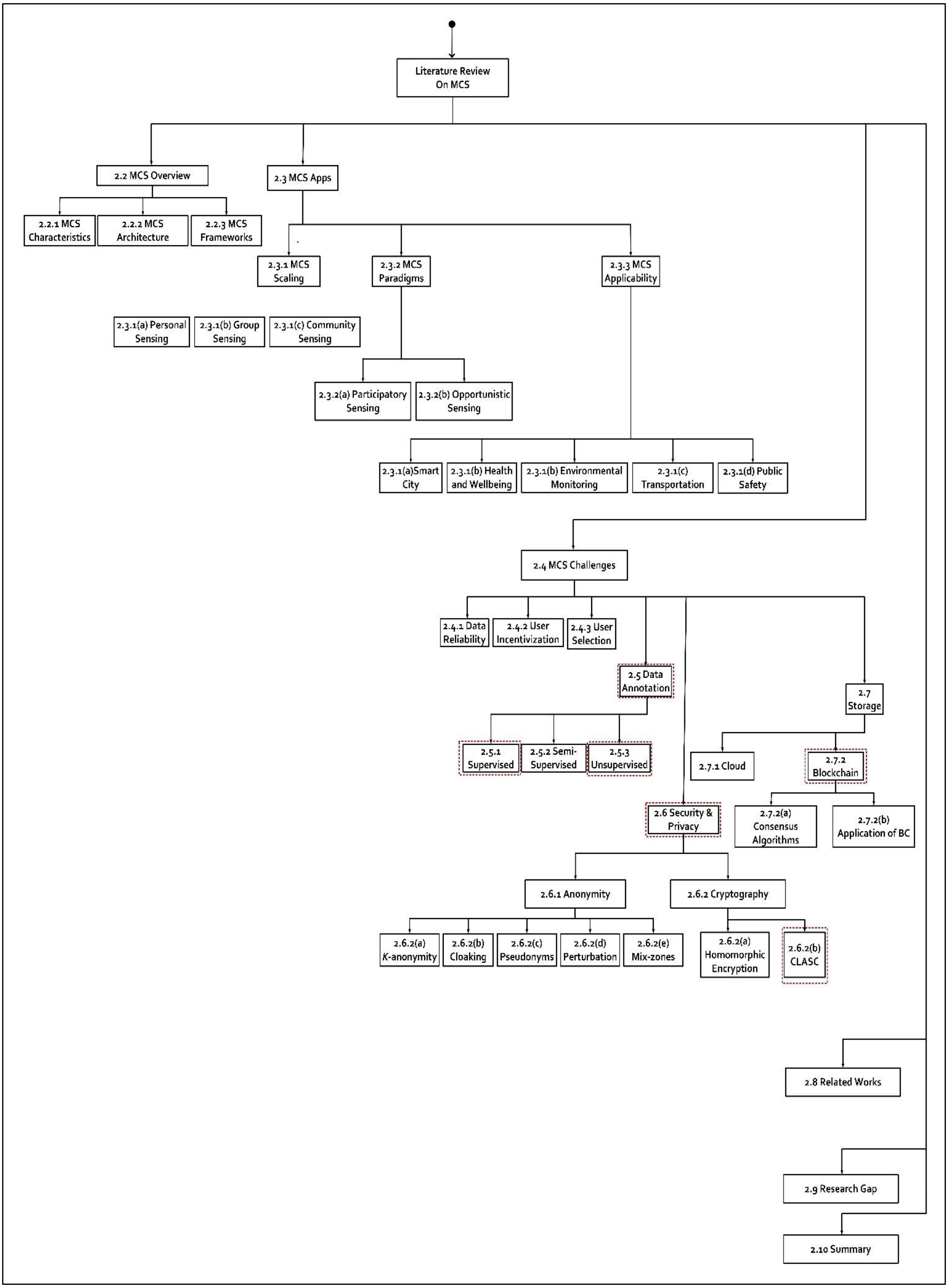

Figure 2.1 Taxonomy of Literature Review

2.2.1 Characteristics of Mobile Crowd Sensing

In wireless sensor networks (WSNs), sensor nodes are organised into a cooperative network of multiple sensor nodes which is controlled by a network administrator (Alswailim et al., 2014). As a result, the sensed data and their corresponding results are owned and controlled by the WSN operator. Conversely, mobile crowd sensing systems do not have a sole data operator, as different participants can contribute to a single application (Alswailim et al., 2014).

Furthermore, mobile crowd sensing systems employ existing sensing and communication infrastructure in smartphones, making their deployment cost near zero compared to WSN (Christin et al., 2011b). Also, the mobility of smartphone users in mobile crowd sensing offers broader coverage in the case of random events; leading to several urban-scale sensing applications (Hu et al., 2013; Antonić et al., 2016). Furthermore, smartphones and tablets used presently as sensing devices have more storage resources than traditional sensors. Sensors in MCS can be employed for multiple applications, whereas in traditional sensing, a sensor is meant for a particular application (Gunasekaran and Rathnamala, 2013). Table 2.1 summaries the differences between MCS and traditional WSN.

Table 2.1 The Differences between MCS and Traditional Sensor Networks (Hu et al., 2013; Antonić et al., 2016)

	Operators	Data Quality	Deployment Cost	Coverage
Mobile Crowd Sensing (MCS)	-Any Smartphone user	-Low, prone to built-in sensor performance issues -Also, issues with data trustworthiness from contributed data	-Low: Makes use of existing infrastructure	-Large-scale: Leverages the mobility of phone carriers to offer unprecedented Spatio-temporal coverage
Wireless Sensor Network (WSN)	-Government agencies, public institutions	-High, sound level sensors	-High: Expensive sensors and infrastructure for network deployment	-Limited coverage area due to static sensor odes

2.2.2 Architecture and Components of Mobile Crowd Sensing

IoT systems have been actualised based on the concept of service-oriented architecture (SOA) and resource-oriented architecture (ROA). The IoT systems architecture can be classified into four layers (Jing et al., 2014), perception, network, middleware and application layers as depicted in Figure 2.2. The perception (or sensing layer) coordinates the sensing and management of physical devices and gathering of sensor data from sensing devices (Tiburski et al., 2015). The network layer offers pervasive network access for devices in the perception layer and connects devices in the perception layer to the upper layers. The application layer serves as a platform where IoT applications can be developed. It also allocates logical and computational resources to sensing devices and applications. Services and interfaces are hosted in this layer. Most importantly, end users can visualize sensing results from the application layer.

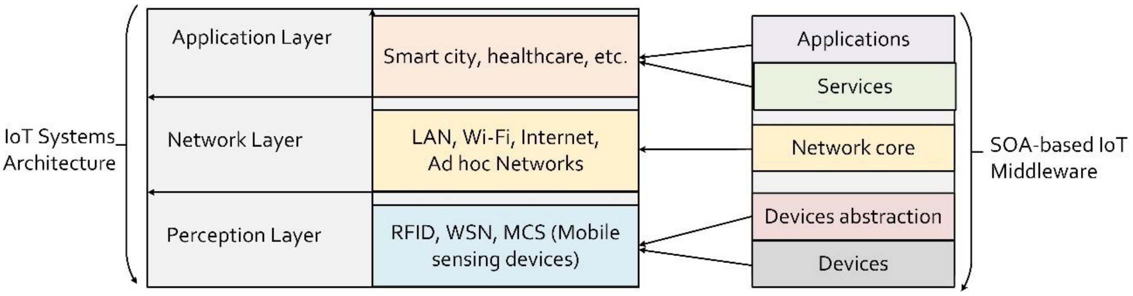

Figure 2.2 The IoT Systems Architecture and SOA-based IoT Middleware (Tiburski et al., 2015)

Though MCS lacks a unified architecture that can solve current problems with the deployment of MCS applications (Ganti et al., 2011), it is mostly implemented based on the IoT SOA-based architecture presented above. The following are some of the proposed architectures in mobile crowd sensing.

A general architecture of MCS that consists of the sense to learn, inform, share and persuasion phases was proposed in Lane et al. (2010). In the MCS sense phase, smartphones acquire sensor data from integrated sensors in the phone. Most of these smartphones are open and can be programmed to provide application interfaces (APIs) and software tools. Novel applications across several domains especially in personal healthcare, are possible with continuous sensing. However, for continuous sensing to be possible, the smartphone must support background processing and multitasking which are attributes of most smartphones today. Also, data sampling must be taken into consideration while designing sensing applications to make the best use of embedded sensors in the phone, as sensors have different sampling rates. However, continuous sensing with mobile devices can be energy demanding. More so, most sensing applications perform optimally under specific sampling contexts (such as specific periods or places). Such sensing apps call for event-triggered sensing which defines triggers that collect data in a context-aware manner (Bao and Roy Choudhury, 2010).

The learning phase of MCS deals with information extraction from sensor data using machine learning and data mining techniques (Lane et al., 2010). These processes take place either in the mobile cloud, directly on the phone or within a bridge between the cloud and the phone. Several factors such as communication cost, security and privacy, available computing resources and sensor fusion, determine where information extraction should be performed. Raw sensor data from the sensing phase are useless without information extraction. Presently, supervised learning approaches are the leading algorithms in developing mobile inference systems. Data in this learning technique are manually annotated (labelled), referred to as training data, are then passed to the learning algorithm (classifier), which fits a model to the classes based on the sensor data. Sensor data are generally passed to the classifier in the form of extracted features, which show the attributes that distinguish classes. Other learning algorithms are semi-supervised (i.e., only some of the data are labelled by the user) and unsupervised (i.e., no labels are given by the user).

In the inform/share/persuasion phase, different processes are involved depending on the scale of the MCS application. For instance, a personal sensing application will only inform the user, while a group or sensing application may share information with a larger population. Visualisation is a standard method of sharing data in MCS (Lane et al., 2010). Sensor data acquired from groups and communities can be employed not just to inform users but to persuade them to embark on positive behavioural changes (e.g., healthcare and fitness apps).

A typical MCS architecture, as shown in Figure 2.3, has several components that interrelate based on the client-server model (Christin et al., 2011b). The first component which performs sensing is mostly on the mobile phones of participants and collect various kinds of sensor data such as sound data, pollution data, pictures, location, time and biometric data. The sensor data can be collected using any of the following modes: manual, automatic and context-aware (Estrin et al., 2010). In the manual mode, MCS participants initiate sensing when they identify significant events, such as traffic congestion. Automatic, on the other hand, is performed opportunistically without the participant’s involvement. In context-aware settings, certain pre-specified conditions trigger the collection of sensor readings.

The second component, which is the tasking component, aids the sensing component by sharing sensing tasks to the mobile phone of participants. Sensing steps are outlined and managed by tasks based on the requirements of each application. Location and time frame of interest are some of the necessary information contained in the task. The reporting component deals with the forwarding of sensor readings gathered by the sensing component to the application server. Communication infrastructures such as Wireless LAN, or GSM/GPRS/3G connectivity are commonly used for the transmission of sensor data. The storage component stores the collected sensor data on the mobile phone and the reported data on the server. Permanent storage of reported data is stored on the server while the mobile phone stores temporary data meant for processing or transmission to the server. Sensor data from MCS applications which fall under Big data are mostly stored in Not only SQL databases (NoSQL) such as MongoDB and senseDB.

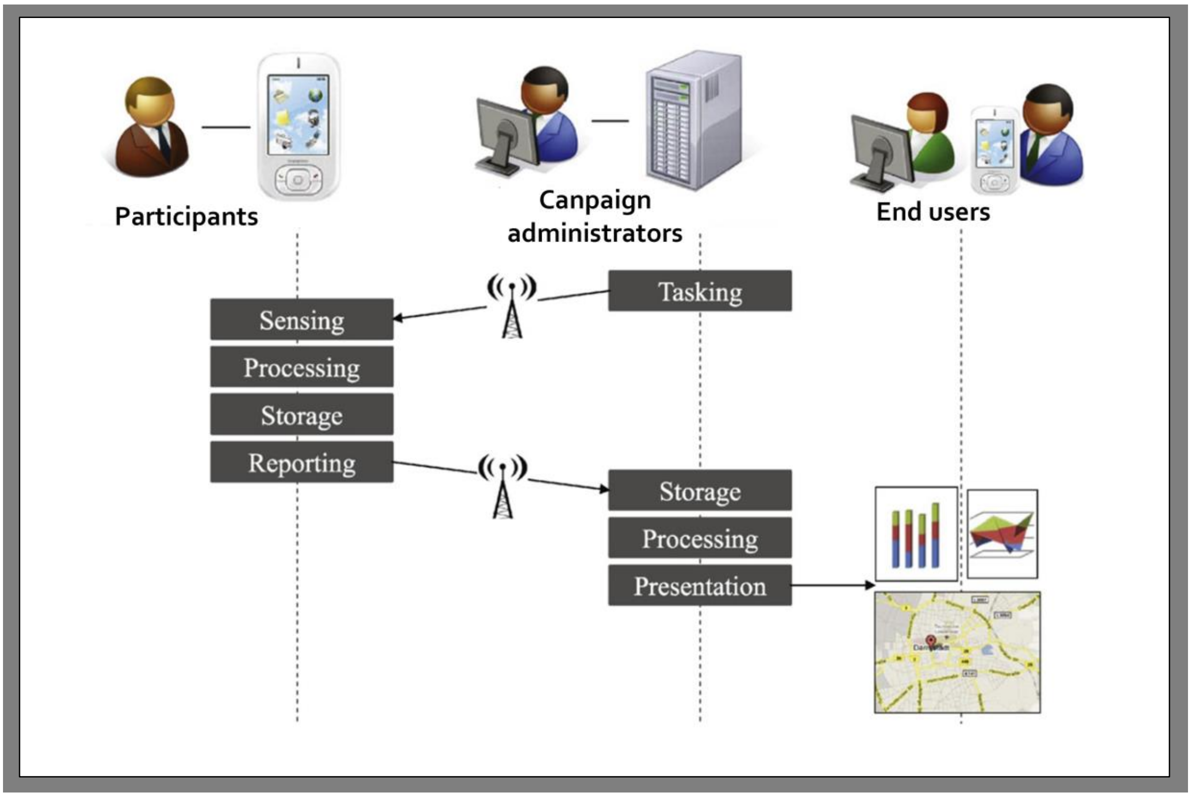

Figure 2.3 Mobile Crowd Sensing Architecture (Christin et al., 2011b)

The processing component either extracts features directly from the sensor data on the mobile phone on a personal scale or the server at a community scale. Furthermore, analyses and synthesis of collected sensor data to be forwarded to the presentation component is also done by the processing component. The last component, which is the presentation component displays the processed results from the processing components to the end-users. The results are either to the mobile phones of participants or via web portals. The results can be displayed as raw data to enable users to perform analysis themselves or presented as maps, graphs and geographic data (Christin et al., 2011b).

As shown Figure 2.4, components in MCS are represented as a platform, a set requester, and a group of mobile users (also known as workers or participants) (Sun et al., 2018). In MCS, a sensing task is published to the platform by a requester aiming to collect events happening in his/her region-of-interest. The MCS platform then recruits an appropriate group of workers (participants) and assign the sensing tasks to the recruited workers. After the raw sensor data from the workers have been collected, the platform provides the requesters with aggregated data (Koutsopoulos, 2013; Zhang et al., 2014a). In order to attract more workers to join and submit quality data in the sensing task, the requester has to incentivise workers by paying the platform. The platform then has to pay the workers when they submit quality raw sensor data in the sensing task. Typically, a platform is based on the client-server architecture, including a client that runs on users' devices (e.g., smartphones) to coordinate sensing activities. The server, on the other hand, analyses, stores and displays results to users (Lane et al., 2010).

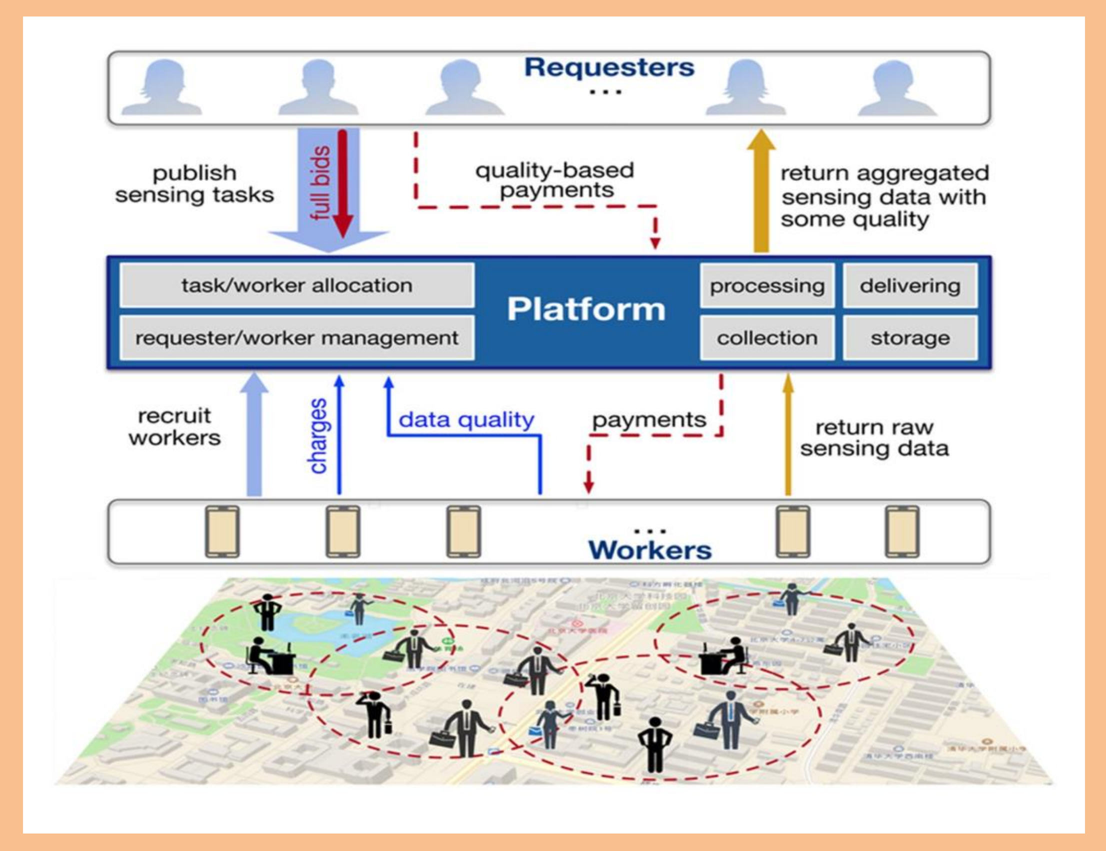

Figure 2.4 Components of Mobile Crowd Sensing (Sun et al., 2018)

At this point, it is important to discuss the different task distribution models in MCS. Task management models are grouped based on their distribution methods among MCS participants. Centralised, decentralised and hybrid models are the three groups of task management in MCS (Pournajaf et al., 2014a). In centralised models, a central server or tasking object offers participants with different tasks to carry out. A drawback with a central model is the possibility of a single point of failure for communications between participants and applications. Decentralized model, on the other hand, allows each participant to become a tasking entity and can choose whether to carry out a task or push it to other MCS participants who might be more suitable to perform the said task. Specific features of other participants such as available computing resources, the location could be used as preferences to decide which participant is most suitable for a current task. The hybrid model consists of both centralised and decentralised models (Pournajaf et al., 2014a). In this type of model, a central server and a set of participants serving as tasking objects form the task management core. The bubble scheme (Lu et al., 2010) is an example of the hybrid model. This model type has a central server that controls the sensing task which is distributed in a decentralised manner. A specific task is published in a certain location-of-interest by a participant, and in this context, referred to as a bubble creator. The server registers the task and informs other participants present in the location of interest to become bubble carriers (Pournajaf et al., 2014a).

In MCS, tasking schemes can be categorised into four groups based on their attributes or tasking objects (Pournajaf et al., 2014a). These categories are: push/pull-based, autonomous/coordinated, event-based/continuous and spatial/non-spatial models.

A. Push/Pull based model: The push or pull model rely on sensing objects that begin a sensing task. In the push model, tasks are triggered by a tasking entity by sending (pushing) tasks to participants' mobile phones. Meanwhile, in the pull model, participants query and download tasks as apps at a given time and location.

B. Autonomous/Coordinated: Allocation scheme can also be used to categorise the distribution of tasks to participants. Autonomous and coordinated task assignment fall under this group (Pournajaf et al., 2014b). In autonomous task selection, MCS participants employ tasks by autonomously selecting one or more tasks to carry out. Selection decisions made by participants do not need to be communicated to the task distributing entity. However, the sensing efficiency declines due to the lack of an optimisation algorithm for task distribution. Furthermore, bias is easily experienced from collected sensor data with autonomous sensing schemes. On the other hand, enhancing sensor data quality through optimisation of participants' recruitment in performing MCS tasks is the goal of coordinated task assignment. This optimisation is built on sensing conditions such as sensing costs, sensing coverage, sensing quality and reliability of sensed data (Pournajaf et al., 2014b).

C. Event-based/Continuous: Data querying frequency is another method of classifying various possible tasks. The frequency could either be event-based or continuous. In event-based tasks, a specific situation triggers an event and could be conditions like the availability of a participant in a particular location. In continuous tasks, information is required from participants at given periods or regularly.