Privacy-Aware Incentivisation for Participatory Sensing

Martin Connolly

A thesis submitted to the University of Dublin, Trinity College in fulfillment of the requirements for the degree of Doctor of Philosophy

2020

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work. I agree to deposit this thesis in the University's open access institutional repository or allow the library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement. I consent to the examiners retaining a copy of the thesis beyond the examining period, should they so wish (EU GDPR May 2018).

Martin Connolly

Dated: 29th March 2019

Permission to lend or copy

I, the undersigned, agree that the Trinity College Library may lend or copy this thesis upon request.

Martin Connolly

Dated: 29th March 2019

Acknowledgements

Firstly, I would like to thank my supervisor Dr. Mélanie Bouroche. Her knowledge, patience and attention to detail have been of enormous help to me over the past five years. Mélanie's supervision has been an education not just in Computer Science but in the nature of academic research and scientific writing. I am also indebted to Dr. Ivana Dusparic, who acted as my supervisor while Mélanie was on leave. Her valuable ideas and thoughts on my research were of great assistance in honing my ideas into the work that is presented here. I must also acknowledge Dr. Georgios Iosifidis, Ussher Assistant Professor in Future Networks at Trinity, whose advice on applying statistical methods to incentivisation deepened my understanding of the area and played a fundamental role in shaping a key contribution of this work.

My employers, Cork Institute of Technology, made this PhD possible. In particular, I wish to thank the registrar's office for their financial support and the Department of Accounting and Information Systems for allocating a substantial proportion of my teaching hours towards my research activity. In addition, I would like to thank Professor Noel Barry, former head of Academic Affairs at CIT, for his positive encouragement throughout the process and wish him a happy and healthy retirement. I am also grateful to Dr. J.B. McCarthy of the Department of Mathematics in CIT for the insightful questions he posed regarding initial drafts of my paper on incentivisation. These were of great assistance in the refinement of my ideas.

I am also grateful to my family and friends who were a constant positive support through this journey. In particular, I would like to thank my very good friends Phil Corcoran and Cathal Hoare who were ever willing sounding boards for my ideas and were a willing and candid audience when practicing my conference presentation skills. Thank you to my family, my Dad, Terry, for being my constant and willing companion on my trips to Dublin and my sister Joanne for her unwavering support, advice and encouragement. Finally, my nieces Freya and Lauren Healy never fail to make life more joyous. I hope in the years to come that they will appreciate the value of education and learning. I dedicate this thesis to them.

Martin Connolly,
University of Dublin,
Trinity College, March 2019

Abstract

Participatory sensing is a paradigm through which mobile device users (or *participants*) collect and share data about their environments. The data captured by participants is typically submitted to an intermediary (the *service provider*) who will build a service based upon this data.

For a participatory sensing system to attract the data submissions it requires, its users often need to be incentivised. Such an incentivisation mechanism typically requires users to at least partially disclose their identity to be able to reward them, and to ensure that they are only rewarded for truthful submissions (called incentive compatibility). This, however, might deter privacy conscious users from participating. Therefore, an incentivisation mechanism needs to support anonymous and unlinkable data submission and untraceable and unlinkable rewarding while also ensuring data truthfulness (An incentivisation scheme is not in and of itself incentive compatible but should be able to facilitate incentive compatibility). Furthermore, as an environment can quickly and suddenly change (for example, an accident causing elevated pollution levels and a buildup of traffic), the value of a given data item to the service provider is likely to change significantly over time, and therefore an incentivisation scheme must be able to adapt the rewards it offers in real-time to match the environmental conditions and current participation rates, thereby optimising the consumption of the service provider's budget.

There are numerous approaches in the state of the art in the areas of identity privacy and incentivisation for participatory sensing. For example, one approach proposes a digital currency to enable participants to make data submissions without disclosing their privacy, while another approach devises tokens to exchange data and rewards in a privacy-preserving fashion. However, while basic identity privacy may be preserved in these approaches, other forms of identity privacy such as behavioural habits and frequent trajectories can be inferred from the submitted data, with some proposed solutions actually increasing the threat of such inference attacks occurring. Furthermore, none of the approaches in the state of the art support adapting the reward to match the environmental conditions.

This thesis presents *Privacy-Aware Incentivisation (PAI)*, which is a decentralised peer-to-peer exchange to enable anonymous and unlinkable data submission, untraceable and unlinkable reward allocation and spending, and adaptive incentive-compatible reward computation. This is achieved through the modification and extension of the concept of decentralised trading for cryptocurrencies to make payments (i.e. rewards) sent to a recipient (i.e. the participant) untraceable. Furthermore, the use of the Diffie-Hellman Exchange Protocol is modified to enable participants to create their own untraceable reward currency in the form of tokens to which the service provider can then assign value. Finally, the Lyapunov Optimisation method is used to create an adaptive reward allocation model that optimises the consumption of the service provider's budget.

The principal contributions of PAI are:

1. A platform for anonymous and unlinkable data submission and untraceable and unlinkable rewarding that is robust to inference attacks from semi-honest service providers and other potential attackers.

2. A privacy-aware adaptive incentive-compatible incentivisation scheme.

PAI is evaluated by proof and by comparing the approach to the most relevant approaches in the state of the art. The privacy robustness of PAI is demonstrated by proofs showing that participants can make anonymous and unlinkable data submissions to the service provider and receive untraceable and unlinkable rewards in return. The incentive compatibility of PAI is also demonstrated by proofs showing that rewards will not be allocated for data submissions that are deemed to be non truthful with the privacy preserving character of the incentive compatibility approach also being proven. The adaptiveness and budget consumption of PAI's adaptive reward allocation method is compared with the most relevant approaches in the state of the art for reward computation using experiments carried out in a simulated participatory sensing environment. The results of these experiments are, in general, favorable with the reward allocation method adapting in a more timely fashion compared to similar approaches. Experiments are also conducted to compare the performance and computational complexity of PAI with the most relevant privacy preserving incentivisation methods proposed in the state of the art. The results of these experiments find that, in general, PAI's energy consumption is less than that of other privacy preserving incentivisation methods while its core algorithms require less resources.

Publications Related to this PhD

- Martin Connolly, Ivana Dusparic, Georgios Iosifidis and Mélanie Bouroche, *Privacy-Aware Incentivization for Participatory Sensing*, In MDPI Sensors, 2019.

- Martin Connolly, Ivana Dusparic, Georgios Iosifidis and Mélanie Bouroche, An Identity Privacy Preserving Incentivization Scheme for Participatory Sensing, In IEEE 11th International Conference on Mobile Computing and Ubiquitous Networking (ICMU), Auckland, New Zealand, October 2018.

- Martin Connolly, Ivana Dusparic, Georgios Iosifidis and Mélanie Bouroche, *Adaptive Reward Allocation for Participatory Sensing*. In Wireless Communications and Mobile Computing, 2018.

Contents

List of Figures

List of Tables

List of Algorithms

Chapter 1

Introduction

This chapter introduces the field of research for this thesis. The background to the thesis, specifically, the concept of participatory sensing, is introduced in Section 1.1. The motivation for the research undertaken in this thesis is presented in Section 1.2 while Section 1.3 discusses its context and briefly presents related work in the area. The challenges to be addressed are discussed in Section 1.4 while Section 1.5 introduces the proposed approach to be taken. The goals of the research and the principal contributions made by the approach are explored in Section 1.6. Section 1.7 outlines the assumptions made as well as the scope of the threat model while a roadmap for the rest of this thesis is presented in Section 1.8. Section 1.9 summarises this chapter.

1.1 Background

This section provides background information in the area of participatory sensing, the domain for which the research in this thesis is being undertaken. Section 1.1.1 explores the concept of participatory sensing while Section 1.1.2 discusses the importance of privacy for participatory sensing campaigns.

1.1.1 Participatory Sensing

Mobile devices such as smart phones are now ubiquitous and have a number of embedded sensor types that enable them to capture, classify and transmit data such as environmental readings, images, acoustic measurements and location, either interactively or autonomously. In addition, the ever increasing processing and storage capacity of such devices gives them the potential to act as sensor nodes and location-aware data collection instruments. As a result, smartphones are becoming very powerful mobile sensor platforms and location-aware data collection instruments that accompany users during their daily lives (Predic et al., 2013). The potential of these devices to enable users to gather, analyse and share data is known as participatory sensing (Burke et al., 2006), a form of crowdsourcing whereby individuals and communities submit scalar and/or multimedia data from mobile devices such as personal smart phones. The submitted data can be GPS coordinates revealing location or trajectory, a sensed data measurement or multimedia content such as photos, sound clips or video.

A typical participatory sensing application (or app) consists of a client on the participatory sensing device (for example, a SmartPhone) which connects to a server that resides in the Cloud (the participatory sensing app service). Participants can submit data (i.e. act as data submitters) and/or consume information derived by the service provider from submitted data (i.e. act as data consumers). One of the goals for a service provider is to ensure that as many participants as possible are data submitters as well as data consumers.

The wide range of data that can be captured by participatory sensing is reflected in the diversity of its potential applications including, among others, smart cities (Szabo et al., 2013), air pollution exposure (Predic et al., 2013), noise pollution (Coulson et al., 2018), health (Clarke and Steele, 2014b), crime reporting (Cilliers and Flowerday, 2014) and agriculture (Mohite et al., 2015). Participatory sensing campaigns are currently being undertaken across the world in, for example, Amsterdam[1] where sensed data is shared among people living in a particular neighbourhood as well as with the police authorities. Participatory sensing campaigns for health monitoring are also common. For example, the city of Louisville, Kentucky[2] used participatory sensing for its (now completed) asthma hotspot monitoring project.

1.1.2 Participatory Sensing & Privacy

Privacy is a key consideration for participatory sensing systems given their reliance upon mobile devices, in particular, their use of a mobile application. Moreover, the privacy consciousness of mobile users has the potential to directly impact upon the use of participatory sensing applications. For example,

[1] See `https://amsterdamsmartcity.com/products/buur`

[2] See `https://www.airlouisville.com`.

the surveys carried out by Boyles et al. [2012] and Brandtzaeg et al. [2018] both found that more than half of surveyed users were deterred from using a mobile app because they did not want to share their information while Hutton et al. [2014] and Almuhimedi et al. [2015] point out user concerns when location access is sought by mobile applications. These findings are of concern to participatory sensing given that the essence of the paradigm is the sharing of readings that may reveal confidential data pertaining to, for example, participant location or behaviour. Some potential participants may thus be deterred by this possibility of their privacy being compromised. and, indeed, privacy concerns have been found to hinder the effectiveness of some participatory sensing campaigns. For example, Ogie [2016] cites the example of a flood reporting participatory sensing service for Jakarta where potential participants living in temporary waterfront dwellings are deterred by fears that they will be traced by the government and punished for living in what are illegal settlements.

Privacy concerns with respect to mobile applications have been found to be particularly acute in the case of participatory sensing. For instance, the survey carried out by Christin et al. [2013a] found that potential users of participatory sensing applications rated the importance of privacy very highly. These findings are borne out by the work undertaken by Shilton and Martin [2013] which highlights the selling of data to an exchange (which is what submitting data to a participatory sensing service provider actually is) as having a negative effect on the privacy perceptions of mobile users. The authors also point out that mobile users expect more privacy from a third party data collector. This finding is particularly pertinent to participatory sensing applications as their operation typically comprises of submissions of data such as location and images to a third party service provider. However, despite these findings, participant privacy is frequently not taken into account in participatory sensing campaigns (Kounadi and Resch, 2018). Furthermore, the issue of privacy, which can be crucial to potential participants, needs to reconciled with that of data accuracy which is identified by potential data collectors and service providers as being of critical importance (Jiang et al., 2018).

1.2 Motivation

The key to the success of a participatory sensing application is attracting a critical mass of relevant data which meets the service provider's quality requirements. To achieve this, the service provider must attract a sufficient number of participants. In the majority of participatory sensing campaigns, participants may be willing to make data submissions but will, in general, expect some form of tangible (monetary or non-monetary) reward in return (Christin et al., 2013a, Christin et al., 2014, Mohite et al., 2015, Zaman et al., 2015, Arakawa and Matsuda, 2016, Restuccia et al., 2016 and Khoi et al., 2018). Moreover, participants expect to be compensated for costs such as battery consumption (Jin et al., 2015). By identifying a number of participatory sensing applications that have failed to attract sufficient participation rates, Xu et al. [2018] highlight the need for an effective incentivisation and reward allocation scheme that enables the service provider to attain a meaningful dataset.

In addition to its role in attracting participants, the issue of incentivisation also has direct implications for the quality of a service provider's dataset. In the case of participatory sensing and other similar data sharing environments, it has been found that proper incentive allocation improves data quality. For example, Wang et al. [2012] point out that content from data sharing is suboptimal without proper incentives, a claim that has its basis in economic theory while Gao et al. [2015a] highlight the scope that exists to improve data quality through proper incentive allocation. It should also be noted that incentivisation plays a key role in ensuring data submissions are timely as well as being of sufficient quality as users who are paid for assigned tasks complete them significantly more quickly than volunteer users (Mao et al., 2013).

While incentivisation schemes are critical to the success of participatory sensing campaigns, they face a number of challenges to ensure that the service provider's dataset is relevant and timely. In particular, the conditions in a participatory sensing environment can suddenly change, for example, a bridge connecting two areas of a city being closed due to high winds would result in a buildup of traffic. As a result of such sudden changes, the utility value of a particular type of data submission to the service provider can also change significantly over time. Moreover, as participation rates will vary over time, the service provider needs the ability to adapt the level of reward it offers to match the current response rate. At the same time, a service provider will have a finite budget and will want to optimise its consumption of this budget. This is not only of benefit to the service provider but also to those participants who want to consume the data and will therefore want it to be as relevant and timely as possible.

Crucially, it must also be noted that the incentivisation scheme, and in particular, the allocation of rewards, is a point of privacy violation as participants then need to disclose their identity to receive this reward, thus potentially deterring privacy conscious users from participating (as discussed in Section 1.1.2). At the same time, privacy preservation should not prevent the service provider from allocating the rewards it needs to offer to attract and incentivise users. Moreover, the service provider must be able to evaluate whether the data submission is a truthful and accurate one, a concept known as incentive compatibility[3] (Koutsopoulos, 2013). Both reward allocation and incentive compatibility are very challenging tasks if the service provider has no access to participant identity. Thus, the principal focus of this thesis is to investigate whether it is possible to provide an incentivisation scheme for participatory sensing campaigns that can allocate rewards in a way that protects participant identity privacy.

1.3 Context & Related Work

As attracting a sufficient level of participation is key to the success of a participatory sensing campaign, incentivisation for participatory sensing has been considered extensively in the state of the art. Incentivisation schemes for participatory sensing are, however, sometimes implemented without regard for participant privacy. For example, the bidding process used in auctions requires access to participant information. In addition, certain approaches require participant information such as past bidding history and geographic location (for example, Li et al., 2018). Other microeconomic-based incentivisation approaches (for example, Game Theory) are also vulnerable to privacy leakage through the use of traceable credit tokens and third party components. Statistical-based methods are also subject to privacy vulnerabilities. For example, access to participants' private information is fundamental to the task assignment mechanism used by CrowdMind (Xiong et al., 2017). Other statistical approaches display shortcomings in considering all facets of an incentivisation scheme. For instance, the approach taken by Yang et al. [2015] does not consider budget consumption

[3] Incentive compatibility, an economic concept that addresses how players in an economic system can achieve the best outcomes (trustworthy data for the service provider and a reward for the participant in the case of participatory sensing) for themselves by acting truthfully (Hurwicz, 1973), is important in interactions in which a player has access to information that is inaccessible to at least one other player. Such interactions need to be structured so that the player with more information is motivated to act in the interest of the other party, resulting in incentive compatibility.

optimisation.

Like incentivisation, privacy preservation for participatory sensing is widely addressed in the state of the art. However, this is sometimes achieved in a way that negatively impacts other aspects pertinent to participatory sensing. Specifically, privacy preservation methods such as perturbation provide participant privacy at the expense of the quality of the service provider's dataset. Moreover, anonymisation methods such as k-Anonymity and Differential Privacy are vulnerable to inference attacks (for example, To et al., 2014). Other methods, for example, the coarsening of the participant's location (Wiesner et al., 2014) also reduce the quality of the data for the service provider.

Other privacy-preserving approaches provide direct identity privacy (i.e. they prevent access to a participant's true identity) without impacting data quality but do not provide full participant privacy. Specifically, they are prone to *inference attacks* that enable the service provider or a third party to track a participant's behaviours and activities. This tracking can be facilitated by, for example, the use of a third party component that is itself a point of privacy vulnerability (for example, De Cristofaro and Soriente, 2013) or pseudonyms (for example, Zhang et al. [2012b]). Those approaches that seek to provide anonymous reward allocation are also prone to such attacks. For example, the credit token system proposed by Li and Cao [2016] incorporates a direct link to a participant's ID while the approach proposed by Niu et al. [2014] also has the potential to be used to trace participants. Other approaches do address privacy preservation when allocating rewards but do not address the potential for reward spending to be tracked, for example, Liu et al., 2018 and Dimitriou, 2018a. Cryptocurrencies such as Bitcoin (Nakamoto, 2008) are not an alternative to address this issue as, despite offering anonymity, they do not prevent the tracking of users.

Providing effective privacy preservation for incentivisation is not a straightforward task because of the need to allocate rewards to participants and facilitate the spending of these rewards. While there are many approaches in the state of the art such as source anonymous message authentication (Li et al., 2015a) that provide privacy preservation, the underlying methods used make it impossible to know which participants should be rewarded. This issue also impacts incentive compatibility as the evaluation of the validity of submitted data becomes more difficult if the credibility of the data submitter cannot be assessed. Indeed, many of the approaches in the state of the art that seek to provide incentive compatibility (for example, Zhou et al., 2017) do not take participant privacy into account with those that do being vulnerable to inference attacks (for example, Tanas et al., 2015 and Wang et al., 2018c)

1.4 Challenges

The key question to be addressed when designing a privacy-preserving incentivisation scheme is resolving how the need to link participant submissions so as to reward them and the need to break this link to ensure privacy preservation can be reconciled (Li et al., 2018). This results in a number of challenges that need to be addressed with respect to privacy preservation as the use of an incentivisation scheme in itself leads to a number of potential points of privacy vulnerability.

The first point of potential privacy vulnerability occurs when participants are making data submissions. To ensure identity privacy for participants, it is necessary to hide their real identity by considering the means by which participants can submit data anonymously. However, while this prevents disclosure of a participant's identity, anonymity in itself is insufficient in preventing the service provider or a third party from carrying out an inference attack to gain further private information about participants such as their habitual behaviour, location and trajectory. For this reason, the pseudonymous monitoring of participant activities must be prevented by ensuring that no links can be made between multiple data submissions made by the same participant. In addition, the service provider must have the ability to reward participants without knowing who they are.

Another potential point of privacy leakage that must be addressed occurs when a participant receives a reward. When seeking to prevent an inference attack from occurring, a significant challenge is how to design an approach that ensures that the service provider cannot identify a participant, cannot link the allocated reward to the participant's data submission, cannot trace participant activity and behaviour and cannot infer further information about that participant from the allocation of a reward. This protection for the participant must also hold when the reward is being spent.

In addition to ensuring privacy preservation, another key challenge is determining how the participatory sensing incentivisation scheme ensures that the service provider's budget is not consumed in a wasteful fashion by, for example, allocating rewards for inaccurate or untruthful data submissions. The scheme should therefore ensure that data submissions received by the service provider are evaluated to determine whether they are truthful ones. Moreover, this has to be achieved without impinging upon the participant's privacy through, for example, linking the participant's data submissions to a reputation score. The dynamic environments in which participatory sensing campaigns may potentially operate should also be taken into account. Specifically, the rewards offered by the service provider must be set at a level that matches current environmental conditions and current participation rates. It is also critical that such a scheme seeks to optimise the consumption of the service provider's budget and gives the service provider the flexibility to tune the importance and utility of the data being sought.

1.5 The Approach

This thesis presents *Privacy-Aware Incentivisation (PAI)*, a privacy-preserving approach to participatory sensing incentivisation and reward allocation. PAI provides a means of allocating untraceable and unlinkable rewards to attract data submissions that reflect the dynamic changes occurring in the participatory sensing environment; preserves the identity privacy of participants; prevents the inference of participant activity such as behavioural habits and frequently visited locations and ensures incentive compatibility for the service provider. The core of the approach is a peer-to-peer decentralised exchange platform that enables participants to anonymously make unlinkable data submissions in exchange for an untraceable and unlinkable reward from the service provider. Offers to which participants can respond are published on the modified equivalent of a cryptocurrency *OrderBook* which lists all offers made by the service provider. Participants who elect to make data submissions in response to these offers publish their acceptance on the OrderBook. As the OrderBook is hosted on multiple peer devices, the platform does not necessitate the use of a Trusted Third Party or other potential means of tracking participant activity and behaviour. As a result, the approach is robust to inference attacks.

To preserve identity privacy, the use of the Diffie-Hellman Exchange Protocol is modified to create the concept of a One-Time Key which is composed of public and private key components. The public key is used as a means of identifying the offer acceptance published by the participant on the OrderBook while the private component is held solely by the participant. As the One-Time Key is only used once in response to a particular offer, the service provider or other third party cannot trace participants using the public key component. In addition, the data included in the offer acceptance is encrypted using the service provider's public key to ensure that the service provider has sole access to the data for which it has paid.

Once the service provider decrypts the data submission, it evaluates it to see if it meets the terms of the offer. To address the potential of rewards being allocated to non-truthful data submissions, the Maximum Likelihood Estimation method is used to provide incentive compatibility. The service provider then indicates whether the participant should receive a reward by publishing an update to this effect on the OrderBook. The OrderBook generates an encrypted spendable reward if the service provider has deemed the data submission to be valid. The participant has sole access to this reward using the private part of the One-Time Key. Rewards are not only untraceable when they are allocated but also when they are spent as the OrderBook holds an identity certificate for the service provider. This ensures that the service provider cannot change its signature to track spendable rewards.

To enable the computation of rewards that reflect changing environmental conditions and participation rates, the incentive mechanism is modeled using the stochastic Lyapunov Optimisation technique. Through the ongoing computation of the reward level, the incentive mechanism can adapt to changes in the sensed environment and participants' response rates by adjusting the reward level whilst balancing budget consumption with the importance of the data being sought.

PAI is evaluated by proof and by comparing the approach to the most relevant approaches in the state of the art. The privacy robustness of PAI, and in particular, robustness to inference attacks is demonstrated by proof. These theorems prove that participants can make anonymous and unlinkable data submissions to the service provider and receive untraceable and unlinkable rewards in return. Similarly, the effectiveness of PAI's incentive compatibility is demonstrated by proof to show that data submissions that are considered to be non-truthful will not receive a reward whilst, at the same time, not violating participant privacy. A simulated participatory sensing environment is used to conduct experiments evaluating the adaptiveness and budget consumption of PAI's adaptive reward allocation method in comparison to the most relevant approaches in the state of art for participatory sensing reward computation. In addition, the approach is evaluated using experiments assessing the performance and computational complexity of the overall approach using the most relevant approaches in the state of the art in privacy preserving reward allocation as baselines.

1.6 Goals & Contribution

As highlighted by research undertaken by authors such as Christin et al. [2013a], an incentivisation scheme whose means of reward allocation ensures privacy for potential participants increases the probability of the service provider attaining a critical mass of participants given that privacy conscious users are more likely to participate. At the same time, participant privacy must not be at the expense of the service provider who will want to have sole access to the data submissions being paid for. In addition, current conditions and activities in the sensing environment need to be taken into account. Specifically, the response rate to previous requests made by the service provider and the utility of the data being submitted need to be considered when computing the reward offered to participants for data submissions. Incentive compatibility must also be considered so that participants are motivated to make truthful data submissions.

The goal of this thesis, therefore, is to investigate whether a participatory sensing service provider can offer incentive-compatible untraceable and unlinkable rewards to encourage anonymous, unlinkable and protected data submissions that reflect relevant changes in the environment, whilst ensuring identity privacy for participants and in particular, preserving behavioural privacy through the prevention of inference attacks.

The principal contributions of this thesis are:

1. A decentralised platform that provides a data submission mechanism for participatory sensing that meets the privacy requirements of the participant by ensuring that data is submitted anonymously and cannot be linked to any data submissions previously made by the same participant. This mechanism also meets the privacy requirements of the service provider by ensuring that the submitted data can only be accessed by this party.

2. A means of allocating rewards to participants without their having to cede their identity privacy. Specifically, rewards allocated to participants cannot be used to trace the activity of participants and cannot be used to link participants to their data submissions. In addition, rewards are also untraceable and unlinkable when participants spend them. This ensures

robustness to inference attacks from service providers and other potential attackers.

3. A privacy preserving reward computation mechanism and incentivisation scheme that adapts to participation rates and environmental conditions, seeks to optimise budget consumption and enables the service provider to tune the scheme so as to balance data capture and budget consumption.

4. The incorporation of a mechanism that assesses the truthfulness of data submissions made without violating participant privacy, thereby demonstrating that the approach facilitates incentive compatibility.

Proofs and experiments are used to evaluate these contributions. Proofs are used to demonstrate that the integrated approach provides anonymous and unlinkable data submission, untraceable and unlinkable reward allocation and spending and privacy preserving incentive compatibility. A simulated participatory sensing environment is also provided to compare the effectiveness of the reward computation mechanism with the state of the art in adapting to changes in the participatory sensing environment and balancing budget consumption with data capture. In addition, a simulation is provided to evaluate the energy consumption and computational complexity of the integrated approach in comparison with the most relevant approaches in the state of the art.

1.7 Assumptions & Scope

Participatory sensing systems have two principal actors, the service provider and the participant. The goal of the service provider is to attract data to build a dataset that it will, for example, disseminate to other users or analyse to learn about the environment being monitored. Participants will capture and submit this data using mobile devices such as smart phones, typically in expectation of a reward. PAI assumes that higher rewards attract a larger number of responses with participants only receiving these rewards when a sensing task has been completed in full. Furthermore, the budget for rewarding the sensing activity is assumed to be finite.

To meet the goals of this thesis, a service provider's rewards must be privacy preserving, both when being allocated and when being spent. PAI thus seeks to achieve a level of privacy preservation that not only prevents direct privacy violation but also prevents inference attacks by a semi-honest service provider i.e. one who will seek sensed data in return for genuine rewards but will seek to use that data and the allocation and spending of rewards to violate participant privacy by attempting to obtain further information about participant behaviour and activity without that person's consent. For this reason, the **Semi-Honest Threat Model** is the principal privacy model that must be addressed so as to prevent the tracking of participant activity and behaviour. To ensure that increased participant privacy is not achieved at the expense of the service provider's budget consumption through the rewarding of false or spurious data, the potential for **False Data Injection Attacks** is also addressed by PAI.

1.8 Roadmap

The remainder of this thesis is organised as follows:

- Chapter 2 discusses the problem to be addressed by PAI in terms of the system and threat models to be used and defines the requirements that need to be fulfilled.

- Chapter 3 considers related work in the areas of incentivisation, privacy preservation and incentive compatibility with research gaps in these areas being identified.

- Chapter 4 introduces PAI, the main contribution of this thesis. It describes the design of the decentralised platform used to allocate incentive compatible, untraceable and unlinkable rewards in return for anonymous and unlinkable data submissions. The design of a reward computation mechanism that facilitates adaptive reward allocation is also considered in this chapter.

- Chapter 5 describes the implementation of PAI. The evaluation of PAI is also discussed in terms of how well the approach addresses the requirements. The performance, computational complexity, privacy robustness and data truthfulness of the approach is also analysed.

- Chapter 6 concludes this thesis and offers possible directions for future work.

1.9 Summary

This chapter introduces the work to be undertaken in this thesis. Participatory sensing and privacy for the area is introduced with the need for a participatory sensing incentivisation scheme that is robust to inference attacks seeking to learn about participant behaviour and activity also being discussed. The context for developing such an incentivisation scheme is then explored with the shortcomings of current approaches in the state of the art being identified. Having identified the challenges that are posed by the research to be undertaken in this thesis, *Privacy-Aware Incentivisation (PAI)*, the approach described in this thesis, is then introduced. The goals and contributions made by this approach are also explored. The scope of the work to be undertaken is then outlined in terms of the assumptions made for the participatory sensing model as well as the scope of the threat model to be addressed. The chapter concludes with a roadmap outlining the remainder of this thesis.

Chapter 2

Problem Definition

This chapter defines the problem to be addressed in this thesis and outlines its scope. Section 2.1 describes the participatory sensing system model used. The participatory sensing threat model is described in Section 2.2 while the threat model to be addressed by the approach is discussed in Section 2.3. The requirements to be fulfilled are then discussed in Section 2.4. Section 2.5 summarises this chapter.

2.1 System Model

This section considers the participatory sensing system model. Section 2.1.1 describes the incentivisation scheme to be used while the assumptions made for the participatory sensing system model are discussed and justified in Section 2.1.2.

2.1.1 Incentivisation Scheme for Participatory Sensing

Figure 2.1 presents the architecture and operation of a typical participatory sensing system. A participatory sensing system comprises two actors, the service provider and the participant. The goal of the former is to capture data. This data can then be used for different purposes, for example, publication for consumption by other users or the building of a data set on which statistical analysis is conducted. The service provider initiates data collection campaigns by issuing offers indicating the type and scope of the sensed data being sought (e.g. air quality levels in a particular area of a city between 5pm and 7pm) and the corresponding reward that participants will receive for making data submissions matching the criteria outlined in the offer.

Participants will typically use mobile devices such as smart phones, tablet computers, wearable devices or smart vehicles which have embedded sensors to capture data. The data captured by these smart devices can be scalar (e.g. temperature, air quality levels or GPS coordinates) or multimedia (e.g. photos or video). Once captured, data is submitted to the service provider in anticipation of a reward. In addition to submitting data, participants may also be consumers of the data captured by a service provider.

Once the service provider issues an offer, participants can then elect whether or not to respond to it. As participants are assumed to incur costs (for example, battery consumption, mobile data consumption) when making data submissions, many participants cannot be expected to respond without incentivisation. Participants are thus modeled as having a reward threshold beyond which they will consider making a data submission. If the reward is greater than or equal to this threshold value, the participant will decide whether to make a data submission in response to this offer having taken issues such as battery consumption into account.

The fundamental problem being addressed by a participatory sensing incentivisation scheme can be considered to be a time average cost minimisation one as the service provider is seeking to set the offered reward and corresponding budget consumption at the minimum level that will attract an acceptable level of relevant and timely responses from participants that meet the quality criteria set by the service provider. To model this problem, it is assumed that the service provider operates in discrete time over slots $t \in 1,2...$ with the reward level being reviewed at the start of each time slot. This review is necessary as the dynamic nature of participatory sensing environments means that participation rates and the data being sought by service providers changes over time. Once the reward level is computed for a particular time slot, t, the service provider can issue one or more offers seeking data submissions. Offers can be categorised by different levels of granularity of the service provider's choosing, for example, location accuracy. A participant only receives a reward on full completion of a sensing task with rewards only being allocated until the service provider has received its desired number of responses.

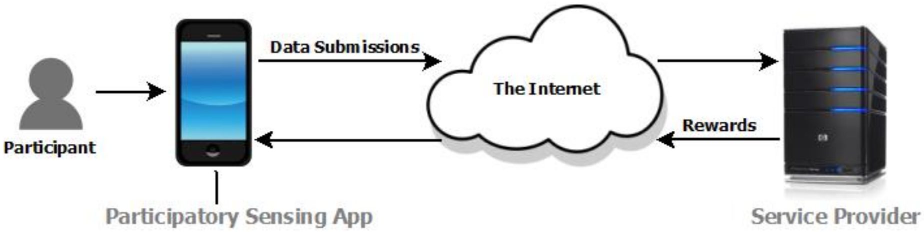

Figure 2.1: A Typical Participatory Sensing System

2.1.2 Assumptions

A number of assumptions are made for the participatory sensing system considered in this thesis. These are listed as follows:

- When considering the issue of reward allocation, it is assumed that participants are rational i.e. the higher the reward offered for a particular type of data, the larger the number of responses (assuming other factors such as privacy perceptions remain constant).
- The incentivisation budget held by the service provider for allocating these rewards is assumed to be finite with this budget either being a monetary one or consisting of tangible rewards (for example, Wi-Fi access).
- The service provider is also assumed to be rational and will not undertake actions that would adversely affect the success of its participatory sensing campaigns.

Other assumptions made pertain to the scalability and scope of the participatory sensing system and do not diminish the core contributions described in this thesis:

- It is assumed that there is only one service provider. This is done so as to simplify implementation and evaluation. Any evaluation results would still hold even if multiple service providers are incorporated.
- Scalability is also restricted through the assumption that participants respond to offers made by the service provider and do not make unsolicited data submissions. While this assumption simplifies the problem to be addressed, it does not diminish the core contribution of the thesis as participants making unsolicited data submissions could do so in the same privacy preserving fashion as they would in response to offers.

Other assumptions define the scope of the participatory sensing system in terms of functionality:

- It is assumed that parameters such as the number of responses to reward are configurable. This can be considered to be reasonable as the service provider may wish to do such configuration itself on the basis of its knowledge of the domain for which it is seeking sensed data. While it would be possible to automate the adaptive reward allocation mechanism to set these parameters on an ongoing basis, this is regarded to be outside the scope of this thesis.

- Only scalar data submissions are considered as the allocation of untraceable and unlinkable rewards would be carried out in the same fashion for multimedia data submissions.

- The potential to illegitimately use privacy-disclosing attributes (such as location or journey trajectory) in the data content of a single data submission to derive further information about a participant is outside the scope of this thesis as there are several approaches in the state of the art to address this issue through the use of, for example, obfuscation (Bettini and Riboni, 2015).

- While the participatory sensing paradigm relies upon the use of distributed computer networks, the addressing of networking issues such as reliability and communication failures is not considered as this would not be a core contribution in meeting the goals of this thesis.

2.2 Participatory Sensing Threat Model

The attack surface in a typical participatory sensing system is a large one with the sensing device, service provider infrastructure, third party components used by the service provider and the Internet communication all being potential points of attack. There are inherent threats, therefore, that have the potential to compromise participants' privacy as well as the integrity of the data held by the service provider. Those parties who threaten the system (known as *adversaries*), the participants or the service provider can be either **malicious** or **semi-honest** in their intent (these terms are defined in work undertaken by, among others, Cramer [1998]). As the term implies, a malicious adversary intends to do harm to the system or a party within the system. On the other hand, a semi-honest adversary will be one of the parties within the system and follows the protocol specification exactly. However, it may try to learn more information than intended by examining data that it receives. Figure 2.2 presents the threats faced by a typical participatory sensing system. It can be seen from the diagram that threats in participatory sensing can be **external** or **internal** to the system.

It should be noted that this diagram format is used as existing formal methods for modeling threats on computer systems (for example, Attack Trees [Schneier, 1999] and STRIDE [Shostack, 2014]) would not reflect the multiple points of vulnerability at which a number of threats exist. For example, the unauthorised disclosure of data could be an internal or external attack that takes place within the participatory sensing app, while the data is in transit or within the service provider's infrastructure.

2.2.1 External Attacks

External attacks by malicious adversaries can take place when submitted data is in transit over an external network to the service provider, when the participatory sensing app is compromised or when attempts are made to gain access to or disrupt the service provider's infrastructure. The many attacks that can be carried out by external third parties include eavesdropping, data disclosure, unauthorized access, data misuse, tampering, spoofing and denial of service (Chang et al., 2013, De Cristofaro and Soriente, 2013 and Qiu et al., 2013). While such attacks can have serious consequences, there are a number of solutions available to prevent them such as end-to-end encryption (see, for example, Li and Cao, 2015) and host hardening (see, for example, Shimeall and Spring, 2013).

2.2.2 Internal Attacks

Internal attacks can be carried out by malicious participants, semi-honest participants or the service provider. Malicious participants can carry out collusion attacks by sharing information from the participatory sensing system, thus enabling them to gain access to more information than they are entitled to (Günther et al., 2014). There is also the potential for malicious participants to submit false or corrupted data to the service provider. The participatory sensing application itself is also a possible source of internal threats. A malicious application could potentially disclose or infer data without the participant's consent;

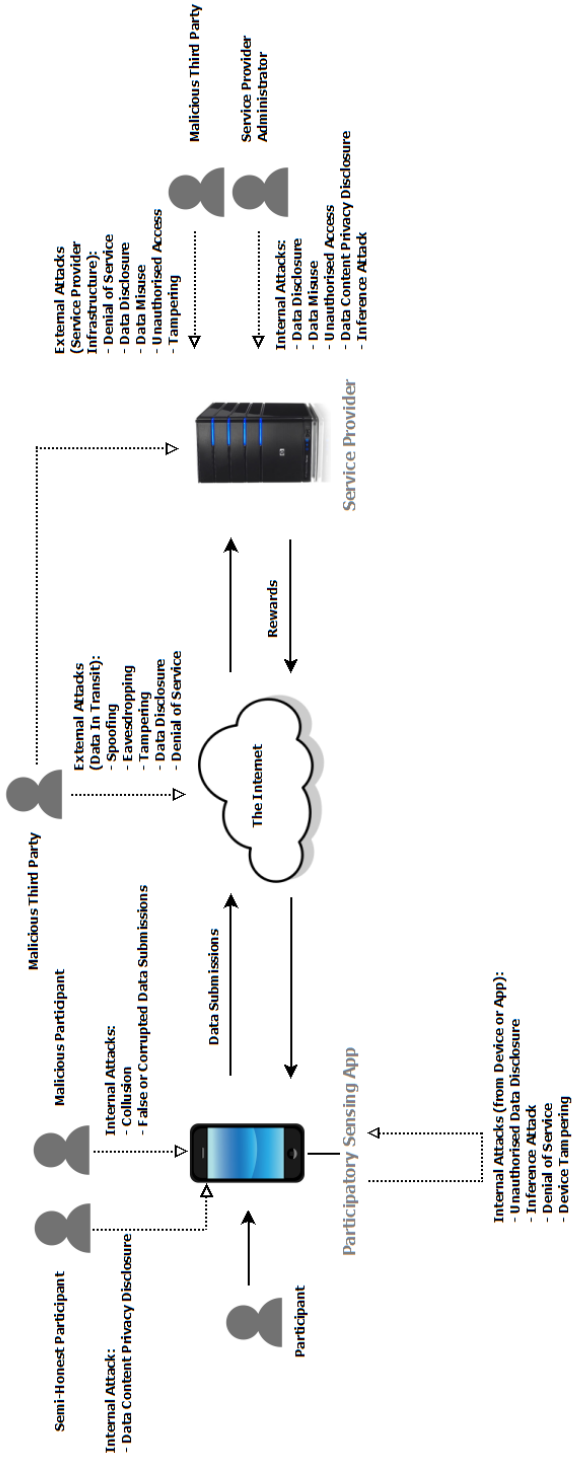

Figure 2.2: Threats to Participatory Sensing Systems

could tamper or damage the device by, for example, encrypting all its data thus denying access to the device owner or could deny service access to the device.

It can be seen from Figure 2.2 that there are many potential internal threats from the service provider. The service provider can potentially misuse data without the participant's consent or could deliberately or accidentally disclose or grant access to that data to an unauthorized third party inside or outside its organisation. When carrying out such actions, the service provider falls under the definition of a malicious adversary. There is also potential for the service provider to carry out an inference attack that analyses the dataset to illegitimately gain further knowledge about the participant without that party's consent. The service provider acts as a semi-honest adversary in this scenario. Similarly, a semi-honest service provider or participant could also use potentially privacy disclosing attributes from the data submission content to illegitimately gain further knowledge about a particular participant.

2.3 Threat Model Addressed

Section 2.2 describes the wide range of threats within a participatory sensing environment. This section identifies those threats that would directly impact upon the issue of privacy preserving incentivisation and defines the threat model to be addressed in this thesis.

The definition of identity privacy used in this thesis is based upon the legal principles of Personally Identifiable Information (PII). PII is any information which relates to an identified or identifiable person. The concept is enshrined in privacy law legislation in many jurisdictions, for example, the European Commission's General Data Protection Regulation[1]. In the context of participatory sensing, PII requires that participants must be protected not only from disclosure of their identity but also attacks that use their participation in the system to discover or infer information about them (i.e. an **inference attack**). Using the concept of PII, identity privacy in this thesis is therefore defined as any data that can disclose who a person is, where that person is located and what that person is doing in terms of their behaviour, activities and habits.

Meeting this definition of identity privacy requires that participant data is secured from unauthorized intrusions by the service provider, a challenge that is more demanding than preventing external attacks (Qiu et al., 2013).

[1] See http://ec.europa.eu/justice/data-protection/reform/files/regulation_oj_en.pdf

Specifically, the fact that the service provider has full access to participants' sensing records means that it can potentially infer or obtain private data such as location, trajectory or identity. The focus of this thesis is therefore on preventing the service provider from potentially using participants' data submissions or allocated rewards to carry out an inference attack to obtain further information which the participant does not wish to disclose, for example, the frequency of sensing activity in a particular location.

The potential for the service provider to act in a semi-honest fashion by using data submissions and rewards to gain unauthorised private information is the key threat to be addressed when considering privacy preserving reward allocation. The **Semi-Honest Service Provider Threat Model** is therefore the privacy model that must be addressed in order to fulfill the goals of this thesis, specifically, to prevent inference attacks tracking participants' activity and behaviour through their data submissions, reward allocations and reward spending. Other threats such as malicious apps and collusion attacks, while serious, can be addressed independently of privacy preserving incentivisation[2]. For this reason, such attacks are not considered to be within the scope of this thesis. In addition, as outlined in Section 1.7, the potential of the content in the data submission to be a point of privacy vulnerability is not within the scope of this thesis.

While the Semi-Honest Service Provider Threat Model addresses participant privacy, it also increases the potential for malicious participants to deceive the service provider and gain unearned rewards by submitting false or spurious data without penalty. This also affects participants who consume the service provider's dataset as its overall quality is degraded. The Semi-Honest Service Provider Threat Model must therefore be addressed in a way that does not open up the possibility of the quality of the service provider's dataset being diminished. For this reason, **False Data Injection Attacks** also fall within the threat model addressed in this thesis.

2.4 Requirements

This section considers the requirements that must be met to fulfill the goals of this thesis. These requirements are organised around the need to address the

[2] For example, the potential for malicious apps to conduct attacks has been addressed in the approach outlined by Wang et al. [2018a] while Amintoosi et al. [2014] and Günther et al. [2014] propose approaches for combating collusion attacks.

Semi-Honest Service Provider Threat Model, False Data Injection Attacks and the dynamic nature of the participatory sensing environment. The requirements outlined in this section stem from the motivation for this thesis as outlined in Section 1.2 as well as the the threat model discussed in Section 2.3 i.e. to provide a privacy preserving incentive compatible incentivisation scheme that addresses the potential for both inference attacks and the submission of false or corrupted data.

To ensure privacy preservation, several potential points of privacy vulnerability must be addressed. The first point of vulnerability under the Semi-Honest Service Provider Threat Model occurs when the participant is making a data submission. Participants must therefore be able to make data submissions anonymously without any scope for the service provider to infer further information about that participant by, for example, linking multiple data submissions from the same participant. At the same time, the data submission should not be devalued for the service provider by making it potentially accessible to other parties. This leads to the first requirement to be addressed in this thesis:

Anonymous, Unlinkable & Protected Data Submission (R1)
Participants must be able to make anonymous, unlinkable and protected data submissions to the service provider that preserve identity privacy. Specifically:

- The service provider cannot identify participants from the data submissions they make.
- The service provider cannot link multiple data submissions made by the same participant.
- The data submission should only be accessible by the service provider.

The next point of privacy vulnerability under the Semi-Honest Service Provider Threat Model occurs when participants are given a reward by the service provider. To meet the level of identity privacy defined in this thesis, reward allocation must not enable the service provider to identify participants or trace their activities and behaviours. This is considered in the second requirement to be addressed in this thesis:

Untraceable & Unlinkable Reward Allocation (R2)
Participants must receive untraceable and unlinkable rewards that preserve their identity privacy. In addition, the service provider or a third party should not be able to conduct an inference attack to gain further private information about participants such as their habitual behaviour, location and trajectory. Specifically:

- The service provider cannot identify a participant through the allocation of a reward.
- The service provider cannot trace participant activity and behaviour, or infer further information about that participant, from the allocation of a reward.

The privacy standard identified for reward allocation must also hold when the reward is being spent i.e. the service provider should not be able to identify participants or trace their activities when they spend the rewards they have been given. This leads to the next requirement to be addressed in this thesis:

Untraceable & Unlinkable Reward Spending (R3)

- The service provider or a third party should not be able to conduct an inference attack when a reward is being spent.

As the potential for False Data Injection Attacks falls within the scope of the threat model defined in Section 2.3, the potential for participants to make non-truthful data submissions must also be addressed:

Incentive Compatibility (R4)
For the participatory sensing system model and threat model identified in Section 2.1 and Section 2.3, incentive compatibility must ensure that:

- only data submissions that are truthful and accurate receive rewards from the service provider.
- the requirement for *Untraceable and Unlinkable Reward Allocation (R2).* is not violated.

The incentivisation scheme discussed in Section 2.1.1 recognises the dynamic nature of the environment in which participatory sensing systems operate. To ensure the timely capture of the most relevant data, therefore, the service provider needs a mechanism that periodically recomputes the level of reward to offer. This is considered in the final requirement to be addressed in this thesis:

Adaptive & Tunable Reward Allocation (R5)

The incentivisation scheme used to motivate participation must:

- be able to adapt the rewards it offers in real-time to match current environmental conditions and current participation rates, thereby optimising the consumption of the service provider's budget.
- be tunable to enable the service provider to balance data capture with budget consumption optimisation and vice versa.
- adhere to the requirement for *Untraceable and Unlinkable Reward Allocation (R2).*
- not impair the service provider's quality requirements[a].

[a] While methods from the state of the art that are used to preserve participant privacy with respect to the content of a single data submission may diminish the quality of this data, the approach in this thesis will not.

2.5 Summary

This chapter defines the participatory sensing model as well as the assumptions pertaining to participant behaviour and service provider incentivisation activity that are made for this model. The attack surface for participatory sensing is then discussed with the need to address the Semi-Honest Service Provider Threat Model and False Data Injection Attacks being identified. Having defined the system and threat models, the requirements for *Anonymous, Unlinkable and Protected Data Submission (R1)*, *Untraceable and Unlinkable Reward Allocation (R2)*, *Untraceable and Unlinkable Reward Spending (R3)*, *Incentive Compatibility (R4)* and *Adaptive and Tunable Reward Allocation (R5)*, all of which need to be fulfilled to meet the goals of this thesis, are then outlined.

Chapter 3

Related Work

Having defined the scope of the problem to be addressed in this thesis, Chapter 2 identified the need to fulfill requirements for *Anonymous, Unlinkable and Protected Data Submission (R1)*, *Untraceable and Unlinkable Reward Allocation (R2)*, *Untraceable and Unlinkable Reward Spending (R3)*, *Incentive Compatibility (R4)* and *Adaptive and Tunable Reward Allocation (R5)*. This chapter evaluates approaches in the state of the art that address incentivisation, privacy preservation and incentive compatibility, the areas most pertinent to the work undertaken in this thesis, in light of these requirements. The incentivisation schemes for participatory sensing that are proposed for the state of the art are discussed in Section 3.1 while the issue of privacy preservation is explored in Section 3.2. Approaches to incentive compatibility are outlined in Section 3.3. Section 3.4 summarises this chapter.

3.1 Incentivisation

The goal of incentivisation is to motivate a sufficient number of participants to make data submissions that meet the service provider's requirements. Balancing the needs of privacy preservation and incentivisation is a challenge as, in some cases, the means of incentivisation is itself a point of privacy leakage. At the same time, any effort to preserve participant privacy must not occur at the expense of pertinent issues such as data utility, response rate and participation rate.

The majority of incentivisation schemes can be categorised according to the academic discipline used as the basis for their design and implementation, specifically, microeconomics, statistics or a combination of both. For this reason, participatory sensing incentivisation schemes are categorised into economic and statistical approaches and are discussed in Section 3.1.1 and Section 3.1.2 respectively.

3.1.1 Economic Approaches to Incentivisation

This section considers approaches to incentivisation that are based upon microeconomic concepts. Section 3.1.1.1 considers those approaches that are based upon auctions, where an *auctioneer* sells some goods to a group of *bidders* who place bids to buy these goods. Incentivisation approaches based upon the microeconomic concept of Contract Theory, the study of how economic actors make contractual arrangements in the presence of asymmetrical information where one party has access to more information than the other, are then explored in Section 3.1.1.2. The use of Game Theory, a microeconomic concept defined as a bargaining game concerned with how to divide surpluses between two players (Luo et al., 2017), is discussed in Section 3.1.1.3 while the use of other microeconomic concepts to design participatory sensing incentivisation schemes is considered in Section 3.1.1.4.

3.1.1.1 Auctions

Auctions are a means of incentivisation that is used extensively in the state of the art (Luo et al., 2017). In the case of participatory sensing, the auctioneer corresponds to the service provider who, rather than selling goods, offers a reward for the completion of a task to sense data. The bidders, in this case, are the participants who place bids to denote the reward they are seeking in exchange for a data submission. A number of approaches in the state of the art use reverse auctions, whose reversal of the traditional roles of buyer and seller is particularly appropriate for participatory sensing systems as a single buyer (i.e. the service provider) can offer out a contract (the sensed data it is seeking) for bidding by multiple sellers (participants). Figure 3.1 presents a high level overview of how an auction would operate for a participatory sensing campaign.

Auctions entail a high level of overhead (Kumar and Feldman, 1998). For a participatory sensing environment, this means that the service provider will typically need to gather all bids before deciding which participants to select. This in turn leaves participants vulnerable to privacy violations as, even if pseudonyms

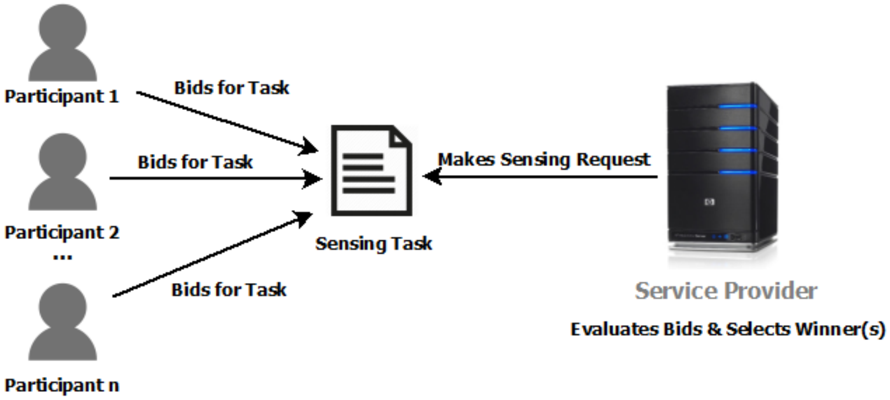

Figure 3.1: Auction in Participatory Sensing

are used, the service provider can monitor participants' bid activity. The bid process is thus a key point of privacy leakage for auction-based approaches (see, for example, Koutsopoulos, 2013, Feng et al., 2014b, Zhang et al., 2014, Jin et al., 2015, Duan et al., 2016, Dai et al., 2018 and Li et al., 2019a). Moreover, certain approaches exacerbate vulnerability to privacy leakage through their requirement for further information as they incorporate mechanisms that require access to additional information such as participant identification, active sensing time and social network contacts (Feng et al., 2014a, Wei et al., 2015, Sun et al., 2016, Guo et al., 2017, Jin et al., 2017b, Mukhopadhyay et al., 2017, Cai et al., 2018, Li et al., 2018, Niu et al., 2018a, Restuccia et al., 2018a and Xu et al., 2018). Other approaches have attributes such as credit tokens and reputation scores that further enable participant activity to be tracked (Jaimes et al., 2015b, Luo et al., 2015, Xu et al., 2017b, Jaimes and Calderon, 2018 and Yu et al., 2019). The combining of auctions with other techniques such as evolutionary algorithms and linear programming does not alleviate these concerns and, in several cases, introduces other avenues for inference attacks because of, for example, the retention of participant selection history (Singla and Krause, 2013a, Kumrai et al., 2014, Gao et al., 2015a, Zheng et al., 2016, Chen et al., 2017a and Shi et al., 2018). In addition, it should also be noted that auctions can be vulnerable to collusion attacks (Sandholm, 2000). In a participatory sensing environment, this means that colluding participants could consume a disproportionate amount of the service provider's budget, thus diminishing the quality of the overall dataset. This violates the requirement for *Adaptive and*

Tunable Reward Allocation (R5), which seeks to ensure that the service provider attracts data that meets its desired quality standards. It should be noted that this is a potential problem for several auction-based approaches in the state of the art which assume that participants will not engage in collusion attacks prior to bidding, for example, the approach proposed by Liu et al. [2019].

Some approaches do claim privacy protection for their auction-based scheme, for example, the approaches proposed by Jin et al. [2016a] and Wang et al. [2016b]. However, in both cases, the privacy protection is intended to enable privacy between participants and, indeed, access to private information such as participants' bid activity is required. In addition, there are a number of auction-based approaches to incentivisation that claim privacy preservation but actually facilitate the trading of privacy and do not address the potential for further private information to be accessed, for example, the approaches taken by Holzbauer and Bulut [2012], Jin et al. [2016b] and Wang et al. [2018c]. Other approaches do seek to provide privacy preservation between the participant and service provider but nevertheless do not meet the requirement for *Anonymous, Unlinkable and Protected Data Submission (R1).* For example, the approach taken by Sun and Ma [2014], a first price sealed bid auction[1] that uses oblivious transfer to preserve privacy, does not address the potential for inference attacks and further exacerbates privacy vulnerabilities by the use of a bulletin board displaying all bids in the auction. Similarly, while Jin and Zhang [2018] claim privacy preservation for their approach, their reverse auction has no privacy preserving attributes that could prevent the tracking of participant activity.

Other auction-based approaches seek to preserve privacy through the use of a third party component (for example, Dimitriou and Krontiris, 2017). However, such a component could itself be a point of privacy violation through, for example, attacks, database leaks or seizure by governments (Ziegeldorf et al., 2017), given the confidential nature of the data that is often stored. For instance, the approach proposed by Li et al. [2017b], who explicitly state that the third party component used in their approach is semi-honest, incorporates a cryptographic key generator that has access to participants' IDs, a potential point of privacy leakage.

There are a number of auction-based approaches that would partially address the requirement for *Anonymous, Unlinkable and Protected Data Submis-*

[1] Easley and Kleinberg [2010] define a first-price sealed-bid auction, also known as a blind auction, as one where each bidder submits a bid to the seller that is hidden from other bidders. The highest bidder wins and pays the bid made for the good.

sion (R1). For example, the use of different pseudonyms in different auctions for the approach proposed by Dimitriou and Krontiris [2017] ensures that bids made by the same participant in multiple auctions cannot be tracked. However, this method still offers scope to monitor participant activity as it is possible to track multiple bids made by a participant during a single auction. As a result, this approach does not fully meet requirement *R1* and, in addition, would not meet the requirement for *Untraceable and Unlinkable Reward Allocation (R2).*

To conclude, while there are many auction-based incentivisation schemes in the state of the art, the bidding process means that such schemes cannot meet requirement *R1.* This also applies to those auction-based schemes that claim privacy preservation as, in general, such approaches do not address the fundamental privacy vulnerabilities presented by the bidding process.

3.1.1.2 Contract Theory

Contract theory defines two players who take very different roles (Luo et al., 2017). For participatory sensing, the **principal**, the player who has all the bargaining power and spells out the terms of the contract, corresponds to the service provider. The **agent**, who can only accept or reject the contract and cannot make a counter-offer, corresponds to the participant.

There are two main contract models. In the adverse selection model, the agent has certain hidden information that the principal tries to elicit. In participatory sensing, this corresponds to the service provider attempting to obtain sensed data (i.e. the hidden information) from the participant. In the moral hazard model, the agent could exert some hidden effort that is of economic value to the principal while the principal tries to induce a desired effort level at a minimal cost. For participatory sensing, the moral hazard model introduces the concept of value for data submissions. In this case, the service provider attempts to obtain the data at the minimal reward level possible.

Contract Theory has merits in terms of enhancing the quality of data submissions and ensuring data truthfulness, particularly through the use of the moral hazard model. Furthermore, the absence of a bidding and participant selection process removes a potential point of privacy vulnerability. However, the nature of the approach, which entails the exchange of contracts between participant and service provider, makes it difficult to achieve a level of privacy preservation that would fulfill the requirement for *Anonymous, Unlinkable and Protected Data Submission (R1).* For example, the moral hazard model of Contract Theory is used by Zhang et al. [2018a] to devise an incentivisation approach that focuses on data quality by ensuring that participants are rewarded through the evaluation of their task completion performance from multiple perspectives (or *dimensions*). However, while linking the participant's reward to task performance has merit in attracting better quality data submissions, the approach gives the service provider direct access to the participant's behaviour and activities. The moral hazard model-based approach proposed by Zhao et al. [2017] also exhibits privacy vulnerabilities as its linking of a contract issued to the participant with a data submission results in the participant being vulnerable to inference attacks. The incentive compatible approach outlined by Chen et al. [2017b] is more robust from a privacy preservation perspective in that it could be extended to incorporate anonymous data submission with some modification of the core algorithm. However, data submissions made by the same participant would be linkable.

The user matching mechanism for the Broker-less Participatory Sensing Scheme proposed by Oide et al. [2016], which uses a Contract Theory-based approach to match consumers and providers of sensing information, is more robust from a privacy preservation perspective as it uses a peer-to-peer approach that removes the need for a central server and thus avoids many of the pitfalls of privacy leakage to the service provider. However, this approach would not fulfill the requirement for *Anonymous, Unlinkable and Protected Data Submission (R1)* as the sensed data cannot be stored by the service provider. While the authors assert that sensed data does not need to be stored anywhere, this will not be the case for many service providers and would not meet the needs of the participatory sensing model presented in Section 2.1.

Other approaches that consider privacy preservation do so from the perspective of privacy trade-offs. For example, the contracts issued between participant and service provider by REAP (Zhang et al., 2018b), while seeking to compensate participants for loss of privacy, gives the service provider the potential to conduct inference attacks by monitoring the nature and frequency of these contracts.

3.1.1.3 Game Theory

In the case of participatory sensing, the players in Game Theory correspond to the participant and service provider. Two main bargaining models have been identified (Luo et al., 2017). The Rubinstein bargaining model takes a strategic approach by modeling the bargaining procedure as a sequential game, in which the two players alternately propose offers until one accepts the offer proposed by the other while the Nash bargaining model focuses on deriving an outcome that satisfies certain mathematical conditions (or *axioms*).

Like Contract Theory-based approaches, the Game Theory-based incentivisation mechanisms outlined in the state of the art for participatory sensing have inherent attributes that make them vulnerable to unauthorised privacy disclosure. For example, the authors of the QUOIN incentivisation scheme (Ota et al., 2018), which uses the Stackelberg Game Theory model to maximise the utility of the data a service provider collects from participants, point out that their method is prone to privacy leakages while Theseus (Jin et al., 2017a), which uses Game Theory for its payment mechanism, retains details tracking the quality of participants' data submissions.

Incentivisation schemes that combine Game Theory with other methods exhibit similar vulnerabilities to privacy leakage. For instance, Stable-GRS (Azzam et al., 2018), which uses both genetic algorithms and Game Theory to recruit and incentivise participants respectively, requires significant access to participants' private information, specifically GPS locations and mobility patterns. Similarly, the Nash bargaining model-based approach outlined by Zhan et al. [2018b], which envisages the use of credits that can be later used by the participant to claim a reward, is vulnerable to inference attacks and, specifically does not meet the requirement for *Untraceable and Unlinkable Reward Allocation (R2)* as allocated rewards can be used to track participants.

Other approaches use Game Theory to offer a different perspective on incentivisation. For example, the approach proposed by Wang et al. [2012] considers the incentive mechanism in networks such as participatory sensing systems as a system rule whose goal is to influence participants to behave in a certain manner. The authors propose what they term evolutionary Game Theory (EGT)-based incentive mechanisms, which, for participatory sensing, would mean that participants could imitate other participants' behaviours so as to increase their rewards. However, such an approach would not be viable from a privacy preservation perspective as it requires access to all participant activity. The approach outlined by Yang et al. [2017a] also uses Game Theory in a different way to other incentivisation approaches in the state of the art. In this case, participants are not offered tangible rewards. Instead, the authors purport to incentivise participation through social relationships and performance ranking, both potential sources of privacy leakage. Privacy is considered by the Stackelberg Game-based incentive mechanism proposed by Koh et al. [2018]. However, this is in terms of privacy trade-offs as participant rewards are determined on the basis of the location granularity of their data submission.

3.1.1.4 Other Microeconomic Concepts

A wide diversity of other microeconomic concepts are used to design incentivisation schemes with the core tenet of supply and demand being used in many approaches. Unfortunately, like incentivisation schemes based upon auctions, Contract Theory and Game Theory, some of these approaches exhibit privacy vulnerabilities that prevent them from meeting the requirement for *Anonymous, Unlinkable and Protected Data Submission (R1).* For example, the supply and demand-based approach outlined by He et al. [2017], which uses Walrasian Equilibrium to devise a vector of rewards, selects and recruits individual participants rather than making a general offer that could be responded to anonymously. Moreover, the privacy preservation claims of the approach proposed by Zhan et al. [2018a], which seeks to maximise social welfare[2], are undermined by the fact that participants must register an ID and their location with the service provider who can also track participant activity through reputational quality scores. Similarly, Pournaras et al. [2016], who use supply and demand to build a market mechanism for participatory sensing, acknowledge that the potential for inference attacks is an open issue for their approach. SEQTGREEDY (Singla and Krause, 2013b), which uses marginal utility to maximise the service provider's marginal gain, also considers privacy preservation. However, the authors assume that, regardless of their privacy concerns, participants are willing to share what they term certain non-sensitive private information (which is not defined). It should also be noted that the use of obfuscation and random perturbation by this approach to devise what is called a 'privacy profile' is not robust against certain types of inference attacks such as distribution analysis[3].

While many of the approaches in the state of the art that use microeconomic concepts exhibit privacy leakage, there is no reason why the use of these concepts should preclude meeting the privacy preserving requirements for *Anonymous, Unlinkable and Protected Data Submission (R1)*, *Untraceable and Unlinkable Reward Allocation (R2)* and *Untraceable and Unlinkable Reward Spending (R3).*

[2] As previously noted, social welfare is an economic concept that, in the case of participatory sensing, measures the benefits accrued by both the participant and the service provider.

[3] A distribution analysis attack entails reconstructing the probability density function of a dataset, which, in some cases, can lead to privacy disclosure (Liu et al., 2008).

For example, SenseUtil (Tsujimori et al., 2014 and Thepvilojanapong et al., 2013), which uses the principles of supply and demand in conjunction with marginal utility to determine the value of sensed data, does not consider privacy but could be integrated with a appropriate method of privacy preservation. It must be noted, however, that SenseUtil cannot be directly adapted for the work addressed in this thesis as it does not meet the requirement for *Adaptive & Tunable Reward Allocation (R5).* Specifically, the approach does not attempt to optimise rewards to determine a level at which data submissions will be made below that value.

3.1.2 Statistical & Machine Learning Approaches to Incentivisation

There are a wide variety of statistical-based incentivisation approaches in the state of the art for participatory sensing. Given the nature of incentivisation schemes, it is unsurprising that optimisation methods are used for computing the reward level in several schemes. These are considered in Section 3.1.2.1. In addition to optimisation, there are also a number of approaches that use a variety of techniques such as probability and machine learning. These are discussed in Section 3.1.2.2.

3.1.2.1 Optimisation

Optimisation is used to design incentive schemes in a wide range of domains in the state of the art (Huang et al., 2016). However, while these approaches have diverse goals such as optimising budget consumption or increasing participation rates, a number of these approaches would not meet the requirement for *Anonymous, Unlinkable and Protected Data Submission (R1)* as they seek to enhance data quality by ensuring the selection of those participants who are best placed to meet the service provider's data requirements and, in many cases, require additional participant information such as trajectories and expertise (for example, Karaliopoulos et al., 2016, Xiong et al., 2016, Back et al., 2017, Xiong et al., 2017 and Sun and Liu, 2018)[4]. Other approaches have goals that are not consistent with those of requirement *R1*, for example, the trading of rather than the preservation of privacy (Alsheikh et al., 2017).

[4] See also the approaches taken by Han and Zhu [2014], Luo et al. [2014], Song et al. [2014], Amintoosi et al. [2015] and Ren et al. [2015].

There are a number of optimisation approaches in the state of the art that do not require participants' private information but nevertheless prevent participant anonymity when submitting data. For example, the approach taken by Wang et al. [2016a], which addresses incentivisation by formulating a multi-objective optimisation problem to maximise both the received data quality and participants' benefits, utilises a reputation framework (that is integral to their approach) which can be used to monitor participant performance and behaviour. Similarly, Wang et al. [2014], who use optimisation to build a stochastic[5] Markov model[6] for urban traffic modeling, state that the potential for inference attacks is an open issue for their approach. Both privacy trade-offs and privacy preservation are taken into account in the case of the approach taken by Messaoud et al. [2016], which applies an optimisation technique to balance the trade-off between privacy leakage and data utility as well as obfuscating the data submission. However, the use of a third party component by this approach can serve as the basis for an inference attack. Similarly, *Anonymous, Unlinkable and Protected Data Submission (R1)* is not possible under the approach taken by Li and Zhu [2018] as participant activity is visible to other participants and the service provider.

While many of the approaches in the state of art would not fulfill the requirement for *Anonymous, Unlinkable and Protected Data Submission (R1)*, it must be noted that many of the underlying methods are indeed suitable for the fulfillment of requirement *R1* and, indeed, the other privacy preserving requirements for *Untraceable and Unlinkable Reward Allocation (R2)* and *Untraceable and Unlinkable Reward Spending (R3)*. This is reflected in several approaches in the state of the art. For instance, the optimisation method used by the Quality Utilization Aware Data Gathering (QUADG) scheme (Ren et al., 2018)[7], the Expectation Maximization Algorithm used by Peng et al. [2018], the budget constrained simulated annealing technique that is used by EPPI (Niu et al., 2014) and the formulation of participant incentivisation as a multi-objective optimisation problem for the Time and Location Correlation Incentive (TLCI) scheme (Ma et al., 2018) would not violate requirements *R1*, *R2* and *R3*. Similarly, Lyapunov Optimisation, which is used by Liu et al. [2017c] to strike a balance between social welfare maximization[8] and the sensed data queues in vehicular

[5] `http://www.businessdictionary.com` defines stochastics as a modeling approach for processes that are continuously evolving over time in a random (i.e. uncertain) fashion.

[6] A Markov model is a stochastic model used to model randomly changing systems.

[7] This approach is influenced by the work carried out by Zhao and Zhu [2014].

[8] The economic concept of social welfare is used to determine the benefits accruing to the

participatory sensing systems, could also serve as the basis for an incentivisation scheme that fulfills these requirements. The approaches taken by Yang et al. [2015] and Han et al. [2014], which use Lyapunov Optimisation in conjunction with other techniques such as Mechanism Design, are also compatible with requirements *R1*, *R2* and *R3*.

Unfortunately, however, while these approaches address the privacy preserving requirements *R1*, *R2* and *R3*, some are incompatible with the participatory sensing model discussed in Section 2.1. For example, the approaches taken by Liu et al. [2017c] and Ren et al. [2018] are only suitable for vehicular participatory sensing systems while the approach taken by Peng et al. [2018] is only suitable for certain categories of sensed data such as noise decibel levels. Similarly, the Distributed Utility-Maximizing Algorithm (Han et al., 2014) places a burden on the participant's mobile device as the correlated scheduling algorithm requires it to detect its context (for example, location) and monitor the queue of sensing tasks to be carried out on behalf of the service provider.

There are other approaches which could fulfill privacy preserving requirements *R1*, *R2* and *R3* but do not meet the requirement for *Adaptive and Tunable Reward Allocation (R5).* For example, EPPI (Niu et al., 2014) and the 'Backpressure Meets Taxes' (BMT) mechanism (Yang et al., 2015) do not fulfill requirement *R5* as they do not consider issues such as the current participation rate (EPPI), the dynamic environmental changes that may occur (EPPI) and budget optimisation (BMT). TLCI (Ma et al., 2018) better addresses requirement *R5* as data quality, response rates and participation rates are all considered by, for example, taking the number of users in different times and locations, the data sensing cost and willingness to participate into account. However, while the approach is *adaptive,* it is not *tunable*, as it is unable to prioritise efficient budget consumption over data capture and vice versa.

3.1.2.2 Other Statistical & Machine Learning Methods

A number of incentivisation schemes are based upon the use of descriptive statistics[9] and/or probability. However, a number of these approaches would not fulfill the requirement for *Anonymous, Unlinkable and Protected Data Submission (R1).* For example, the incentivisation scheme outlined by Sun and Tham

participatory sensing system.

[9] http://www.businessdictionary.com defines descriptive statistics as a set of mathematical quantities (such as mean, median and standard deviation) that summarise and interpret some of the properties of a sample dataset but do not infer the properties of the population from which the sample was drawn.

[2015a] (also described in Sun and Tham, 2015b), which uses descriptive statistics and probability distributions to evaluate data contributions from two categories of participants (those who contribute data from specific locations and those who contribute aggregated sensed data that is captured throughout the monitored area), not only necessitates participant selection but also records participant reputation scores, which can be a potential source of privacy leakage. Descriptive statistics are also used by the NoiseMap mobile application (Schweizer et al., 2012), in this case to address participant retention by giving participants motivational feedback and publicly ranking their performance but at the expense of their privacy. The descriptive statistics used by the approach proposed by Ji et al. [2017] to outline a series of varied incentivisation strategies for participatory sensing that seek to increase participation rates do not consider privacy preservation. For example, the ranking system used to measure participant contributions requires privacy disclosure. Approaches that use other methods also exhibit privacy vulnerabilities. For example, the approach taken by Liu et al. [2016b], which uses the Minimum Cut of a graph from graph theory and the machine learning concept of support vector machine (SVM)-based pattern recognition, to determine the utility of a particular participant's potential sensed data, does not provide anonymous and unlinkable data submission as participant utility is evaluated by grouping those in similar geographical locations.

However, while many of the statistical based incentivisation approaches in the state of the art have privacy vulnerabilities that render them unsuitable in meeting the requirement for *Anonymous, Unlinkable and Protected Data Submission (R1)*, the use of statistical methods in and of itself does not preclude the fulfilling of the privacy preserving requirements for *Anonymous, Unlinkable and Protected Data Submission (R1)*, *Untraceable and Unlinkable Reward Allocation (R2)* and *Untraceable and Unlinkable Reward Spending (R3)*. Probabilistic based methods such as those advocated by the Bayesian Truth Serum used by Radanovic and Faltings [2015] to evaluate the data submitted using a probabilistic scoring system and the Gur Game-based approach[10] proposed by Liu et al. [2011] would not violate these requirements. This is also true for the binary search and the Multi-Armed Bandit (MAB) Framework[11] used by the STOC-PISCES algorithm (Biswas et al., 2015). Machine learning approaches

[10] This is a mathematical modeling of what is termed reward and punishment.

[11] The Multi-Armed Bandit (MAB) Framework is a probabilistic method of resource allocation.

could also potentially be used to fulfill requirements *R1*, *R2* and *R3*. For example, the supervised machine learning approach proposed by Sun et al. [2018] could be adapted to meet these requirements.

Adapting these approaches from the state of the art for the work to be undertaken in this thesis would not, however, address the requirement for *Adaptive and Tunable Reward Allocation (R5).* The approaches taken by Liu et al. [2011] and Radanovic and Faltings [2015] do not address optimal budget consumption or reward level adaptiveness. While STOC-PISCES (Biswas et al., 2015) does adapt the reward level, it does not take budget constraints into account. The approach taken by Sun et al. [2018] is also unsuitable as it is specifically designed for the multi-label classification problem and cannot be used for other types of participatory sensing activity.

3.1.3 Incentivisation: Summary

While there are a wide variety of economic-based incentivisation schemes proposed in the state of the art, the nature of approaches such as auctions, Contract Theory and Game Theory means that they cannot fulfill the requirement for *Anonymous, Unlinkable and Protected Data Submission (R1).* For example, the bidding process used by auctions grants the service provider access to participant activities. It should be noted, however, that other microeconomic concepts such as supply and demand could be used without violating the privacy preserving requirements for *Anonymous, Unlinkable and Protected Data Submission (R1)*, *Untraceable and Unlinkable Reward Allocation (R2)* and *Untraceable and Unlinkable Reward Spending (R3).* For example, the adoption of SenseUtil (Thepvilojanapong et al., 2013 and Tsujimori et al., 2014) would not violate these requirements (though, as noted in Section 3.1.1.4, it does not address requirement *R5*). Statistical methods are of more promise in meeting the needs of the question addressed in this thesis. Although statistical-based methods such as the STOC-PISCES algorithm (Biswas et al., 2015) at best only partially address the requirement for *Adaptive and Tunable Reward Allocation (R5)*, the inherent nature of methods such as Lyapunov Optimisation does not preclude their use in meeting requirements *R1*, *R2* and *R3*. These requirements could therefore be facilitated through the use of a statistical method that computes and allocates rewards without impinging upon a participant's private information, for example, location. Given that many participatory sensing systems operate in fast changing dynamic environments, the modeling of the incentivisation scheme as a stochastic process (i.e. one which randomly changes over time) that facilitates privacy preservation, adapts to environmental changes such as participation rates, optimises budget consumption and enables the prioritisation of data collection or budget consumption optimisation to be chosen would be of particular relevance to the requirements to be addressed in this thesis and is therefore a very promising approach.

3.2 Privacy Preservation

The inherent conflict to be addressed when designing a privacy-preserving incentivisation scheme is, on the one hand, the need to link participant submissions so as to reward them, and on the other, the need to break this link to ensure privacy preservation (Christin, 2015). Reconciling this conflict is a challenge because, as seen in Section 3.1, the attributes of many of the underlying methods used in the design of incentivisation schemes mean that they cannot fulfill the privacy preserving requirements for *Anonymous, Unlinkable and Protected Data Submission (R1)*, *Untraceable and Unlinkable Reward Allocation (R2)* and *Untraceable and Unlinkable Reward Spending (R3).* At the same time, any privacy preservation method should diminish data quality as little as possible so that the requirement for *Adaptive and Tunable Reward Allocation (R5)* is met[12]. This section explores the means by which privacy preservation for incentivisation can be addressed and discusses approaches in the state of the art that are pertinent to the area.

As outlined in Section 3.1.3, statistical methods in and of themselves do not impinge upon identity privacy. This raises the question of whether the privacy preservation approaches in the state of the art could be adapted in conjunction with the use of a statistical method for incentivisation to address the work to be undertaken in this thesis. To this end, Section 3.2.1 explores the appropriateness of the privacy preserving methods used in the state of the art in addressing requirements *R1*, *R2* and *R3*, their potential for adoption for a privacy-aware incentivisation scheme and their compatibility with the requirement for *Adaptive and Tunable Reward Allocation (R5).* The other key challenge to be addressed for a privacy-aware incentivisation scheme is to ensure that the medium for reward allocation is not itself a point of privacy violation. The state of the art

[12] This importance of data quality being unimpaired by the privacy preservation method used is also identified in work undertaken by Vergara-Laurens et al. [2013] and Jaimes et al. [2015a].

in this area is discussed in Section 3.2.2.

3.2.1 Privacy Preserving Methods for Incentivisation

There are a number of approaches that offer a potential architecture for privacy preserving incentivisation but do not present a fully fledged incentivisation scheme, for example, Saremi and Abdelzaher [2016] whose work considers how varying the nature of the incentivisation scheme can serve to increase participation rates. However, several of these approaches have inherent goals and attributes that mean that they cannot meet the privacy preserving requirements for *Anonymous, Unlinkable and Protected Data Submission (R1)*, *Untraceable and Unlinkable Reward Allocation (R2)* and *Untraceable and Unlinkable Reward Spending (R3)*. For example, some approaches address privacy trade-offs rather than privacy preservation (Wang et al., 2014, Katsomallos et al., 2017, Liu et al., 2017b and Chi et al., 2018). In addition, some approaches do not provide full participant privacy. For instance, Chen et al. [2014] point out that their approach has some degree of location privacy leakage in the interests of data quality while the approach taken by Gao et al. [2015b], an anonymisation-based location privacy method, envisages the selection of partners by the participant.

This section considers the different categories of privacy preservation that are used in the state of the art for participatory sensing. The use of pseudonyms and third party components in privacy preservation are discussed in Section 3.2.1.1 and Section 3.2.1.2 respectively while the use of anonymisation is explored in Section 3.2.1.3. Section 3.2.1.4 considers the use of encryption and statistical-based methods for privacy preservation.

3.2.1.1 Privacy Preservation using Pseudonyms

While the use of pseudonyms prevents direct access to participants' identities, the use of pseudonyms would not provide the level of identity privacy defined in Section 2.3 as they can be used by the service provider to track activity and behaviour (see, for example, Zhang et al., 2012b, Clarke and Steele, 2014b, Gisdakis et al., 2014, Lim and Abumuhfouz, 2015 and Yao et al., 2015). As a result, the requirement for *Anonymous, Unlinkable and Protected Data Submission (R1)* cannot be met. Moreover, there are some methods that claim to preserve identity privacy but actually seek to monitor participant activities and behaviours. For example, while Niu et al. [2018b] correctly assert that their privacy-preserving identification mechanism does not reveal the actual identity of a participant, the two-layer neural network used in their approach not only has access to participants' pseudonyms and/or IDs, it is used to learn participants' behaviours and activities with the objective of generating an identity feature database.

Other privacy preserving methods used in the state of the art such as mix-networks[13] also require the use of pseudonyms. While the mix-network based TrPF (Gao et al., 2013) uses two pseudonyms for participants on entering and leaving a region, this necessitates the use of a third party component which contains a listing of all participants and their pseudonyms. The approach proposed by Clarke and Steele [2014a], which uses mix networks to collect aggregated health data, is more robust from a privacy perspective and could be adapted to meet the requirement for *Anonymous, Unlinkable and Protected Data Submission (R1).* However, it cannot be extended to reward participants and thus cannot fulfill the requirement for *Untraceable and Unlinkable Reward Allocation (R2).* Other privacy preserving approaches, while not using pseudonyms per se, have attributes that enable pseudonymous tracking of participants. For example, the use of participant selection and ranking by Tian et al. [2017] in their Secure Multi Party Computation-based approach and the use of a bulletin board to tag locations by Techu (Agadakos et al., 2017) enables tracking of behaviour and activity.

3.2.1.2 Use of Third Party Components

While, as noted in Section 3.1.1.1, third party components can themselves be points of privacy vulnerability, there are several approaches that seek to offer privacy through the use of trusted third party components and publicly available third party software that, for example, register both participants and the service provider and can track when participants join and leave the system. (for example, De Cristofaro and Soriente, 2013, Li et al., 2017c, Kim et al., 2017, Xu et al., 2017a, Zhuo, 2017, Zhuo et al., 2017 and Chen et al., 2018)[14]. FIDES (Restuccia and Das, 2014) is itself a third party component that accesses all participant data while the PAMPAS (Privacy-Aware Mobile Participatory Sensing) approach (That et al., 2016), not only uses a third party component but also requires the enhancement of participants' sensing devices with bespoke secure

[13] Mix networks facilitate the transmission of data anonymously.

[14] Other examples of approaches that use third party components include those proposed by Chakraborty et al. [2012], Wang and Ku [2012], Xiao et al. [2012], Zhang et al. [2012b], Günther et al. [2014], Haderer et al. [2014], Saleem et al. [2014], Krontiris and Dimitriou [2015], Li et al. [2015b] and Zeng et al. [2016].

hardware. This is not consistent with the participatory sensing model outlined in Section 2.1 which only assumes that, at most, a mobile app is installed on participant devices.

It must be noted that while, in general, many of the third party components outlined for approaches in the state of the art mean that the requirement for *Anonymous, Unlinkable and Protected Data Submission (R1)* cannot be met, there are a number of privacy preserving approaches for data aggregation that incorporate third party components that have no access to a participant's private information (Erfani et al., 2013, Li and Cao, 2013a, Chen and Ma, 2014, Li et al., 2015a, Zhang et al., 2016b, Zhang et al., 2016a and Zhang et al., 2017) . Such approaches, therefore, could potentially meet the requirement. However, the use of, for example, homomorphic encryption by these approaches ensures that they cannot reward participants. As a result, the requirement for *Untraceable and Unlinkable Reward Allocation (R2)* cannot be met.

3.2.1.3 Anonymisation

Anonymisation is widely used in the state of the art, in particular, k-Anonymity and Differential Privacy. k-Anonymity, which eliminates the uniqueness of participants' information by merging the information for k (i.e. a number of) participants, is used in several privacy preserving approaches for participatory sensing (Rodhe et al., 2012, Vu et al., 2012, Alswailim et al., 2014, Lakshmi et al., 2017 and Wang et al., 2018b). Differential Privacy extends the k-Anonymity model by offering a formal technique to ensure that a computation does not reveal whether any one person participated in the input to the computation or not and is used in the approaches proposed by To et al. [2014] and Han et al. [2018]. However, neither k-Anonymity nor Differential Privacy are robust to inference attacks (Sun et al., 2014 and Liu et al., 2016a respectively) meaning that participant activity can be pseudonymously monitored, a violation of the requirement for *Anonymous, Unlinkable and Protected Data Submission (R1).*

Some approaches use anonymisation in conjunction with other methods, For example, in SLICER (Qiu et al., 2013, also discussed in Qiu et al., 2014), k-Anonymity is used to ensure that the service provider cannot identify the generator of the sensing record from at least k participants while encryption is used to secure data submission. In addition to using anonymisation, the approach taken by Liu et al. [2012] uses machine learning to classify sensed data. However, despite the use of additional methods, these approaches are still prone to inference attacks and would thus fail to meet privacy preserving requirements *R1*, *R2* and *R3*.

3.2.1.4 Encryption & Statistics

Encryption-based approaches have the potential to meet the requirement for *Anonymous, Unlinkable and Protected Data Submission (R1)*, However, they would fulfill requirement *R1* at the expense of *Untraceable and Unlinkable Reward Allocation (R2)* as the one-way communication used means that it is not possible to allocate rewards to participants (Biswas and Vidyasankar, 2012, Liu et al., 2013, Liu et al., 2017a, Shen et al., 2017, Xing et al., 2017, Yan et al., 2017, Perez and Zeadally, 2018 and Wang and Huang, 2018). Other, statistical-based, approaches also fulfill requirement *R1* but cannot be adapted to provide reward allocation (Sabrina and Murshed, 2012, Drosatos et al., 2012, Xing et al., 2013, Drosatos et al., 2014, Tan et al., 2016, Xiao et al., 2017 and Ziegeldorf et al., 2017) .

In addition, there are a number of approaches that achieve privacy preservation at the expense of *Adaptive and Tunable Reward Allocation (R5)* given this requirement's expectation that the quality of the service provider's dataset is unaffected by the privacy preservation method used. For example, approaches that use perturbation cannot fulfill requirement *R5* given that this method entails the modification of data to preserve its submitter's privacy (Zhang et al., 2012a, Lyu et al., 2016 and Lyu et al., 2018). The use of negative surveys[15] in the approaches taken by Aoki et al. [2012] and Groat et al. [2013], also has the potential to impair the fulfillment of requirement *R5* as perturbation is integral to the method. Similarly, privacy preservation through the coarsening and/or disguising of the participant's location using methods such as obfuscation, Laplace noise (a statistical method that modifies data) and dummy locations is not consistent with the data quality goals of the requirement (Boutsis and Kalogeraki, 2013, Gao et al., 2013, Agir et al., 2014, Mun et al., 2014, Wei et al., 2014, Wiesner et al., 2014, Bettini and Riboni, 2015 and Li et al., 2017a).

3.2.1.5 Methods for Privacy Preservation: Summary

This section has considered the wide variety of privacy preservation methods in the state of the art with a view to evaluating their potential appropriateness for

[15] This is a privacy-aware probabilistic method that keeps the target data undisclosed by asking participants to instead make a series of decisions with the data in mind (Esponda, 2006).

the work undertaken in this thesis. However, while many of these approaches offer anonymity, the vulnerability of pseudonyms, third party components and anonymisation to inference attacks means that such methods cannot provide the level of identity privacy defined in Section 2.3 and hence cannot fulfill the privacy preserving requirements for *Anonymous, Unlinkable and Protected Data Submission (R1)*, *Untraceable and Unlinkable Reward Allocation (R2)* and *Untraceable and Unlinkable Reward Spending (R3).* Other approaches that use encryption and statistics facilitate the requirement for *Anonymous, Unlinkable and Protected Data Submission (R1)* but make it impossible to allocate rewards to participants. Similarly, approaches using techniques such as perturbation have the potential to meet requirement *R1*, but in this case, at the expense of data quality, a violation of the requirement for *Adaptive and Tunable Reward Allocation (R5).* It must therefore be concluded that there is no current privacy preserving approach in the state of the art that could be adapted to meet the requirements pertaining to privacy preservation and data quality that have been identified for the work to be undertaken in this thesis.

3.2.2 Medium for Reward Allocation

The medium for reward allocation is a crucial question to be addressed when designing a privacy-preserving incentivisation scheme for participatory sensing. Reward allocation is typically addressed in the state of the art through the use of existing cryptocurrencies or the creation of reward tokens. These are discussed in Section 3.2.2.1 and Section 3.2.2.2 respectively.

3.2.2.1 Cryptocurrencies

Some approaches in the state of the art seek to achieve privacy-preserving incentivisation by using cryptocurrencies to allocate rewards. Cryptocurrencies, electronic forms of value exchange, have the potential to be used for online purchases, trading and transactions. They use cryptographic methods to protect the integrity of transactions and of the currency itself. However, while cryptocurrencies such as Bitcoin (Nakamoto, 2008) were developed to protect the privacy of those engaging in transactions, the creators of Bitcoin point out that the cryptocurrency offers anonymity but does not prevent its users from being pseudonymously tracked. Specifically, the address at which a payee receives Bitcoins acts as a pseudonym with every transaction involving that address being stored in the BlockChain. For example, PaySense (Tanas et al., 2015) uses the participant's Bitcoin address as a pseudonym. While cryptocurrency 'mixer' services can be used to make Bitcoins impossible to trace by enabling users to swap Bitcoins with each other, this necessitates the trusting of what is often an anonymous third party service. The approach proposed by Spathoulas et al. [2017], whose architecture necessitates the use of a third party component, also uses Bitcoin to reward participants. The use of Bitcoin, therefore, does not provide the level of identity privacy defined in Section 2.2 and will not fulfill the requirements for *Untraceable and Unlinkable Reward Allocation (R2)* and *Untraceable and Unlinkable Reward Spending (R3)*.

Alternatives to Bitcoin such as Monero[16] and DASH[17] also have weaknesses that would prevent rewards from being untraceable. Miller et al. [2017] indicate that there are weaknesses in Monero in its use of fake coins, called mixins, to obscure transaction behaviour but which, in fact, make transactions linkable under certain conditions as the mixins are sampled from a distribution that does not resemble real transaction inputs. In the case of DASH, an anonymous paper[18], approved by the currency's promoters, points out that it is necessary to anonymise network traffic over the Tor anonymity network in order to ensure that DASH can be used securely and anonymously. In addition, the third party MasterNode mixing service provided by a third party component increases the probability of tracing a payment, especially if there are few other users to swap coins with.

3.2.2.2 Reward Tokens

There are several approaches in the state of the art that use tokens, which are mappable to a tangible monetary or non-monetary item of value, to allocate rewards. However, this is sometimes at the cost of privacy preservation through the use of third parties (Zhang et al., 2012b and Li et al., 2017c) and pseudonyms (Zhang et al., 2012b). The credit token system proposed by Li and Cao [2016] (also described in Li and Cao, 2013b and Li and Cao, 2015) does attempt to address the challenges of anonymous reward allocation by using a blind signature to break the link between the credit token and what is termed the pseudo-credit so as to ensure that the service provider does not know the data submission for which the credit is earned. However, this approach does not address inference attacks as each credit token is directly linked to the participant's ID. This is

[16] See `https://getmonero.org`

[17] See `https://www.dash.org`

[18] See `https://dashpay.atlassian.net/wiki/display/DOC/Dash+Security-Privacy+Paper`

acknowledged by the authors themselves who point out that the approach is vulnerable to a credit-based inference attack as the service provider may infer if participants have submitted data for a task from the number of credits that they have, a violation of the requirement for *Untraceable and Unlinkable Reward Allocation (R2).* EPPI (Niu et al., 2014), which allocates rewards using token-based E-Cents, an exchangeable and untraceable unit bearer currency, also fails to meet the requirement for *Untraceable and Unlinkable Reward Allocation (R2).* This is because the approach's 'mix zone', which is used to enable participants to anonymously exchange E-Cents so as to ensure untraceability, requires the use of a pseudonym on the part of the participant and is itself a potential source of privacy violations if it is compromised. The privacy evaluation experiments carried out by the authors also indicate that the approach is, in certain circumstances, vulnerable to inference attacks. For example, a tracing probability of 23% is reported when the participant pledge, which functions as a motivation for the participant to submit truthful data (as the participant will forfeit the E-Cents if they submit false data), is set to 20 E-Cents. An evolution of EPPI (Niu et al., 2018a) also notes the same vulnerability in terms of tracing probability. In addition, as highlighted in Section 3.2.1.1, this approach introduces other potential points of privacy vulnerability.

The credit token scheme proposed by Dimitriou [2018a] is more robust as participants are issued with a single token that accumulates rewards and is not linkable to a particular data submission. However, participants are required to reveal their identity when redeeming rewards which means that their spending can be tracked. While an extension to this approach (Dimitriou, 2018b) does not appear to require participants to reveal their identity, it must be noted that the user ID is embedded in the single token, a point of privacy violation if the token is illegitimately accessed. Moreover, the user ID will be disclosed in the case of double spending, even in the case when this is accidental on the participant's part. Similarly, while the credit tokens issued by the scheme proposed by Liu et al. [2018] cannot be linked to data submissions, participants are required to reveal their identity when depositing them. Indeed, the authors admit that there is a possibility of linkages with data submissions being made if a participant deposits multiple credit tokens simultaneously. As a result, while these approaches do address the requirement for *Untraceable and Unlinkable Reward Allocation (R2)* to a certain extent, they do not fulfill the requirement for *Untraceable and Unlinkable Reward Spending (R3).*

3.2.2.3 Medium for Reward Allocation: Summary

The options for reward allocation in participatory sensing incentivisation schemes can be categorised into cryptocurrencies and credit tokens. Cryptocurrencies are unsuitable in meeting the goals of this thesis as their pseudonymous attributes preclude them from fulfilling the requirement for *Untraceable and Unlinkable Reward Allocation (R2)*. Similarly, the approaches in the state of the art that use reward tokens at best only partially address this requirement while there is no approach that meets the requirement for *Untraceable and Unlinkable Reward Spending (R3)*. Nevertheless, tokens represent the most promising means of addressing the privacy preserving requirements of this thesis if they are used in a way that does not necessitate the use of third party components and pseudonyms and, crucially, provide a level of untraceability that meets the level of identity privacy defined in this thesis.

3.3 Incentive Compatibility

In the area of participatory sensing, incentive compatibility seeks to address the question of whether the service provider can trust the data it receives from participants. In order to meet the privacy preserving requirements for *Anonymous, Unlinkable and Protected Data Submission (R1)*, *Untraceable and Unlinkable Reward Allocation (R2)* and *Untraceable and Unlinkable Reward Spending (R3)*, this must be achieved without violating participant privacy. Ensuring both privacy preservation and data trustworthiness is a key challenge of participatory sensing systems as the implementation of one can compromise the other (Marusic et al., 2014). The incentive compatibility approaches taken in the state of the art are either reputation management systems that evaluate how trustworthy participants are, considered in Section 3.3.1, or methods to evaluate data truthfulness independently of who submitted it, discussed in Section 3.3.2.

3.3.1 Trust & Reputation Management

There are several approaches in the state of the art that claim incentive compatibility through the design of reputation management systems. Such systems function by assigning a reputational value to each participant as a measure of the trust placed on data submitted by that participant. Reputation management is thus used to evaluate the trustworthiness of data submitters. However, the computation of a participant's reputation score is often at the expense of participant privacy as access to, for example, participants' past performance and ability is required (Yang et al., 2011, Alswailim et al., 2016, Guo et al., 2016, Lu et al., 2017, Mousa et al. 2017, Xiang et al., 2017, Zenonos et al., 2017 and Zhou et al., 2017). Other reputation management systems incorporate mechanisms that increase privacy vulnerabilities through, for example, the use of third party components that can monitor participant activity or the granting of access of private information to other participants (Amintoosi et al., 2014, Ren et al., 2015, Haider et al., 2016, Gao et al., 2017, Mihaita et al., 2017, Pouryazdan et al., 2017, Sun et al., 2017 and Restuccia et al., 2018b).

Several approaches claim privacy-preserving incentive compatibility through the design of privacy-preserving reputation management systems. However, while many of these privacy-preserving trust and reputation management systems offer anonymity, several approaches enable pseudonymous tracking of participant activity (Christin et al., 2014, Michalas and Komninos, 2014, Tanas et al., 2015 and Wang et al., 2018c). Some approaches do seek to address the vulnerability of pseudonyms to tracking through for example, the generation of a unique pseudonym for each data submission. However, these necessitate the use of third party components that could themselves be points of privacy vulnerability (Huang et al., 2012, Chang et al., 2013 and Christin et al., 2013b). For instance, the third party component used in the approach proposed by Chang et al. [2013] holds a list of mappings between the participant's real identity and all the participant's pseudonyms. Other approaches seek to protect the participant against inference attacks but do so at the expense of other aspects of privacy by, for example, assuming that participants are willing to share location information with each other (Kalui et al., 2016 and Hu et al., 2018). As a result, none of these approaches meet the requirement for *Anonymous, Unlinkable and Protected Data Submission (R1).*

There are a number of approaches that go further in addressing the potential for inference attacks in reputation management systems. For example, the TAPAS (Trustworthy Privacy-Aware Participatory Sensing scheme) protocol (Kazemi and Shahabi, 2013) addresses the problem of participants being identified by their location but will only meet meet the requirement for *Anonymous, Unlinkable and Protected Data Submission (R1)* if there is a critical mass of participants in a particular location. Similarly, ARTSense (Wang et al., 2013), which separates the data reporting and reputation update processes and uses a different unlinkable blinded ID for each submission, also requires a critical mass of participants to ensure privacy.

3.3.2 Data Truthfulness

Rather than recording participants' reputations and trustworthiness, a number of approaches in the state of the art seek to address incentive compatibility by evaluating data truthfulness. For example, the approach proposed by Farokhi et al. [2015] uses Game Theory to model interactions between participants and service providers as a strategic game that ultimately encourages truthful data submissions, albeit without addressing how to actually evaluate the truthfulness of the submitted data.

Some approaches claim data truthfulness through the incentivisation mechanism used. This particularly applies for auction-based approaches (for example, the sealed-bid online auction used by Sun and Ma [2014]) although in some cases this pertains to participants' bids rather than the data submissions they make (Zhang et al., 2014). However, while claims that auctions are inherently incentive compatible are supported by some in the field of economics (Smith, 1977), there is not unanimous agreement on this point (for example, Brubaker, 1980). Crucially, as previously noted in Section 3.1.1.1, the use of auctions does not meet the requirement for *Anonymous, Unlinkable and Protected Data Submission (R1)*. Other approaches that evaluate data truthfulness are also in conflict with requirement *R1* through the need for reputation scores (Jin et al., 2017a), third party components (Gisdakis et al., 2015, Miao et al., 2017 and Zhang et al., 2018b) and direct access to private participant information such as location and sensing activity (Bhattacharjee et al., 2017, Cheng et al., 2017, Gong and Shroff, 2017 and Gong and Shroff, 2018). While there are a number of approaches that do provide direct identity privacy, they do not address the potential for inference attacks. For example, Yang et al. [2017b] envisage one of their approach's core algorithms being used to reward and reprimand users without addressing how this can be done in a privacy preserving manner.

There are a number of approaches that do ensure data truthfulness without violating participant privacy. For example, incentive compatibility is claimed by the approach proposed by Yang et al. [2015] through the use of mechanism design as participants ultimately make losses if they make untruthful data submissions. However, this is at the expense of the service provider as there are losses in the meantime through the rewarding of untruthful data which is not initially detected, a violation of the requirement for *Adaptive and Tunable Re-*

ward Allocation (R5) which seeks to optimise the service provider's budget. On the other hand, the approach outlined by Xiang et al. [2015] (also considered in Xiang et al., 2013) would not conflict with requirement *R5.* Furthermore, while the focus of this approach on calibrating sensing devices for the purpose of monitoring pollution sources does not meet the needs of the participatory sensing model presented in Section 2.1, the Expectation Maximisation statistical method it uses to create a model that evaluates data truthfulness would not violate the privacy preserving requirements for *Anonymous, Unlinkable and Protected Data Submission (R1)*, *Untraceable and Unlinkable Reward Allocation (R2)* and *Untraceable and Unlinkable Reward Spending (R3).* CLOR (Zhou et al., 2016) also facilitates privacy preservation. However, the focus of this approach on clustering data solely on the basis of location would make it difficult to adapt or extend to meet the needs of the participatory sensing model presented in Section 2.1.

3.3.3 Incentive Compatibility: Summary

The majority of approaches in the state of the art for incentive compatibility do not preserve participant privacy (Feng et al., 2017). The monitoring of participant behaviour and activity in most reputation management systems violates the requirement for *Anonymous, Unlinkable and Protected Data Submission (R1).* Those reputation management systems that do seek to address privacy preservation are also incompatible with requirement *R1* as, for example, the facilitation of pseudonymous participant tracking results in linkable data submissions and traceable and linkable reward allocation. Many methods that seek to evaluate data truthfulness exhibit similar shortcomings from a privacy preservation perspective through their use of auctions, reputation scores and third party components. However, statistical based approaches have the potential to meet the requirement for *Incentive Compatibility (R4)* and would not hinder the requirements for *Anonymous, Unlinkable and Protected Data Submission (R1)*, *Untraceable and Unlinkable Reward Allocation (R2)*, *Untraceable and Unlinkable Reward Spending (R3)* and *Adaptive and Tunable Reward Allocation (R5).* For example, the use of the Expectation Maximisation statistical method in the approach proposed by Xiang et al. [2015] would be appropriate for the work to be undertaken in this thesis.

3.4 Summary

This chapter evaluates related work in the areas of incentivisation, privacy preservation and incentive compatibility for participatory sensing in terms of meeting the requirements of this thesis for *Anonymous, Unlinkable and Protected Data Submission (R1)*, *Untraceable and Unlinkable Reward Allocation (R2)*, *Untraceable and Unlinkable Reward Spending (R3), Incentive Compatibility (R4)* and *Adaptive and Tunable Reward Allocation (R5).* While there has been work in the area of privacy-preserving incentivisation, the approaches to date do not address the potential for inference attacks to be carried out by the service provider and thus do not provide the level of identity privacy defined in this thesis. Therefore, while the privacy preserving incentivisation mechanisms outlined by Niu et al. [2014], Li and Cao [2016] and Dimitriou [2018b] come closest to addressing requirements *R1*, *R2* and *R3,* there is no scheme in the state of the art that fully addresses all of the privacy preserving requirements. In addition, it should also be noted that, although the STOC-PISCES algorithm (Biswas et al., 2015) and SenseUtil (Thepvilojanapong et al., 2013 and Tsujimori et al., 2014) partially address the requirement for *Adaptive and Tunable Reward Allocation (R5)*, there is no incentivisation approach that fully addresses requirement *R5.* There is therefore a need for a privacy-preserving incentivisation approach that is adaptable to current environmental conditions and participation rates; can be tuned by the service provider; is robust with respect to inference attacks; is not itself a point of privacy vulnerability and facilitates incentive compatibility in a privacy-preserving manner.

The use of stochastics to model the incentivisation scheme is of interest as such methods model random changes in an environment. Lyapunov Optimisation which is, as was outlined in Section 3.1.2.1, used by some incentivisation schemes in the state of the art, is suitable for modeling rapid changes over time in an environment and can be tuned to prioritise different goals. It therefore has particular promise as a method that would meet the requirement for *Adaptive and Tunable Reward Allocation (R5).*

Stochastic methods do not have inherent privacy vulnerabilities but in and of themselves do not address the challenge of *Untraceable and Unlinkable Reward Allocation (R2)* and *Untraceable and Unlinkable Reward Spending (R3).* While cryptocurrencies do not meet these requirements, the underlying BlockChain technology and decentralised exchanges used by these digital assets offer interesting possibilities in meeting both these requirements and the requirement for

Anonymous and Unlinkable Data Submission (R1). Lastly, as noted in Section 3.3.3, there are a number of statistical methods that could be used to meet the requirement for *Incentive Compatibility*, for example, an Expectation Maximisation method such as Maximum Likelihood Estimation.

Chapter 4

Design

This chapter discusses the design of the *Privacy-Aware Incentivisation (PAI)* approach. As outlined in Chapter 2, PAI must address five requirements to meet the goal of providing privacy preserving reward allocation and spending that is adaptive to the environment and the needs of the service provider. To achieve privacy preservation, PAI meets the requirements for *Anonymous,Unlinkable and Protected Data Submission (R1)*, *Untraceable & Unlinkable Reward Allocation (R2)* and *Untraceable & Unlinkable Reward Spending (R3)* by using a decentralised platform, referred to as *Identity Privacy Preserving Incentivisation (IPPI).* Section 4.1 describes how IPPI addresses requirements *R1*, *R2* and *R3*. The requirement for *Incentive Compatibility (R4)*, achieved through the development of a *Data Truthfulness Estimation (DTE)* algorithm, is discussed in Section 4.2. Finally, the requirement for *Adaptive and Tunable Reward Allocation (R5)* is met through the development of a Lyapunov Optimisation-based model, referred to as *Adaptive Reward Allocation (ARA)* and is described in Section 4.3. Section 4.4 summarises this chapter. Notations used in the chapter are defined as they are introduced as well as in the nomenclature as the end of this thesis.

Figure 4.1 presents the PAI platform and the core components (IPPI, DTE and ARA) that are used to meet the requirements for privacy preservation, reward allocation and incentive compatibility respectively.

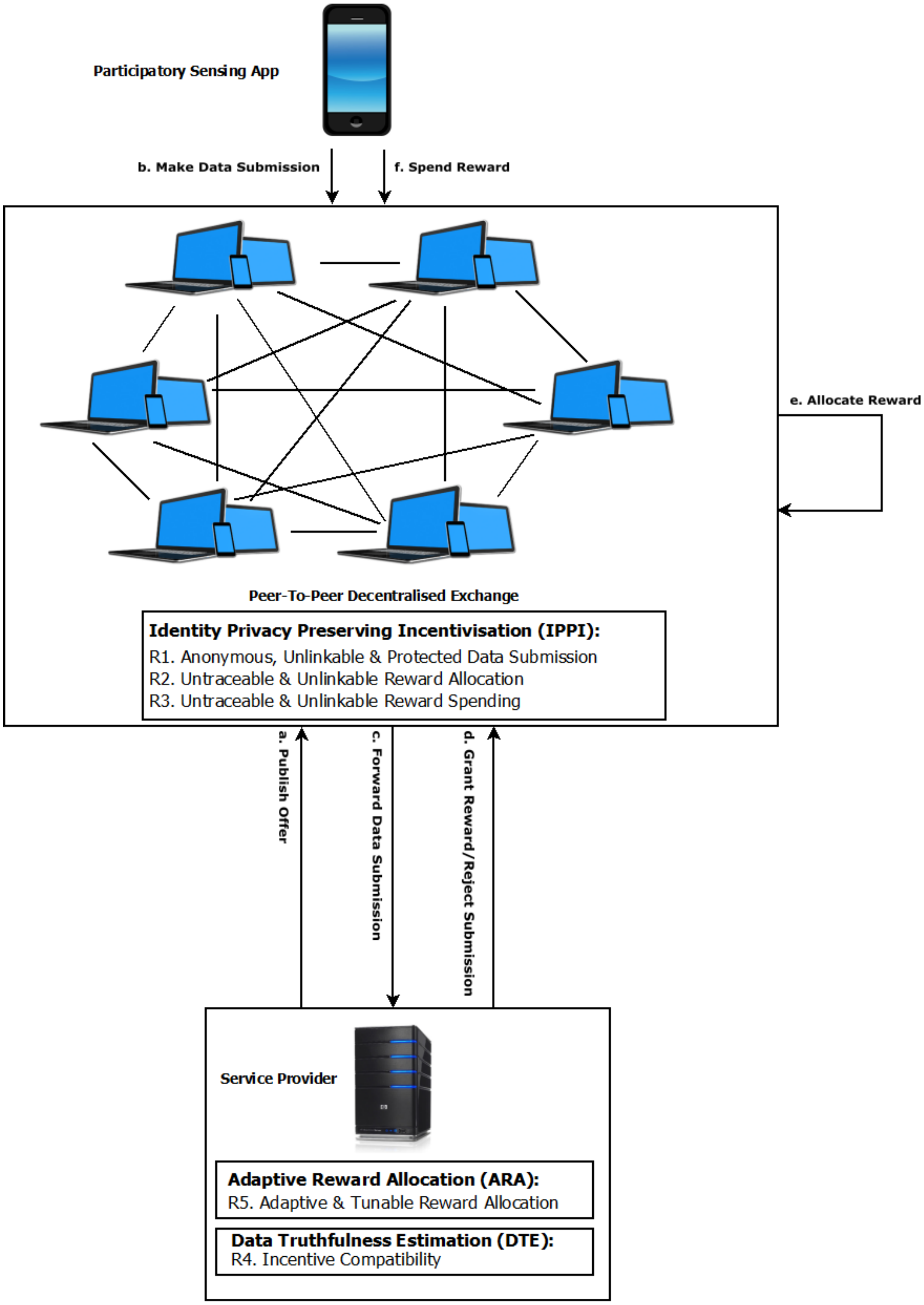
Participatory Sensing App
b. Make Data Submission
f. Spend Reward
e. Allocate Reward
Peer-To-Peer Decentralised Exchange
Identity Privacy Preserving Incentivisation (IPPI):
R1. Anonymous, Unlinkable & Protected Data Submission
R2. Untraceable & Unlinkable Reward Allocation
R3. Untraceable & Unlinkable Reward Spending
a. Publish Offer
c. Forward Data Submission
d. Grant Reward/Reject Submission
Service Provider
Adaptive Reward Allocation (ARA):
R5. Adaptive & Tunable Reward Allocation
Data Truthfulness Estimation (DTE):
R4. Incentive Compatibility

Figure 4.1: The PAI Platform

4.1 Privacy Preservation for the PAI Platform

This section discusses how PAI's privacy preserving requirements are addressed. Section 4.1.1 discusses the suitability of different privacy preserving methods used in other domains[1] in meeting the requirements for *Anonymous, Unlinkable and Protected Data Submission (R1)*, *Untraceable and Unlinkable Reward Allocation (R2)* and *Untraceable and Unlinkable Reward Spending (R3)*. The design of the *Identity Privacy Preserving Incentivisation (IPPI)* decentralised exchange model used to address requirement *R1* is discussed in Section 4.1.2. Section 4.1.3 introduces the concept of the One-Time Key and how this is used to fulfill requirement *R2*. Finally, Section 4.1.4 discusses how IPPI meets requirement *R3*.

4.1.1 Meeting the Privacy Requirements

As discussed in Chapter 3, many of the underlying approaches in the state of the art have attributes that fail to meet PAI's privacy preserving requirements, *R1*, *R2* and *R3*. due to, for instance, the use of pseudonyms, third party components and techniques that are vulnerable to inference attacks such as anonymisation. There are a number of approaches considered in the state of the art in other domains that could preserve participant privacy, for example, Virtual Private Networks (Sharma and Yadav, 2015). However, like some approaches in the state of the art for participatory sensing (for example, the source anonymous message authentication used by Li et al., 2015a), these approaches would disguise the provenance of a data submission but would make it impossible to actually allocate rewards to participants, thus in turn making it impossible to fulfill requirements *R2* and *R3*. Similarly, the use of an encryption method would address the Semi-Honest Threat Model discussed in Section 2.3 but would prevent reward allocation unless that use is modified.

Other approaches such as data obfuscation can also disguise data but do not address the privacy preservation of data submissions in transit, leading to the potential viewing of submissions by third parties other than the service provider. This violates requirement *R1*, which seeks to protect data submissions so that they can only be viewed by the service provider. Similarly, steganography is vulnerable to stegananalysis attacks (Mishra and Bhanodiya, 2015) that, in the

[1] As discussed in Chapter 3, the privacy preserving methods from the state of the art for participatory sensing do not meet the goals of this thesis.

case of participatory sensing, would potentially reveal both the data submissions and the rewards.

As pseudonyms and third party components cannot be used to meet PAI's privacy preserving requirements, a **decentralised** rather than **centralised** approach would be of promise in ensuring that there is no central server, third party component or other means of tracing and linking participants' data submissions and rewards. While, as discussed in Section 3.2.2.1, the use of cryptocurrencies is unsuitable for PAI, the fundamental architecture used by decentralised cryptocurrency exchanges is of interest given that such approaches do not use any central components. Decentralized cryptocurrency exchanges such as CryptoNote (See `https://cryptonote.org`) are peer-to-peer (P2P) networks consisting of computers known as *nodes* with all exchange users sharing responsibility for payment processing and recording. There is no central authority, coordinating entity or middlemen[2].

The fundamental operation of decentralised exchanges is the peer-to-peer trading of cryptocurrencies i.e. trading one cryptocurrency for another. Transactions are listed on a distributed ledger called the *OrderBook* (Hileman and Rauchs, 2017), a database that resides on multiple peer devices (Mills et al., 2016). In addition to recording completed transactions, the OrderBook is used to validate and/or authenticate these transactions. Furthermore, as this database is a distributed one, there is no single point of failure and, in particular, there is no central third party component that could be the target of an inference attack. Typically, individual transactions are stored as blocks that are linked using cryptography with multiple blocks together forming a *blockchain*. Transactions cannot be modified or removed once they are recorded in the blockchain.

Figure 4.2 presents the operation of a typical decentralised cryptocurrency exchange. When a party (the sender in the diagram) makes a request to trade one cryptocurrency for another, that trade is broadcast to the P2P network. The request can be broadcast to all potential parties who would potentially be willing to make the trade or forwarded to a recipient (for simplicity this scenario is depicted in Figure 4.2 with the recipient being referred to as the receiver) who then accepts or rejects this transaction. On acceptance of a trading request, the receiver broadcasts confirmation to this effect to the P2P network. The P2P network then validates the transaction and, typically, combines the transaction

[2] While decentralised approaches for participatory sensing have been considered in the state of the art (for example, Tsolovos et al., 2018), the focus of these approaches is on securing the sensed data from potential attacks rather than privacy preserving reward allocation.

with other transactions to create a new block of data that is added to the existing blockchain on the OrderBook.

While the peer-to-peer trading of cryptocurrencies means that two parties can interact directly with each other without the involvement of a central authority, the fundamental trading operation in participatory sensing (the rewarding of participants by the service provider in exchange for a data submission) must be conducted without leakage of the participant's identity privacy. Therefore, while this architecture can serve as a basis for the IPPI platform, it must be modified to fulfill the properties of untraceability and unlinkability sought by PAI.

In particular, the role of the OrderBook is solely for publishing transactions, with the service provider (equivalent to the receiver in Figure 4.2) being responsible for validating data submissions. While these listed transactions contain encrypted data, they do not necessarily have to be stored in a blockchain as a distributed database would suffice in meeting the privacy preserving requirements. In addition, all data submission rejections as well as acceptances from the service provider are published on the OrderBook so as to indicate whether participants (the equivalent to the sender in Figure 4.2) are to be rewarded or not. Furthermore, while the OrderBook plays a key role in privacy preservation in removing the need for direct communication between the participant and the service provider and in providing untraceable and unlinkable rewards, cryptographic operations must also be conducted by these parties to meet the privacy preserving requirements of PAI. Finally, it should be noted that, unlike decentralised cryptocurrency exchanges, the service provider must not have access to any pseudonym identifying the participant (while still being able to allocate a reward).

A decentralised exchange would play a fundamental role in addressing the requirement for *Anonymous, Unlinkable and Protected Data Submission (R1)*. However, it would not in and of itself address the requirements for *Untraceable and Unlinkable Reward Allocation (R2)* and *Untraceable and Unlinkable Reward Spending (R3)*. As discussed in Chapter 3, crypocurrencies are not a suitable medium of allocation in addressing the goals of this thesis as their pseudonymous nature means that they can be used to track participant behaviour and activity. While the token-based approaches for reward allocation in the state of the art have similar tracing vulnerabilities, a reward token that is both untraceable and unlinkable would fulfill requirements *R2* and *R3*. By being unlinkable, such a token would also be robust to side information attacks (defined by, for

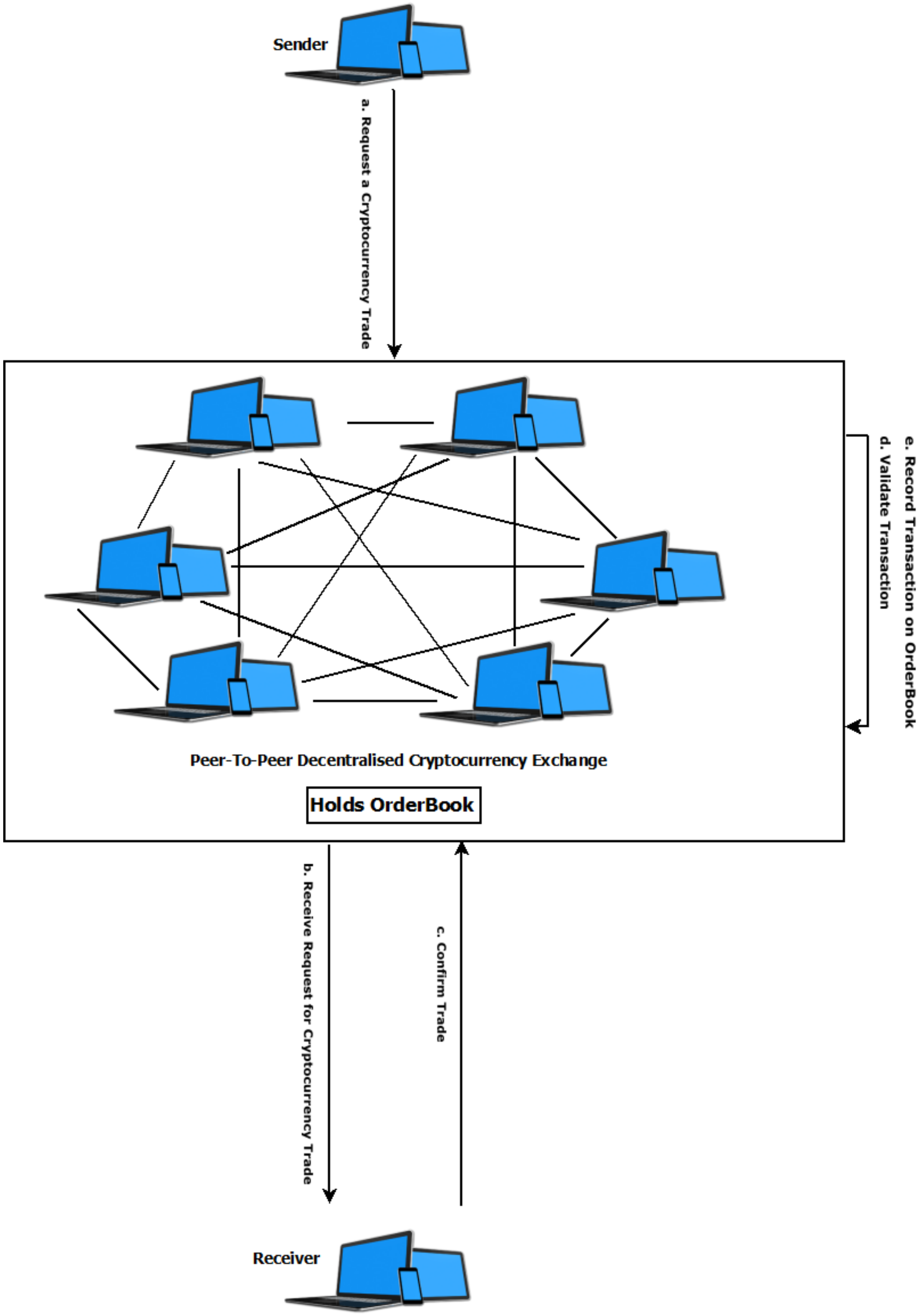
Sender
a. Request a Cryptocurrency Trade
Peer-To-Peer Decentralised Cryptocurrency Exchange
Holds OrderBook
e. Record Transaction on OrderBook
d. Validate Transaction
b. Receive Request for Cryptocurrency Trade
c. Confirm Trade
Receiver

Figure 4.2: Operation of a Typical Decentralised Cryptocurrency Exchange

example, Tang and Ren [2015]) as the absence of an ID tied to the participant's identity means that potential attackers cannot use side information to link data submissions to a particular participant.

4.1.2 Anonymous, Unlinkable & Protected Data Submission (R1)

The concept of a decentralised exchange presents a number of attributes that facilitate IPPI in its privacy preserving approach. In particular, the peer-to-peer architecture facilitates the requirement for *Anonymous, Unlinkable and Protected Data Submission (R1)* without the need to hold any private data pertaining to participants by removing the need for direct communication between participants and the service provider. This prevents the service provider from tracing a participant through, for example, an IP address. Moreover, this is achieved without the introduction of any potentially privacy compromising third party components[3]. As IPPI's architecture necessitates the use of peers, these peers can be incentivised to host the service provider's data submission and reward information by receiving a payment or being granted access to the service provided. It should be noted that, as outlined in Section 2.1.2, the addressing of networking issues is regarded as being outside the scope of this thesis. For this reason, potential issues such as ensuring data consistency between peer devices are not considered.

While the basic network architecture remains the same, the other core concepts used for decentralised cryptocurrency exchanges are modified for the purposes of IPPI to ensure anonymous and unlinkable data submission as well as untraceable reward allocation. The OrderBook, which is used to record expressions of interest by both buyers and sellers of currency trades, is used by the service provider to publish requests for data submissions and allocate rewards and by participants to make anonymous data submissions and receive untraceable rewards. Offers, which contain details of the data requested and the reward being offered, are published on the OrderBook by the service provider. All participants are aware of the existence of an offer when it is generated and can elect to respond to it by submitting data and a reward token. Rewards are allocated until the desired number of responses is achieved or when the offer expires.

[3] This absence of a central server also diminishes the potential for other external attacks such as Distributed Denial of Service (DDOS) as all nodes hosting the network would have to be targets.

Figure 4.3 presents an overview of how the IPPI platform provides privacy preservation.

4.1.3 Untraceable & Unlinkable Reward Allocation (R2)

The concept of the One-Time Key, which is used in cryptocurrency exchanges to ensure that multiple payments received by the same payee cannot be linked[4] is modified by IPPI to ensure that the service provider (the equivalent of the payer in a cryptocurrency exchange) does not have access to the participant's real or pseudonymous identity, thereby providing untraceable rewards to participants and preventing inference attacks.

The One-Time Key used in decentralised cryptocurrency exchanges is based on the Diffie-Hellman Key Exchange Protocol (Diffie and Hellman, 1976). Diffie-Hellman is a cryptographic algorithm that allows two parties to produce a common secret key derived from their public keys. This shared secret is then used to exchange cryptography keys for use in symmetric encryption algorithms such as AES. In the modified use of the Diffie-Hellman Key Exchange adopted in decentralised cryptocurrency exchanges, the sender (i.e. the payer) uses the receiver's (i.e. the payee's) public address to compute a One-Time Key for the payment. As the sender and receiver can compute only the public and private parts of this key respectively, only the receiver can access and transfer the funds after the transaction is committed with the private part of the One-Time Key being used to confirm that the transaction indeed belongs to this receiver. As the receiver is the only party with access to the private key component, no other party can confirm the transaction and hence link the One-Time Key with the receiver's unique public address.

As used in decentralised cryptocurrency exchanges, the concept of the One-Time Key is not suitable for IPPI as the sender requires access to the receiver's public address. In the equivalent participatory sensing scenario, this would mean that the service provider would need access to the participant's identity (or pseudonym) for reward allocation. Therefore, IPPI modifies the use of the One-Time Key's underlying Diffie-Hellman Key Exchange Protocol to create a One-Time Key to provide untraceable rewards. The use of the Diffie-Hellman protocol is modified for IPPI to enable the participant to publish the public component of the One-Time Key on the OrderBook whilst retaining the private component i.e. the participant holds both parts of the One-Time Key.

[4] See, for example, `https://cryptonote.org`.

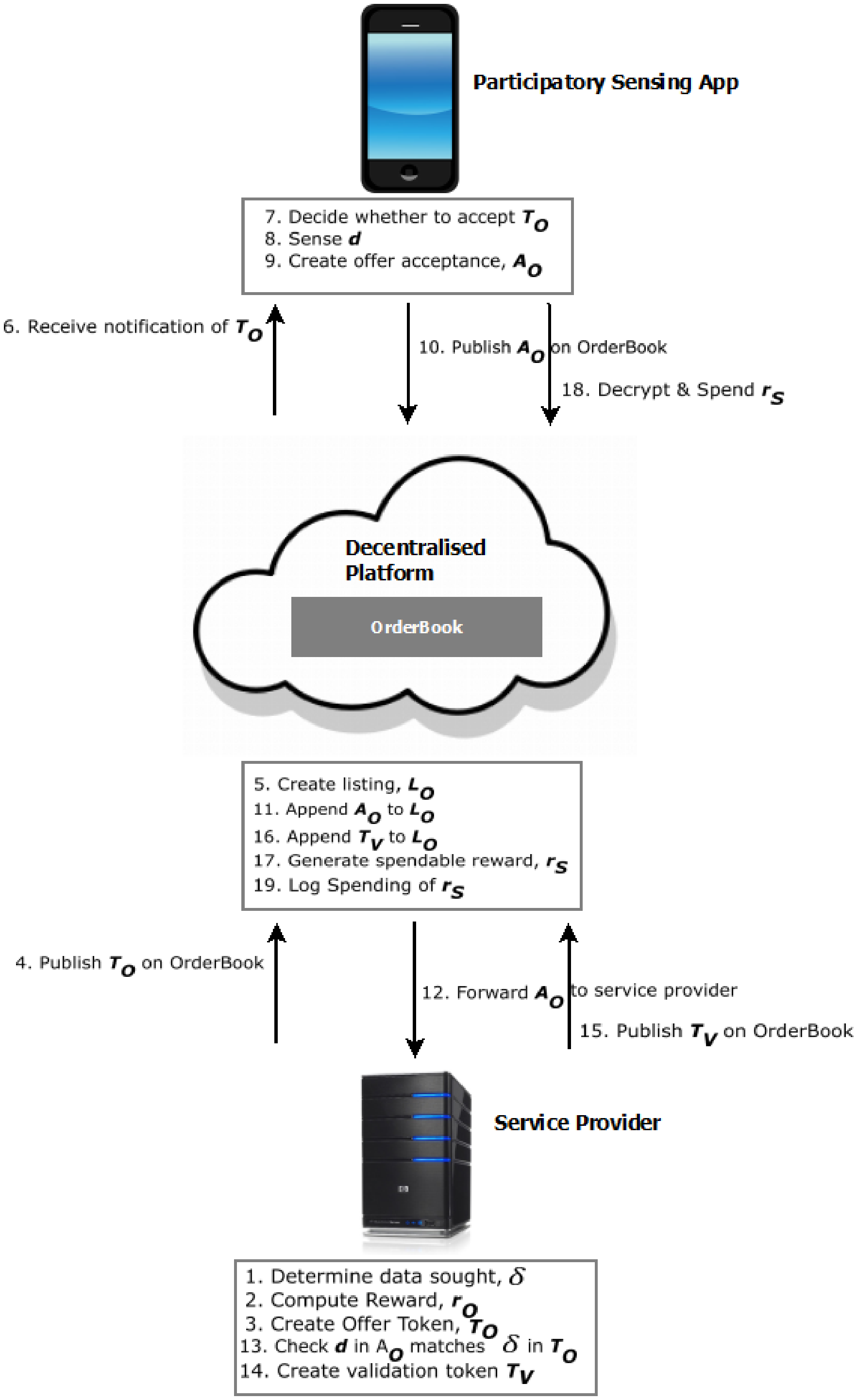
Participatory Sensing App
7. Decide whether to accept T_O
8. Sense d
9. Create offer acceptance, A_O
6. Receive notification of T_O
10. Publish A_O on OrderBook
18. Decrypt & Spend r_S
Decentralised Platform
OrderBook
5. Create listing, L_O
11. Append A_O to L_O
16. Append T_V to L_O
17. Generate spendable reward, r_S
19. Log Spending of r_S
4. Publish T_O on OrderBook
12. Forward A_O to service provider
15. Publish T_V on OrderBook
Service Provider
1. Determine data sought, δ
2. Compute Reward, r_O
3. Create Offer Token, T_O
13. Check d in A_O matches δ in T_O
14. Create validation token T_V

Figure 4.3: Privacy Preservation in the IPPI Platform

The service provider makes participants aware of its offer by publishing an **offer token,** T_{O} on the OrderBook.

$$T_{\mathrm{O}} = \{\delta, r_{\mathrm{O}}, i_{\mathrm{SP}}, i_{\mathrm{O}}\} \tag{4.1}$$

where δ comprises the type and granularity of the data being sought as well as other conditions such as location, the number of data submissions sought and when the offer expires, r_{O} is the amount of the reward offered, i_{SP} is the ID of the service provider and i_{O} is the ID of the offer token. The offer token, T_{O} is published on the OrderBook as part of a listing, L_{O}, to which participants' responses are appended.

A participant who accepts the offer token, T_{O}, then generates a One-Time Key, K_{O} (consisting of $a_{\mathrm{K_O}}$ and $a^*_{\mathrm{K_O}}$) and an offer acceptance

$$A_{\mathrm{O}} = \{\{d\}_{\mathrm{b_{SP}}}, a_{\mathrm{K_O}}, i_{\mathrm{O}}, i_{\mathrm{A_O}}\} \tag{4.2}$$

where $a_{\mathrm{K_O}}$[5] is the public part of the generated key K_{O}. As the service provider may not want a peer, or indeed another potential service provider, to view the data it is paying for, the participant is required to take the service provider's privacy requirements into account by encrypting the data submission d in the offer acceptance using the service provider's public key b_{SP}. A_{O} also contains the ID of the corresponding offer token, i_{O}. $i_{\mathrm{A_O}}$is A_{O}'s unique ID and is assigned by the OrderBook on receipt of A_{O}. To ensure untraceability, the participant does not assign any ID to A_{O}.

A_{O} is appended to the offer listing, L_{O}, on the OrderBook and is then forwarded to the service provider who, after decrypting $\{d\}_{\mathrm{b_{SP}}}$ using its private key b^*_{SP}, determines whether the data submission merits a reward. This entails verifying that the data submission matches the criteria set out in the offer token, T_{O}, and assessing the truthfulness of the data submitted (as will be discussed in Section 4.2). The service provider has no means of determining the identity of the data submitter when this evaluation is taking place.

Having evaluated the data, d, in A_{O}, the service provider generates a validation token

$$T_{\mathrm{V}} = \{i_{\mathrm{V}}, i_{\mathrm{A_O}}, v\}_{\mathrm{b^*_{SP}}} \tag{4.3}$$

[5] The symbols used correspond to those used in the approach proposed by Diffie and Hellman [1976].

which includes T_{V}'s unique ID, $i_{\mathrm{V},}$, the offer acceptance ID $i_{\mathrm{A_O}}$and a flag v which denotes whether the data submission merits a reward. The service provider signs T_{V} using its private key, b^*_{SP}, and publishes it on the OrderBook. It is assumed that the service provider is not a malicious one and will honestly allocate rewards to those data submissions that merit one. This is consistent with the assumption that the service provider is rational as, while depriving legitimate data submissions of rewards may lead to some short terms savings in terms of budget consumption, it will ultimately demotivate participants and impair the service provider in attracting the quality of data submissions it requires.

4.1.4 Untraceable & Unlinkable Reward Spending (R3)

As discussed in Section 3.2.2, the service provider may seek to track participants not only through the allocation of a reward but also the spending of that reward. To prevent this potentiality and fulfill the requirement for *Untraceable and Unlinkable Reward Spending (R3)*, the OrderBook first generates a unique spendable reward ID i_{S}, and uses the public part of the One-Time Key, $a_{\mathrm{K_O}}$, to encrypt a spendable reward token

$$r_{\mathrm{S}} = \{i_{\mathrm{S}}, i_{\mathrm{V}}, r_{\mathrm{O}}\}_{\mathrm{a_{K_O}}} \tag{4.4}$$

which is comprised of i_{S}, the associated ID of the validation token, i_{V} and the reward value, r_{O}. It then appends the validation token T_{V} and the encrypted spendable reward r_{s} to the offer listing L_{O} on the order book.

When a participant wants to spend a reward, it retrieves the encrypted spendable reward r_{s} from the offer listing, L_{O}, decrypts it thanks to $a^*_{\mathrm{K}_O}$, the private part of the One-Time Key, K_{O}, and sends it to the OrderBook. The OrderBook checks the validity of the ID provided, i_{S}, checks that it has not been spent previously and then verifies the signature of the associated validation token, T_{V}, using the service provider's public key b_{SP}, to ensure that the validation token was indeed generated by the service provider. It then permits spending of the reward and, to prevent the problem of double spending, logs it as spent.

While all participants can see that a data submission has been given a reward, only the participant who made the data submission can spend the reward allocated by decrypting the spendable reward, r_{s}, using the private part of the One-Time Key, $a^*_{\mathrm{K_O}}$. Other participants are unable to forge this verification. Moreover, so as to ensure that the service provider cannot change its signature to track spendable rewards, the OrderBook holds an identity certificate signed by a peer to confirm that the service provider is the owner of the public key, b_{SP}, used to verify all rewards issued by the service provider.

Algorithm 1 presents the algorithm used to ensure that participants are allocated untraceable rewards in exchange for an anonymous data submission. The algorithm is initiated when a service provider publishes an offer token and a participant accepts this offer. Algorithm 2 presents the algorithm used when spending the reward.

4.2 Incentive Compatibility (R4)

Economic theory states that, when resources are being allocated among a group, individuals may find it in their interest to distort the information they provide so that they can acquire more of the resources than they should be entitled to (Ledyard, 1977). These distortions in turn may lead to a suboptimal situation for the group as a whole as resources are inappropriately allocated. For participatory sensing, such a situation would occur if the service provider allocates rewards to participants who submit false or inaccurate data submissions. The service provider suffers under this situation as its budget is wasted and the quality of its dataset is diminished. Other honest participants suffer as the budget for legitimate data submissions is effectively reduced meaning that they could be deprived of rewards they would otherwise receive.

As outlined in Chapter 3, the potential for dishonest behaviour is addressed by the economic concept of incentive compatibility. As discussed in that chapter, the use of incentive compatibility for participatory sensing seeks to ensure that rewards are only allocated by the service provider in return for truthful data. In this context, a truthful data submission is defined as one which accurately reflects the environmental measurement(s) being sought. An 'untruthful' data submission might not necessarily reflect dishonesty on the data submitter's part (for example, the submission could contain inaccurate readings due to a hardware problem in the participant's device) but nonetheless would not merit a reward from the service provider.

The requirement for *Incentive Compatibility (R4)* is achieved for PAI by estimating the truthfulness of the data submitted to the service provider. A data submission can contain one or more categories of measurement values. As the majority of participatory sensing data are scalar numeric readings, it is

```
[Service Provider publishes an offer token $T_{\mathrm{O}}$]
// OrderBook operation.
Append $T_{\mathrm{O}}$ to offer listing, $L_{\mathrm{O}}$.

[On acceptance of $T_{\mathrm{O}}$ by a participant]
Capture $d$
// Generate One-Time Key's public and private parts
Generate $a_{\mathrm{K_O}}$and $a^*_{\mathrm{K_O}}$
Encrypt $d$ using service provider's public key $b_{\mathrm{SP}}$
// Create offer acceptance, $A_{\mathrm{O}}$.
$A_{\mathrm{O}} = \{\{d\}_{\mathrm{b_{SP}}}, a_{\mathrm{K_O}}, i_{\mathrm{O}}, i_{\mathrm{A_O}}\}$
// Participant retains One-Time Key as the private key, $a^*_{\mathrm{K_O}}$, is used to
// claim reward.
// $[a_{\mathrm{K_O}}, a^*_{\mathrm{K_O}}]$ denotes the set of One-Time Keys held.
$[a_{\mathrm{K_O}}, \mathrm{a}^*_{\mathrm{K_O}}] += \{\mathrm{a}_{K_{\mathrm{O}}}, a^*_{\mathrm{K_O}}\}$
Publish $A_{\mathrm{O}}$ on OrderBook

// OrderBook Operation.
// Append $A_{\mathrm{O}}$ to $L_{\mathrm{O}}$
$L_{\mathrm{O}} += A_{\mathrm{O}}$
Forward $A_{\mathrm{O}}$ to Service Provider

// Service Provider Operation.
[On receipt of $A_{\mathrm{O}}$]
// Decrypt the data submission.
Decrypt $\{d\}_{\mathrm{b_{SP}}}$using $b^*_{\mathrm{SP}}$
$v$= Validate $A_{\mathrm{O}}$
if $v$ then
        // Allocate the reward
        Log allocation of $r_{\mathrm{O}}$
        $T_{\mathrm{V}} = \{i_{\mathrm{V}}, i_{\mathrm{A}_O}, v\}_{\mathrm{b^*_{SP}}}$
        // Publish $T_V$ on OrderBook by appending it to $L_{\mathrm{O}}$.
        $L_{\mathrm{O}} + = T_{\mathrm{V}}$
        // OrderBook generates encrypted spendable reward.
        $r_{\mathrm{S}} = \{i_{\mathrm{S}}, i_{\mathrm{V}}, r_{\mathrm{O}}\}_{\mathrm{a_{K_O}}}$
end if
```

Algorithm 1: Allocating the Reward

```
[Participant wants to spend the reward]
Decrypt $r_\mathrm{s}$ using $a^*_{\mathrm{K_O}}$
Forward $r_\mathrm{s}$ to OrderBook

[OrderBook operation]
// Verify that the associated validation token was signed by the service
// provider.
Verify signature of $T_\mathrm{V}$ (identified by $i_\mathrm{V}$ entry in $r_\mathrm{s}$)
if verification passes then
        Check that $i_\mathrm{s}$ is not already recorded as spent
        if $i_\mathrm{s}$ is not recorded as spent then
            Permit spending
            Record $i_\mathrm{s}$ as spent
        end if
end if
```

Algorithm 2: Spending the Reward

this category of data that is addressed by the proposed incentive compatibility method. There are a number of statistical methods that can be used to estimate data truthfulness for scalar data. Section 4.2.1 discusses the choice of approach for estimating data truthfulness while Section 4.2.2 describes the design and implementation of this approach.

4.2.1 Choosing an Approach for Estimating Data Truthfulness

As noted in Section 3.3, the use of statistical techniques could ensure data truthfulness without impinging upon the privacy requirements of the participant. In statistical terms, data truthfulness can be considered as a case of incomplete data i.e. the truthfulness of the data submission cannot be known with 100% certainty. There are a number of statistical methods that can be used to make estimations for incomplete probabilistic models. For instance, the Gradient Descent algorithm, an optimisation algorithm which has been used to train neural networks (Bottou, 2012) could be used for this purpose. However, this method is reported to be slow (Johnson and Zhang, 2013) and would hinder both the scalability and performance of PAI. While the Newton-Raphson method, a well-established technique for solving non-linear algebraic equations (Ypma, 1995), could be adapted for the purposes of estimating data truthfulness, its effectiveness is dependent on the accuracy of an initial 'guess'. The

Method of Moments, another possible approach for estimating population parameters such as the mean or standard deviation, has been found to be less precise than the Maximum Likelihood Estimation (MLE) method (Eisenhauer et al., 2015). For this reason, MLE is used to quantify the correctness of data submission measurements.

Economic theory acknowledges that full incentive compatibility cannot be guaranteed for all categories of exchanges between parties (Roberts and Postlewaite, 1976). In this case of **limiting incentive compatibility**, there is the potential for a party to gain from misrepresentation. This is the case here as the MLE method computes a range that does not provide a definitive evaluation of data truthfulness but rather is used to estimate whether the data submission is truthful or not. Hence, there is the potential for participants to make false data submissions within the defined range i.e. false positives could be rewarded. In addition, there is also a potential for false negatives. For example, a valid submission from a participant may be deemed to be untruthful as it falls outside the current range.

While the absence of full incentive compatibility means that the approach can never fully guarantee that a data submission is truthful, this is unavoidable as economic theory states that it is impossible to implement full incentive compatibility in finite economies (Groves and Ledyard, 1987) i.e. a market where there is a limited number of participants. This is the case for participatory sensing where there is a finite number of participants willing to make data submissions. However, the presence of an incentive compatibility approach does reduce the probability of untruthful data submissions being rewarded. This probability further diminishes as the number of participants increases and the service provider's dataset grows and improves in quality. This is because the latter will be able to calculate narrower acceptable ranges for data truthfulness, making it more difficult for dishonest participants to submit spurious data. The narrower range will also ensure PAI robustness with respect to incentive compatibility even in those participatory sensing campaigns where there is a high number of malicious participants submitting untruthful data.

It must be acknowledged that, given that the core aim of the work undertaken in this thesis is to preserve identity privacy for participants when receiving and spending rewards, a dishonest participant who is rewarded for a data submission that is subsequently found to be untruthful will not be prevented from future participation in the system. While this is indeed an unfortunate side effect of providing a high level of identity privacy, it should also be noted that precluding a malicious participant from, for example, a participatory sensing system that use pseudonyms, does not guarantee that the participant will not subsequently rejoin using a new pseudonym. In addition, there are a number of strategies a service provider could further adopt to prevent a high level of false data submissions from being rewarded. For example, like the approach undertaken by Luo et al. [2019], the service provider could cross validate data by publishing another offer seeking confirmation of the submitted data. Depending on the service provider's needs, other approaches such as managing the areas of interest as neighbouring grids [Kong et al., 2019] could also be used to further address the potential for dishonest participants.

4.2.2 Adapting an Approach for Estimating Data Truthfulness

This section describes the operation of the *Data Truthfulness Estimation (DTE)* algorithm that is used by PAI to estimate data truthfulness. Data truthfulness is assessed for each category of measurement value, c, received in the participant's data submission, d. The algorithm exits if it considers any measurement value for a category, m_{c}, to be untruthful. If this occurs, the validation status, v, is assigned a value of *false* and the validation token, T_{V}, published on the OrderBook denotes that the data submission is an invalid one. This of course then means that the participant will not receive a reward. The high level operation of the DTE approach is presented in Figure 4.4.

As shown in Figure 4.4., once a measurement value for c, m_c, is read, it is then checked to see whether it falls within the minimum and maximum threshold limits, $m_{\mathrm{c_{min}}}$ and $m_{\mathrm{c_{max}}}$, set by the service provider for c. If it is not, v is assigned a value of *false*. The next step in this process is to read the relevant data, $[d_{\mathrm{c}}]$, that is held by the service provider for c. Depending on the measurement category, this could be the entire dataset, a subset of the dataset from a recent time period as determined by the service provider (for example, readings in the last hour), the last n number of readings where n is a number determined by the service provider or the last n number of readings at, for instance, a particular time and/or location, The height, h, of the probability density function (PDF) for $[d_{\mathrm{c}}]$ is then computed to assess how close the data values are to each other and is then used to formulate the natural logarithm to be used for the Maximum Likelihood Estimation (MLE) method, known as the Log Likelihood Function (LLF). Once initial estimates are set for the mean and standard deviation of the data value set (μ_{e} and σ_{e} respectively), the MLE method is applied using these initial estimates and the LLF. The application of MLE results in the computation of an estimate for the mean, μ, and the standard deviation, σ, for which the normal distribution best describes this set of data values.

To prevent outliers and other potentially interesting (and valid) data being miscategorised as non-truthful data submissions and thus being discarded, the service provider can configure a scaling factor, f_σ, that is applied to σ to create σ_{scaled}. By adding and subtracting σ_{scaled} to and from μ respectively, the scaled limits, $l_{\min}$ and $l_{\max}$, are computed for m_{c}, as highlighted in Figure 4.4. m_{c} is then evaluated and if it is not between $l_{\min}$ and $l_{\max}$, v is assigned a value of *false*. This process is repeated for all measurement categories $[c]$. If each m_{c} is considered to be truthful, d is considered to be truthful and v is assigned a value of *true*.

Algorithm 3 presents the *Data Truthfulness Estimation (DTE)* algorithm. This algorithm is applied for every measurement value contained in a data submission made to the service provider. The dynamic participatory sensing environment means that the data will be ever changing and evolving. As the dataset is changing with each data submission, the MLE parameters are thus either recomputed every time a change in the dataset occurs or periodically with the service provider setting the recomputation interval in the latter case. This ensures that the incentive compatibility approach not only provides a means of estimating data truthfulness but does so in a way that reflects changes that have been captured in the service provider's dataset. The configurable scaling factor also ensures that potentially interesting patterns and outliers that are reflected in incoming data streams are not inadvertently disregarded by the service provider. This further ensures that the requirement for the dataset to be reflective of changes in the dynamic participatory sensing environment is met. It should also be noted that the estimation of data truthfulness does not require any disclosure of identity privacy by the participant as the data submitted to the service provider contains no reference to the participant who made the submission. Therefore, while the participant who makes a non-truthful data submission does not receive a reward, the requirement for identity privacy is not violated.

It should be noted that the underlying MLE method has a number of limitations. For instance, it is only suitable for scalar data and cannot be used to evaluate the truthfulness of multimedia data content. This is unsurprising given that many statistical methods are only appropriate for scalar data. In addition, as noted in Section 2.1.2, sensed multimedia data is not considered in this thesis. The effectiveness of MLE has also been found to be limited in a number of situations, for example, when the percentage of censored data (i.e. when the value of a measurement is only partially known) is large and the sample size is small (Jain and Wang, 2008). Nevertheless, it can be concluded that DTE does fulfill the requirement for *Incentive Compatibility (R4)* as it seeks to demonstrate how incentive compatibility can be facilitated without impinging upon identity privacy. The goal of requirement *R4* is to demonstrate that PAI can facilitate incentive compatibility in a privacy preserving manner and the use of the MLE method has achieved this. It should be noted that it is possible to adopt a variant of MLE without violating the requirement. For example, the modified MLE methods proposed in Li et al., 2019b and Wang and Chan, 2018 have the potential to serve as the basis for an MLE method that could cater for censored data.

4.3 Adaptive & Tunable Reward Allocation (R5)

This section describes how PAI meets the requirement for *Adaptive and Tunable Reward Allocation (R5)* through the development of the Adaptive Reward Allocation (ARA) model. The options for modeling ARA as a stochastic process and the rationale for adopting Lyapunov Optimisation are discussed in Section 4.3.1. In order to formulate the optimal trade-off between budget consumption and reward level, Lyapunov Optimisation requires a dataset consisting of the number of responses given in return for the different reward levels. Section 4.3.1 therefore discusses the design of the prediction model used to estimate the number of responses for the different reward levels and then models the environment for the purposes of reward determination and budget optimisation. To establish benchmarks for evaluating ARA, the formulation of the offline budget optimisation problem is presented in Section 4.3.2 with the time average budget consumption of online reward allocation policy being defined in Section 4.3.3. The design of the reward algorithm to minimise this time average budget consumption is then presented in Section 4.3.4. Lastly, the incorporation of data utility in the model is discussed in Section 4.3.5.

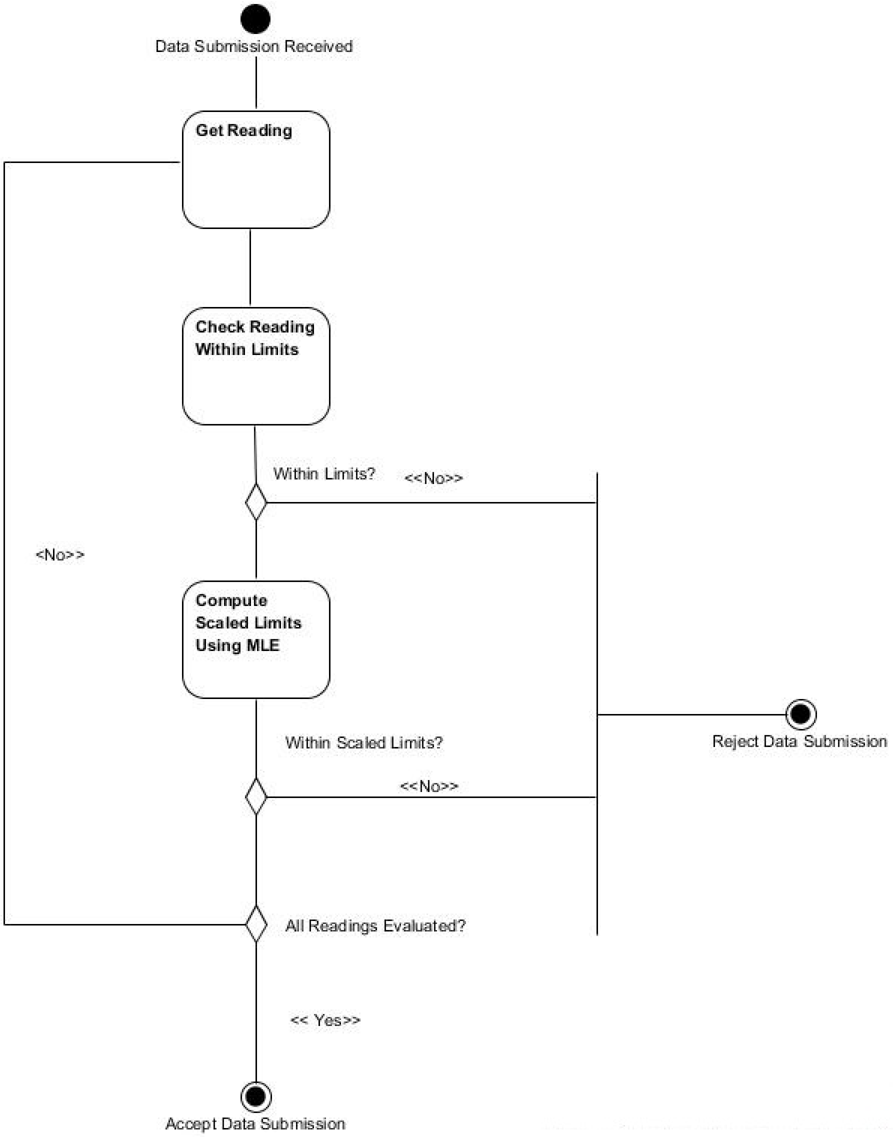

Figure 4.4: The Data Truthfulness Estimation Approach

```
[Service Provider receives a data submission $d$]
// Evaluate data truthfulness for each category of measurement
foreach $c$ in $d$
        // Check if the measurement value for this category is within the
        // limits set by the service provider. If not, set the validation
        // status to false and exit the algorithm.
        if $m_{\mathrm{c}} < m_{\mathrm{c}_{min}}$ or $m_{\mathrm{c}} > m_{\mathrm{c}_{\max}}$ then
            $v = false$
            return $v$
        end if
        // Read the previously held data for this category of
        // measurement.
        Read data $[d_{\mathrm{c}}]$ for $c$
        // Compute the height of the dataset's PDF.
        Get h from $[d_{\mathrm{c}}]$'s PDF
        // Compute the Log Likelihood Function.
        $LLF = h - sum(log(h))$
        // Use initial estimates for the mean and standard deviation.
        Use initial estimates $\mu_{\mathrm{e}}$ and $\sigma_{\mathrm{e}}$
        // Use MLE to estimate the two parameters (mean and standard
        // deviation) for which the normal distribution best describes the
        // data.
        Compute $\mu$ and $\sigma$ by applying MLE using LLF, $\mu_{\mathrm{e}}$ and $\sigma_e$
        // Read the standard deviation factor for this measurement
        // category, compute the scaled standard deviation setting
        // and set the threshold limits.
        Read $f_\sigma$
        $\sigma_{\mathrm{scaled}} = \sigma * f_\sigma$
        $l_{\mathrm{min}} = \mu - \sigma_{\mathrm{scaled}}$
        $l_{\mathrm{max}} = \mu + \sigma_{\mathrm{scaled}}$
        // Check if the measurement value is within the threshold limits.
        // If not, set the validation status to false and exit the algorithm.
        if $m_{\mathrm{c}} < l_{\mathrm{min}}$ or $m_{\mathrm{c}} > l_{\mathrm{max}}$ then
            $v = false$
            return $v$
        end if
end foreach
// If the loop has finished without exiting the algorithm, the data
// submission is a valid one.
$v = true$
return $v$
```

Algorithm 3: Estimating Data Truthfulness

4.3.1 Modeling ARA as a Stochastic Process

As discussed in Chapter 3, the modeling of ARA as a stochastic process is of particular interest in addressing requirement *R5* as such a method would not impinge upon participant privacy and would take account of the fast changing dynamic environments in which participatory sensing systems operate. In addition, stochastic optimisation processes can be used to optimise the trade-off of conflicting objectives. While there are many multi-objective optimisation methods that can be used to optimise the trade-off between conflicting objectives, these have limitations that make them inappropriate for addressing requirement *R5*. For example, Tsai and Chen [2014] note that mathematical programming methods can be limited in terms of their scalability while other methods such as the Order-Weighted Average (OWA) optimisation technique have been found to lead to non-optimal solutions. Similarly, evolutionary multi-objective optimisation methods have also been found to have issues that would lead to requirement *R5* not being fulfilled (Emmerich and Deutz, 2018). For example, it can be difficult to achieve regular spacing of solutions (i.e. a consistent set of solutions for the optimisation of a particular problem) using Pareto-based optimisation methods.

There are a number of stochastic modeling techniques that have been used for incentive design in the state of the art in other fields. For instance, the approach proposed by Huang et al. [2016] uses a Markov Decision Process model to design an incentive scheme for shopping coupons that reflects users' privacy sensitivities while Stochastic Resource Auctions are used for pricing wind power (Tang and Jain, 2015). However, the Markov Decision Process is used for modeling problems where the outcome is partly under the control of the decision maker. This is not the case in participatory sensing as the service provider has no control over whether a participant accepts an offer or not. While the option of using a Stochastic Resource Auction has more promise in that it facilitates dynamic and scalable incentive computation, it necessitates the use of an auction which, as was noted in Chapter 3, would lead to privacy vulnerabilities that would violate PAI's privacy requirements. Stochastic Submodular Maximisation, which is used by Singla and Krause [2013b] for privacy trade-offs in participatory sensing, has an inherent assumption of diminishing returns (i.e. incremental returns are lower over time). This is not appropriate for meeting requirement *R5* as the service provider expects equal or better data submissions over time at a lower cost per submission.

In contrast, Lyapunov Optimisation, a stochastic technique that seeks to push the backlog of queues in dynamic systems toward a lower congestion state with a view to achieving network (i.e. system) stability (Neely, 2010), is a method that is particularly suitable for the controlling of dynamic systems. It is used for the computation of incentives and pricing in communication networks and has been previously used for incentive design for participatory sensing, though not for reward computation (for example, Han et al., 2014). It can be used to minimise dynamic costs (Liu et al., 2015) and is suitable for rapid changes over time in the environment in which it is applied (Urgaonkar et al., 2010). These attributes are directly relevant given the desire by service providers that budget consumption be optimised. For this reason, the ARA model of reward allocation is based upon this method.

Lyapunov Optimisation is particularly appropriate for ARA as the approach seeks to dynamically adapt rewards so as to respond to sudden and rapid changes in an environment with the nature, accuracy, quality and level of detail of the data varying depending on the circumstances. Furthermore, the fact that a Lyapunov Optimisation solution at any one time affects the constraint to be applied the next time the optimisation is carried out is important for ARA as the service provider's budget is being consumed with each optimisation solution that results in accepted offers. Finally, the use of Lyapunov Optimisation does not require future knowledge of the rate of response to offers made by the service provider. This is crucial for ARA's reward model.

As Lyapunov Optimisation is principally used for resource allocation problems in domains such as computer networking (Lee and Heo, 2016), its use must be modified for the problem PAI is seeking to address. This is principally because there are a number of differentiating attributes of an economic market in participatory sensing. In particular, the data being sought by the service provider (equivalent to the product in other economic markets) can potentially change suddenly and its value to the service provider will change depending on that party's needs at a particular point in time. While demand may change in other price optimisation scenarios such as wind power or cloud infrastructure rental, the product does not. In participatory sensing, the 'product' (type of data sought) not only changes over time but is time sensitive and needs to match the information sought by the service provider (Tham and Luo, 2015). It is thus an appropriate candidate for a market-based model.

There are three facets to modeling ARA as a stochastic process. The estimation of the number of responses for the different reward levels and the formulation of the statistical prediction model to use is described in Section 4.3.1.1. Having defined the prediction model, the problem to be addressed, and, in particular, the trade-off to be optimised, is defined in Section 4.3.1.2. This trade-off is then used to model the environment for budget optimisation in Section 4.3.1.3.

Figure 4.5 presents a high level overview of the operation of the ARA model. The reward computation for a particular category of data submission takes place at the beginning of each *timeslot*, t, and is applied to all offers during that timeslot.

4.3.1.1 Estimating the Number of Responses

The reward included in the offer published by the service provider is a key factor in determining the number of responses in a particular timeslot, $N_{\mathrm{O}}(t)$, for each offer O. It is therefore assumed that $N_{\mathrm{O}}(t)$ is a function f of the offered reward in a particular timeslot, denoted $r_{\mathrm{O}}(t)$:

$$N_{\mathrm{O}}(t) = f(r_{\mathrm{O}}(t)) \tag{4.5}$$

To estimate $N_{\mathrm{O}}(t)$, ARA requires a dataset that it can use to compute the appropriate value for $r_{\mathrm{O}}(t)$. In microeconomic terms, this is the reservation price at which the participant is willing to 'sell' data. While the reservation price is typically computed by methods such as the Conjoint Analysis (Kalish and Nelson, 1991) and Contingent Valuation methods (Lee and Heo, 2016), these methods are dependent upon surveying potential customers (or participants in this case) which is not a practical option for meeting the requirement for *Adaptive and Tunable Reward Allocation (R5)* in a participatory sensing environment. Instead, ARA builds up a picture of participants' willingness to accept offers at particular rates from supply curves that use previous data submissions from the service provider's existing dataset. Previous data submissions thus act as a substitute for a survey to present an ongoing evolving picture of the willingness to accept offers at particular levels of reward.

As the level of reward set by the service provider is a key determinant of the number of data submissions it obtains in response to an offer, the above function can be modeled using the microeconomic concept of a supply curve. The formal definition of a supply curve is a graphic representation of the relationship between product price and the quantity of the product that a seller is willing and able to supply. In terms of the model used for ARA, a number of

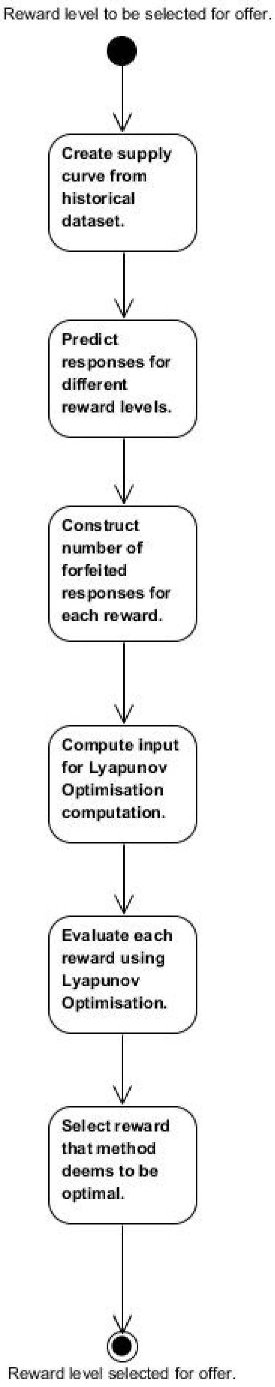
Reward level to be selected for offer.
Create supply curve from historical dataset.
Predict responses for different reward levels.
Construct number of forfeited responses for each reward.
Compute input for Lyapunov Optimisation computation.
Evaluate each reward using Lyapunov Optimisation.
Select reward that method deems to be optimal.
Reward level selected for offer.

Figure 4.5: ARA Operation

supply curves are used to estimate the relationship between the reward offered and the number of responses different categories of offers attract from participants. These supply curves evolve over time as more offers are made by the service provider and more responses to offers are received. The relationship between the number of responses and the reward level thus serves as ever evolving training data (a set of data used to discover relationships) to enable the service provider to more accurately estimate the reward that will generate its desired number of responses.

Each supply curve is modeled using regression analysis to predict the willingness of participants to accept offers at different reward levels. Typically, both demand and supply are modeled as a function of price and cost respectively using linear regression in the field of Econometrics (Hill, 2011, see also, for example Labandeira et al., 2017). However, to facilitate the incorporation of other predictors that will not necessarily have a linear relationship (for example, the effort involved in capturing the data), a non-linear multiple regression method is used to predict the number of responses, N_{predict}. Specifically, a rolling window time series regression model is used to construct the prediction model so that only the most recent data is taken into account. The size of the rolling window used can be altered depending on the circumstances in the participatory sensing environment without impacting the algorithm. Indeed, any form of predictive modeling technique can be used to update the supply curves, thus allowing the service provider to evaluate which is the best predictive model to use (Martini and Spezzaferri, 1984).

As noted in Section 2.1, the participant will incur costs when submitting data resulting in different willingness to make data submissions. These costs can be considered as random effects that are summarised as a cost parameter C, which is random, i.i.d (independent and identically distributed) and varies between time slots. When the cost is high (for example, the smartphone is required for the user's own needs; the battery is low), the user needs a higher reward to participate. When C is low (for example, the device is idle; the user has time to complete the task) then even a low reward might be enough. While the service provider does not have access to each individual participant's circumstances during a particular time slot, it can nevertheless estimate C in terms of, for example, battery consumption, data transmission costs and latency i.e. the time taken to accept a task, carry out a task, make a data submission and receive the reward for the completion of the task.

The number of current active participants P in each time slot t is another parameter of interest when predicting the number of responses. For example, when there are many active participants, a small reward that can motivate only 10% of these users might be enough in order to ensure the required number of responses. On the other hand, a higher per user reward is necessary for a participatory system with less active participants.

Therefore, N_{predict} can be defined in terms of the rewards offered r, the cost of carrying out the task, $-C$, and the ratio of the number of responses sought to the current number of participants, P_{ratio}. Using X to denote this set of predictors as a vector and β to denote a vector of parameter coefficients, N_{predict} can be expressed as follows:

$$N_{\text{predict}} = f(X, \beta) + \varepsilon \tag{4.6}$$

where ε is an error term.

Equation 4.6 can be expanded to incorporate r, $-C$ and P_{ratio}. In addition, while the problem is non-linear, it can be expressed in epigraph form as follows:

$$N_{\text{predict}} = \beta_0 r - \beta_1 C + \beta_2 P_{\text{ratio}} \tag{4.7}$$

where β_0 is the regression coefficient for r.
β_1 is the regression coefficient for –C.
β_2 is the regression coefficient for P_{ratio}.

Equation 4.7 can be extended by the service provider to incorporate other coefficients if there are other factors that determine the number of responses, for example, the level of privacy to be ceded. In addition, the service provider can remove what it deems to be irrelevant predictors without impacting the underlying reward model. For example, a service provider who is only seeking scalar data such as temperature might consider the task cost to be broadly similar between time slots.

It should be noted that if the number of responses is greater than that desired by the service provider in a particular timeslot, $N_{\text{desired}}(t)$, it will only be desirable from the service provider's perspective to reward some of the responses to an offer. Moreover, while it may be possible to attract $N_{\text{desired}}(t)$, this might necessitate a reward level that is not consistent with optimal consumption of the service provider's budget. Hence, while the supply curves can be used to determine reward levels, the trade-off between achieving $N_{\text{desired}}(t)$, and budget consumption must be addressed. It is thus necessary to model this trade-off for the participatory sensing environment.

4.3.1.2 Modeling the Environment for Reward Determination

The relationship assumed by Equation 4.5 is used to build up a picture of the (estimated) number of participant responses to a particular reward. However, there will be a point at which increasing the reward will not lead to an increase in the number of responses even if parameters such as C remain unchanged. This is because the maximum number of responses is equal to the number of participants in the participatory sensing system, $P(t)$, and varies over time as participants join and leave the system (either by formally deregistering or ceasing to participate). Thus for every time slot t:

$$0 \leq N_{\mathrm{O}}(t) \leq P(t) \tag{4.8}$$

$r_{\mathrm{P}}(t)$ denotes the reward level when the number of responses equals the number of participants i.e. when demand equals supply in economic terms:

$$N_{\mathrm{O}}(t) = P(t) \tag{4.9}$$

$P(t)$ is upper bounded by a constant, $P_{\max}$, which corresponds to the maximum number of participants potentially active on the system. This leads to the following constraint for every time slot t:

$$0 \leq P(t) \leq P_{\max} \tag{4.10}$$

Using the supply curves, PAI can estimate the number of responses that should be received at different levels of rewards for different categories of data. For example, the service provider estimates that it will receive N_{O} number of responses when the reward level is set to r_{O}. Taking Equation 4.9 and Equation 4.10 into account, r_{O} should not exceed $r_{\mathrm{P}}(t)$ as exceeding $r_{\mathrm{p}}(t)$ will not increase the number of responses:

$$0 \leq r_{\mathrm{O}} \leq r_{\mathrm{P}}(t) \tag{4.11}$$

As the supply curves evolve over time, the process of updating each curve is undertaken at the beginning of each time slot when reviewing the reward level. The service provider uses the reward-response data it has observed over previous time periods and, accordingly, updates the supply curve for this time slot.

The problem PAI is seeking to address can thus be defined as follows:

Problem Definition

For a given number of responses in a time slot, t, that follows an i.i.d. process with mean cost C, and for a certain level of minimum participants that the system should recruit, design a dynamic algorithm that finds the optimal level of reward so as to satisfy the above constraints while minimising the budget consumption of the service provider.

To achieve a trade-off between minimising the number of offers forfeited due to too low a reward and optimising budget consumption, the former is defined as a queue for a time slot t, $Z_{\text{forfeit}}(t)$ [6]. The number of forfeited responses is what is termed a 'virtual queue'. As the name implies, virtual queues do not exist in reality and are only implemented in software to facilitate the definition of the Lyapunov Optimisation-based model (Neely, 2010).

$Z_{\text{forfeit}}(t)$ is computed in terms of the number of responses desired by the service provider, $N_{\text{desired}}(t)$. Thus, in any time slot, t, $\mathrm{Z}_{\text{forfeit}}(t)$ is the difference between the actual number of responses received, $N_{\text{received}}(t)$, and $N_{\text{desired}}(t)$[7]:

$$Z_{\text{forfeit}}(t) = \begin{cases} 0 & \text{if } N_{\text{received}}(t) \geq N_{\text{desired}}(t) \\ \min(N_{\text{desired}}(t), P(t)) - N_{\text{received}}(t) & N_{\text{received}}(t) < N_{\text{desired}}(t) \end{cases} \tag{4.12}$$

4.3.1.3 Modeling the Environment for Budget Optimisation

As originally formulated, Lyapunov Optimisation is used to minimise the backlog of a queue for the purposes of optimising resource allocation [Neely, 2010]. In mathematical terms, the method is the sum of squares of the queue (multiplied by $\frac{\mathbf{1}}{\mathbf{2}}$) arising from a resourcing problem:

$$L(t) = (\frac{1}{2})\sum_{i} Z_{\text{i}}(t)^2 \tag{4.13}$$

Equation 4.13 measures the queue backlog for the system model, the queue being the number of forfeited responses as defined by Equation 4.12.

The computation of $N_{\text{desired}}(t)$ is dependent upon the requirements of the service provider. While the nature of participatory sensing campaigns means that different service providers could have diverse needs, two options are consid-

[6] Z is used to denote a virtual queue as this notation corresponds to that used in the work undertaken by Neely [2010].

[7] Or $P(t)$ where this is less than $N_{\text{desired}}(t)$.

ered in this thesis. Firstly, in a fast changing environment, the service provider could decide that its desired number of responses is determined by its needs at a particular time, i.e. for every time slot t, the desired number of responses is independent of previous timeslots:

$$N_{\text{desired}}(t) \perp N_{\text{desired}}(t-1) \tag{4.14}$$

In such a case, it is assumed that $N_{\text{desired}}(t)$ is i.i.d. over the time slots. Furthermore, unlike other scenarios typically modeled using Lyapunov Optimisation (for example, Liu et al., 2015), $Z_{\text{forfeit}}(t)$ is, for every time slot t, independent of queue backlogs from previous timeslots:

$$\mathrm{Z}_{\text{forfeit}}(t) \perp Z_{\text{forfeit}}(t-1) \tag{4.15}$$

Alternatively, the second option is that the service provider may decide that, if, for a previous timeslot $t-1$, $N_{\text{desired}}(t-1) < N_{\text{received}}(t-1)$, $N_{\text{desired}}(t)$ is determined by $N_{\text{desired}}(t-1)$ i.e. for every time slot t:

$$\begin{aligned} N_{\text{desired}}(t) = N_{\text{desired}}(t-1) - N_{\text{received}}(t-1) \\ s.t.\ N_{\text{desired}}(t-1) > N_{\text{received}}(t-1) \end{aligned} \tag{4.16}$$

This then implies that the value of $Z_{\text{forfeit}}(t)$ is determined by $Z_{\text{forfeit}}(t-1)$ i.e.

$$Z_{\text{forfeit}}(t) = f(Z_{\text{forfeit}}(t-1)) \tag{4.17}$$

While the underlying probability distribution and other statistical characteristics of $N_{\text{desired}}(t)$ are not known by the service provider and are not required for Lyapunov Optimisation, it must be assumed that its maximum value is finite:

$$0 \leq N_{\text{desired}}(t) \leq N_{\text{desired}_{\max}} \tag{4.18}$$

Moreover, a further assumption is that the number of received responses to offers is bounded by the number of potentially active participants in the system. Thus, the expected values (the long run average values) of $N_{\text{desired}}(t+1)$ and $P(t)$ adhere to the following rule:

$$\mathbf{E}(N_{desired}(t+1)){\leq}\mathbf{E}(P(t)) \tag{4.19}$$

This inequality ensures that there is a reward allocation schedule that ensures the stability of $Z_{\text{forfeit}}(t)$. Using the rate stability theorem (Neely, 2010), $\bar{Z}_{\text{forfeit}_\text{k}}$ is used to denote the time average queue backlog for the forfeited responses. The stability of the queue is equal by definition as follows:

$$\bar{Z}_{\text{forfeit}_\text{k}} \triangleq \underset{t\to\infty}{lim}\frac{1}{t}\sum_{t=0}^{t-1}E\{Z_{\text{forfeit}_\text{k}}(t)\} < \infty \tag{4.20}$$

It is assumed that the reward is upper bounded by a constant $r_{\max}$. This means that for all time slots t:

$$0{\leq}r(t){\leq}r_{\max} \tag{4.21}$$

In addition, the service provider can also set a maximum value for the proportion of the budget, $B_{\text{proportion}_{\max}}$, that can be consumed for an offer in a given time slot[8]:

$$0 \leq N(t)r(t) \leq B_{\text{proportion}_{\max}} \tag{4.22}$$

4.3.2 Formulating the Offline Problem

Before modeling the budget consumption problem for ARA, it is necessary to establish benchmarks to evaluate the approach. This section formulates the problem of reward allocation as two offline problems with complete future information (Section 4.3.2.1) and stochastic information (Section 4.3.2.2) respectively as benchmarks. These benchmark cases assume information symmetry i.e. the service provider knows the response rate for a particular reward in the case of full information and knows the budget consumption under different scenarios in the case of stochastic future information. Analysis of the benchmarks is presented in Section 4.3.2.3.

4.3.2.1 Complete Future Information

With complete future information, the service provider can determine the response rate jointly in all time slots to minimise budget consumption. To for-

[8] To simplify the model, it is assumed that this proportion is not altered between timeslots. This could be done without impacting the core contribution of the ARA model.

mulate the offline budget consumption problem, T is defined as the set of all time slots $1..t_n$ during the sensing period where t_n represents the final time slot. As no linear relationship is assumed between the number of responses and the reward offered, the problem is a nonlinear convex optimisation problem and can be formulated as follows for an individual timeslot t for the reward range:

$$\begin{aligned}
\min_{t \in T} \{N_{\min}(t) r_{\min}(t) \ldots N_{\max}(t) r_{\max}(t)\} \\
s.t. \\
r(t) \geq r_{\min} \\
r(t) \leq r_{\max} \\
N(t) \geq N_{\text{desired}} \\
N(t) r(t) \leq B_{\text{remain}}
\end{aligned} \tag{4.23}$$

where $r_{\min}$ is the minimum reward.
$r_{\max}$ is the maximum reward.
$N_{\min}$ is the number of responses received for $r_{\min}$.
$N_{\max}$ is the number of responses received for $r_{\max}$.
B_{remain} is the remaining budget.

The problem of minimising the budget consumption over the entire set of time slots T, is subject to the same constraints and is formulated as follows:

$$\frac{1}{t} \sum_{t \in T} N(t) r(t) \tag{4.24}$$

The offline reward allocation problem solved in Equation 4.24 incorporates the explicit response rate of every time slot in advance. There are a wide range of optimisation methods that can be used to solve Equation 4.24, for example, the first fit and best fit algorithms, nonlinear programming methods, mixed integer linear programming methods (by formulating the problem in linear epigraph form) or, by using linear programming relaxation, the simplex method or KKT analysis (Gao et al., 2015a).

The formulation and solving of Equation 4.24 requires complete knowledge of the future response rate in every time slot t, which is obviously impractical. For this reason, a model which only requires certain future information is defined.

4.3.2.2 Stochastic Future Information

This section proposes a benchmark based on stochastic future information where the response rate for each time slot follows the same probability space. With stochastic information only, the service provider cannot decide the reward for a timeslot in advance as it does not have complete future information. This case focuses on the expected budget consumption optimisation based on stochastic information. Θ defines the set of possible scenarios (or information realisations) that can occur when a service provider makes an offer at a particular reward level, r. $r(\theta)$ and $N(\theta)$ respectively denote the reward level and the number of expected responses to that reward under a particular information realization θ. Budget consumption under θ is $N(\theta)r(\theta)$. Therefore, the expected budget optimisation problem can be defined as follows:

$$\min \int_{\theta \in \Theta} N(\theta)r(\theta) \tag{4.25}$$

Like Equation 4.24, Equation 4.25 is an offline problem subject to the same constraints that in this case defines a contingency plan which specifies the budget consumption under each information realisation, θ. It is a nonlinear programming problem with an infinite number of variables as θ is continuous (Gao et al., 2015a). It should be noted that formulating and solving Equation 4.24 requires certain (stochastic) future information, which may not be available in practice.

4.3.2.3 Analysing the Benchmarks

The next step is to analyse the gap between the minimum budget consumption with complete future information derived from Equation 4.24 and the minimum budget consumption with stochastic future information derived from Equation 4.25. These are denoted by B^o and B^* respectively. As indicated in the state of the art (Gao et al., 2015a), this can be expressed formally as follows:

Lemma 1

$$\text{if } T \to \infty, \text{ then } B^o \to B^* \tag{4.26}$$

Lemma 1 indicates that, as long as the total sensing period T is of sufficient length, the diminution in budget consumption optimality caused by the loss of complete future information is negligible. Hence, both B^o and B^* can serve as the same benchmark for an online policy that does not require future information. An online policy is necessary as the stochastic future information required by Equation 4.25 may not be available in practice. ARA's reward mechanism is thus modeled as an online problem of reward allocation i.e. with no future information. The offline problem serves as a benchmark only.

4.3.3 Online Budget Consumption Optimisation Problem

The Lyapunov Optimisation-based budget optimisation problem formulated in this section relies only on past response rates to particular rewards and does not require any future information. The goal of the service provider is to minimise the time average reward and hence optimise its budget consumption. The service provider's budget (B) consumed in time slot t is given by:

$$B(t) = N(t)r(t) \tag{4.27}$$

Lyapunov Optimisation requires a control decision. For ARA, the control decision refers to the setting of an optimal reward level, $r(t)$, for a particular time slot t. Thus, $r(t)$ is the control decision made in time slot t. The resultant reward allocation policy arising from $r(t)$ must meet the constraints presented in Equations 4.18, 4.19, 4.20 and 4.21.

The time average budget consumption of this policy can then be defined as:

$$B_{\mathrm{AV}} \triangleq \lim_{t \to \infty} \frac{1}{t} \sum_{t=0}^{t-1} E\{N(t)r(t)\} \tag{4.28}$$

The goal of ARA's reward model is to determine a reward level $r(t)$ that minimises the time average budget consumption subject to the constraints presented in Equations 4.18, 4.19, 4.21 and 4.22.

4.3.4 Designing the Reward Algorithm

The virtual queue, $Z_{\mathrm{forfeit}}(t)$, in the modeled system is the dimension that has to be considered to achieve an optimal reward for a time slot t. As a result, from Equation 4.13, the Lyapunov Function for t can then be defined as:

$$L(t) \triangleq \frac{1}{2} Z_{\mathrm{forfeit}}(t)^2 \tag{4.29}$$

Equation 4.29 is a quadratic Lyapunov function, a scalar measure of the total queue backlog in the participatory sensing system. The expected change in the Lyapunov function over one time slot t is referred to as the one-slot conditional

Lyapunov drift and is defined as:

$$\triangle(t) \triangleq L(t+1) - L(t) \tag{4.30}$$

To achieve adaptive reward allocation that minimises the reward offered for a data submission (and thus optimises budget consumption) and still obtain meaningful and timely responses for the service provider's dataset, Equation 4.30 must be greedily minimised for each timeslot t (i.e. the solution that is the best for the current timeslot is chosen) so as to minimise the queue backlog. In queuing theory terms, this means that the queue backlogs are pushed towards a lower congestion state on an ongoing basis with the goal of achieving queue stability. Therefore the budget consumption term $B(t)$ is incorporated into Equation 4.30 to produce a *drift-plus-penalty expression*:

$$\triangle(t) + V\{B(t)\} \tag{4.31}$$

Given that the overall objective is to minimise budget consumption, it should be minimised at the same time as the queue backlog is being minimised. This minimisation objective is known as a *penalty* under Lyapunov Optimisation. The fundamental objective of Lyapunov Optimisation is to minimise the bound (limit) on the drift-plus-penalty expression (Neely, 2010). V is a non-negative control parameter that is used to incorporate the weighted budget consumption term in the control decision. This facilitates the trade-off required by the service provider between reducing the backlog of $Z_{\text{forfeit}}(t)$ and minimising $B(t)$. Thus, in statistical terms, the goal is to find the upper bound for Equation 4.31, which will then be minimised to determine the optimal trade-off between the number of forfeited responses (i.e. the queue backlog) and the budget optimisation.

The drift-plus-penalty bound for a general case (Neely, 2010) can be extended for the environment in which PAI operates. For the purposes of this model, the number of responses received for an offer, $N_{\text{received}}(t)$, is assumed to be i.i.d. over time slots. Therefore, under any control algorithm that seeks to minimise the reward allocated, $r(t)$, the drift-plus-penalty expression used for Lyapunov Optimisation (Neely, 2010) can be formulated for ARA with the following upper bound:

$$\triangle(t)+V\{B(t)\}\leq B_{\text{constant}}(t)+V\mathbf{E}\{B(t)\,\big|\,Z(t)\}+Z(t)\mathbf{E}\{N_{\text{received}}(t-1)-N_{\text{desired}}(t)\,\big|\,Z(t)\} \tag{4.32}$$

It should be noted that $B_{\text{constant}}(t)$ is a positive number used in the Lyapunov Optimisation computation and is defined by:

$$B_{\text{constant}}(t) \triangleq \frac{1}{2}(N_{\text{received}}(t-1) - N_{\text{desired}}(t))^2 \tag{4.33}$$

Like other Lyapunov Optimisation-based models (Neely, 2010), the objective of the reward allocation algorithm presented for PAI is not to directly minimise Equation 4.31. The goal rather is to minimise the upper bound on the right hand side of Equation 4.32. Therefore, the reward allocation algorithm observes the queue backlog $Z(t)$ in every time slot t and adapts the Lyapunov Optimisation approach (Neely, 2010) to choose the budget consumption $B(t)$ as the solution to the following problem:

$$\min N(t)(Vr(t) + Z(t)) \tag{4.34}$$

As was noted in Equation 4.5, $N(t)$ is a function of $r(t)$. This constraint is ensured by the supply curves and thus the solution to Equation 4.34 must be one of the rewards depicted on the relevant curve for the current time slot. This means that the reward to be allocated, $r(t)$, can only be one of a number of possible values for each time slot t i.e. it is a discrete variable. The algorithm evaluates Equation 4.34 for all possible levels of budget consumption and selects the reward corresponding to the optimal level of consumption. After this reward is selected, the responses are processed and rewarded by the service provider. The appropriate supply curve is then updated to reflect $N(t)$, the number of responses obtained. The execution of the algorithm is repeated for every time slot in which an offer is made.

A typical Lyapunov Optimisation model only requires the current system state. This is modified for PAI as the algorithm determines the reward $r(t)$ to offer on the basis of the number of responses received in previous timeslots. In other words, the algorithm offers higher rewards when the backlog for Z_{forfeit} is large and lowers the level of reward to offer when the backlog for Z_{forfeit} is small.

The optimality of Equation 4.34 can be proven using standard Lyapunov Optimisation theory (Neely, 2010). $B^{\dagger}(t)$ denotes the budget consumption for the online model in a timeslot t. Using B^{*}, the budget consumption benchmark that assumes stochastic future information, the following theorem can be presented.

Theorem 1 (Adapted from Neely, 2010)

$$\underset{t\to\infty}{lim}\sum_{t\in\mathrm{T}}\mathbf{E}(B^{\dagger}(t)) \;\geq\; B^{*} - \frac{B_{\text{constant}}(t)}{\varPhi} \tag{4.35}$$

Equation 4.35 implies that the formulation for the online budget consumption optimisation converges to the minimum budget consumption asymptotically (as time tends towards infinity), with a controllable error bound $O(\frac{1}{\varPhi})$.

4.3.5 Incorporating Data Utility

The value of V is a key factor in devising an optimal budget consumption policy (Urgaonkar et al., 2010). Specifically, if B^{*}_{av} is the objective value of the time average maximisation problem under an optimal policy the following theorem holds (Neely and Urgaonkar, 2008):

Theorem 2 (Adapted from Neely and Urgaonkar, 2008)

Suppose the number of responses received in a previous timeslot, $N_{\text{received}}(t-1)$, and the number of desired responses, $N_{\text{desired}}(t)$, are i.i.d. for each time slot. If there exists an $\gamma > 0$ such that:

$$\mathbf{E}\{N_{\text{received}}(t-1)\}\leq\mathbf{E}\{N_{\text{desired}}(t)\} - \gamma \tag{4.36}$$

The following performance guarantees are then realised:

$$\begin{gathered}\underset{t\to\infty}{lim}\frac{1}{t}\sum_{t=0}^{t-1}\mathbf{E}\{p(t)N(t)\} < B^{*}_{\text{av}} + \frac{B_{\text{constant}}(t)}{V}\\ \bar{Z}\triangleq\underset{t\to\infty}{lim}\frac{1}{t}\sum_{t=0}^{t-1}\mathbf{E}\{Z(t)\}\leq B_{\text{constant}}(t) + \frac{B_{\text{constant}}(t) + Vr_{\max}P_{\max}}{\epsilon} + N_{\max}\end{gathered} \tag{4.37}$$

$p(t)$ is the penalty used for achieving queue stability in Lyapunov Optimisation (budget consumption in this case) while ϵ represents a constant > 0.

Theorem 2 indicates that, by choosing a large value for V, the budget consumption can be arbitrarily close to the optimal solution. However, the average queue backlogs increase as the value of V is increased. This means that there is a trade-off between budget consumption and the size of $Z(t)$ that can be tuned by the service provider depending on the significance of the data it is seeking in a particular timeslot, t.

As the importance of data being sought will vary for the service provider, it can set a utility weighting, U, for these data submissions. The utility weighting increases with the importance of the data to the service provider and can be used to capture dynamic changes in the participatory sensing environment. To reflect the importance of the data being sought, the value of U is mapped to that of V. Specifically, the value of V is increased in accordance with the data utility weighting so as to prioritise attracting data submissions over budget consumption i.e.

$$U \propto \frac{1}{V} \tag{4.38}$$

Utility weighting can thus be used to capture dynamic changes in the participatory sensing environment. It should be noted that the predictive model could also be used to tune the utility of the sought data submission without the need to modify the PAI algorithm. For example, the most recent data received could be weighted when constructing the predictive model if data is being sought on the basis of the most recent submissions.

Algorithm 4 presents the algorithm for reward computation. Table 4.1 presents the additional notations used in this algorithm.

4.4 Summary

This chapter describes the design of PAI and presents the algorithms for the approach. The fundamental architecture of the decentralised cryptocurrency exchange has been modified to meet PAI's privacy preservation requirements. Specifically, the concept of the OrderBook is modified in the *Identity Privacy Preserving Incentivisation (IPPI)* platform to enable participants to make anonymous data submissions that are only accessible by the service provider, thus meeting the requirement for *Anonymous, Unlinkable and Protected Data Submission (R1).* The concept of the One-Time Key, and the use of the underlying Diffie-Hellman Exchange Protocol, is also modified to enable participants to assign once-off, untraceable and linkable IDs to their data submissions in order to receive rewards. Through the use of the One-Time Key, the requirement for *Untraceable and Unlinkable Reward Allocation (R2)* has been met. The One-Time Key is also used to ensure that the participant can spend rewards in a privacy preserving manner, thus meeting the requirement for *Untraceable and Unlinkable Reward Spending (R3).*

```
// Construct a Linear Regression model from the supply curve.
Create Supply Curve for $\{[N_{\text{actual}}], r\}$ from Historical Dataset
$M_{\text{predict}}$= Linear Regression Model for Supply Curve $\{[N_{\text{actual}}], r\}$
// Predict the responses for different reward levels for this timeslot $t$.
$[r] = 0..r_{\max}$
$[N_{\text{predict}}(t), r] = \mathit{predict}(M_{\text{predict}}, [r])$
// Construct the queuing state variables for each reward.
foreach $r$
        $\mathrm{Z}_{\text{forfeit}}(t) = N_{\text{predict}}(t) - N_{\text{desired}}(t)$
        $\{[r, N_{\text{predict}}(t), Z_{\text{forfeit}}(t))]\} + = [r, N_{\text{predict}}(t), Z_{\text{forfeit}}(t)]$
end foreach
// Compute the constant used for Lyapunov Optimisation.
$B(t)_{\text{constant}} = \frac{1}{2} * (N_{\text{received}}(t-1) - N_{\text{desired}}(t))^2$
// Map the data utility weighting $U$ to the value of $V$.
$V = [U, V]$
// Evaluate each reward using Lyapunov Optimisation.
foreach $[r, N_{\text{predict}}, Z_{\text{forfeit}}(t)]$
        // Compute the Budget used by this reward.
        $B(t) = r * N_{\text{predict}}(t)$
        // Check that the budget consumption does not exceed the set
        // maximum.
        if $B(t) > B_{\text{proportion}_{\max}}$ then
            break
        end if
        // Carry out the Lyapunov Optimisation computation.
        $L = \frac{1}{2} * Z_{\text{forfeit}}(t)^2$
        // Compute the one slot conditional Lyapunov drift.
        $\Delta(t) = L - L_{\text{last}}(r)$
        // Evaluate the drift plus penalty expression.
        $DPP_{\text{LHS}} = \Delta(t) + (V * B)$
        $DPP_{\text{RHS}} = B(t)_{\text{constant}} + (V * B) + Z(t)(N_{\text{received}}(t-1)$
                    $-N_{\text{desired}}(t)))$
        if $DPP_{\text{LHS}} >= DPP_{\text{RHS}}$then
            continue
        end if
        // Evaluate the current optimisation computation.
        $OPT_{\text{current}} = N_{\text{predict}} * ((V * r) + Z(t))$
        if $OPT_{\text{current}} > OPT_{\text{solution}}$ then
            $OPT_{\text{solution}} = OPT_{\text{current}}$
            $r_{\text{optimal}} = r$
        end if
end foreach
```

Algorithm 4: Reward Computation

Notation	**Meaning**
DPP_{LHS}	Left hand side of Drift Plus Penalty expression.
DPP_{RHS}	Right hand side of Drift Plus Penalty expression.
$L_{\text{last}}(r)$	Last Lyapunov function calculated for a particular reward.
OPT_{current}	Current Lyapunov Optimisation calculation.
OPT_{solution}	Lyapunov Optimisation solution.
M_{predict}	Linear Regression prediction model.
r_{optimal}	The optimal value for the reward.
$[r]$	Set of possible reward values.
$r(t)$	The optimal reward to offer in a particular timeslot, t.
$[r, N_{\text{predict}}, Z_{\text{forfeit}}(t)]$	Reward, number of predicted responses and queuing state variable for this reward.
$\{[r, N_{\text{predict}}, Z_{\text{forfeit}}(t)]\}$	The set of rewards, their respective queuing state variables and number of predicted responses.
$\{[N_{\text{actual}}], r\}$	Actual responses for the different reward levels.
$[N_{\text{predict}}, r]$	The number of predicted responses for the different reward levels.
$[U, V]$	A map of data utility weightings and the constant V used for computing the Lyapunov Drift.
$\Delta(t)$	One-slot conditional Lyapunov Drift.

Table 4.1: Additional Notations for Reward Computation Algorithm

PAI also facilitates limiting *Incentive Compatibility (R4)* in a privacy preserving manner. Through the adoption of the Maximum Likelihood Estimation method, the *Data Truthfulness Estimation (DTE)* algorithm estimates whether a data submission is truthful or not, thus ensuring that only those rewards that were deemed to be valid receive a reward from the service provider. This facilitates the service provider in optimising the consumption of the budget set aside for the participatory sensing campaign as does the *Adaptive and Tunable Reward Allocation (R5)* provided by PAI. The meeting of this requirement ensures that the reward given to participants is reflective of the current response rate. The *Adaptive Reward Allocation (ARA)* model can also be tuned by the service provider to prioritise data capture over budget consumption and vice versa. The reward allocation mechanism uses linear regression and microeconomic supply curves to build up a picture of the response rate for different data categories. Using Lyapunov Optimisation, the method calculates the reward that is most likely to garner the required number of responses. This process is repeated for each time slot. The service provider then publishes its offer on the OrderBook at this reward level.

Chapter 5

Evaluation

This chapter evaluates the Privacy-Aware Incentivisation (PAI) approach proposed in this thesis. The implementation carried out to evaluate PAI is described in Section 5.1. PAI is evaluated by proof and by comparing the approach to the most relevant approaches in the state of the art with Section 5.2 discussing how the design proposed in Chapter 4 meets the requirements for *Anonymous, Unlinkable and Protected Data Submission (R1)*, *Untraceable and Unlinkable Reward Allocation (R2)*, *Untraceable and Unlinkable Reward Spending (R3)*, *Incentive Compatibility (R4)* and *Adaptive and Tunable Reward Allocation (R5)*. Proofs are presented to illustrate how PAI meets privacy preserving requirements *R1*, *R2* and *R3* as well as the approach's facilitation of the requirement for *Incentive Compatibility (R4)* in a privacy preserving manner. Requirement *R5* uses a simulated participatory sensing environment to carry out experiments evaluating the adaptiveness and tunability of PAI's adaptive reward allocation method in comparison to the most relevant approaches in the state of art for participatory sensing reward computation. Experiments are also carried out to evaluate the overall performance of the system using the most relevant approaches in the state of the art in privacy preserving reward allocation as baselines, described in Section 5.3. Section 5.4 discusses the results of the evaluation and summarises the chapter.

5.1 Implementation

PAI is implemented using the C++ and statistical R programming languages. The goal of this implementation is to validate the design of PAI and the presented theorems and conduct the experiments evaluating requirement *R5*. The implementation is organised around three core components representing the participant (the *Participant Component (CC1)*), the OrderBook (the *OrderBook Component (CC2)*) and the service provider (the *Service Provider Component (CC3)*). Each core component is implemented as a C++ class.

As discussed in Section 4.1.1, PAI is a decentralised exchange platform that uses peer devices to host the contents of the OrderBook (*CC2* in the implementation). The number of peers on which *CC2* resides is configurable for different runs of the implementation. The service provider is also configurable with the initial reward to offer, the maximum reward to offer, the number of responses sought, the initial response rate, the rolling regression window to use for forecasting the different response rates, the service provider's budget and the data utility all being adjustable. These are set in and accessed from the system's *Configuration*.

The core components use four supporting components for token creation (the *Token Library (SC1)*), cryptographic functionality (the *Cryptographic Library (SC2)*), statistical operations (the *Statistical Library (SC3)*) and the storage of the number of responses for different reward levels as well as, of course, the data submissions themselves (*Storage (SC4)*). The supporting components are organised into libraries containing C++ structures and classes. In addition, the statistical library, *SC3*, also contains a number of programs written in the statistical R programming language[1] while the storage, *SC4*, is implemented using MongoDB database software[2]. While *SC1* is directly accessible by the core components, *SC2, SC3* and *SC4* are accessed using a C++ interface[3]. This is done so as to separate the implementation from the third party components used.

Figure 5.1 presents the components used to implement PAI and identifies the dependencies and associations between these different components. The role of these components in implementing the privacy preserving requirements

[1] Statistical R is chosen to implement *SC3* as the toolkits it provides address the majority of techniques available in the statistics domain, including all the methods used in this thesis. See `https://www.r-project.org`

[2] See `https://www.mongodb.com`

[3] An interface written in Statistical R is also used for requirements *R4* and *R5*.

for *Anonymous, Unlinkable and Protected Data Submission (R1)*, *Untraceable and Unlinkable Reward Allocation (R2)* and *Untraceable and Unlinkable Reward Spending (R3)* is described in Section 5.1.1 while the implementation of the requirement for *Incentive Compatibility (R4)* and *Adaptive and Tunable Reward Allocation (R5)* is explored in Section 5.1.2 and Section 5.1.3 respectively. The simulated participatory sensing environment that is used to evaluate PAI is described in Section 5.1.4.

5.1.1 Implementing Privacy Preservation (R1, R2 & R3)

The *Token Library*, (*SC1*) contains the tokens and structures used by PAI, as outlined in Section 4.1.2, Section 4.1.3 and Section 4.1.4. These artifacts are required to achieve the privacy preserving requirements for *Anonymous, Unlinkable and Protected Data Submission (R1)*, *Untraceable and Unlinkable Reward Allocation (R2)* and *Untraceable and Unlinkable Reward Spending (R3).* Structures written in C++ are used to define the offer token, T_{O}, used by the *Service Provider Component (CC3)* to publish an offer, the offer listing, L_{O}, published by the *OrderBook Component (CC2)*, the offer acceptance, A_{O}, used by the *Participant Component (CC1)*, the validation token, T_{V}, published by *CC3* on *CC2* to denote whether the offer is to receive a reward or not and the spendable reward token, r_{S}, allocated to *CC1.* Those structures defined in *SC1*, T_{O}, L_{O}, A_{O}, T_{V} and r_{S}, are identified in Figure 5.1 as *SC1.1*, *SC1.2, SC1.3*, *SC1.4* and *SC1.5* respectively.

Similarly, the *Cryptographic Library (SC2)* is also required to implement requirements *R1*, *R2* and *R3.* The cryptographic primitives required to implement the modified use of the Diffie-Hellman Key Exchange and generate the One-Time Key, K_{O}, use CryptoPP[4], a C++ library of cryptographic schemes. This generation of K_{O} is carried out by a *One-Time Key Generator* C++ class (*SC1.2*) which is invoked by the Participant Component, *CC1.* The other cryptographic operations used for PAI are also reliant upon this library. RSA[5] encryption is used for the encryption of the participant's data submission, d, in A_{O} using the service provider's public key, b_{SP}, and the encryption of the spendable reward token, r_{S}, by the OrderBook using the public part of the participant's One-Time Key, $a_{K_{\mathrm{O}}}$. The corresponding decryption operations, the decryption of d by the service provider using its private key, b^{*}_{SP}, and the decryption of r_{S} by the par-

[4] See `https://www.cryptopp.com`

[5] RSA is named after its creators, Ron **R**ivest, Adi **S**hamir and Leonard **A**dleman.

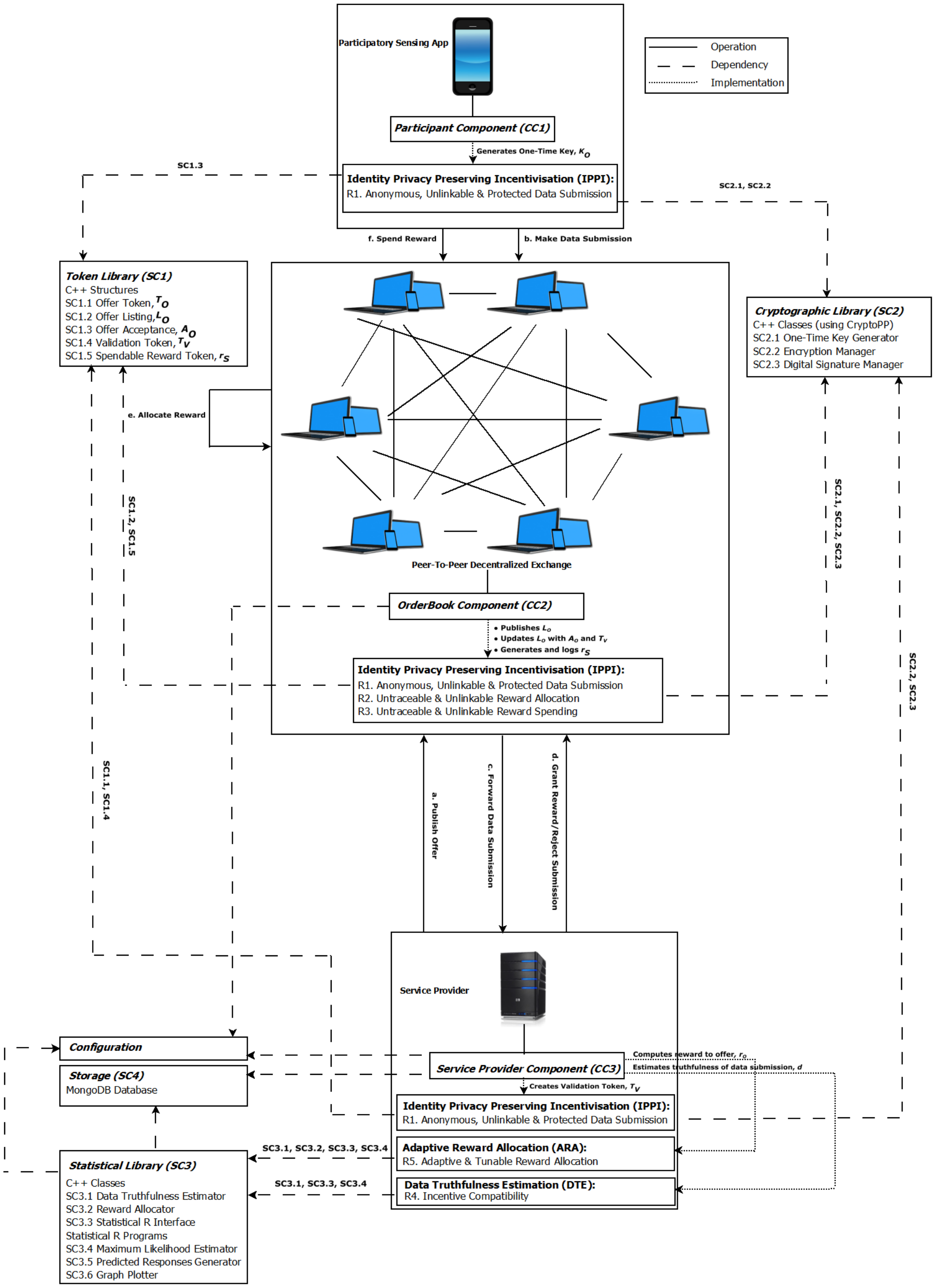
Participatory Sensing App
Operation
Dependency
Implementation
Participant Component (CC1)
Generates One-Time Key, K_O
Identity Privacy Preserving Incentivisation (IPPI):
R1. Anonymous, Unlinkable & Protected Data Submission
SC1.3
SC2.1, SC2.2
f. Spend Reward
b. Make Data Submission
Token Library (SC1)
C++ Structures
SC1.1 Offer Token, T_O
SC1.2 Offer Listing, L_O
SC1.3 Offer Acceptance, A_O
SC1.4 Validation Token, T_V
SC1.5 Spendable Reward Token, r_S
Cryptographic Library (SC2)
C++ Classes (using CryptoPP)
SC2.1 One-Time Key Generator
SC2.2 Encryption Manager
SC2.3 Digital Signature Manager
e. Allocate Reward
SC1.2, SC1.5
SC2.1, SC2.2, SC2.3
SC2.2, SC2.3
Peer-To-Peer Decentralized Exchange
OrderBook Component (CC2)
• Publishes L_o
• Updates L_o with A_o and T_v
• Generates and logs r_S
Identity Privacy Preserving Incentivisation (IPPI):
R1. Anonymous, Unlinkable & Protected Data Submission
R2. Untraceable & Unlinkable Reward Allocation
R3. Untraceable & Unlinkable Reward Spending
SC1.1, SC1.4
a. Publish Offer
c. Forward Data Submission
d. Grant Reward/Reject Submission
Service Provider
Configuration
Storage (SC4)
MongoDB Database
Service Provider Component (CC3)
Computes reward to offer, r_o
Estimates truthfulness of data submission, d
Creates Validation Token, T_V
Identity Privacy Preserving Incentivisation (IPPI):
R1. Anonymous, Unlinkable & Protected Data Submission
Adaptive Reward Allocation (ARA):
R5. Adaptive & Tunable Reward Allocation
Data Truthfulness Estimation (DTE):
R4. Incentive Compatibility
SC3.1, SC3.2, SC3.3, SC3.4
SC3.1, SC3.3, SC3.4
Statistical Library (SC3)
C++ Classes
SC3.1 Data Truthfulness Estimator
SC3.2 Reward Allocator
SC3.3 Statistical R Interface
Statistical R Programs
SC3.4 Maximum Likelihood Estimator
SC3.5 Predicted Responses Generator
SC3.6 Graph Plotter

Figure 5.1: Implementation of the PAI Approach

ticipant using the private part of the One-Time Key, $a^*_{\mathrm{K_O}}$, are also implemented using this library with both encryption and decryption being undertaken by the *Encryption Manager* C++ class (*SC2.2*). RSA is also used for digital signature operations, specifically, the digital signing of T_{V} by the service provider using its private key, b^*_{SP}, and the verification of this signature by the OrderBook using the corresponding public key, b_{SP}. *The Digital Signature Manager* C++ class (*SC2.3*) is responsible for these operations.

The steps taken by Algorithm 1[6] to allocate rewards to participants in a privacy preserving fashion are implemented by the *Participant Component (CC1)*, the *OrderBook Component (CC2)* and the *Service Provider Component (CC3)*. *CC1* generates K_{O} and publishes offer acceptances, $[A_{\mathrm{O}}]$, on the OrderBook, *CC2*. *CC2*, which publishes the service provider's offer listing, L_{O}, updates L_{O} with $[A_{\mathrm{O}}]$ and the validation tokens for each A_{O}, $[T_{\mathrm{V}}]$, with each T_{V} being created by *CC3*. Algorithm 2, which ensures that rewards are spent in a privacy preserving manner, is implemented in *CC2*, which generates r_{S} and logs its spending.

5.1.2 Implementing Incentive Compatibility (R4)

Having decrypted the data submission, d, in A_{O}, the service provider must then estimate the truthfulness of this data submission. To meet the requirement for *Incentive Compatibility (R4),* a program written in the statistical R programming language, named the *Maximum Likelihood Estimator (SC3.4)*, is used to implement the Maximum Likelihood Estimation (MLE) method that was identified in Section 4.2 as the basis for this requirement. This program is invoked as needed by the *Service Provider Component (CC3)* through the *Data Truthfulness Estimator (SC3.1)*, a C++ class that has been implemented to estimate data truthfulness.

SC3.1 in turn uses the *Statistical R Interface* C++ Class (*SC3.3*) to invoke the R program implementing Algorithm 3, the algorithm proposed in Section 4.2.2 for estimating data truthfulness. This program initially connects to the MongoDB *Storage (SC4)* via an interface, retrieves the measurement values for the pertinent data categories and stores this data in a dataframe structure. This data is then formatted into a vector array which is used to formulate the log-likelihood function, *LLF*, whose natural logarithm output is used in turn as a parameter when calling statistical R's *mle* function. As the name implies, this

[6] All algorithms are described and presented in Chapter 4.

function is used to estimate the mean, μ, and standard deviation, σ, for which the normal distribution best describes the data. Once computed, the values are returned to *SC3.1*. *SC3.1* then uses σ to compute the scaled standard deviation, σ_{scaled}, using the scaling factor, f_σ, set in the *Configuration*. σ_{scaled} and μ are then used to determine whether the submitted data value falls within the scaled limits and thus estimate whether the data submission is a truthful one.

5.1.3 Implementing Adaptive & Tunable Reward Allocation (R5)

The core operation for a participatory sensing environment is the publishing of offers on the OrderBook by the service provider and the response to these offers by participants. For the service provider, the key decision to be made when publishing an offer is the level of reward to allocate. This requirement for *Adaptive and Tunable Reward Allocation (R5)* is implemented in the *Statistical Library, (SC3)*, using both the C++ and R programming languages. The reward allocator is implemented as a C++ class, named *Reward Allocator (SC3.2)*, with Algorithm 4, the Lyapunov Optimisation-based reward computation algorithm, being the key function of this class.

To generate the predicted responses for the different reward levels for a particular category of measurement, *SC3.2* uses the *Statistical R Interface* C++ Class (*SC3.4*) to call the *Predicted Responses Generator (SC3.5)*, an R program that connects to the MongoDB *Storage (SC4)*. Having connected to the database, the program then reads the last n number of rows of reward levels and corresponding responses for this measurement (n corresponds to the value set for the rolling regression window in the system *Configuration*). After using the *Graph Plotter (SC3.6)*, to plot the supply curves depicting the relationship between the number of responses and the different reward levels, the program then uses statistical R's *nlsLM* function to construct a non-linear regression model. This model serves as a parameter for statistical R's *predict* function. In conjunction with a dataframe containing the different reward levels, this function is used to predict the number of responses for different reward levels, up to the maximum reward set in the *Configuration*.

The predicted output is written to a CSV[7] file that is read by the *Reward Allocator (SC3.2)*. These values are then used to compute $Z_{\text{forfeit}}(t)$. As discussed in Section 4.3.1.2, $Z_{\text{forfeit}}(t)$, the number of forfeited responses at a particular

[7] CSV, a file format, stands for **C**omma **S**eparated **V**alues.

reward level, is the queue backlog for the Lyapunov Optimisation system for a timeslot, t. Having computed $Z_{\text{forfeit}}(t)$ for the different reward levels, *SC3.2* then implements the rest of Algorithm 4 to determine the reward that represents the optimal trade-off between response rate and budget consumption. This process is repeated for each timeslot t, the duration of which is defined in the configuration. The reward for a particular timeslot, $r(t)$, is published as part of any offer token, T_{O}, listed on the *OrderBook Component (CC2)* during that timeslot.

Algorithm 4, which computes the reward in an adaptive and tunable fashion, is implemented by the *Service Provider Component (CC3).*

5.1.4 Implementing a Simulated Environment for Participatory Sensing

In addition to the core and supporting components implemented for requirements *R1*, *R2*, *R3*, *R4* and *R5*, a simulated participatory sensing environment has been implemented in C++ to evaluate these components and the overall PAI approach. The simulator models a typical participatory sensing scenario where a service provider publishes its offers on the OrderBook with participants responding to these offers. The goal of this simulator is to validate the algorithms (Algorithms 1, 2 and 3) that have been proposed in Chapter 4 to meet the requirements (*R1*, *R2*, *R3*, *R4*) pertaining to privacy preservation and incentive compatibility. The performance of the *Adaptive Reward Allocation (ARA)* model that is proposed in Chapter 4 to meet requirement *R5* is also evaluated using the simulator.

The simulator is configurable as the duration of the simulation and the number of participants and OrderBook hosts can be set prior to runtime. These values, like those for the *OrderBook Component (CC2)* and the *Service Provider Component (CC3)*, are accessed from the system *Configuration.* On initial startup, a simulation creates the desired number of *Participant Components (CC1)*, the desired number of distributed hosts for the *OrderBook Component (CC2)* and the *Service Provider Component (CC3).*

The simulator can also be used to populate the *Storage (SC4).* On initial startup. *SC4* may contain no data. As training data is required to facilitate the ARA model in the learning it needs to predict response rates for the different reward levels, the absence of such data would hinder the ability of the approach in making these predictions. For this reason, a configurable option is used to denote whether data pertaining to the number of responses for different reward levels is to be generated. This is achieved by using a uniform discrete probability distribution to generate the number of responses for different reward levels up to the maximum reward level.

5.2 Analysis & Validation

This section analyses PAI to evaluate whether it meets the requirements identified in Chapter 2. Requirements *R1*, *R2*, *R3* and *R4* are evaluated by proof in Sections 5.2.1, 5.2.2, 5.2.3 and 5.2.4 respectively. Requirement *R5* is evaluated in a simulated participatory sensing environment. The simulation setup and the experiments conducted for this requirement are described in Section 5.2.5.

5.2.1 Anonymous, Unlinkable & Protected Data Submission (R1)

Requirement *R1* is demonstrated through showing that participants make data submissions to the service provider anonymously (Theorem 3), illustrating that the data submissions made by participants are unlinkable (Theorem 4) and showing that only the service provider can access these data submissions (Theorem 5).

Theorem 3 Participants make data submissions to the service provider anonymously.

Proof To obtain the data it needs, the service provider publishes a series of offer tokens $[T_{\mathrm{O}}]$ on PAI's decentralised OrderBook denoting the data being sought, δ, and the reward being offered, r_{O}. Each offer token, T_{O}, is published as a listing, L_{O}, to which the acceptances of the participants, $[A_{\mathrm{O}}]$, are appended. The OrderBook is accessible to all participants and service providers. A participant can then choose to make a data submission, d, in return for the offered reward, r_{O}.

When issuing the offer token, T_{O}, the service provider has no direct communication with any of the participants. Similarly, a participant does not communicate directly with the service provider when publishing the offer acceptance, A_{O}, which contains the data submission, d. Instead, A_{O} is appended to the offer listing, L_{O}, and published on the OrderBook. The offer acceptance, A_{O}, and its constituent components, the ID of the offer being responded to, i_{O}, the data submission, d (unless the participant willingly submits privacy-ceding attributes) and the public part of the generated key, $a_{\mathrm{K_O}}$, do not contain any link to the participant's identity or any anonymised ID or pseudonym that could be used to identify the participant. In addition, as the offer acceptance's unique ID, $i_{\mathrm{A_O}}$, is only assigned on receipt of the acceptance by the OrderBook, it cannot be traced back to the participant.

The service provider is notified of the data submission when A_{O} is forwarded to it. While it can access and evaluate the data, d, contained therein, it has no means of linking the submission to the participant's identity. Therefore, the participant makes its data submission to the service provider anonymously without disclosure of identity[8]. ■

Theorem 4 Participants make unlinkable data submissions to the service provider.

Proof The public part of the One-Time Key, $a_{\mathrm{K_O}}$, is used to identify the data submission made by a participant. In addition to containing no link to the participant's identity, a different $a_{\mathrm{K_O}}$ is generated for each data submission. As each data submission is an independent transaction with a different ID, the service provider has no means of linking data submissions made by the same participant. It should also be noted that data submission unlinkability is facilitated by the fact that data can be submitted anonymously, as demonstrated by Theorem 3. ■

Theorem 5 Data Submissions can only be accessed by the Service Provider

Proof When a participant publishes an offer acceptance, A_{O}, on the OrderBook, the data submitted is encrypted using the service provider's public key, b_{SP}. This encrypted data, $\{d\}_{\mathrm{b_{SP}}}$, cannot be accessed by any peer hosting the OrderBook or any other participant or service provider as it can only be accessed through decryption using the service provider's private key, b^{*}_{SP}. ■

[8] Similarly, a peer listing the data submission cannot determine the participant's identity from A_{O}. The participant can also take further steps to hide their device ID or, if necessary, their IP address through the use of, for example, a Virtual Private Network (VPN).

5.2.2 Untraceable & Unlinkable Reward Allocation (R2)

Requirement *R2* is demonstrated through showing that participants receive rewards that are untraceable (Theorem 6) and unlinkable (Theorem 7).

Theorem 6 Participants receive untraceable rewards.

Proof The public part of the One-Time Key, $a_{\mathrm{K_O}}$, is published on the OrderBook as part of the offer acceptance, A_{O}. A participant's $a_{\mathrm{K_O}}$ cannot be traced back to the participant who generated the corresponding One-Time Key, K_{O}, as it has no relationship with the participant's identity.

The service provider publishes a flag v for A_{O} on the OrderBook as part of a validation token, T_{V}, indicating its decision with respect to whether A_{O} should be rewarded. $i_{\mathrm{A_O}}$, the ID of the offer acceptance, is used to indicate the reward's recipient. As $i_{\mathrm{A_O}}$ is generated by the OrderBook, it cannot be linked to the participant. Once T_{V} is published on the OrderBook, an encrypted spendable reward r_{s} is created for A_{O} using $a_{\mathrm{K_O}}$. Only the owner of A_{O} can access and consume r_{s} as it is the only party that holds the private part of the One-Time Key, a^*_{K}. Thus, a participant can make a valid data submission and receive an untraceable reward without disclosure of identity. ■

Theorem 7 Participants receive unlinkable rewards.

Proof A One-Time Key, K_{O}, is used only for one offer acceptance, A_{O}, so cannot be used to link a participant's set of offer acceptances, $[A_{\mathrm{O}}]$. In addition, the ID of the spendable reward, i_{S}, which is generated by the OrderBook, is neither linkable to the participant nor to the public part of the One-Time Key itself, $a_{\mathrm{K_O}}$. Moreover, the service provider cannot connect r_{s} and its ID, i_{S}, to a participant or to $a_{\mathrm{K_O}}$. This is because r_{s} is only decrypted when it is being spent. Therefore, neither $a_{\mathrm{K_O}}$ nor the r_{s} encrypted using $a_{\mathrm{K_O}}$ can be used to establish linkages between the data submissions of a particular participant.

The absence of any linkable ID in a participant's set of offer acceptances, $[A_{\mathrm{O}}]$, set of public One-Time Key components, $[a_{\mathrm{K_O}}]$ or set of spendable rewards, $[r_{\mathrm{s}}]$ therefore means that the service provider has no means of inferring any data about that participant's behavior and activity beyond what is contained in the data submission itself. ■

5.2.3 Untraceable & Unlinkable Reward Spending (R3)

Requirement *R3* is demonstrated by Theorem 8, which shows that participants' reward spending is untraceable and unlinkable.

Theorem 8 Participants' reward spending is untraceable and unlinkable.

Proof The service provider has no role in the publishing of the offer acceptance, A_{O}, or the allocation of the reward offered, r_{O}. Specifically, the fact that the service provider cannot assign traceable IDs to A_{O} means it cannot trace participant activity on the participatory sensing system subsequent to reward allocation through the assignment of r_{O}.

A_{O} has no fixed ID. While the OrderBook assigns a unique ID, $i_{\mathrm{A_O}}$, on receipt of A_{O}, this ID cannot be used to trace the participant. In addition, as the One-Time Key, K_{O}, is only used once and then discarded, the service provider has no means of linking participant activity through the publication of the latter's offer acceptances, $[A_{\mathrm{O}}]$. Moreover, the spendable reward, r_{s}, cannot be connected to A_{O} as it is only decrypted when it is being spent. Therefore, a participant's set of spendable rewards, $[r_{\mathrm{s}}]$, is untraceable and unlinkable as is the spending of these rewards. ■

5.2.4 Incentive Compatibility (R4)

Requirement *R4* is demonstrated through showing that data submissions that are considered to be non-truthful do not receive a reward (Theorem 9) and illustrating that the incentive compatibility method is privacy preserving (Theorem 10).

Theorem 9 Data submissions that are considered to be non-truthful do not receive a reward.

Proof The service provider sets the minimum and maximum threshold limits, $m_{\mathrm{c_{min}}}$ and $m_{\mathrm{c_{max}}}$ and also configures a scaling factor, f_σ, that is used to compute the interval between the scaled limits, $l_{\min}$ and $l_{\max}$, that is deemed to contain valid values for a measurement category, m_{c}. Participants have no role in determining these parameters. Any measurement category in a sensed data submission, d, that fails either of these two tests results in d as a whole being considered untruthful and, consequently, not receiving a reward. ■

Theorem 10 The incentive compatibility method is privacy preserving.

Proof The service provider receives the sensed data, d, as part of the offer acceptance, A_{O}. Prior to approving the reward allocation, the service provider ensures that d matches the criteria of the offer represented by i_{O} and can then evaluate the truthfulness of d using the *Data Truthfulness Estimation (DTE)* algorithm described in Chapter 4. While this ensures that the service provider does not allocate rewards for non-truthful data submissions, it does not violate identity privacy as the service provider has no access to the participant's identity. At the same time, the fact that the validation token, T_{V}, is published for every offer acceptance ID, $i_{\mathrm{A_O}}$, ensures that participants can see that a non-truthful data submission has been rejected. ■

5.2.5 Adaptive & Tunable Reward Allocation (R5)

This section evaluates PAI's requirement for *Adaptive and Tunable Reward Allocation (R5)*. Section 5.2.5.1 describes the experimental setup for the simulated participatory sensing environment that is used to evaluate the adaptiveness, utility and budget consumption of PAI's reward allocation model. The experiments and results evaluating adaptiveness and utility are presented in Section 5.2.5.2 with Section 5.2.5.3 presenting the experiments and results evaluating budget consumption. It should be noted that requirement *R5*'s specification that data quality not be impaired is met through the fulfillment of requirements *R1*, *R2*, *R3* and *R4* (as these requirements do not alter, or otherwise diminish, the quality of the submitted data).

As outlined in Chapter 3, incentivisation has been extensively considered in the state of the art with SenseUtil (Tsujimori et al., 2014) and the STOC-PISCES (Biswas et al., 2015) approaches in particular addressing the need to consider participation rates and data utility. Given their similarity in intent to PAI's ARA component, these approaches are used as baselines for comparison in the evaluation of requirement *R5*.

5.2.5.1 Experimental Setup

The *Adaptive Reward Allocation (ARA)* component that has been designed to meet PAI's requirement for *Adaptive and Tunable Reward Allocation (R5)* is evaluated in a simulated participatory sensing environment. The service provider makes a series of offers over the duration of the simulation with 100 responses being sought for each offer. This figure is chosen to facilitate ease of response rate computation and to reflect a participatory sensing environment where the service provider is seeking to build a large dataset. Furthermore, while the number of responses sought will vary among participatory sensing applications as well as over time, this figure is also chosen to clearly determine whether the reward level is adapting to the response rate and the utility of the data sought. The minimum and maximum reward to be set for an offer are 10 and 200 units respectively as this range should produce diverse reward levels that will demonstrate the adaptiveness and tunability of the ARA model. To generate a comprehensive dataset, each simulation runs for one hour with offers being generated every 5 seconds[9]. Each simulation thus generates reward and response rate pairings for over 700 offers on average. For the purposes of the simulation, it is assumed that each offer corresponds to one timeslot t (i.e. one offer is produced per timeslot) with the reward being re-evaluated with each offer.

The simulation is run in two types of environment, one with a high initial number of responses (referred to as the 'High Response Environment') and one with a low initial number of responses (referred to as the 'Low Response Environment'). The high response environment is modeled so as to ascertain whether ARA takes the high response rate into account by lowering the reward offered while the low response environment is modeled to determine whether ARA responds to a low response rate by increasing the offered reward in order to attract a higher number of responses, The initial response rate ranges between 70% and 200% and 10% and 50% respectively for these environments.

The participant response rate is generated using a continuous uniform distribution. The simulation model varies this response rate using a randomly generated increment to evaluate how the reward adapts to these changes in the response rate. This randomness is incorporated to reflect other factors in the participatory sensing environment that may affect the response rate. The range for a randomly generated increment (also using a continuous uniform distribution) is set between 20 and 40 responses for a high response environment and 5 and 10 responses for a low response environment in the simulation configuration. These ranges are chosen to ascertain the ARA model's effectiveness in detecting the type of participatory sensing environment in which it is currently operating. For the purposes of the simulation, the response rate is calculated simply as the

[9] The same simulation setup parameters are used for the baselines, STOC-PISCES (Biswas et al., 2015) and SenseUtil (Tsujimori et al., 2014).

ratio of the number of responses submitted to the number of responses sought. However, the response rate could be defined using other metrics such as the coverage of a particular area (Girolami et al., 2016 and Girolami et al., 2017) without impacting the underlying algorithm.

As noted in Section 4.3.5, the value of V, the non-negative control parameter that is used to tune the prioritisation of the budget consumption over data capture and vice versa for the Lyapunov Optimisation model, can be tuned to reflect the data attributes that are of most interest to a service provider at a particular point in time. For the purposes of the evaluation, the value of V for each offer is set either to 0 to prioritise attracting data submissions or 1000 to prioritise budget consumption. The figure of 1000 is chosen as it should be sufficiently high enough to clearly demonstrate the effect of prioritising budget optimisation over data capture. As V is set to 0 for the majority of experiments, the value of V is only indicated when its value is 1000. Finally, the size of the rolling regression window used for predicting the number of responses is set to the last 100 responses for the majority of the experiments carried out. In addition, the last 50 responses are used for a number of experiments to determine the effect of changing the value of the rolling regression window. These values are chosen so as to model a dynamic environment where the most recent responses to a service provider's offers are the most relevant in determining the reward level. For example, this would be the case for an environmental sensing application. The service provider can, of course, configure different regression rolling window values for other types of participatory sensing environments, for example, a lower figure may be appropriate when the data collection is more infrequent and/or sparse. Finally, the budget is set to 50,000 units for those experiments that evaluate budget consumption. This figure, which is 250 times the value of the maximum reward of 200 units, is used so as to assess the effectiveness of the ARA model's budget consumption over time compared with similar approaches from the state of the art. Table 5.1 presents the parameters used for the simulation.

While the similar objectives of ARA and STOC-PISCES (Biswas et al., 2015) means that the latter can be integrated in the modeled participatory sensing environment without customization of the underlying algorithm, this is not the case for SenseUtil (Tsujimori et al., 2014) as the latter does not adapt rewards to the response rate. Rather, SenseUtil determines the reward to offer on the basis of the number of potential participants. To ensure a valid comparison, the SenseUtil model is simulated in a participatory sensing environment with the number of potential participants set to 50 (the figure used by the authors) and 100 (which, according to the authors of the approach, should lead to lower rewards) respectively. The authors assume that these figures remain constant for the duration of the participatory sensing campaign. The computed utility, as is the case for the simulation used by the authors, is mapped on a one-to-one basis to an economic point system. Using this system, points accumulated by participants can be used to request sensed data from the service provider[10]. Alternatively, the authors state that any kind of monetary or virtual currency can be applied to SenseUtil for payments and rewards. A one-to-one mapping between the economic point system and the reward to be offered is used for the simulation with the utility range being set from 10 to 200 to correspond to the reward range used for ARA and STOC-PISCES. As SenseUtil is designed for location based participatory sensing, the approach computes the utility of a data submission on the basis of its distance from the location requested by the service provider. The distance threshold used to compute this location utility in the simulation is (like the authors' simulation) set to 50m and 100m. As SenseUtil differs from ARA in its reliance upon distance in its reward computation mechanism, it is only used as a basis of comparison when comparing the average reward offered by the approaches.

5.2.5.2 Adaptiveness & Utility

Figure 5.2 presents the adaptiveness of ARA to the response rate. It can be seen from the graph that the reward is increased so as to attract more data submissions at low response rates while the reward is reduced where the response rate approaches or exceeds 100%. Moreover, the reward settles on a value over time that generates a response rate close to 100%. It should be noted that the reward is not always immediately adapted after a change in the response rate for a particular offer as the regression model used for the supply curves ensures that the focus is on changes that occur over time rather than sudden changes that may be outliers, thus ensuring that the budget is not needlessly consumed. Figures 5.3-5.8 compare the adaptiveness of ARA with the STOC-PISCES algorithm. The number of initial trials used by STOC-PISCES is set to 10, which is the figure used by the approach's authors in their evaluation. As the STOC-PISCES algorithm initially runs a number of trials offering an initially higher reward at the median (105 for a range of 10-200), the initial reward for ARA is set to an initial value of 105 units to ensure a fair comparison. It can be

[10] i.e. data sensed by other participants.

Parameter	Value/Range
No. Of Simulations	10
Simulation Duration	3600 seconds
Offer Interval	5 seconds
No. Of Responding Participants	10-200
Initial Reward	105
No Of Responses Sought	100
Initial Response Rate	
- High Response Environment	70%-200%
- Low Response Environment	10%-50%
Response Increment	
- High Response Environment	20-40
- Low Response Environment	5-10
Minimum Reward for an Offer	10
Maximum Reward for an Offer	200
Service Provider Budget	50,000
V	0, 1000
Rolling Window for Regression	100, 50

Table 5.1: Simulation Parameters for Requirement R5

seen from Figure 5.4 and Figure 5.7 that STOC-PISCES adapts to the response rate at a much slower rate than ARA in a high response environment. This is reflected in the average reward offered by STOC-PISCES which, at 102.87 units is almost four times higher than the figure of 26.33 units for ARA. It should also be noted that the error bars in Figure 5.4, reflecting standard deviation, show that the variability of the data becomes steady over time as ARA determines a reward with a response rate at or approaching 100%.

The findings for a low response environment for ARA and STOC-PISCES are presented in Figure 5.5 and Figure 5.8 respectively. In this environment, STOC-PISCES rapidly and substantially increases the reward it offers so as to attract more submissions. This leads to much higher rewards being offered for the equivalent response rate received by ARA. ARA offers a substantially higher average reward of 48.26 units in this environment compared to the high response environment but this figure is, nonetheless, still three and a half times lower than the average of 175.25 units offered by STOC-PISCES. The error bars in Figure 5.5 indicate that the reward eventually settles at a level that attracts a response rate above 80%, ultimately approaching and exceeding a 100% response rate.

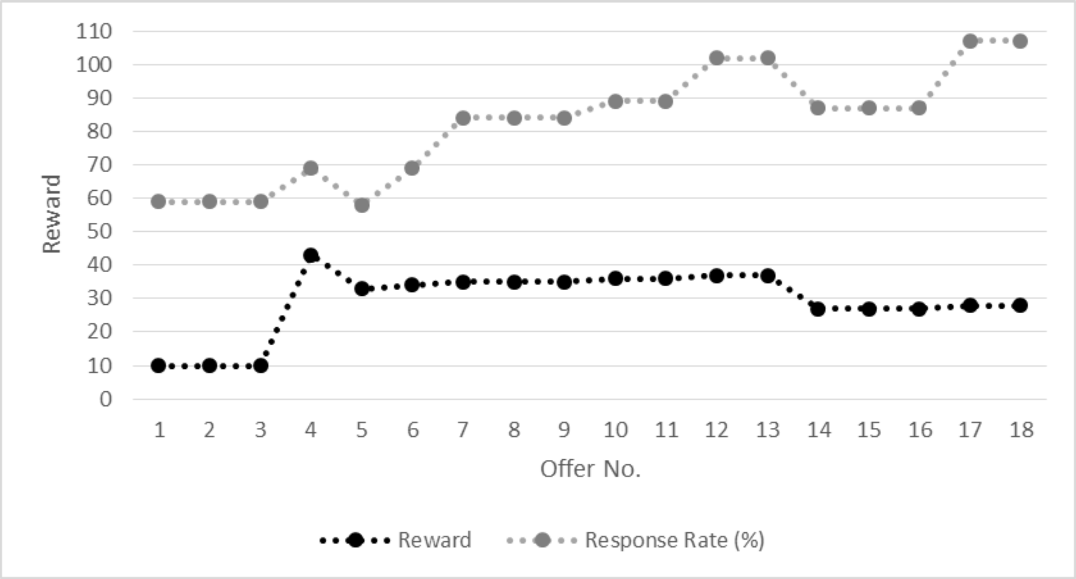

Figure 5.2: Adapting the Reward to the Response Rate

Figure 5.6 shows that the average reward is significantly lower (11.08 units) when, by setting the value of V to 1000, budget consumption is prioritized over attracting data submissions in a low response environment. This highlights how ARA not only adapts its reward to response rates but also uses the value of V to take data utility into account.

It should be noted that the reward level increases noticeably on a consistent basis at and around offer number 100 for both Figure 5.3 and Figure 5.4 which depict a single simulation and the average results over 10 simulations respectively. This appears to occur as a result of the rolling regression window being set to 100, as the point at which the reward level increases roughly corresponds to the value of this window. Figure 5.9 and Figure 5.10 depict a high response environment over one and ten simulations respectively with the rolling regression window being set to 50. In this case, the reward level increases noticeably at and around offer 50. It should be noted that the higher average reward of 41.09 reflects a more volatile environment in which the regression window is set to a smaller value. This would be appropriate in, for example, a traffic monitoring application in a congested city.

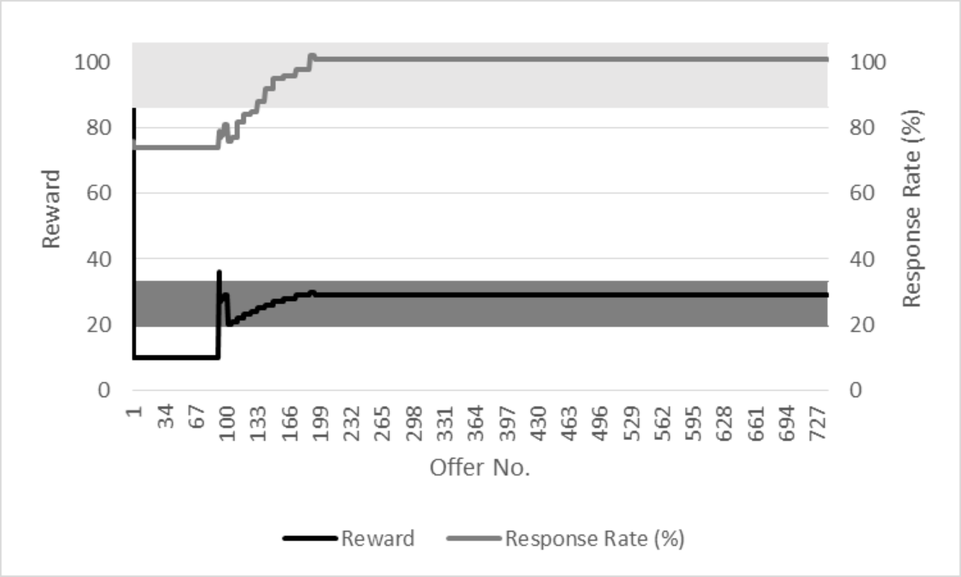

Figure 5.3: ARA Adaptiveness (High Response Environment, Rolling Regression Window=100, No. Simulations=1)

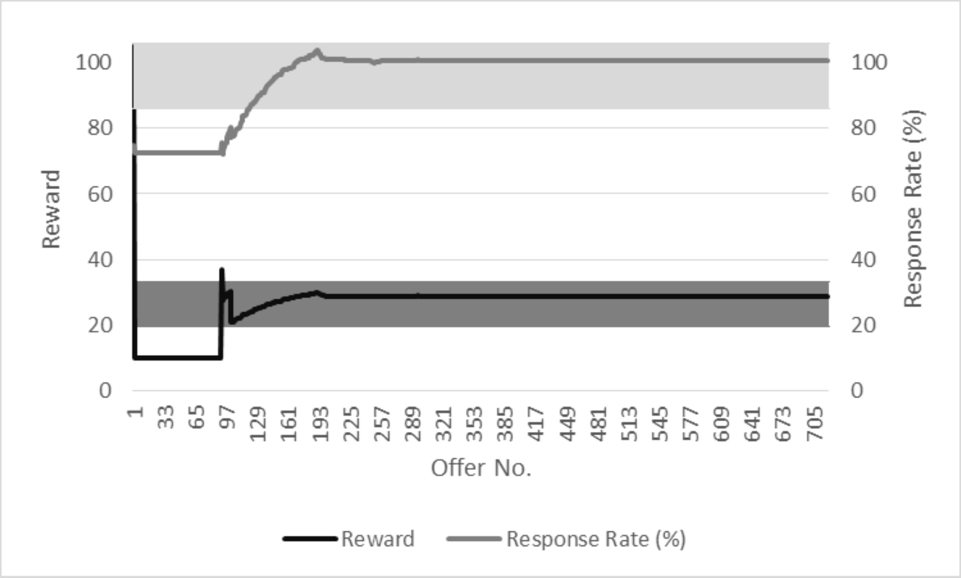

Figure 5.4: ARA Adaptiveness (High Response Environment, Rolling Regression Window=100, No. Simulations=10)

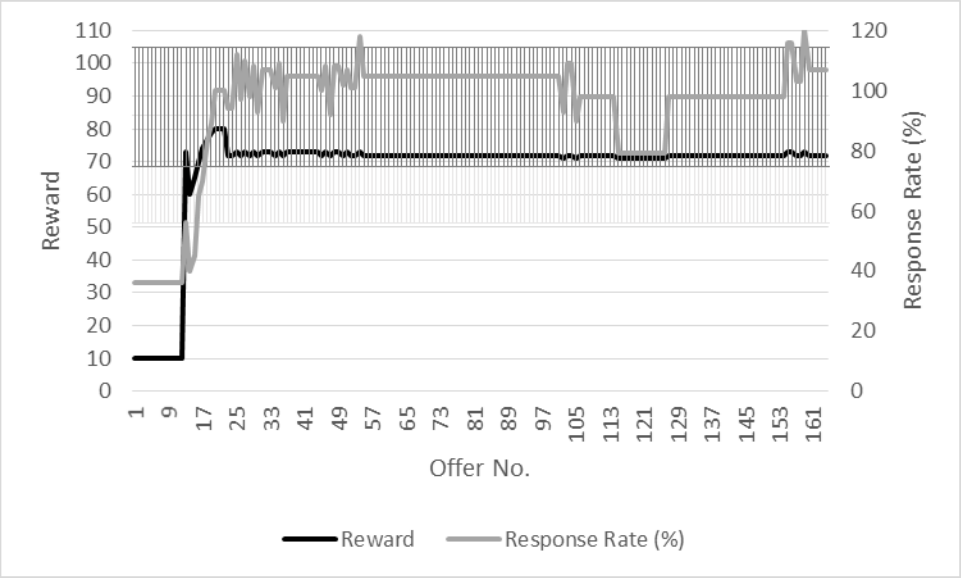

Figure 5.5: ARA Adaptiveness (Low Response Environment)

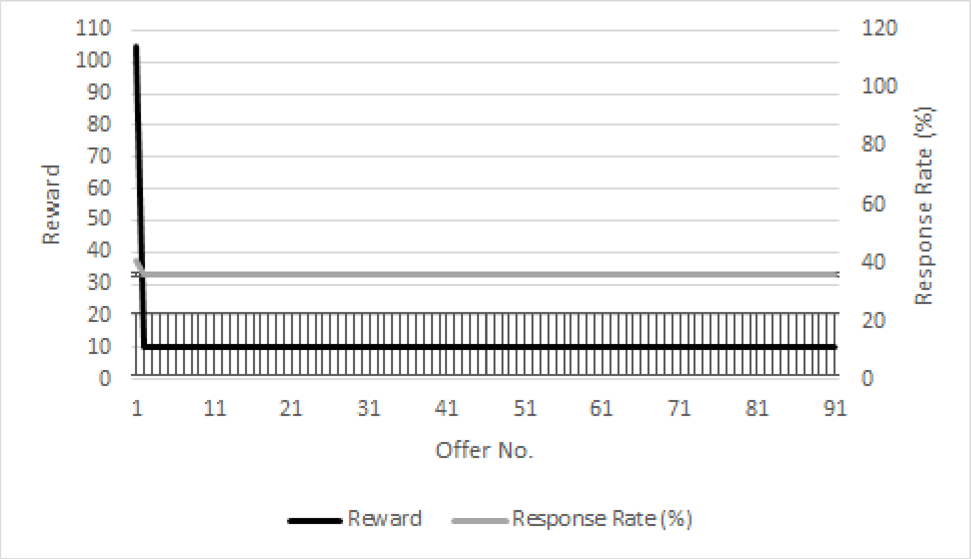

Figure 5.6: ARA Adaptiveness (Low Response Environment, V = 1000)

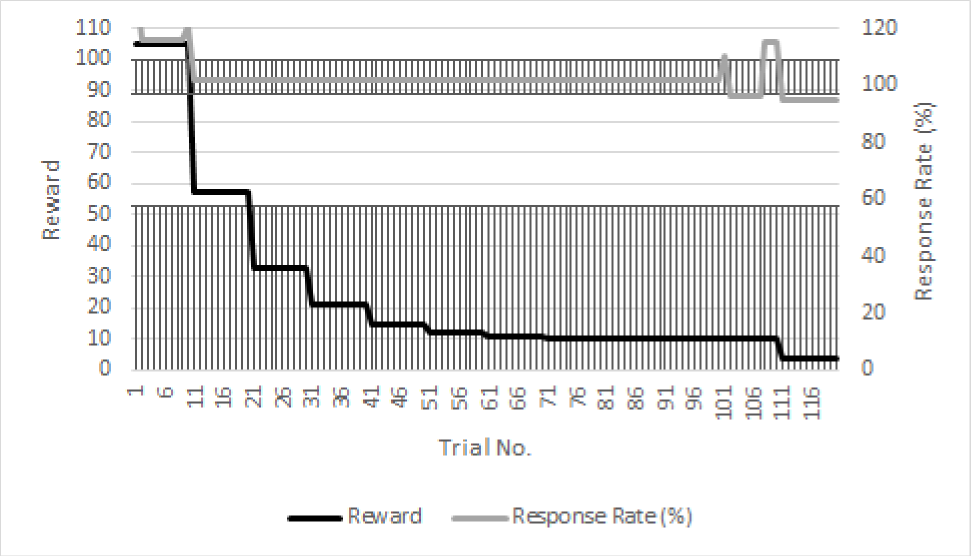

Figure 5.7: STOC-PISCES Adaptiveness (High Response Environment)

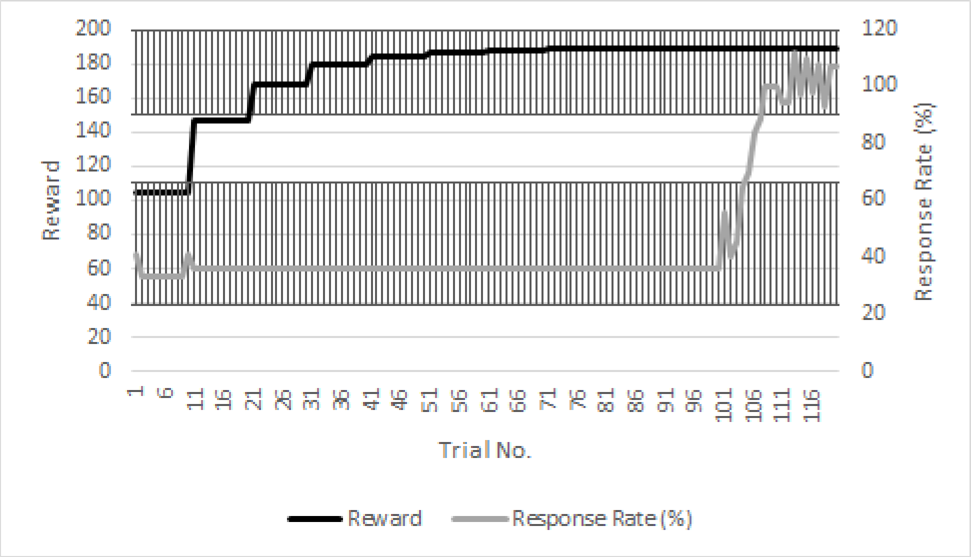

Figure 5.8: STOC-PISCES Adaptiveness (Low Response Environment)

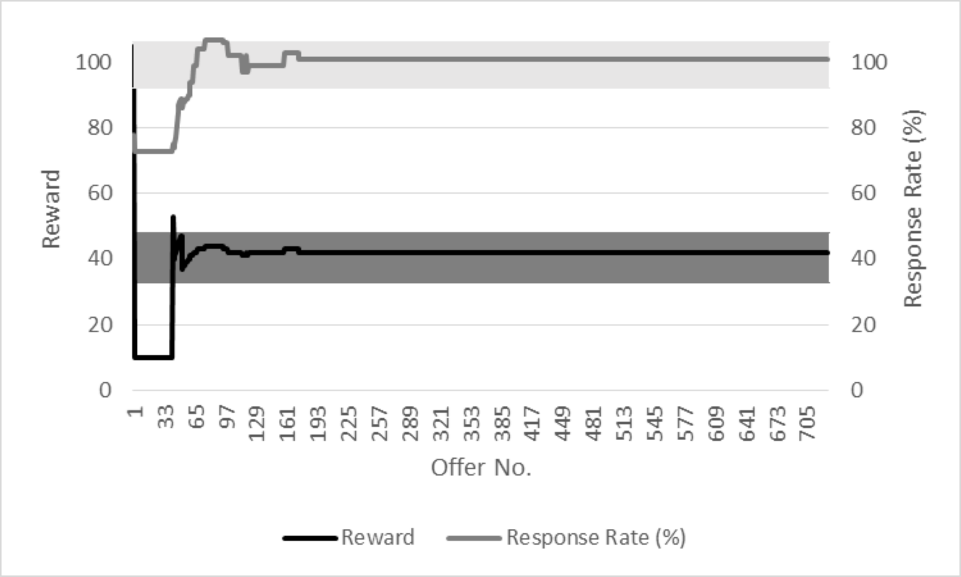

Figure 5.9: ARA Adaptiveness (High Response Environment, Rolling Regression Window=50, No. Simulations=1)

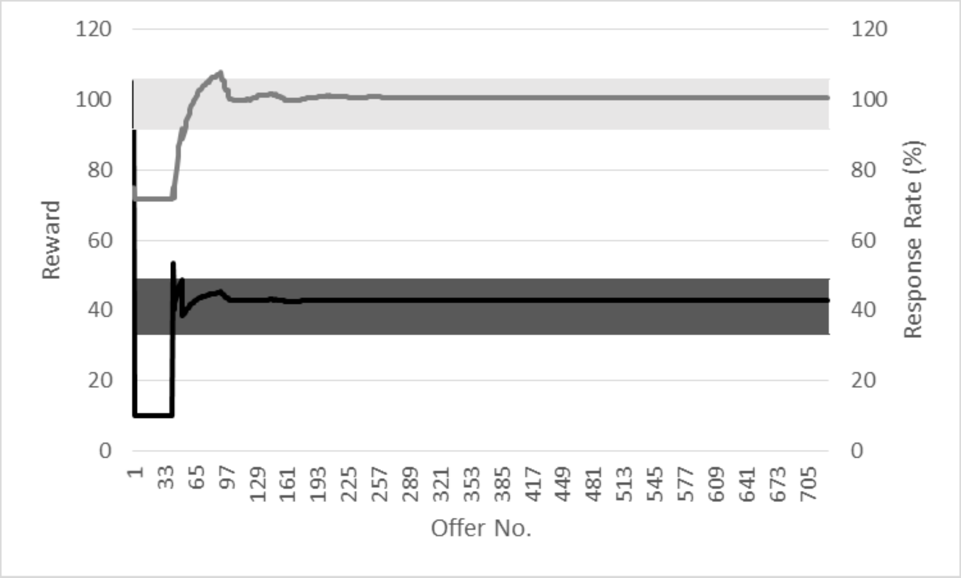

Figure 5.10: ARA Adaptiveness (High Response Environment, Rolling Regression Window=50, No. Simulations=10)

5.2.5.3 Budget Consumption

This section evaluates the budget consumption of PAI compared to that of STOC-PISCES and SenseUtil. As the level of reward offered is a key factor in the rate of budget consumption, the average reward is assessed for each approach in both a high and low response environment. As the service provider will wish to generate as many offers and resultant responses as possible, experiments are also conducted to determine whether PAI generates more offers and responses than STOC-PISCES[11] for the same budget. Finally, the rate of budget consumption is evaluated to see whether ARA uses its budget more effectively than STOC-PISCES.

Figure 5.11 presents the average reward for ARA, STOC-PISCES and SenseUtil. It can be seen that the average reward for ARA is lower than that computed by the other two approaches. This is the case even in a low response environment with V being set to zero so as to attract as many responses as possible. Figures 5.12, 5.13, 5.14 and 5.15 present a comparison of the budget consumption by ARA and the STOC-PISCES algorithm. The SenseUtil approach is not used for this experiment as integrating adaptiveness to the response rate would require modification and extension of the underlying algorithm. Each simulation has been run until the allocated budget has been consumed. It can be seen from Figure 5.12 that, with a budget of 50,000 units, ARA generates 3325 responses in a high response environment. This figure is over five and a half times larger than that for STOC-PISCES at 598. Moreover, in a low response environment, ARA generates over 3682 responses, albeit with a much higher number of offers than in the high response environment. It should also be noted that, in both the low and high response environment, the budget optimisation of ARA is superior to that of STOC-PISCES as the former generates a higher number of offers with the same budget. As shown in Figure 5.13, the number of offers is 31 and 102 respectively for ARA; the number of offers is 6 (five times less) and 13 (almost eight times less) respectively for STOC-PISCES. The higher number of responses generated by ARA appears to result from STOC-PISCES not taking budget consumption into account. Specifically, the offering of the same reward for a number of trials regardless of the response rate results in a higher overall average reward and more rapid budget consumption. In contrast, ARA reduces the reward it offers more quickly in a high response environment and while it does increase its reward in a low response environment, it does so more

[11] The goals of SenseUtil mean that a similar comparison cannot be made.

prudently than STOC-PISCES which tends to raise the level of reward offered close to the maximum reward more quickly than ARA. This is borne out by the budget consumption which is much steadier for ARA in both a high and a low response environment as shown in Figure 5.14 and Figure 5.15.

It should be noted that the budget does not incorporate the cost of provisioning the service provider for any of the approaches discussed in this section, including PAI. In addition, the cost of incentivising the peer devices is not considered for PAI. However, this would potentially be offset by the savings made by the reduced infrastructure requirements for the service provider when using the approach.

5.2.5.4 Summary

Section 5.2.5 evaluates the performance of the approach used to implement the requirement for *Adaptive and Tunable Reward Allocation (R5).* The ability of ARA to adapt to changes in the response rate is demonstrated in Section 5.2.5.2 for both a low and high response environment. The ability to tune ARA to prioritise budget consumption over data capture is also explored in this section. The STOC-PISCES approach is used as a basis of comparison with experiments showing that ARA adapts more rapidly to the response rate in both a low and high response environment. In addition, the average reward offered by ARA is lower than that of both STOC-PISCES and SenseUtil in both a high and low response environment. Budget consumption is further explored in Section 5.2.5.3 which shows that ARA generates a higher number of offers and receives a higher number of responses than STOC-PISCES. As a result, the rate of budget consumption in both a high and low response environment is superior to that of STOC-PISCES[12]. The average reward offered by ARA compared to that of STOC-PISCES has been found to be 73.92% and 72.46% lower in a high and low response environment respectively. In addition, ARA generates 82.02% and 83.76% more offers using the same budget as STOC-PISCES in these respective environments. Overall, the metrics used to evaluate the two approaches find that the performance of ARA is, on average, 79.16% better than that of STOC-PISCES.

[12] The performance of STOC-PISCES was not found to be equal to or better than ARA in any of the experiments carried out for this thesis.

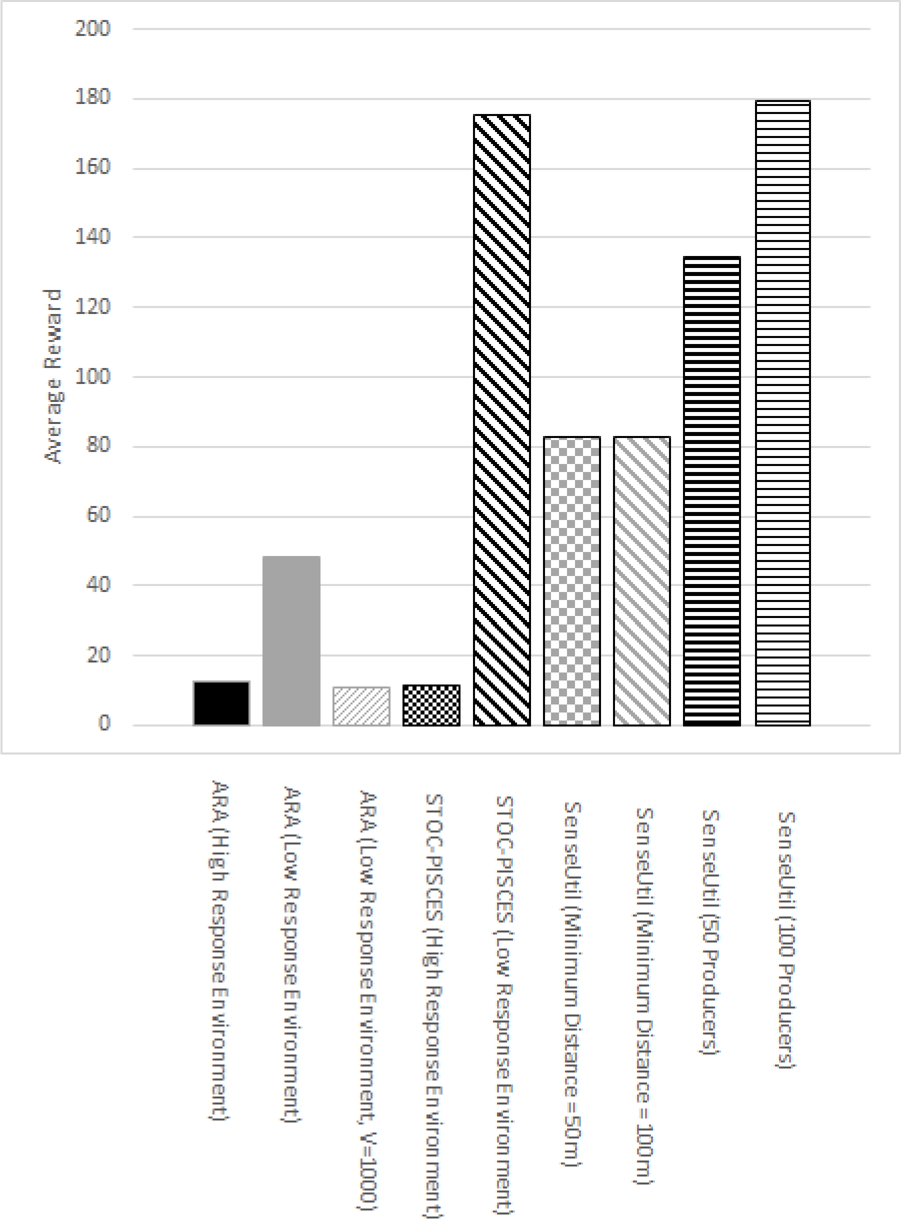

Figure 5.11: Average Reward

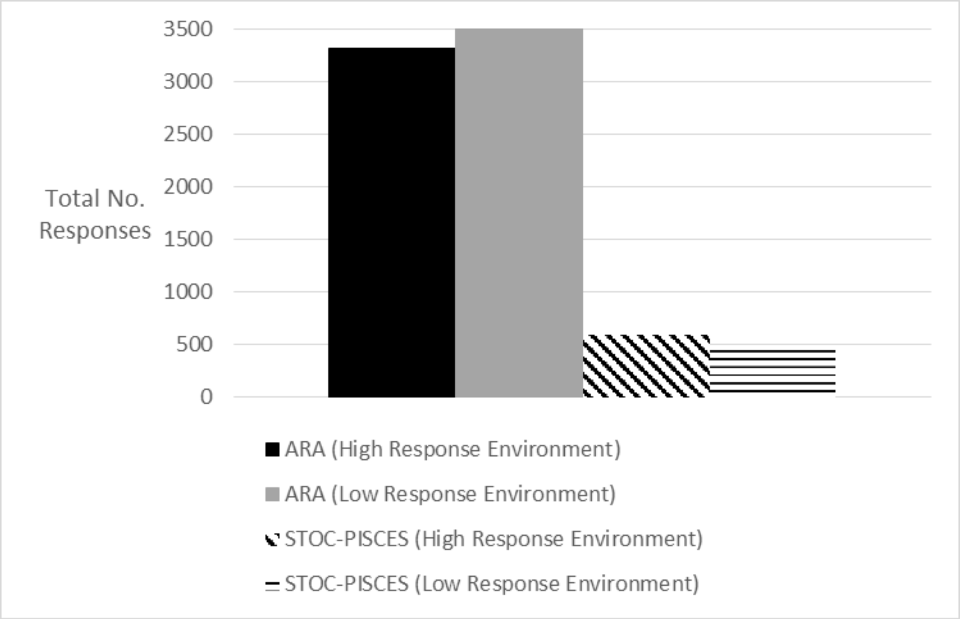

Figure 5.12: Total No. Responses

5.3 Performance Evaluation

This section evaluates the performance of PAI, focusing on resource consumption in terms of time, energy and processing requirements. The most likely potential resource consumption bottleneck for any privacy preservation approach pertains to its use of cryptographic primitives for the privacy preserving requirements for *Anonymous, Unlinkable and Protected Data Submission (R1)*, *Untraceable and Unlinkable Reward Allocation (R2)* and *Untraceable and Unlinkable Reward Spending (R3)*. Thus, Section 5.3.1 evaluates the performance of those primitives used by PAI. To analyse the level of processing resources required by the algorithms as a whole, Section 5.3.2 assesses the computational complexity of PAI's four algorithms.

The approaches taken by Li and Cao [2016] and Dimitriou [2018b], both of which seek to provide anonymous reward allocation, are the approaches in the state of the art that are most similar in intent to PAI with respect to privacy preservation. These approaches, unlike STOC-PISCES (Biswas et al., 2015) and SenseUtil (Tsujimori et al., 2014), seek to provide privacy preserving reward allocation. However, they do not propose a mechanism to compute such rewards.

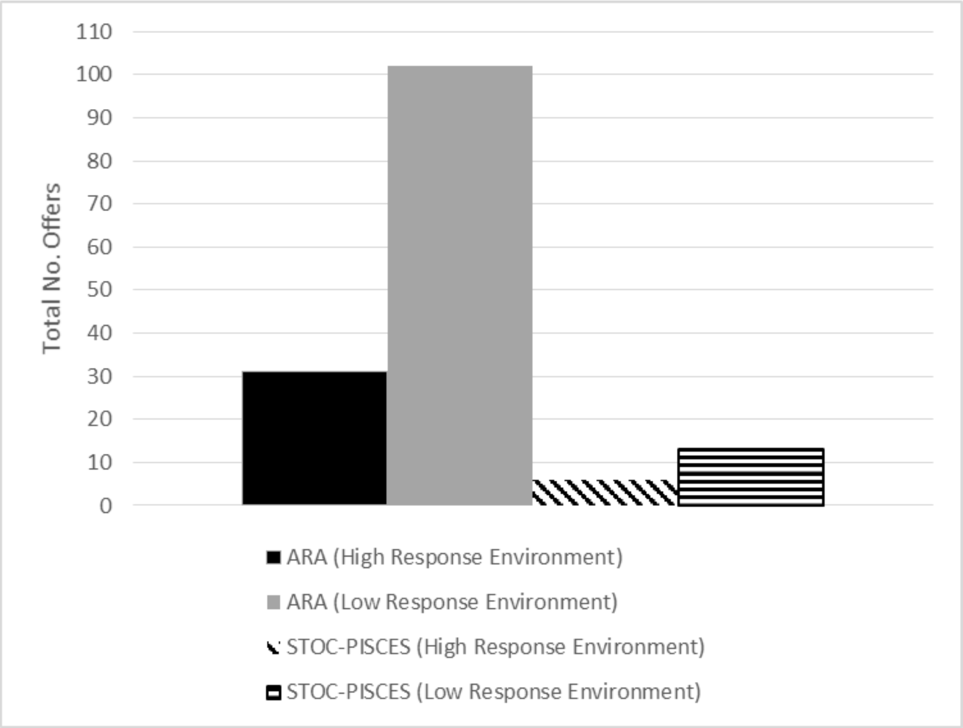

Figure 5.13: Total No. Offers

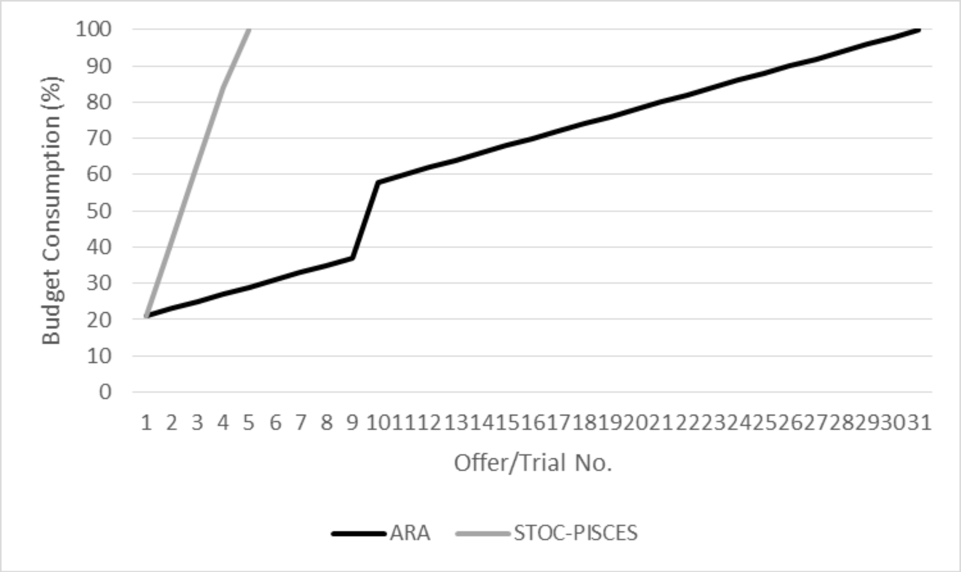

Figure 5.14: Rate of Budget Consumption (High Response Environment)

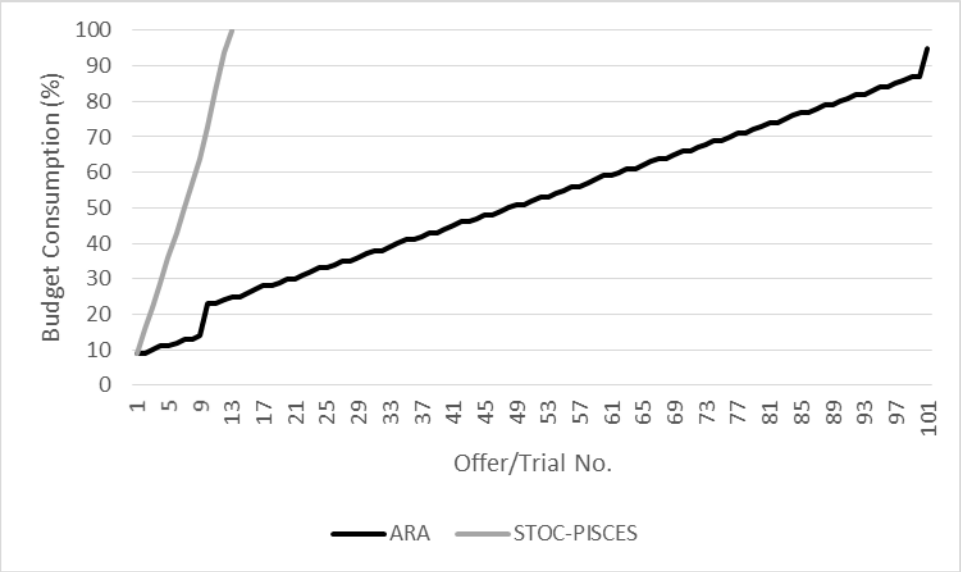

Figure 5.15: Rate of Budget Consumption (Low Response Environment)

While PAI seeks to progress beyond these approaches in providing untraceable and unlinkable reward allocation and spending, this should not come at an excessive performance cost. For this reason, these approaches will be used as the bases of comparison when evaluating PAI's performance. EPPI (Niu et al., 2014) is also used as a basis of comparison when evaluating the computational complexity of PAI. However, this approach is not considered when assessing energy consumption as the authors do not discuss the cryptographic algorithms used.

5.3.1 Cryptographic Primitives used for Privacy Preserving Requirements R1, R2 & R3

In order to evaluate the energy and resource consumption of PAI, the privacy preserving requirements *(R1, R2* and *R3)* have also been implemented for the Android mobile operating system using the Java programming language[13] with the cryptographic primitives for the participant and OrderBook peers being implemented using the SpongyCastle API[14]. The DSA[15] algorithm using the SHA-1 message digest algorithm is used to specify the digital signature. The peer's verification and decryption primitives are also implemented for the Windows 10 operating system using the Java Programming language and the BouncyCastle API[16]. The latter implementation is carried out as a peer may elect to support the OrderBook on fixed nodes such as a Laptop or PC Server rather than a mobile device. Energy consumption is not measured for this latter implementation as it tends not to be a critical concern for such devices.

Using these implementations, the running time and power consumption of the cryptographic primitives for the submitter and a typical peer are measured on a Samsung Galaxy S7 Edge Android Smartphone (Android 7.0, 4GB RAM, Quadcore 2.3 and Quadcore 1.6 GHz CPU) and, in the case of a peer using a fixed device, on an 8GB Lenovo T450s ThinkPad Laptop computer (8GB RAM, Intel Core i7-5600 2.6GHz CPU). The results of these experiments are presented in Table 5.2.

The cryptographic primitives used by the data submitter and the peer (both of which have to use resources when data submissions are made) in PAI's algorithms pertain to generation of the One-Time Key, the encryption of the data

[13] This is in addition to the C++ simulated environment discussed in Section 5.1.

[14] See `https://rtyley.github.io/spongycastle`

[15] DSA denotes digital signature algorithm.

[16] See `https://www.bouncycastle.org`

submission, the verification of the reward token by the peer when the user wants to spend the reward and the decryption of the spendable reward. The running time of the cryptographic primitives for PAI is evaluated by executing the associated algorithm over 100 times and computing the average time taken. The time taken for the submitter when generating the One-Time Key and encrypting the data submission is 4.12ms on average while the time taken for peer verification (when the user wants to spend the reward) and ID decryption operations is under 1ms. This compares favorably to participant resource consumption for the token-based approach used by Li and Cao [2016] which, on average, takes 12.5% longer. The overall time taken by the cryptographic primitives used by PAI is also substantially lower than that taken for participants in the approach taken by Dimitriou [2018b], whether the peer hosts the OrderBook on a Laptop or a SmartPhone. The time taken by this approach is 257.5 and 3.8 times more than PAI when the peer hosts the OrderBook on a Laptop and SmartPhone respectively.

However, the time taken for the verification of the reward token by the peer when it is being spent is more expensive than the approach taken by Li and Cao [2016] in terms of time (339ms) when the peer hosts the OrderBook on a SmartPhone. With the data encryption and ID decryption taking 0.120ms and 0.146ms on average respectively, the total running time for the cryptographic primitives of 343.266ms is 57 times that of the token-based approach taken by Li and Cao [2016]. The reward token verification is responsible for the majority of this cost which is nevertheless well under half a second, a figure that is substantially lower than the 2 seconds commonly cited as the upper limit users expect for response time (Nah, 2004 and Hong et al., 2018). This verification process is a core part of the mechanism to decouple the reward allocation and reward spending process and thus ensures that, unlike the approach used by Li and Cao [2016], the data submitter cannot be the victim of inference attacks. Moreover, this only applies in the case of SmartPhone hosting of peer operations. If the service provider wants to encourage hosting on fixed nodes, it has the option to offer higher reward levels for those peers who do so.

PAI's SmartPhone power consumption for the data submitter is 71% lower than the approach taken by Li and Cao [2016]. Crucially, it should be noted that the resource consumption totals of 6.004ms and 0.29J for the method used by Li and Cao [2016] pertain solely to the data submitter. In these terms, the resource consumption of 4.12ms and 0.025J is much less for the data submitter under PAI as the majority of the cost is borne by the peer. Energy consumption is also

	Time (ms)	Power (J)
Submitter (Android Phone)		
- One-Time Key	4.000	0.023
- Data Encryption	0.120	0.002
Peer (Laptop)		
- Verification	0.944	N/A
- Decryption	0.005	N/A
Peer (SmartPhone)		
- Verification	339.000	3.290
- Decryption	0.146	0.002
Total (Laptop)	5.069	0.025
Total (SmartPhone)	343.266	3.317
Total (Li and Cao, 2016, Laptop)	5.704	N/A
Total (Li and Cao, 2016, SmartPhone)	6.004	0.290
Total (Dimitriou, 2018b)	1305.500	12.612

Table 5.2: Resource Consumption of Cryptographic Primitives

substantially lower than the approach taken by Dimitriou [2018b], 3.8 and 504.48 times lower for a SmartPhone-hosted and Laptop-hosted peer respectively.

5.3.2 Computational Complexity

Computational complexity involves the study of the efficiency of algorithms based on the time and memory space required to solve a problem of a particular size (Rosen, 2007). Complexities are expressed using the Big O notation.

The majority of the computation for Algorithm 1 (Reward Allocation) is of the order $O(1)$ i.e. the cost of these operations are independent of the input. The generation of the digital signature by the service provider is dependent on the size of the service provider's private key, b^*_{SP}, used to sign the offered reward, r_{O}. Hence, using k to denote the size of b^*_{SP}, the complexity of this operation is $O(k)$. Similarly, the generation of r_{s} (the encrypted spendable reward) and $\{d\}_{\mathrm{b_{SP}}}$ (the encryption of the data submitted) is dependent on the size of the message block, m, to be encrypted so the complexity for these operations can be expressed as $O(m)$. It should be noted that, in both cases, neither k nor m would be of a significant size as they entail digital signing using the service provider's private key and the encryption of the spendable reward or data submission respectively.

The runtime of the majority of the operations for generating the One-Time-

Key is $O(1)$. The generation of the key pair used for the reward ID depends on the size of the keys. Assuming both keys are of the same size k, computational complexity can be expressed as $O(2k)$. The most expensive part of the operation is the use of modular exponentiation which entails the use of two digits of size n and an exponent of size k bits. A multiplication algorithm M is also used. The computational complexity of the modular exponentiation operation can be expressed as $O(M(n)k)$ (Knuth, 1997). As this is the most expensive part of the operation, the Big O notation for Algorithm 1 can thus be expressed as $O(M(n)k)$.

The most computationally expensive parts of Algorithm 2 (Reward Spending) pertain to the verification of r_{O}, K_{O} (the One-Time Key), the size of b_{SP} (the service provider's public key) and $a^{*}_{\mathrm{K_O}}$ (the private part of the One-Time Key) respectively. Assuming the size of b_{SP} and $a^{*}_{\mathrm{K_O}}$ are k_1 and k_2, computational complexity can be expressed as $O(k_1)$ and $O(k_2)$ respectively. Likewise, the computational complexity when decrypting the spendable reward, r_{s}, depends on the size of the message block m and so can be expressed as $O(m)$. As this is the most expensive part of the operation, the Big O notation for Algorithm 2 can thus be expressed as $O(m)$.

Algorithm 3, which is used to estimate data truthfulness, is potentially a computationally expensive algorithm as its computational complexity is dependent on the number of measurement categories $|c|$ and the number of items read from the dataset pertaining to this measurement category $|[d_c]|$. For a single category, therefore, the computational complexity can be expressed as $O(|[d_c]|)$. Assuming the total number of items read from the dataset is $|[d]|$, the overall computational complexity of the algorithm can be expressed as $O(|[d]|)$. The potential computational expense of Algorithm 3 is due to the need to read at least a subset of the dataset $[d_c]$ for each measurement category c so as to estimate the mean and standard deviation using the Maximum Likelihood Estimation (MLE) method. This expense can be offset by carrying out this computation periodically and by reducing the size of the subset to be read from the dataset.

The computational complexity of the core operations for Algorithm 4 (Reward Computation) depends on the range of rewards to be offered by the service provider i.e. from zero to the maximum reward level, r_{max}. Assuming the number of possible rewards the service provider could offer is $|r|$[17], the Big O notation for this aspect of Algorithm 4 can be expressed as $O(|r|)$. The most

[17] As outlined in Section 4.3.4, r is a discrete variable.

expensive part of Algorithm 4 pertains to the regression method used to construct the supply curves for each category of measurement, c. Assuming that the number of elements for a particular c is $|n_{\mathrm{c}}|$ and the number of coefficients for the regression equation is $|l|$, the computational complexity for the equation is $O(|l|^2 * (|n_c| + |l|))$. This is because the underlying operation of regression analysis entails matrix multiplication. Given that most service providers should have a relatively small number of predictors (for example, cost) to include as coefficients when using a regression model to predict the number of responses[18], $|l|$ should be a comparatively low number.

The computational complexity of PAI compares favourably to that of the approach taken by Li and Cao [2016] as the modular exponentiation scheme used for the RSA[19]-based blind signature algorithm used by this approach has computationally expensive operations pertaining to the public key, private key and key generation that can be expressed as $O(k^2)$, $O(k^3)$ and $O(k^4)$ respectively[20]. Thus, the Big O notation for the approach taken by Li and Cao [2016] can be expressed as $O(k^4)$ which is much larger than those for PAI's Algorithm 1 and Algorithm 2.

The computational complexity of EPPI (Niu et al., 2014) can be expressed in terms of its public key encryption and private key signature. Assuming the size of the E-Cent to be encrypted is s and that of the private key is k, the most expensive parts of the operation of this scheme can be expressed as $O(k)$ and $O(s)$ respectively. This compares favourably to both the approach proposed by Li and Cao [2016] and PAI. However, depending on the asymmetric encryption scheme used for EPPI (for example, RSA), computational complexity can be up to $O(k^4)$, which is much larger than the worst case for PAI's algorithms. Similarly, computational complexity of PAI is favourable when compared with the approach outlined by Dimitriou [2018b]. This approach relies upon zero knowledge proofs which can have computational complexity of $O(|x|n)$ assuming a problem instance x and error probability $2^{-\mathrm{n}}$ (Cramer and Damgard, 2009).

It should be noted that, while none of the approaches used for comparison have an algorithm that is potentially as computationally expensive as Algorithm 3, these approaches do not address the issue of incentive compatibility and data truthfulness evaluation.

[18] See Section 4.3.1.1 for considerations typically used as regression coefficients.

[19] As previously noted, RSA is named after its authors **R**ivest, **S**hamir and **A**dleman.

[20] See `http://www.rsasecurity.com`

5.4 Summary & Discussion of Results

This chapter evaluates how PAI meets its requirements. PAI's privacy preserving requirements for *Anonymous, Unlinkable and Protected Data Submission (R1)*, *Untraceable and Unlinkable Reward Allocation (R2)* and *Untraceable and Unlinkable Reward Spending (R3)* are evaluated by proof in Section 5.2 with theorems being presented to demonstrate that participants make unlinkable data submissions to the service provider anonymously *(R1)*, receive untraceable and unlinkable rewards *(R2)* and can spend these rewards in an untraceable and unlinkable manner *(R3)*. The requirement for *Incentive Compatibility (R4)* is also evaluated by proof with two theorems being presented to demonstrate that data submissions that are considered to be non-truthful do not receive a reward and that the incentive compatibility method is privacy preserving.

The requirement for *Adaptive and Tunable Reward Allocation (R5)* is evaluated by means of experiments conducted in a simulated participatory sensing environment. Using STOC-PISCES (Biswas et al., 2015) as a baseline for comparison, it is demonstrated in Section 5.2.5.2 that PAI's Adaptive Reward Allocation (ARA) component adapts more rapidly to conditions in the participatory sensing environment and offers a lower average reward in return for the same or better response rate. In addition, Section 5.2.5.2 also shows how ARA's Lyapunov Optimisation model can be tuned by the service provider to prioritise data capture over budget consumption and vice versa. The budget consumption of ARA is compared to STOC-PISCES and SenseUtil (Tsujimori et al. [2014]) in Section 5.2.5.3. This section illustrates how the average reward for ARA is lower than that computed by the other two approaches, even in a low response environment. The superior budget consumption of ARA is also demonstrated given its ability to generate a far larger number of responses in both a high response and low response participatory sensing environment.

The overall performance of PAI is evaluated in Section 5.3. Section 5.3.1 describes how the time taken and energy consumed by PAI's cryptographic primitives is superior to the results found for the approaches taken by Li and Cao [2016] and Dimitriou [2018b], which are used as the baselines for comparison. These two approaches are also used as bases for comparison when evaluating PAI's computational complexity in Section 5.3.2. In general, the computational complexity of the four algorithms designed for PAI compares favourably to the approaches taken by not only Li and Cao [2016] and Dimitriou [2018b] but also Niu et al. [2014].

Unlike other approaches, PAI offers incentive compatibility. This has consequences in terms of computational complexity as Algorithm 3 for estimating data truthfulness is potentially more expensive than any of the operations in these other approaches. However, it should be noted that, by reducing the frequency of its computation and the size of the dataset used, the computational complexity of this algorithm can be reduced.

Chapter 6

Conclusions & Future Work

This chapter concludes this thesis. The contributions made by this thesis are outlined in Section 6.1 while possible directions for future work are discussed in Section 6.2. Section 6.3 summarises this chapter.

6.1 Contributions

The motivation for the work described in this thesis arose from the need to provide privacy preserving rewards for participatory sensing tasks by making these rewards untraceable and unlinkable, thus preventing inference attacks. From the service provider's perspective, this privacy preservation needs to be done in a way that still enables the truthfulness of submitted data to be evaluated and computes the rewards to offer in a way that takes budget consumption and participation rates into account.

To address the problem of providing untraceable and unlinkable rewards, five key requirements have been identified in this thesis. The requirement for *Anonymous, Unlinkable and Protected Data Submission (R1)* specifies that any data submission made to a service provider can only be accessible by that party and cannot be used to identify a participant's identity, activities or behaviours. The vulnerability of participants to identification when receiving or spending rewards is addressed by the requirements for *Untraceable and Unlinkable Reward Allocation (R2)* and *Untraceable and Unlinkable Reward Spending (R3).* In addition, privacy preservation must be carried out without impinging upon the service provider's ability to attract data submissions of sufficient quality.

This is acknowledged by the requirements for *Incentive Compatibility (R4)* and *Adaptive and Tunable Reward Allocation (R5)*, which address the need for evaluating data truthfulness and computing rewards that reflect current environmental conditions and current participation rates respectively. A review of existing work has shown that none of these requirements have been fully met. Moreover, there is no existing approach in the state of the art that can be directly adapted to meet these requirements. For example, many of the methods for privacy preservation proposed in the state of the art would make it impossible to reward participants.

To meet these requirements, this thesis proposes *Privacy-Aware Incentivisation (PAI)*, a decentralised platform that allocates rewards to participants without compromising their identity privacy. The platform consists of three key components. Firstly, *Identity Privacy Preservation (IPPI)* is a privacy preserving mechanism that ensures data submission anonymity as well as reward allocation and spending untraceability and unlinkability. This is principally achieved through the use of a One-Time Key, a cryptographic public and private key pair that is generated by the participant. The public component of the One-Time Key serves as an identifier for any data submissions made by the participant and is used to assign rewards to that party. Identity privacy is ensured as this key is generated randomly, used only once and cannot be connected to the participant. Rewards can only be spent by the participant using the private component of the One-Time Key, solely held by that party. All transactions including offers made by the service provider and data submissions made by the participant are published on the OrderBook, a distributed ledger that is hosted by multiple peer devices, thus ensuring that there is no single point of failure or privacy vulnerability.

The needs of the service provider for incentive compatibility and an incentivisation scheme are served through the provision of the *Data Truthfulness Estimation (DTE)* and *Adaptive Reward Allocation (ARA)* components, based on the Maximum Likelihood and Lyapunov Optimisation statistical methods respectively. Four algorithms are proposed based on the design of these components. Specifically, these algorithms are for *Untraceable and Unlinkable Reward Allocation*, *Untraceable and Unlinkable Reward Spending*, *Data Truthfulness Estimation* and *Reward Computation*.

The evaluation of the approach is in two parts, proof and experiment. Proofs are presented to assess how the approach meets the privacy preserving requirements for *Anonymous, Unlinkable and Protected Data Submission (R1)*, *Un-*

traceable and Unlinkable Reward Allocation (R2) and *Untraceable and Unlinkable Reward Spending (R3)*. Similarly, the requirement for *Incentive Compatibility (R4)* is also evaluated by proof while experiments are used to evaluate the implementation of the requirement for *Adaptive and Tunable Reward Allocation (R5)* using a simulated participatory sensing environment written in the C++ and statistical R programming languages. The experiments carried out to evaluate the adaptiveness and budget consumption of ARA, the component implementing requirement *R5*, demonstrate a faster rate of adaptiveness in response to participation rates and better budget consumption compared with similar work in the state of the art. The overall performance of PAI is evaluated using an implementation written for the Android Operating System, with the performance of the cryptographic primitives used for the privacy preserving requirements and the computational complexity of the proposed algorithms both being favourable compared with other approaches in the state of the art that are similar in intent to PAI. To conclude, PAI not only progresses beyond the current state of the art in terms of providing privacy preserving incentivisation but also demonstrates a level of performance that is superior to existing approaches.

6.2 Possible Directions for Future Work

There are a number of possible directions that can be taken in further developing PAI. Currently, PAI is suitable for scalar data submissions only. As multimedia data submissions such as images or videos would be of potential interest for some participatory sensing campaigns, one possible avenue for future work is the extension of PAI to facilitate multimedia data submissions. The key challenge for such work would be ensuring that PAI's requirements are both met and adhered to for multimedia data. Specifically, the requirement for *Incentive Compatibility (R4)* would have to be expanded to incorporate a method for estimating data truthfulness for multimedia data submissions. Challenges here range from the detection of content manipulation to similarity comparisons of different multimedia content to evaluate their consistency. Similarly, the requirement for *Adaptive and Tunable Reward Allocation (R5)* would have to be extended to incorporate a method for categorising the different forms of multimedia content sought by a service provider. It should also be noted that while the underlying Maximum Likelihood Estimation method used for requirement *R4* demonstrates that PAI facilitates incentive compatibility in a privacy preserving manner, there is scope to enhance this method, for example, to better cater for censored data.

Further challenges to be addressed include the evaluation of alternative methods to the regression method used for predicting the response rate for a particular reward. While regression analysis is an established technique for forecasting, there are many other methods from the fields of statistics and Artificial Intelligence that could be used. For instance, Kalman Filtering (also known as Linear Quadratic Estimation) is used in many domains to make predictions using historical data. The evaluation of the effectiveness of these methods for predicting the response rate at different reward levels and their appropriateness for different participatory sensing environments and different categories of multimedia and scalar data would be an avenue of further research in determining whether different prediction methods would be appropriate for *Adaptive and Tunable Reward Allocation (R5)* in different types of participatory sensing environments. Other challenges to be addressed in this area include the predicting of response rates when the service provider has a small dataset and investigating whether different forecasting methods are appropriate at a particular point, for example, when a particular volume of data is held by the service provider.

This thesis has evaluated PAI in a simulated participatory sensing environment. The evaluation of the implementation in a real-world environment would be an interesting area of future work as this may identify participant behaviours and performance issues that would not be apparent in a simulation. For example, it could be the case that some types of requested data might attract a very low response rate no matter what level of reward is offered. The different means by which peer devices could be motivated to host the OrderBook could also be investigated. By extending PAI to address these challenges, it is possible that the approach could serve as the basis for a privacy preserving marketplace for participatory sensing and, potentially, other forms of crowdsourcing.

6.3 Summary

This chapter summarises the motivation for, and the significant contributions of, the work described in this thesis. In particular, the design, implementation and evaluation of the requirements identified to provide privacy preserving reward allocation and spending for participatory sensing are discussed in Section 6.1. Possible directions for future work, specifically, the facilitation of multimedia data submissions, further research into response rate prediction and the deployment of the approach in a real-world environment, are explored in Section 6.2.

Nomenclature

$[c]$ Set of measurement categories

$[d_{\mathrm{c}}]$ Dataset for a measurement category

$[N_{\mathrm{predict}}, r]$ Number of predicted responses for the different reward levels

$[r, N_{\mathrm{predict}}, Z_{\mathrm{forfeit}}(t)]$ Reward, number of predicted responses and queuing state variable for this reward

$[r]$ Set of possible reward values

$[U, V]$ A map of data utility weightings and the constant used for computing the Lyapunov drift

$\bar{Z}_{\mathrm{forfeit}_{\mathrm{k}}}$ Average queue backlog for the number of forfeited responses

β_0 Regression coefficient for the reward

β_1 Regression coefficient for the costs of making a data submission

β_2 Regression coefficient for the ratio of the number of responses sought to the current number of participants

β Vector of parameter coefficients

δ Data Type and Granularity and Offer Conditions

ϵ Positive number used in Equation 3 (which defines performance guarantees)

γ Positive number used in Theorem 2

∞ Infinity

μ_{e} Initial estimated mean (used for Maximum Likelihood Estimation method)

μ Estimated mean (used for Maximum Likelihood Estimation method)

σ Estimated standard deviation (used for Maximum Likelihood Estimation method)

σ_{e} Initial estimated standard deviation (used for Maximum Likelihood Estimation method)

σ_{scaled} Scaled mean (used for Maximum Likelihood Estimation method)

Θ Set of possible scenarios that can occur when a service provider makes an offer at a particular reward level

θ Information realisation (scenario)

$\triangle(t)$ One-Slot Conditional Lyapunov Drift

ε Error term used when formulating number of predicted responses

$\varPhi$ Used for controllable error bound in Theorem 1

$\{[N_{\mathrm{actual}}], r\}$ Actual responses for the different reward levels

$\{[r, N_{\mathrm{predict}}, Z_{\mathrm{forfeit}}(t)]\}$ Set of rewards, their respective queuing state variables and number of predicted responses

$\{d\}_{\mathrm{b}_{SP}}$ Data Submission encrypted using Service Provider's public key

$a_{\mathrm{K_O}}$ Public part of One-Time Key

$a^{*}_{\mathrm{K_O}}$ Private part of One-Time Key

A_{O} Offer Acceptance

$B(t)$ Budget consumption in a particular timeslot

$B^{\dagger}(t)$ Budget consumption for the online model in a particular timeslot (used for Theorem 1)

B^{*} Minimum budget consumption with stochastic future information

B^{o} Minimum budget consumption with complete future information

B^{*}_{av} Objective value of the time average maximisation problem under an optimal budget policy

B_{AV} Time average budget consumption

$B_{\mathrm{constant}}(t)$ Positive number used in the Lyapunov Optimisation computation

$B_{\mathrm{proportion}_{\mathrm{max}}}$ Maximum proportion of the budget that can be consumed for an offer in a particular timeslot

B_{remain} Remaining budget

b^*_{SP} Service provider's private key

b_{SP} Service provider's public key.

B Service provider's budget

C Cost parameter summarising the costs incurred by participants when making data submissions.

c Measurement category

DPP_{LHS} Left hand side of Drift Plus Penalty Expression

DPP_{RHS} Right hand side of Drift Plus Penalty Expression

d Data Submission

f_{σ} Scaling factor (used for Maximum Likelihood Estimation method)

f Denotes a function

i_{O} Offer Token ID

i_{SP} Service Provider ID

$i_{\mathrm{A_O}}$ Offer Acceptance ID

i_{S} Spendable reward ID

$i_{\mathrm{V},}$ Validation Token ID

K_{O} One-Time Key

$L(t)$ Lyapunov Function

$L(t+1)$ Lyapunov functon in the next timeslot

$L_{\mathrm{last}}(r)$ Last Lyapunov function calculated for a particular reward

l_{max} Maximum scaled limit (used for Maximum Likelihood Estimation method)

l_{min} Minimum scaled limit (used for Maximum Likelihood Estimation method)

L_{O} Offer Listing

$m_{\mathrm{c}_{max}}$ Maximum threshold value for a measurement category

$m_{\mathrm{c}_{min}}$ Minimum threshold value for a measurement category

m_{c} Value for a measurement category

M_{predict} Linear regression prediction model

$N(\theta)$ Number of responses under a particular information realisation/scenario

$N_{\mathrm{desired}_{\mathrm{max}}}$ Maximum for desired number of responses

$N_{\mathrm{desired}}(t)$ Number of desired responses for a particular timeslot

$N_{\mathrm{desired}}(t+1)$ Number of desired responses in the next timeslot

$N_{\mathrm{desired}}(t-1)$ Number of desired responses in the previous timeslot

N_{max} Number of responses received for the maximum reward level

N_{min} Number of responses received for the mnimum reward level

$N_{\mathrm{O}}(t)$ Number of responses to an offer in a particular timeslot

N_{predict} Predicted number of responses

$N_{\mathrm{received}}(t)$ Number of responses received in a particular timeslot

$N_{\mathrm{received}}(t-1)$ Number of responses received in the previous timeslot

n Number of measurement readings

OPT_{current} Current Lyapunov Optimisation calculation

OPT_{solution} Lyapunov Optimisation solution

O Offer

$P(t)$ Number of active participants in a particular timeslot

$p(t)$ Penalty used for achieving queue stabillity in Lyapunov Optimisation

P_{max} Maximum number of participants potentially active on the system

P_{ratio} Ratio of the number of responses sought to the current number of participants

P Number of current active participants

$r(\theta)$ Reward level under a particular information realisation/scenario

r_{O} Reward amount offered

r_{max} Maximum reward level

r_{min} Minimum reward level

$r_{\text{O}}(t)$ Reward set for an offer in a particular timeslot

$r_{\text{P}}(t)$ Reward level in a particular timeslot when the number of responses equals the number of participants

r_{s} Encrypted spendable reward

T_{O} Offer Token.

T_{V} Validation Token

t_n Final timeslot

T Set of all timeslots

t Timeslot

U Utility weighting for data submission

V Non-negative control parameter used to incorporate the weighted budget terms in the Lyapunov Optimisation control decision

v Validation Flag

X Set of predictors used to predict the number of responses

$Z_{\text{forfeit}}(t)$ Number of forfeited responses in a particular timeslot

$Z_{\text{forfeit}}(t-1)$ Number of forfeited repsonses in the previous timeslot

Z_{forfeit} Virtual queue representing the number of forfeited responses

Bibliography

I. Agadakos, J. Polakis, and G. Portokalidis. Techu: Open and privacy-preserving crowdsourced gps for the masses. In *In Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*, pages pp. 475–487. ACM, June 2017.

B. Agir, T.G. Papaioannou, R. Narendula, K. Aberer, and J-P. Hubaux. User-side adaptive protection of location privacy in participatory sensing. *GeoInformatica*, 18(1):165–191, January 2014.

H. Almuhimedi, F. Schaub, N. Sadeh, I. Adjerid, A. Acquisti, J. Gluck, L.F. Cranor, and Y. Agarwal. Your location has been shared 5,398 times!: A field study on mobile app privacy nudging. *Proceedings of the 33rd annual ACM conference on human factors in computing systems, ACM*, pages pp. 787–796, April 2015.

M.A. Alsheikh, D. Niyato, D. Leong, P. Wang, and Z. Han. Privacy management and optimal pricing in people-centric sensing. *IEEE Journal on Selected Areas in Communications.*, 2017.

M.A. Alswailim, M. Zulkernine, and H.S. Hassanein. Classification of participatory sensing privacy schemes. In *IEEE 39th Conference on Local Computer Networks Workshops (LCN Workshops)*, pages pp. 761–767. IEEE, September 2014.

M.A. Alswailim, H.S. Hassanein, and M. Zulkernine. A reputation system to evaluate participants for participatory sensing. In *IEEE Global Communications Conference (GLOBECOM)*, pages pp.1–6, December 2016.

H. Amintoosi, Kanhere S.S., and M. Allahbakhsh. Trust-based privacy-aware participant selection in social participatory sensing. *Journal of Information Security and Applications*, pages 1–15, 2014.

H. Amintoosi, S.S. Kanhere, and M. N. Torshiz. A socially-aware incentive scheme for social participatory sensing. In *IEEE Tenth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*, pages 1 – 6, 7-9 April 2015.

S. Aoki, M. Iwaiy, and K. Sezakiyz. Limited negative surveys: Privacy-preserving participatory sensing. In *IEEE 1st International Conference on Cloud Networking (CLOUDNET)*, pages 158 – 160, 28-30 November 2012.

Y. Arakawa and Y. Matsuda. Gamification mechanism for enhancing a participatory urban sensing: Survey and practical results. *Journal of Information Processing*, 24(1):pp.31–38, 2016.

R. Azzam, R. Mizouni, H. Otrok, S. Singh, and A. Ouali. A stability-based group recruitment system for continuous mobile crowd sensing. *Computer Communications.*, 2018.

D. Back, B.J. Choi, and J. Chen. Small profits and quick returns: A practical social welfare maximizing incentive mechanism for deadline-sensitive tasks in crowdsourcing. arXiv preprint arXiv:1707.00018., 2017.

C. Bettini and D. Riboni. Privacy protection in pervasive systems: State of the art and technical challenges. *Pervasive and Mobile Computing*, pages 159–174, 2015.

S. Bhattacharjee, N. Ghosh, V.K. Shah, and S.K. Das. W2q: A dual weighted qoi scoring mechanism in social sensing using community confidence. In *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, pages pp. 375–380. IEEE, March 2017.

A.. Biswas, D. Chander, K. Dasgupta, K. Mukherjee, M. Singh, and T. Mukherjee. Pisces: participatory incentive strategies for effective community engagement in smart cities. In *Third AAAI Conference on Human Computation and Crowdsourcing.*, September 2015.

D. Biswas and K. Vidyasankar. Privacy preserving profiling for mobile services. In *The 9th International Conference on Mobile Web Information Systems (MobiWIS)*, pages 569–576, 2012.

L. Bottou. Stochastic gradient descent tricks. In *Neural networks: Tricks of the trade*, pages (pp. 421–436). Springer, Berlin, Heidelberg., 2012.

I. Boutsis and V. Kalogeraki. Privacy preservation for participatory sensing data. In *IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pages 103 – 113, 18-22 March 2013.

J. L. Boyles, A. Smith, and M. Madden. Privacy and data management on mobile devices. Technical report, Pew Research Center Internet And American Life Project., 2012.

P.B. Brandtzaeg, A. Pultier, and G.M. Moen. Losing control to data-hungry apps: A mixed-methods approach to mobile app privacy. *Social Science Computer Review*, 2018.

E.R. Brubaker. On the auction mechanism and its incentive compatibility. *Journal of Political Economy*, 88(3):pp.617–619, 1980.

J.A. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and M.B. Srivastava. Participatory sensing. 2006.

H. Cai, Y. Zhu, Z. Feng, H. Zhu, J. Yu, and J. Cao. Truthful incentive mechanisms for mobile crowd sensing with dynamic smartphones. *Computer Networks*, 2018.

S. Chakraborty, Z. Charbiwala, H. Choi, K. Rangan Raghavan, and M.B. Srivastava. Balancing behavioral privacy and information utility in sensory data flows. *Pervasive and Mobile Computing*, 8:331–345, 2012.

S-H. Chang, Y-S. Chen, and S-M Cheng. Detection of sybil attacks in participatory sensing using cloud based trust management system. In *International Symposium on Wireless and Pervasive Computing (ISWPC)*, pages 1 – 6, 20-22 November 2013.

J. Chen and H. Ma. Privacy-preserving aggregation for participatory sensing with efficient group management. In *IEEE Global Communications Conference (GLOBECOM)*, pages 2757–2762, 8-12 December 2014.

J. Chen, H. Ma, D. Zhao, and D.S. Wei. Participant density-independent location privacy protection for data aggregation in mobile crowd-sensing. *Wireless Personal Communications*, 98(1):pp.699–723, 2018.

X. Chen, X. Wu, X-Y. Li, Y. He, and Y. Liu. Privacy-preserving high-quality map generation with participatory sensing. In *IEEE INFOCOM 2014*, pages 2310–2318, 2014.

Y. Chen, H. Chen, S. Yang, X. Gao, and F. Wu. Jump-start crowdsensing: A three-layer incentive framework for mobile crowdsensing. In *IEEE/ACM 25th International Symposium On Quality of Service (IWQoS)*, pages pp.1–6, June 2017a.

Z. Chen, Y. Lin, X. Feng, H. Zheng, and Y. Xu. Incentive mechanism for participatory sensing: A contract-based approach. In *IEEE Congress On Evolutionary Computation (CEC)*, pages pp.325–332, June 2017b.

L. Cheng, J. Niu, L. Kong, C. Luo, Y. Gu, W. He, and S.K. Das. Compressive sensing based data quality improvement for crowd-sensing applications. *Journal of Network and Computer Applications*, (77):pp.123–134, 2017.

Z. Chi, Y Wang, Y. Huang, and X. Tong. The novel location privacy-preserving ckd for mobile crowdsourcing systems. *IEEE Access*, 6:pp.5678–5687, 2018.

D. Christin. Privacy in mobile participatory sensing: Current trends and future challenges. *Journal of Systems and Software*, March 2015.

D. Christin, C. Buchner, and N. Leibecke. What's the value of your privacy? exploring factors that influence privacy-sensitive contributions to participatory sensing applications. In *IEEE 38th Conference on Local Computer Networks Workshops (LCN Workshops)*, 2013a.

D. Christin, C. Rosskop, M. Hollick, L.A. Martucci, and S.S. Kanherec. Incognisense: An anonymity-preserving reputation framework for participatory sensing applications. *Pervasive and Mobile Computing*, 9:353–371, 2013b.

D. Christin, D.R. Pons-Sorolla, M Hollick, and S.S. Kanhere. Trustmeter: A trust assessment scheme for collaborative privacy mechanisms in participatory sensing applications. In *IEEE Ninth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*, pages 1–6, 21-23 April 2014.

L. Cilliers and S. Flowerday. Information security in a public safety, participatory crowdsourcing smart city project. In *2014 World Congress on Internet Security (WorldCIS)*, 2014.

A. Clarke and R. Steele. Local processing to achieve anonymity in a participatory health e-research system. *Procedia - Social and Behavioral Sciences*, 147:284–292, 25 August 2014a.

A. Clarke and R. Steele. Targeted and anonymized smartphone-based public health interventions in a participatory sensing system. In *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 3678–3682, 26-30 August 2014b.

S. Coulson, M. Woods, M. Scott, D. Hemment, and M. Balestrini. Stop the noise! enhancing meaningfulness in participatory sensing with community level indicators. In *Proceedings of the Designing Interactive Systems Conference*, pages pp. 1183–1192. ACM, June 2018.

R Cramer. Introduction to secure computation. In *School organized by the European Educational Forum*, pages pp. 16–62. Springer, Berlin, Heidelberg, June 1998.

R. Cramer and I. Damgard. On the amortized complexity of zero-knowledge protocols. In *Advances in Cryptology-CRYPTO*, pages 177–191. Springer, Berlin, Heidelberg., 2009.

W. Dai, Y. Wang, Q Jin, and J. Ma. Geo-qti: A quality aware truthful incentive mechanism for cyber-physical enabled geographic crowdsensing. *Future Generation Computer Systems*, (79):pp.447–459, 2018.

E. De Cristofaro and C. Soriente. Extended capabilities for a privacy-enhanced participatory sensing infrastructure (pepsi). *IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY*, 8(12):2021–2033, December 2013.

W. Diffie and M. Hellman. New directions in cryptography. *IEEE transactions on Information Theory*, 22(6):644–654, 1976.

T. Dimitriou. Privacy-respecting rewards for participatory sensing applications. In *Wireless Communications and Networking Conference (WCNC)*, pages pp. 1–6. IEEE, April 2018a.

T. Dimitriou. Privacy-respecting reward generation and accumulation for participatory sensing applications. *Pervasive and Mobile Computing*, (49):139–152, 2018b.

T. Dimitriou and I. Krontiris. Privacy-respecting auctions and rewarding mechanisms in mobile crowd-sensing applications. *Journal of Network and Computer Applications*, (100):pp.24–34, 2017.

G. Drosatos, P.S. Efraimidis, I.N. Athanasiadis, E. D'Hondt, and M. Stevens. A privacy-preserving cloud computing system for creating participatory noise maps. In *IEEE 36th Annual Computer Software and Applications Conference (COMPSAC)*, pages 581 – 586, 16-20 July 2012 2012.

G. Drosatos, P.S. Efraimidis, I.N. Athanasiadis, M. Stevens, and E. D'Hondt. Privacy-preserving computation of participatory noise maps in the cloud. *The Journal of Systems and Software*, pages 170–183, 2014.

Z. Duan, M. Yan, Z. Cai, X. Wang, M. Han, and Y. Li. Truthful incentive mechanisms for social cost minimization in mobile crowdsourcing systems. *Sensors*, 16(4):p.481., 2016.

D. Easley and J. Kleinberg. *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*. Cambridge University Press, 2010.

P. Eisenhauer, J.J. Heckman, and S. Mosso. Estimation of dynamic discrete choice models by maximum likelihood and the simulated method of moments. *International Economic Review*, 56(2):331–357, May 2015.

M.T. Emmerich and A.H. Deutz. A tutorial on multiobjective optimization: fundamentals and evolutionary methods. *Natural computing*, 17(3):pp.585–609, 2018.

S.M. Erfani, S. Karunasekera, C. Leckie, and U. Parampalli. Privacy-preserving data aggregation in participatory sensing networks. In *IEEE Eighth International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, pages 165–170, 2-5 April 2013.

F. Esponda. Negative surveys,. *ArXiv Mathematics e-Prints arXiv:math/0608176.*, 2006.

F. Farokhi, I. Shames, and M. Cantoni. Promoting truthful behavior in participatory-sensing mechanisms. *IEEE Signal Processing Letters*, 22(10): 1538 – 1542, October 2015.

W. Feng, Z. Yan, H. Zhang, K. Zeng, Y. Xiao, and Y.T. Hou. A survey on security, privacy and trust in mobile crowdsourcing. *IEEE Internet of Things Journal*, (770), 2017.

Z. Feng, Y. Zhu, Q. Zhang, L.M. Ni, and A.V. Vasilakos. Trac: Truthful auction for location-aware collaborative sensing in mobile crowdsourcing. In *Proceedings of IEEE INFOCOM*, pages pp. 1231–1239. IEEE, April 2014a.

Z. Feng, Y. Zhu, Q. Zhang, H. Zhu, J. Yu, J. Cao, and L.M. Ni. Towards truthful mechanisms for mobile crowdsourcing with dynamic smartphones. In *IEEE 34th International Conference on Distributed Computing Systems (ICDCS)*, pages 11 – 20, June 30 - July 3 2014b.

L. Gao, F. Hou, and J. Huang. Providing long-term participation incentive in participatory sensing. In *IEEE International Conference on Computer Communications (INFOCOM)*, 2015a.

S. Gao, J. Ma, W. Shi, G. Zhan, and C. Sun. Trpf: A trajectory privacy-preserving framework for participatory sensing. *IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY*, 8(6):874–887, June 2013.

S. Gao, J. Ma, W. Shi, and G. Zhan. Ltppm: a location and trajectory privacy protection mechanism in participatory sensing. *WIRELESS COMMUNICATIONS AND MOBILE COMPUTING*, 15(1):155–169, January 2015b.

Y. Gao, X. Li, J. Li, and Y. Gao. Dtrf: A dynamic-trust-based recruitment framework for mobile crowd sensing system. In *IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*, pages pp. 632–635. IEEE, May 2017.

M. Girolami, S. Chessa, M. Dragone, M. Bouroche, and V. Cahill. Using spatial interpolation in the design of a coverage metric for mobile crowdsensing systems. In *IEEE Symposium on Computers and Communication (ISCC)*, pages pp. 147–152. IEEE, June 2016.

M. Girolami, S. Chessa, G. Adami, M. Dragone, and L. Foschini. Sensing interpolation strategies for a mobile crowdsensing platform. In *5th IEEE International Conference on Mobile Cloud Computing, Services, and Engineering (MobileCloud)*, April 2017.

S. Gisdakis, T. Giannetsos, and P. Papadimitratos. Sppear: Security & privacy-preserving architecture for participatory-sensing applications. In *Proceedings of the 2014 ACM conference on Security and privacy in wireless & mobile networks (WiSec '14)*, pages 39–50, 2014.

S. Gisdakis, T. Giannetsos, and P. Papadimitratos. Shield: A data verification framework for participatory sensing systems. In *Proceedings of the 8th ACM Conference on Security & Privacy in Wireless and Mobile Networks*, page pp. 16. ACM, June 2015.

F. Günther, M. Manulis, and A. Peter. Privacy-enhanced participatory sensing with collusion resistance and data aggregation. In *International Conference on Cryptology and Network Security (CANS)*, pages 321–336, October 22-24 2014.

X. Gong and N. Shroff. Incentivizing truthful data quality for quality-aware mobile data crowdsourcing. In *Proceedings of the Eighteenth ACM International Symposium on Mobile Ad Hoc Networking and Computing*, pages pp. 161–170. ACM, June 2018.

X. Gong and N.B. Shroff. Truthful mobile crowdsensing for strategic users with private qualities. In *International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt).*, May 2017.

M.M. Groat, B. Edwards, J. Horey, W. He, and S. Forrest. Application and analysis of multidimensional negative surveys in participatory sensing applications. *Pervasive and Mobile Computing*, 9:372–391, 2013.

T. Groves and J. Ledyard. Incentive compatibility since 1972. *Information, incentives, and economic mechanisms: Essays in honor of Leonid Hurwicz*, pages pp.48–111, 1987.

B. Guo, H. Chen, Z. Yu, W. Nan, X. Xie, D. Zhang, and X. Zhou. Taskme: toward a dynamic and quality-enhanced incentive mechanism for mobile crowd sensing. *International Journal of Human-Computer Studies*, (102):pp.14–26, 2017.

J. Guo, I.R. Chen, J.J. Tsai, and H. Al-Hamadi. Trust-based iot participatory sensing for hazard detection and response. In *In Proceedings of 1st Int. Workshop IoT Systems Provisioning and Management of Cloud Computing*, pages pp.1–6, October 2016.

N. Haderer, V. Primault, P. Raveneau, C. Ribeiro, R. Rouvoy, and S. Ben Mokhtar. Towards a practical deployment of privacy-preserving crowdsensing tasks. In *Proceedings of the Posters & Demos Session, ACM.*, pages pp. 43–44, December 2014.

I. Haider, M. Hoberl, and B. Rinner. Trusted sensors for participatory sensing and iot applications based on physically unclonable functions. In *Proceedings of the 2nd ACM International Workshop on IoT Privacy, Trust, and Security*, pages pp. 14–21. ACM, May 2016.

K. Han, H. Liu, S. Tang, M. Xiao, and J. Luo. Differentially private mechanisms for budget limited mobile crowdsourcing. *IEEE Transactions on Mobile Computing*, 2018.

Y. Han and Y. Zhu. Profit-maximizing stochastic control for mobile crowd sensing platforms. In *11th International Conference on Mobile Ad Hoc and Sensor Systems (MASS)*, pages pp. 145–153. IEEE, October 2014.

Y. Han, Y. Zhu, and J. Yu. A distributed utility-maximizing algorithm for data collection in mobile crowd sensing. In *Proceedings of IEEE GLOBECOM*, 2014.

S. He, D.H. Shin, J. Zhang, C.H.E.N. Jiming, and P. Lin. An exchange market approach to mobile crowdsensing: Pricing, task allocation and walrasian equilibrium. *IEEE Journal on Selected Areas in Communications*, 2017.

G. Hileman and M. Rauchs. 2017. global cryptocurrency benchmarking study. Technical Report 33, Cambridge Centre for Alternative Finance, 2017.

R. Carter Hill. *Principles of Econometrics.* Wiley, 2011.

B.O. Szymanski B. K. Holzbauer and E. Bulut. Socially-aware market mechanism for participatory sensing. In *Proceedings of the first ACM international workshop on Mission-oriented wireless sensor networking (MiSeNet '12)*, pages 9 – 14, 2012.

S. Hong, K. Kim, and T. Kim. A prefetching scheme for improving the web page loading time with nvram. *Journal of Semiconductor Technology and Science*, 18(1):pp.20–28, 2018.

J. Hu, H. Lin, X. Guo, and J. Yang. Dtcs: An integrated strategy for enhancing data trustworthiness in mobile crowdsourcing. *IEEE Internet of Things Journal.*, 2018.

C. Huang, L. Sankar, and A.D. Sarwate. Designing incentive schemes for privacy-sensitive users. *Journal of Privacy and Confidentiality*, 7(1), March 2016. URL `http://repository.cmu.edu/jpc/vol7/iss1/5/`.

K.L. Huang, Kanhere S.S., and W. Hu. A privacy-preserving reputation system for participatory sensing. In *37th Annual IEEE Conference on Local Computer Networks*, pages 10–18, 2012.

L. Hurwicz. The design of mechanisms for resource allocation. *The American Economic Review*, 63(2):pp.1–30, 1973.

L. Hutton, T. Henderson, and A. Kapadia. "here i am, now pay me!": privacy concerns in incentivised location-sharing systems. In *Proceedings of the 7th ACM Conference on Security and Privacy in Wireless and Mobile Networks (WiSec)*, 2014.

L. G. Jaimes, I.J. Vergara-Laurens, and A. Raij. A survey of incentive techniques for mobile crowd sensing. *IEEE Internet of Things Journal*, 2(5):370 – 380, October 2015a.

L.G. Jaimes and J.M. Calderon. Gaussian mixture model for crowdsensing incentivization. In *IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*, pages pp. 685–689. IEEE, January 2018.

L.G. Jaimes, A. Chakeri, J. Lopez, and A. Raij. A cooperative incentive mechanism for recurrent crowd sensing. In *Proceedings of the IEEE SoutheastCon 2015*, pages 1 – 5, 9 - 12 April 2015b.

R.B. Jain and R.Y. Wang. Limitations of maximum likelihood estimation procedures when a majority of the observations are below the limit of detection. *Analytical chemistry*, 80(12):pp.4767–4772, 2008.

X. Ji, D. Zhao, H. Yang, and L. Liu. Exploring diversified incentive strategies for long-term participatory sensing data collections. In *3rd International Conference on Big Data Computing and Communications (BIGCOM)*, pages pp. 15–22. IEEE, August 2017.

Q. Jiang, A.K. Bregt, and L. Kooistra. Formal and informal environmental sensing data and integration potential: Perceptions of citizens and experts. *Science of the Total Environment*, (619):pp.1133–1142, 2018.

H. Jin, L. Su, D. Chen, K. Nahrstedt, and J. Xu. Quality of information aware incentive mechanisms for mobile crowd sensing systems. In *Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, pages pp. 167–176. ACM, June 2015.

H. Jin, L. Su, B. Ding, K. Nahrstedt, and N. Borisov. Enabling privacy-preserving incentives for mobile crowd sensing systems. In *IEEE 36th International Conference on Distributed Computing Systems (ICDCS)*, pages pp. 344–353, 2016a.

H. Jin, L. Su, H. Xiao, and K. Nahrstedt. Inception: incentivizing privacy-preserving data aggregation for mobile crowd sensing systems. *MobiHoc*, pages pp. 341–350, July 2016b.

H. Jin, L. Su, and K. Nahrstedt. Theseus: Incentivizing truth discovery in mobile crowd sensing systems. In *Proceedings of the 18th ACM International Symposium on Mobile Ad Hoc Networking and Computing*. ACM, July 2017a.

H. Jin, L. Su, and K. Nahrstedt. Centurion: Incentivizing multi-requester mobile crowd sensing. In *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*. IEEE, May 2017b.

X. Jin and Y. Zhang. Privacy-preserving crowdsourced spectrum sensing. *IEEE/ACM Transactions on Networking*, 2018.

R. Johnson and T. Zhang. Accelerating stochastic gradient descent using predictive variance reduction. In *Twenty-seventh Annual Conference on Neural Information Processing Systems (NIPS)*, 5-10 December 2013.

S. Kalish and P. Nelson. A comparison of ranking, rating and reservation price measurement in conjoint analysis. *Marketing Letters*, 2(4):pp.327–335., 1991.

D. Kalui, X. Guo, D. Zhang, Y. Xie, and X. Zhang. Trust assurance privacy preserving framework for moving objects in participatory sensing. In *IEEE International Conference on In Big Data Analysis (ICBDA), , Vancouver*, pages pp. 1–6, March 2016.

M. Karaliopoulos, I. Koutsopoulos, and M. Titsias. First learn then earn: Optimizing mobile crowdsensing campaigns through data-driven user profiling. In *Proceedings of the 17th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, pages pp. 271–280. ACM, July 2016.

M. Katsomallos, S. Lalis, T. Papaioannou, and G. Theodorakopoulos. An open framework for flexible plug-in privacy mechanisms in crowdsensing applications. In *IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, pages pp. 237–242, March 2017.

L. Kazemi and C. Shahabi. Tapas: Trustworthy privacy-aware participatory sensing. *Knowledge and Information Systems*, 37(1):105–128, October 2013.

N.M. Khoi, S. Casteleyn, M.M. Moradi, and E. Pebesma. Do monetary incentives influence users' behavior in participatory sensing?. *Sensors (Basel, Switzerland)*, 18(5), 2018.

M. Kim, J. Lim, H. Yu, K. Kim, Y. Kim, and S.B. Lee. Viewmap: Sharing private in-vehicle dashcam videos. In *14th USENIX Symposium on Networked Systems Design and Implementation (NSDI 17)*, pages pp. 163–176. USENIX Association, March 2017.

D. Knuth. *The Art of Computer Programming*, volume Volume 2. Addison-Wesley, third edition edition, 1997.

J.Y. Koh, G. Peters, I. Nevat, and D. Leong. Spatial stackelberg incentive mechanism for privacy-aware mobile crowd sensing. *papers.ssrn.com*, 2018.

L. Kong, B. Wang, and G. Chen. *When Compressive Sensing Meets Mobile Crowdsensing*. Springer, 2019.

O. Kounadi and B. Resch. A geoprivacy by design guideline for research campaigns that use participatory sensing data. *Journal of Empirical Research on Human Research Ethics*, 13(3):pp.203–222., 2018.

I. Koutsopoulos. Optimal incentive-driven design of participatory sensing systems. In *Proceedings of IEEE INFOCOM*, pages 1402 – 1410, 14-19 April 2013.

I. Krontiris and T. Dimitriou. A platform for privacy protection of data requesters and data providers in mobile sensing. *Computer Communications*, 2015.

M. Kumar and S.I. Feldman. Internet auctions. In *USENIX Workshop on Electronic Commerce*, volume 3, pages pp. 49–60, August 1998.

T. Kumrai, K. Ota, M. Dong, and P. Champrasert. An incentive-based evolutionary algorithm for participatory sensing. In *IEEE Global Communications Conference (GLOBECOM)*, pages 5021 – 5025, 8-12 December 2014.

X. Labandeira, J.M. Labeaga, and X. Lopez-Otero. A meta-analysis on the price elasticity of energy demand. *Energy Policy*, 102:pp.549–568, 2017.

K. Lakshmi, J. Hemalatha, and F. Basha. K-anonymous privacy preserving technique for participatory sensing with multimedia data over cloud computing. *International Journal*, 4(2):pp.48–52., 2017.

J.O. Ledyard. Incentive compatibility and incomplete information. 1977.

C.Y. Lee and H. Heo. Estimating willingness to pay for renewable energy in south korea using the contingent valuation method. *Energy Policy*, .94: pp.150–156., 2016.

H. Li, K. Jia, H. Yang, D. Liu, and L. Zhou. Practical blacklist-based anonymous authentication scheme for mobile crowd sensing. *Peer-to-Peer Networking and Applications*, 2015a.

J. Li, Z. Cai, J. Wang, M. Han, and Y. Li. Truthful incentive mechanisms for geographical position conflicting mobile crowdsensing systems. *IEEE Transactions on Computational Social Systems.*, 2018.

J. Li, Y. Zhu, and J. Yu. Redundancy-aware and budget-feasible incentive mechanism in crowd sensing. *The Computer Journal*, 2019a.

M. Li, F. Wu, G. Chen, L. Zhu, and Z. Zhang. How to protect query and report privacy without sacrificing service quality in participatory sensing. In *IEEE 34th International Performance In Computing and Communications Conference (IPCCC)*, pages pp.1–7, December 2015b.

M. Li, L. Zhu, Z. Zhang, and R. Xu. Achieving differential privacy of trajectory data publishing in participatory sensing. *Information Sciences*, (400):pp. 1–13, 2017a.

Q. Li and G. Cao. Efficient privacy-preserving stream aggregation in mobile sensing with low aggregation error. *Lecture Notes in Computer Science: Privacy Enhancing Technologies*, 7981:60–81, 2013a.

Q. Li and G. Cao. Providing privacy-aware incentives for mobile sensing. In *IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pages pp. 76–84. IEEE, March 2013b.

Q. Li and G. Cao. Privacy-preserving participatory sensing. *IEEE Communications Magazine*, 53(8):68 – 74, August 2015.

Q. Li and G. Cao. Providing privacy-aware incentives in mobile sensing systems. *IEEE Transactions on Mobile Computing*, 15(6):pp.1485–1498, 2016.

T. Li, T. Jung, H. Li, L. Cao, W. Wang, X.Y. Li, and Y. Wang. Scalable privacy-preserving participant selection in mobile crowd sensing. In *IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pages pp. 59–68, March 2017b.

W. Li, Q. Duan, A. Ye, and C. Miao. An improved meta-gaussian distribution model for post-processing of precipitation forecasts by censored maximum likelihood estimation. *Journal of Hydrology*, (574):pp.801–810, 2019b.

X. Li and Q. Zhu. Social incentive mechanism based multi-user sensing time optimization in co-operative spectrum sensing with mobile crowd sensing. *Sensors*, 18(1):p. 250, 2018.

Y. Li, Y. Zhao, S. Ishak, H. Song, N. Wang, and N. Yao. An anonymous data reporting strategy with ensuring incentives for mobile crowd-sensing. *Journal of Ambient Intelligence and Humanized Computing*, pages 1–15, 2017c.

K. Lim and I.M. Abumuhfouz. Stors: secure token reward system for vehicular clouds. In *SoutheastCon 2015*, pages pp. 1–2. IEEE, April 2015.

B. Liu, Y. Jiang, F. Sha, and R. Govindan. Cloud-enabled privacy-preserving collaborative learning for mobile sensing. In *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems*, pages 57–70, 2012.

B. Liu, B. Zhao, B. Liu, and C. Wu. An attribute based private data sharing scheme for people-centric sensing networks. *Lecture Notes in Computer Science: Security Engineering and Intelligence Informatics*, 8128:393–407, 2013.

C. Liu, S. Chakraborty, and P. Mittal. Dependence makes you vulnberable: Differential privacy under dependent tuples. *NDSS*, 16:pp. 21–24, February 2016a.

C. H. Liu, P. Hui, J.W. Branch, C. Bisdikian, and B. Yang. Efficient network management for context-aware participatory sensing. In *8th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON)*, pages 16 – 124, 27-30 June 2011.

C.H. Liu, J. Zhao, H. Zhang, S. Guo, K.K. Leung, and J. Crowcroft. Energy-efficient event detection by participatory sensing under budget constraints. *IEEE Systems Journal*, 2016b.

J. Liu, F. Cai, L. Wu, R. Sun, L. Zhu, and X. Du. Epda: Enhancing privacy-preserving data authentication for mobile crowd sensing. In *IEEE Global Communications Conference (GLOBECOM)*, pages pp. 1–6. IEEE, December 2017a.

J. Liu, X. Li, R. Sun, X. Du, and P. Ratazzi. An efficient privacy-preserving incentive scheme without ttp in participatory sensing network. In *IEEE International Conference on Communications (ICC)*, pages pp. 1–6. IEEE, May 2018.

K. Liu, C. Giannella, and H. Kargupta. A survey of attack techniques on privacy-preserving data perturbation methods. *Privacy-Preserving Data Mining, Springer, Boston, MA.*, pages pp. 359–381, 2008.

R. Liu, J. Liang, W. Gao, and R. Yu. Privacy-based recommendation mechanism in mobile participatory sensing systems using crowdsourced users' preferences. *Future Generation Computer Systems.*, 2017b.

T. Liu, Y. Zhu, R. Jiang, and Q. Zhao. Distributed social welfare maximization in urban vehicular participatory sensing systems. *IEEE Transactions on Mobile Computing*, 2017c.

T. Liu, Y. Zhu, and L. Huang. Tgba: A two-phase group buying based auction mechanism for recruiting workers in mobile crowd sensing. *Computer Networks*, (149):pp.56–75, 2019.

Y. Liu, C. Yuen, N. U. Hassan, S. Huang, R. Yu, and S. Xie. Electricity cost minimization for a microgrid with distributed energy resource under different information availability. *IEEE Transactions on Industrial Electronics*, 62(4): pp.2571–2583., 2015.

J. Lu, Q. Dai, J. Han, H. Peng, and Y. Xin. A binary rating protocol for crowdsensing applications. In *36th Chinese Control Conference (CCC)*, pages pp. 9020–9024. IEEE, July 2017.

T. Luo, S.S. Kanhere, and H-P. Tan. Sew-ing a simple endorsement web to incentivize trustworthy participatory sensing. In *Eleventh Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, pages 636 – 644, June 30 - July 3 2014.

T. Luo, S. S. Kanhere, S.K. Das, and H. P. Tan. Incentive mechanism design for heterogeneous crowdsourcing using all-pay contests. *IEEE Transactions on Mobile Computing*, (99):1 – 13, 02 October 2015.

T. Luo, S.S. Kanhere, J. Huang, S.K. Das, and F. Wu. Sustainable incentives for mobile crowdsensing: Auctions, lotteries, and trust and reputation systems. *IEEE Communications Magazine*, 55(3):pp.68–74, 2017.

T. Luo, J. Huang, Kanhere S.S., J. Zhang, and S.K. Das. Improving iot data quality in mobile crowd sensing: A cross validation approach. *IEEE Internet of Things Journal.*, 2019.

L. Lyu, Y.W. Law, S.M. Erfani, C. Leckie, and M. Palaniswami. An improved scheme for privacy-preserving collaborative anomaly detection. In *IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*, pages 1–6, 14-18 March 2016.

L. Lyu, J.C. Bezdek, Y.W. Law, X. He, and M. Palaniswami. Privacy-preserving collaborative fuzzy clustering. *Data & Knowledge Engineering*, 2018.

F. Ma, X. Liu, A. Liu, M. Zhao, C. Huang, and T. Wang. A time and location correlation incentive scheme for deep data gathering in crowdsourcing networks. *Wireless Communications and Mobile Computing*, 2018.

A. Mao, E. Kamar, Y. Chen, E. Horvitz, M.E. Schwamb, C.J. Lintott, and A.M. Smith. Volunteering versus work for pay: Incentives and tradeoffs in crowdsourcing. In *First AAAI Conference on Human Computation and Crowdsourcing*, 2013.

A. San Martini and F. Spezzaferri. A predictive model selection criterion. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages pp.296–303., 1984.

S. Marusic, J. Gubbi, H. Sullivan, Y.W. Lawand, and M. Palaniswami. Participatory sensing, privacy, and trust management for interactive local government. *IEEE Technology and Society Magazine*, 33(3):pp.62–70, 2014.

R.B. Messaoud, N. Sghaier, M.A. Moussa, and Y. Ghamri-Doudane. On the privacy-utility tradeoff in participatory sensing systems. In *IEEE 15th International Symposium on Network Computing and Applications (NCA)*, pages pp. 294–301, October 2016.

C. Miao, L. Su, W. Jiang, Y. Li, and M. Tian. A lightweight privacy-preserving truth discovery framework for mobile crowd sensing systems. In *IEEE Conference on Computer Communications (INFOCOM 2017)*, pages pp. 1–9. IEEE, May 2017.

A. Michalas and N. Komninos. The lord of the sense: A privacy preserving reputation system for participatory sensing applications. In *IEEE Symposium on Computers and Communication (ISCC)*, pages 1–6, 23-26 June 2014.

A. Mihaita, C. Dobre, F. Pop, B. Mocanu, V. Cristea, and C. Esposito. A trust application in participatory sensing: Elder reintegration. In *International Conference on Green, Pervasive, and Cloud Computing*, pages pp. 596–610. Springer, May 2017.

A. Miller, M. Moser, K. Lee, and A. Narayanan. An empirical analysis of linkability in the monero blockchain. 2017 arXiv preprint arXiv:1704.04299., 2017.

D.C. Mills, K. Wang, B. Malone, A. Ravi, J. Marquardt, A.I. Badev, T. Brezinski, L. Fahy, K. Liao, V. Kargenian, and M. Ellithorpe. Distributed ledger technology in payments, clearing, and settlement. *Finance and Economics Discussion Series, Divisions of Research & Statistics and Monetary Affairs, Federal Reserve Board, Washington, D.C.*, 2016.

R. Mishra and P. Bhanodiya. A review on steganography and cryptography. In *International Conference on Advances in Computer Engineering and Applications (ICACEA)*, pages 119–122, 19-20 March 2015.

J. Mohite, Y. Karale, P. Gupta, S. Kulkarni, B. Jagyasi, and A. Zape. Rups: Rural participatory sensing with rewarding mechanisms for crop monitoring. In *IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*, pages 378–383, 2015.

H. Mousa, O. Mokhtar, S.B. andHasan, L. Brunie, O.S. Youness, and H. Mohiy. A reputation system resilient against colluding and malicious adversaries in mobile participatory sensing applications. In *The 14th Annual IEEE Consumer Communications & Networking Conference (CCNC 2017)*, January 2017.

J. Mukhopadhyay, V.K. Singh, S. Mukhopadhyay, and A. Pal. Online participatory sensing in double auction environment with location information. 2017.

M.Y. Mun, D.H. Kim, K. Shilton, D. Estrin, M. Hansen, and R. Govindan. Pdvloc: A personal data vault for controlled location data sharing. *ACM Transactions on Sensor Networks (TOSN)*, 10(4), June 2014.

F.F.H. Nah. A study on tolerable waiting time: how long are web users willing to wait? *Behaviour & Information Technology*, Vol. 23:pp.153–163, May 2004.

S. Nakamoto. Bitcoin: A peer-to-peer electronic cash system. Technical report, 2008.

M. J. Neely. *Stochastic Network Optimization with Application to Communication and Queueing Systems.* Morgan Claypool (Synthesis Lectures on Communication Networks), 2010.

M.J. Neely and R. Urgaonkar. Opportunism, backpressure, and stochastic optimization with the wireless broadcast advantage. In *42nd Asilomar Conference on Signals, Systems and Computers*, pages 2152 – 2158, 26-29 October 2008.

X. Niu, M. Li, Q. Chen, Q. Cao, and H. Wang. Eppi: An e-cent-based privacy-preserving incentive mechanism for participatory sensing systems. *IEEE International Performance Computing and Communications Conference (IPCCC)*, pages 1–8, 5-7 December 2014.

X. Niu, J. Wang, Q. Ye, and Y. Zhang. A privacy-preserving incentive mechanism for participatory sensing systems. *Security and Communication Networks*, 2018a.

X. Niu, Q. Ye, Y. Zhang, and D. Ye. A privacy-preserving identification mechanism for mobile sensing systems. *IEEE Access*, 2018b.

R.I. Ogie. Adopting incentive mechanisms for large-scale participation in mobile crowdsensing: from literature review to a conceptual framework. *Human-centric Computing and Information Sciences*, 6(1):p. 24, 2016.

T. Oide, T. Abe, and T. Suganuma. A broker-less participatory sensing scheme by user matching mechanism based on market price approach. In *IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*, pages pp. 1–6. IEEE, March 2016.

K. Ota, M. Dong, J. Gui, and A. Liu. Quoin: incentive mechanisms for crowd sensing networks. *IEEE Network*, 2018.

D. Peng, F. Wu, and G. Chen. Data quality guided incentive mechanism design for crowdsensing. *IEEE Transactions on Mobile Computing,*, 17(2):pp.307–319, 2018.

A.J. Perez and S. Zeadally. Design and evaluation of a privacy architecture for crowdsensing applications. *ACM SIGAPP Applied Computing Review*, 18(1): pp.7–18, 2018.

E. Pournaras, J. Nikolic, P. Velasquez, M. Trovati, N. Bessis, and D. Helbing. Self-regulatory information sharing in participatory social sensing. *EPJ Data Science*, December 2016.

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

B. Predic, Y. Zhixian, J. Eberle, D. Stojanovic, and K. Aberer. Exposuresense: Integrating daily activities with air quality using mobile participatory sensing. In *IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)*, pages 303–305, 18-22 March 2013.

F. Qiu, F. Wu, and G. Chen. Slicer: A slicing-based k-anonymous privacy preserving scheme for participatory sensing. In *IEEE 10th International Conference on Mobile Ad-Hoc and Sensor Systems (MASS)*, pages 113 – 121, 14-16 October 2013.

F. Qiu, F. Wu, and G. Chen. Privacy and quality preserving multimedia data aggregation for participatory sensing systems. *IEEE Transactions on Mobile Computing*, (99), August 2014.

G. Radanovic and B. Faltings. Incentive schemes for participatory sensing. In *Proceedings of the International Conference on Autonomous Agents and Multiagent Systems (AAMAS '15)*, pages 1081–1089, 2015.

J. Ren, Y. Zhang, K. K. Zhang, and X.S. Shen. Sacrm: Social aware crowdsourcing with reputation management in mobile sensing. *Computer Communications*, 65:pp.55–65, 2015.

Y. Ren, A. Liu, M. Zhao, C. Huang, and T. Wang. A quality utilization aware based data gathering for vehicular communication networks. *Wireless Communications and Mobile Computing*, 2018.

F. Restuccia and S.K. Das. Fides: A trust-based framework for secure user incentivization in participatory sensing. In *IEEE 15th International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, pages pp. 1–10. IEEE, June 2014.

F. Restuccia, S.K. Das, and J. Payton. Incentive mechanisms for participatory sensing: Survey and research challenges. *ACM Transactions on Sensor Networks (TOSN)*, 2016.

F. Restuccia, P. Ferraro, Silvestri S., S.K. Das, and G.L. Re. Incentme: Effective mechanism design to stimulate crowdsensing participants with uncertain mobility. *arXiv preprint arXiv:1804.11150*, 2018a.

F. Restuccia, A. Saracino, and F. Martinelli. Practical location validation in participatory sensing through mobile wifi hotspots. *arXiv preprint arXiv:1805.07600.*, 2018b.

D.J. Roberts and A. Postlewaite. The incentives for price-taking behavior in large exchange economies. *Econometrica: Journal of the Econometric Society*, pages pp.115–127, 1976.

I. Rodhe, C. Rohner, and Ngai E. C-H. On location privacy and quality of information in participatory sensing. In *Proceedings of the 8th ACM Symposium on QoS and Security for Wireless and Mobile Networks (Q2SWinet '12)*, pages 55–62, 2012.

K. H. Rosen. *Discrete Mathematics and Its Applications.* McGraw Hill, San Francisco, CA, 6th edition edition, 2007.

T. Sabrina and M. Murshed. Analysis of location privacy risk in a plain-text communication based participatory sensing system using subset coding and mix network. In *International Symposium on Communications and Information Technologies (ISCIT)*, pages 718–723, 2-5 October 2012.

N. Saleem, S. lrabaee, F.A. Khasawneh, and M. Khasawneh. Aggregation function using homomorphic encryption in participating sensing application. In *6th International Conference on Computer Science and Information Technology (CSIT)*, pages 166 – 171, 26-27 March 2014.

T Sandholm. Issues in computational vickrey auctions. *International Journal of Electronic Commerce*, 4(3):107–129, 2000.

F. Saremi and T. Abdelzaher. Model transition planning in participatory sensing cold start. In *International Conference on Distributed Computing in Sensor Systems (DCOSS)*, pages pp.99–101. IEEE, May 2016.

B. Schneier. Attack trees. *Dr. DobbÕs journal*, 24(12):pp.21–29, 1999.

I. Schweizer, C. Meurisch, J. Gedeon, R. Bartl, and M. Munhlhauser. Noisemap: multi-tier incentive mechanisms for participative urban sensing. In *In Proceedings of the Third International Workshop on Sensing Applications on Mobile Phones*, pages p. 9–14. ACM, November 2012.

T. Sharma and R. Yadav. Security in virtual private network. *International Journal of Innovations & Advancement in Computer Science*, 4, 2015.

W. Shen, B. Yin, Y. Cheng, X. Cao, and Q. Li. Privacy-preserving mobile crowd sensing for big data applications. In *IEEE International Conference on Communications*, pages pp. 1–6. IEEE, May 2017.

F. Shi, Z. Qin, D. Wu, and J. McCann. Mpcstoken: Smart contract enabled fault-tolerant incentivisation for mobile p2p crowd services. In *IEEE 38th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, July 2018.

K. Shilton and K.E. Martin. Mobile privacy expectations in context. In *TPRC, Research Conference on Communications, Information and Internet Policy*, 2013.

T. Shimeall and J. Spring. *Introduction to Information Security: A Strategic-Based Approach*. Syngress, 2013.

A. Shostack. *Threat modeling: Designing for security*. John Wiley & Sons., 2014.

A. Singla and A. Krause. Truthful incentives in crowdsourcing tasks using regret minimization mechanisms. In *Proceedings of the 22nd international conference on World Wide Web*, pages 1–8, 4-7 August 2013a.

A. Singla and A. Krause. Incentives for privacy tradeoff in community sensing. In *1st AAAI Conference on Human Computation and Crowdsourcing (HCOMP-13)*, 2013b.

V.L. Smith. The principle of unanimity and voluntary consent in social choice. *Journal of Political Economy,*, 85(6):pp.1125–1139, 1977.

Z. Song, E. Ngai, J. Ma, and W. Wang. A novel incentive negotiation mechanism for participatory sensing under budget constraints. In *IEEE 22nd International Symposium of Quality of Service (IWQoS)*, pages 326 – 331, 26-27 May 2014.

G. Spathoulas, P. Vennou, and A. Loukidis. Privacy preserving platform for profitable mobile crowd sensing and users' adoption. In *Proceedings of the 21st Pan-Hellenic Conference on Informatics*, page p. 30. ACM, September 2017.

J. Sun and N. Liu. Frugal incentive mechanism in periodic mobile crowdsensing. *International Journal of Communication Systems*, 31(5), 2018.

J. Sun and H. Ma. Privacy-preserving verifiable incentive mechanism for online crowdsourcing markets. In *23rd International Conference on Computer Communication and Networks (ICCCN)*, 2014.

J. Sun, Y. Pei, F. Hou, and S. Ma. Reputation-aware incentive mechanism for participatory sensing. *IET Communications*, 2017.

J. Sun, N. Liu, and D. Wu. Budget-constraint mechanism for incremental multilabeling crowdsensing. *Telecommunication Systems*, 67(2):pp.297–307., 2018.

W. Sun and C-K. Tham. A spatio-temporal incentive scheme with consumer demand awareness for participatory sensing. In *IEEE International Conference on Communications (ICC)*, pages 6363 – 6369, 8-12 June 2015a.

W. Sun and C-K. Tham. An information-driven incentive scheme with consumer demand awareness for participatory sensing. In *12th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, pages 319 – 326, 22-25 June 2015b.

X. Sun, J. Li, W. Zheng, and H. Liu. Towards a sustainable incentive mechanism for participatory sensing. In *IEEE First International Conference on Internet-of-Things Design and Implementation (IoTDI)*, pages pp. 49–60. IEEE, April 2016.

Y. Sun, L. Yin, L. Liu, and S. Xin. Toward inference attacks for k-anonymity. *Personal and ubiquitous computing*, 18(8):pp.1871–1880, 2014.

R. Szabo, K. Farkas, M. Ispany, A.A. Bencur, N. Batfai, P. Jeszenszky, S. Laki, A. Vagner, L. Kollar, Cs. Sidlo, R. Besenczi, M. Smajda, G. Kover, T. Szincsak, T. Kadek, M. Kosa, A. Adamo, I. Lendak, B. Wiandt, T. Tomas, A. Zs. Nagy, and G. Feher. Framework for smart city applications based on participatory sensing. In *IEEE 4th International Conference on Cognitive Infocommunications (CogInfoCom)*, pages 295–300, 2-5 December 2013.

L. Tan, H. Fan, W. Rui, Z. Xu, S. Zhang, J. Xu, and K. Xing. Mining myself in the community: privacy preserved crowd sensing and computing. In *International Conference on Wireless Algorithms, Systems, and Applications*, pages pp. 272–282. Springer International Publishing, August 2016.

C. Tanas, S. Delgado-Segura, and J. Herrera-Joancomarti. An integrated reward and reputation mechanism for mcs preserving usersÕ privacy. *Data Privacy Management and Security Assurance. Springer, Cham.*, pages pp. 83–99, 2015.

D. Tang and J. Ren. A novel delay-aware and privacy-preserving data-forwarding scheme for urban sensing network. *IEEE Transactions on Vehicular Technology*, 65(4):pp.2578–2588., 2015.

W. Tang and R. Jain. Market mechanisms for buying random wind. *IEEE Transactions on Sustainable Energy*, 6(4):1615–1623, October 2015.

C.K. Tham and T. Luo. Quality of contributed service and market equilibrium for participatory sensing. *IEEE Transactions on Mobile Computing*, 14(4): pp.829–842, 2015.

D.H.T. That, I.S. Popa, K. Zeitouni, and C. Borcea. Pampas: Privacy-aware mobile participatory sensing using secure probes. In *Proceedings of the 28th International Conference on Scientific and Statistical Database Management*, July 2016.

N. Thepvilojanapong, T. Tsujimori, H. Wang, Y. Ohta, Y. Zhao, and Y. Tobe. Impact of incentive mechanism in participatory sensing environment. In *Proceedings of the 2nd International Conference on Smart Systems, Devices and Technologies, SMART*, pages 87–92, 2013.

Y. Tian, X. Li, A.K. Sangaiah, E. Ngai, Z. Song, L. Zhang, and W. Wang. Privacy-preserving scheme in social participatory sensing based on secure multiparty cooperation. *Computer Communications*, 2017.

H. To, G. Ghinita, and C. Shahabi. A framework for protecting worker location privacy in spatial crowdsourcing. *Journal Proceedings of the VLDB Endowment*, 7(10):919–930, June 2014.

H.R. Tsai and T. Chen. Enhancing the sustainability of a location-aware service through optimization. *Sustainability*, 6(12):pp.9441–9455, 2014.

D. Tsolovos, N. Anciaux, and V. Issarny. A privacy aware approach for participatory sensing systems. In *34th Conference on Data Management - Principles, Technologies and Applications*, October 2018.

T. Tsujimori, N. Thepvilojanapong, Y. Ohta, Y. Zhao, and Y. Tobe. History-based incentive for crowd sensing. In *Proceedings of the 2014 International Workshop on Web Intelligence and Smart Sensing (IWWISS '14)*, pages 1 – 6, 2014.

R. Urgaonkar, U.C. Kozat, K. Igarashi, and M.J. Neely. Dynamic resource allocation and power management in virtualized data centers. In *IEEE Network Operations and Management Symposium (NOMS)*, pages pp. 479–486, April 2010.

I. J. Vergara-Laurens, D. Mendez-Chaves, and M. A. Labrador. On the interactions between privacy-preserving, incentive, and inference mechanisms in participatory sensing systems. In *7th International Conference on Network and System Security (NSS 2013)*, pages 614 – 620, 2013.

K. Vu, R. Zheng, and L. Gao. Efficient algorithms for k-anonymous location privacy in participatory sensing. In *Proceedings IEEE INFOCOM*, pages 2399–2407, 2012.

C. Wang and K.S. Chan. Quasi-likelihood estimation of a censored autoregressive model with exogenous variables. *Journal of the American Statistical Association*, 113(523):pp.1135–1145, 2018.

C. Wang, H. Liu, K-L. Wright, B. Krishnamachari, and M. Annavaram. A privacy mechanism for mobile-based urban traffic monitoring. *Pervasive and Mobile Computing*, 2014.

C-J. Wang and W-S. Ku. Anonymous sensory data collection approach for mobile participatory sensing. In *IEEE 28th International Conference on Data Engineering Workshops (ICDEW)*, pages 220–227, 1-5 April 2012.

W. Wang, H. Gao, C.H. Liu, and K.K.Leung. Credible and energy-aware participant selection with limited task budget for mobile crowd sensing. *Ad-Hoc Networks*, 43:pp.56–70, 2016a.

W. Wang, Y. Li, X. Wang, J. Liu, and X. Zhang. Detecting android malicious apps and categorizing benign apps with ensemble of classifiers. *Future generation computer systems,*, (78):pp.987–994, 2018a.

X. Wang, W. Cheng, P. Mohapatra, and T. Abdelzaher. Artsense: Anonymous reputation and trust in participatory sensing. In *Proceedings IEEE INFOCOM*, pages 2517–2525, 2013.

Y. Wang, A. Nakao, and A.V. Vasilakos. Heterogeneity playing key role: Modeling and analyzing the dynamics of incentive mechanisms in autonomous networks. *ACM Transactions on Autonomous and Adaptive Systems (TAAS)*, 2012.

Y. Wang, Z. Cai, G. Yin, Y. Gao, X. Tong, and G. Wu. An incentive mechanism with privacy protection in mobile crowdsourcing systems. *Computer Networks*, 102:157–171, 2016b.

Y. Wang, Z. Cai, Z. Chi, X. Tong, and L. Li. A differentially k-anonymity-based location privacy-preserving for mobile crowdsourcing systems. *Procedia Computer Science*, (129):pp.28–34, 2018b.

Y. Wang, Z. Cai, X. Tong, Y. Gao, and G. Yin. Computer networks. *Truthful incentive mechanism with location privacy-preserving for mobile crowdsourcing systems*, 2018c.

Z. Wang and D. Huang. Privacy-preserving mobile crowd sensing in ad hoc networks. *Ad Hoc Networks*, (73):pp.14–26, 2018.

Y. Wei, Y. Zhu, H. Zhu, Q. Zhang, and G. Xue. Truthful online double auctions for dynamic mobile crowdsourcing. In *IEEE Conference on Computer Communications (INFOCOM)*, pages pp. 2074–2082. IEEE, April 2015.

Z. Wei, B. Zhao, and J. Su. Sps: A novel semantics-aware scheme for location privacy in people-centric sensing network. *Lecture Notes in Computer Science Wireless Algorithms, Systems, and Applications*, 8491:324–335, 2014.

K. Wiesner, S. Feld, F. Dorfmeister, and C. Linnhoff-Popien. Right to silence: Establishing map-based silent zones for participatory sensing. In *IEEE Ninth*

International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), pages 1–6, 21-24 April 2014.

C. Xiang, P. Yang, C. Tian, Y. Yan, X. Wu, and Y. Liu. Passfit: Participatory sensing and filtering for identifying truthful urban pollution sources. *IEEE Sensors Journal*, 13(10):3721 – 3732, October 2013.

C. Xiang, P. Yang, C. Tian, H. Cai, and Y. Liu. Calibrate without calibrating: An iterative approach in participatory sensing network. *IEEE Transactions on Parallel and Distributed Systems*, 26(2):351 – 361, February 2015.

Q. Xiang, J. Zhang, I. Nevat, and P. Zhang. A trust-based mixture of gaussian processes model for robust participatory sensing. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*, pages pp.1760–1762. International Foundation for Autonomous Agents and Multiagent Systems., May 2017.

C. Xiao, W. Jia, H. Zhu, S. Du, and Z. Cao. Leveraging cloud computing for privacy preserving aggregation in multi-domain wireless networks. *Lecture Notes in Computer Science: Wireless Algorithms, Systems, and Applications*, 7405:733–744, 2012.

Z. Xiao, J.J. Yang, M. Huang, L. Ponnambalam, X. Fu, and R.S.M. Goh. Qlds: A novel design scheme for trajectory privacy protection with utility guarantee in participatory sensing. *IEEE Transactions on Mobile Computing*, 2017.

K. Xing, Z. Wan, P. Hu, H. Zhu, Y. Wang, X. Chen, Y. Wang, and L. Huang. Mutual privacy-preserving regression modeling in participatory sensing. In *Proceedings IEEE INFOCOM 2013*, 2013.

K. Xing, C. Hu, J. Yu, X. Cheng, and F. Zhang. Mutual privacy preserving k-means clustering in social participatory sensing. *IEEE Transactions on Industrial Informatics.*, 2017.

H. Xiong, D. Zhang, G. Chen, L. Wang, V. Gauthier, and L.E. Barnes. icrowd: Near-optimal task allocation for piggyback crowdsensing. *EEE Transactions on Mobile Computing*, 15(8):pp.2010–2022, 2016.

H. Xiong, D. Zhang, Z. Guo, G. Chen, and L.E. Barnes. Near-optimal incentive allocation for piggyback crowdsensing. *IEEE Communications Magazine*, 55 (6):pp.120–125, 2017.

C. Xu, X. Shen, L. Zhu, and Y. Zhang. A collusion-resistant and privacy-preserving data aggregation protocol in crowdsensing system. *Mobile Information Systems*, 2017a.

J. Xu, C. Guan, H. Wu, D. Yang, L. Xu, and T. Li. Online incentive mechanism for mobile crowdsourcing based on two-tiered social crowdsourcing architecture. In *15th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*. IEEE, June 2018.

Y. Xu, Y. Zhou, Y. Mao, X. Chen, and X. Li. Can early joining participants contribute more?-timeliness sensitive incentivization for crowdsensing. arXiv preprint arXiv:1710.01918., 2017b.

Y. Yan, D. Han, and T. Shu. Privacy preserving optimization of participatory sensing in mobile cloud computing. In *IEEE 37th International Conference on Distributed Computing Systems (ICDCS*, pages pp.1084–1093, June 2017.

G. Yang, S. He, Z. Shi, and J. Chen. Promoting cooperation by the social incentive mechanism in mobile crowdsensing. *IEEE Communications Magazine*, 55 (3):pp.86–92, 2017a.

H. Yang, J. Zhang, and P. Roe. Using reputation management in participatory sensing for data classification. *Procedia Computer Science*, 5:pp.190–197, 2011.

S. Yang, U. Adeel, and J. McCann. Backpressure meets taxes: Faithful data collection in stochastic mobile phone sensing systems. In *IEEE Conference on Computer Communications (INFOCOM)*, pages 1490 – 1498, April 26 - May 1 2015.

X. Yang, C. Zhao, W. Yu, X. Yao, and X. Fu. A user incentive-based scheme against dishonest reporting in privacy-preserving mobile crowdsensing systems. In *International Conference on Wireless Algorithms, Systems, and Applications*, pages pp. 755–767. Springer, Cham., June 2017b.

Y. Yao, L.T. Yang, and N.N. Xiong. Anonymity-based privacy-preserving data reporting for participatory sensing. *IEEE Internet of Things Journal*, 2(5): 381 – 390, October 2015.

T.J. Ypma. Historical development of the newtonĐraphson method. *SIAM review*, 37(4):pp.531–551, 1995.

R. Yu, J. Cao, R. Liu, W. Gao, X. Wang, and J Liang. Participant incentive mechanism toward quality-oriented sensing: Understanding and application. *ACM Transactions on Sensor Networks (TOSN)*, 15(2):p.21, 2019.

S. Zaman, N. Abrar, and A. Iqbal. Incentive model design for participatory sensing: Technologies and challenges. In *International Conference on Networking Systems and Security (NSysS)*, pages 1 – 6, 5-7 January 2015.

J. Zeng, Y. Wu, Y. Wu, H. Chen, C. Li, and S. Wang. Energy-efficient and privacy-preserving range query in participatory sensing. In *Trustcom/BigDataSE/ISPA*, pages pp. 876–883, August 2016.

A. Zenonos, S. Stein, and N. Jennings. A trust-based coordination system for participatory sensing applications. 2017.

Y. Zhan, Y. Xia, and J. Zhang. Quality-aware incentive mechanism based on payoff maximization for mobile crowdsensing. *Ad Hoc Networks*, (72):pp. 44–55, 2018a.

Y. Zhan, Y. Xia, J. Zhang, and Y. Wang. Incentive mechanism design in mobile opportunistic data collection with time sensitivity. *IEEE Internet of Things Journal*, 5(1):pp.246–256, 2018b.

F. Zhang, L. He, W. He, and X. Liu. Data perturbation with state-dependent noise for participatory sensing. In *Proceedings IEEE INFOCOM*, pages pp. 2246–2254, 2012a.

J. Zhang, J. Ma, W. Wang, and Y. Liu. A novel privacy protection scheme for participatory sensing with incentives. In *IEEE 2nd International Conference on Cloud Computing and Intelligent Systems (CCIS)*, volume 3, pages 1017 – 1021, October 30-November 1 2012b.

L. Zhang, X. Wang, J. Lu, P. Li, and Z. Cai. An efficient privacy preserving data aggregation approach for mobile sensing. *Security and Communication Networks*, 9(16):pp.3844–3853, 2016a.

X. Zhang, Z. Yang, Z. Zhou, H. Cai, L. Chen, and X. Li. Free market of crowdsourcing: Incentive mechanism design for mobile sensing. *IEEE Transactions on Parallel and Distributed Systems*, pages 3190 – 3200, December 2014.

Y. Zhang, Q. Chen, and S. Zhong. Privacy-preserving data aggregation in mobile phone sensing. *IEEE Transactions on Information Forensics and Security*, 11 (5):pp.980–992, 2016b.

Y. Zhang, Q. Chen, and S. Zhong. Efficient and privacy-preserving min and kth min computations in mobile sensing systems. *IEEE Transactions on Dependable and Secure Computing*, 14(1):pp.9–21, 2017.

Y. Zhang, Y. Gu, M. Pan, N.H. Tran, Z. Dawy, and Z. Z. Han. Multi-dimensional incentive mechanism in mobile crowdsourcing with moral hazard. *IEEE Transactions on Mobile Computing*, 17(3):pp.604–616, 2018a.

Z. Zhang, S. He, J. Chen, and J. Zhang. Reap: An efficient incentive mechanism for reconciling aggregation accuracy and individual privacy in crowdsensing. *IEEE Transactions on Information Forensics and Security*, 2018b.

N. Zhao, M. Fan, C. Tian, and P. Fan. Contract-based incentive mechanism for mobile crowdsourcing networks. *Algorithms*, 10(3):p.104., 2017.

Q. Zhao and Y. Zhu. Distributed social welfare maximization in vehicular participatory sensing systems. In *IEEE 22nd International Symposium of Quality of Service (IWQoS)*, pages 332 – 337, 26-27 May 2014.

Z. Zheng, F. Wu, X. Gao, H. Zhu, G. Chen, and S. Tang. A budget feasible incentive mechanism for weighted coverage maximization in mobile crowdsensing. *IEEE Transactions on Mobile Computing.*, 2016.

T. Zhou, Z. Cai, Y. Chen, and M. Xu. Improving data credibility for mobile crowdsensing with clustering and logical reasoning. In *International Conference on Cloud Computing and Security*, pages pp. 138–150. Springer, Cham., July 2016.

T. Zhou, Z. Cai, K. Wu, Y. Chen, and M. Xu. Fidc: A framework for improving data credibility in mobile crowdsensing. *Computer Networks*, (120):pp.157–169, 2017.

G. Zhuo. Privacy-preserving and fine-grained data aggregation framework for crowdsourcing. In *.Tenth International Conference on Mobile Computing and Ubiquitous Network (ICMU)*, pages pp. 1–6. IEE, October 2017.

G. Zhuo, Q. Jia, L. Guo, M. Li, and P. Li. Privacy-preserving verifiable set operation in big data for cloud-assisted mobile crowdsourcing. *IEEE Internet of Things Journal*, 4(2):pp.572–582, 2017.

J.H. Ziegeldorf, M. Henze, J. Bavendiek, and K. Wehrle. Tracemixer: Privacy-preserving crowd-sensing sans trusted third party. In *IEEE 13th Annual Conference on In Wireless On-demand Network Systems and Services (WONS)*, pages pp. 17–24, February 2017.