M.Sc. Engg. Thesis

Development of A Multidimensional Anonymization Technique for Preserving Privacy in Participatory Sensing System

by

Nafeez Abrar

Submitted to

Department of Computer Science and Engineering

in partial fulfillment of the requirements for the degree of Master of Science in Computer Science and Engineering

Department of Computer Science and Engineering

Bangladesh University of Engineering and Technology (BUET)

Dhaka 1000

June 2016

The thesis titled "Development of A Multidimensional Anonymization Technique for Preserving Privacy in Participatory Sensing System", submitted by Nafeez Abrar, Roll No. **0413052084 P**, Session April 2013, to the Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology, has been accepted as satisfactory in partial fulfillment of the requirements for the degree of Master of Science in Computer Science and Engineering and approved as to its style and contents. Examination held on June 12, 2016.

Board of Examiners

1. Dr. Anindya Iqbal 12.6.16
Assistant Professor (Supervisor)
Department of Computer Science and Engineering, BUET, Dhaka, Bangladesh Chairman

2. Dr. M. Sohel Rahman Member
Professor and Head (Ex-Officio)
Department of Computer Science and Engineering, BUET, Dhaka, Bangladesh

3. Dr. Md. Yusuf Sarwar Uddin 12/6/2016 Member
Associate Professor
Department of Computer Science and Engineering, BUET, Dhaka, Bangladesh

4. Dr. A. B. M. Alim Al Islam 12.6.16 Member
Assistant Professor
Department of Computer Science and Engineering, BUET, Dhaka, Bangladesh

5. Dr. Salekul Islam 12-6-16 Member
Associate Professor (External)
School of Science and Engineering,
United International University, Dhaka, Bangladesh

Candidate's Declaration

This is hereby declared that the work titled "Development of A Multidimensional Anonymization Technique for Preserving Privacy in Participatory Sensing System" is the outcome of research carried out by me under the supervision of Dr. Anindya Iqbal, in the Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology, Dhaka. It is also declared that this thesis or any part of it has not been submitted elsewhere for the award of any degree or diploma.

Nafeez Abrar
Candidate

Acknowledgment

First of all I would like to thank my supervisor, Dr. Anindya Iqbal, for assisting me throughout the thesis. Without his continuous inspiring enthusiasm, encouragement, supervision, guidance and advice it would not have been possible to complete this thesis. I am especially grateful to him for giving me his valuable time whenever I needed, and always providing continuous support, motivation and endless patience towards the completion of the thesis.

I would like to take this opportunity to thank Professor Manzur Murshed for introducing me this amazingly interesting topic and giving me guidelines and encouragement. I would also like to extend thanks to my research group member Shaolin Zaman especially for helping me in writing and implementation. I also want to thank the other members of my thesis committee: Dr. Md. Yusuf Sarwar Uddin, Dr. A. B. M. Alim Al Islam and specially the external member Dr. Salekul Islam for their valuable suggestions.

Last but not the least, I am grateful to my guardians, families and friends for their patience, cooperation and inspiration during this period.

Abstract

Participatory sensing technology is designed to facilitate community people collect, analyze, and share information for their mutual benefit in a cost-effective way using smart-phones, camera or other ad-hoc sensing devices. The apparently insensitive information transmitted in plaintext through a lightweight infrastructure of participatory sensing system can be used by an eavesdropper to infer some sensitive information and threaten the privacy of the observer. Sufficient number of participants is imperative for the success of participatory sensing. Participation depends a great deal on the assurance of privacy protection. Existing techniques add some uncertainty to the actual observation to achieve anonymity of the participants which, however, diminishes data integrity to an unacceptable extent. A subset-coding based anonymization technique was proposed in [1] to safeguard observers' location privacy from adversaries while preserving almost loss-less data integrity at the destination server. However, the high computational complexity of that technique $O(N!)$ allowed its use at limited level. In this thesis, we develop an $O(N)$ technique to overcome this limitation. The new technique accommodate variable degree of desired anonymization for the users which eventually enables designing flexible incentive schemes for the users. Finally, to the best of our knowledge, we present the first multi-dimensional privacy preserving scheme that can protect users' privacy at different dimensions simultaneously. For example, both location and product association of an observer can be protected. Comprehensive simulation and Android prototype based experiments are carried out to establish the applicability of the proposed schemes.

Contents

List of Figures

List of Tables

List of Algorithms

Chapter 1

Introduction

Participatory sensing system (PSS) is a framework that facilitates community people sense, collect, analyze, and share information obtained from their surroundings for mutual benefit of themselves. This is evolving as a cost-effective alternative for reliable and impartial data collection, processing and dissemination. Smart-phones equipped with high precision localization capability and camera or other ad-hoc sensing devices may be used to record objects/events of interest by the people in their daily walk of life. The captured data are sent to specific servers via some lightweight inexpensive wireless communication networks. The collective reports from a large number of participating users help the server generate useful information and reply to the queries of the user on-demand.

1.1 Significance

PSS has emerged as a promising alternative to dedicated deployment of Wireless Sensor Network (WSN) due to the ever-increasing availability of sensor-rich and geo-localized mobile devices among the common people. It may be compared to the popularity of citizen journalism, blogs or wikileaks over traditional main-stream media. It has a wide range of real-world applications including consumer price sharing [2–5], measuring safety in localities [6], variation in elevation along bike routes monitoring [7], vehicular transportation monitoring [8–10], public health such as monitoring the effectiveness of diet programs [11], environmental impact and exposure [12], urban planning [13], sound events [14], earthquakes [15], parking availabilities [16], comfort management of building [17], and prediction of bus arrival time [18].

To achieve better quality of service using the PSS collected data, extent of participation is the most important factor. However, sensing activities with smart-phones require time,

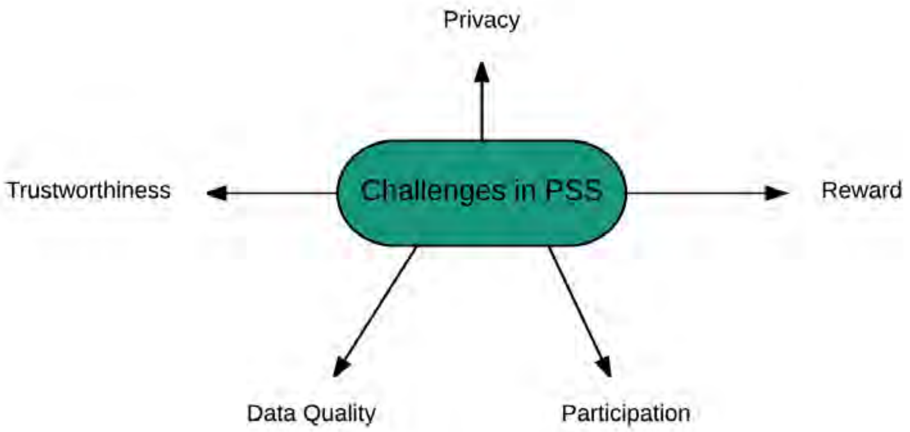


Figure 1.1: Challenges in participatory sensing system.

power consumption (battery) and monetary cost for data transfer to the service provider. Most importantly, user has to share their private sensitive data (like location) and activities to application server which imposes a great threat to their privacy. To encourage participation, PSS system must ensure the privacy of the participants.

To protect the privacy of the participants, hiding identity or obfuscating the shared data are the most common techniques. However, without identity disclosure, it is not possible to build up user's reputation scheme. Moreover, as sharing data requires time and battery power of the user, incentives are required to encourage the users' participation. Providing reward or incentive to the users in proportion to their participation also becomes impossible without preserving identity of every observation. These contradictory challenges are depicted in Figure 1.1. Indeed, this is very challenging to protect user privacy in PSS and concomitantly design incentive schemes or check reputation of source.

1.2 Motivation

The apparently insensitive information transmitted in plain-text in PSS can be used by an eavesdropper to infer some sensitive information to threaten privacy of the observer. In this context, privacy of the user refers to her association with an OOI (Object of Intereset), such as location or purchased product or both. As discussed in Section 1.1, hiding identity cannot be a solution as identity is a requirement for assessing reliability of shared data and developing incentive mechanisms [19]. Hence, the pseudonym based schemes such as hot-potato-privacy-protection (HP^3) [20], where the report is sent via a friend network, cannot provide a feasible solution in this context. Another straightforward solution to the problem would be encrypting the data before transmitting. However, encryption is not sufficient at all. As the sensed data is usually very small in size, intelligent adversary may encrypt predicted message with known public key of application server and then match these against the received message. Moreover, the communication infrastructure of PSS uses the public data network. Hence, encrypted data transmission may unnecessarily raise concerns within the law enforcement agencies and they may ban this as well in many countries.

The existing location privacy protection mechanisms, where location information is transmitted with some anonymity or by adding Gaussian noise or at reduced precision, cannot be used here as the destination expects data accuracy at individual level. For example, PetrolWatch [2] assists drivers to find the cheapest fuel station in the neighborhood. Unless sufficient data quality/accuracy is achieved, users will not be encouraged to use the system. If PetrolWatch recommends a station with higher price for a user looking for cheapest one, its reputation would be destroyed. Therefore, a technique is needed such that each observation from a participant can be transmitted with sufficient anonymity and, at the same time, the data collector can de-anonymize individual data with acceptable accuracy. It is also expected that so long an adversary is unable to intercept reasonably high number of transmissions from the participants, any de-anonymization attempt to infer sensitive information fails [1].

A solution to address this significantly challenging problem was proposed in [1] based on a novel technique called subset-coding. The privacy of the participants who report their observations about an object/event was ensured by k-anonymization. The concept of k-anonymity states that an observation is k-anonymous if the observed object/event is indistinguishable from $k-1$ other objects. The feasibility of the technique relies on an efficient joint de-anonymization technique to obtain sufficient data quality at the desired end. However, that work was limited in application as the de-anonymization would take $O(N!)$ time where N implies the number of objects in the PSS. Consequently, for application scenarios with high value of N (such as $N > 7$) which is quite common in real world, the technique would not work. For example, if the PSS is interested to know about Indian food-shop, it may work. However, for a burger-shop, which has much higher density of occurrence, we have to use a computationally efficient alternative. The goal of this project is to design this efficient solution as well as extend the dimension of privacy protection as elaborated next.

1.3 Aims

The aim of this project is to design a privacy protecting scheme which can simultaneously achieve data integrity with acceptable computational overhead, adoptable to incentive schemes and extend the scheme for multi-dimensional scenario. Specifically, they are elaborated as follows:

- To design an efficient anonymization and decoding technique for protecting privacy and simultaneously preserving high data integrity in participatory sensing system.
- To develop algorithms that provides anonymization in multiple dimensions. For example, anonymization of both the location of purchase and product purchased concomitantly in the scenario of consumer information sharing will be designed.
- To support variable-length anonymity that allows user to prefer different anonymity for different dimensions. It eventually facilitates the design of flexible incentive mechanisms for participating users.
- To investigate the optimization issues of the proposed technique and their feasibility in real time.
- To test the proposed schemes using simulation and real-world experiments.

1.4 Contribution

The most important contribution of this project is designing a privacy scheme that can simultaneously anonymize multiple properties of the user. For example, in some consumer sharing applications, both the location privacy and association with product need to be protected. The presence near a cancer hospital as well as association with sensitive purchased product (e.g., drug for a particular disease) may also threaten privacy of the participant. To the best of our knowledge, our proposed multi-dimensional anonymization to be presented in this thesis is the first technique that provides this solution. The other contributions of the thesis are as follows:

- We present an $O(N)$ anonymization anglorithm and $O(N)$ de-anonymization algorithms in the context of PSS.

- We extend the anonymization technique for multi-dimensional scenario which is the first technique with this solution to the best of our knowledge. The order of anonymization and de-anonymization algorithm for multi-dimensional scenario is $O(d \times N)$ and $O(d \times N)$ respectively.

- Our proposed technique supports variable anonymity k i.e. different user preference of anonymity. This approach provides the user the control on their desired level of privacy. Moreover, this is useful to design incentive schemes considering users' sacrifice of privacy.

- We implemented a prototype of PSS with our proposed technique and simulated with synthesized data. We have also developed an Android Prototype as a proof of concept of our proposed scheme.

1.5 Organization

The organization of the rest of the thesis is as follows.

Chapter 2: System Model This chapter briefly describes the system model of PSS and its related terminologies. We also discuss the adversary model encountered by PSS in this chapter.

Chapter 3: Literature Review This chapter includes a brief discussion of the previously proposed privacy schemes on PSS and their shortcomings. Finally we also analyze the existing subset-coding technique and its limitations for real world applications.

Chapter 4: Single-Dimensional Scenario The proposed privacy-protecting scheme with efficient anonymization and decoding technique for single dimension is briefly described. In this chapter, first we show an example of single dimension scenario of PSS and describe our proposed techniques with the help of an example. Next we include the pseudo-code for implementing them.

Chapter 5: Multi-Dimensional Scenario This chapter describes our proposed scheme's technique for multi-dimensional scenario. In this chapter we also discuss the basic concept, example and pseudo-code for both anonymization and decoding technique in multi-dimensional scenario.

Chapter 6: Experimental Results and Discussion For evaluating the performance of our proposed schemes, we have conducted experiment on synthesized data and also made an Android prototype for evaluating its feasibility in real world. This chapter briefly describes the experimental setups and their results.

Chapter 7: Conclusion and Future Works This chapter concludes the thesis and presents some future research direction.

Chapter 2

System Model

The objective of this chapter is to explain the system model of PSS by defining some frequently used terms and system entities. This chapter also presents the Adversary Model in the context of PSS. The organization of this chapter is as follows. In Section 2.1, some important terms in context of PSS are defined followed by the presentation of the basic PSS Model. Section 2.2 presents the Anonymization Model of PSS with an example. The next Section 2.3 briefly discusses about the adversary model of PSS.

2.1 Basic PSS Model

In participatory sensing system (PSS), users (mobile nodes) report their observations to Application Server (ApS) in exchange of incentive. Example of an observation might be the petrol price in different location or temperature in different cities. We define the term Objects of Interest here to represent objects/events that a PSS is interested in.:

Definition 1. *(Object of Interest): An Object of Interest (OOI) is the object/event whose attributes/properties are observed and reported by the participants of PSS.*

Figure 2.1 shows the scenario of PetrolWatch. Here, petrol stations are OOIs and price of petrol is the relevant attribute. People with smartphone captures fuel price and report it to the Application Server (ApS). To encourage participation, ApS can give incentive/reward back to the user in exchange of his/her shared data. Thus by analyzing all the collected reports of fuel price, the ApS can answer any user's query regarding fuel price, e.g., cheapest fuel station in neighborhood.

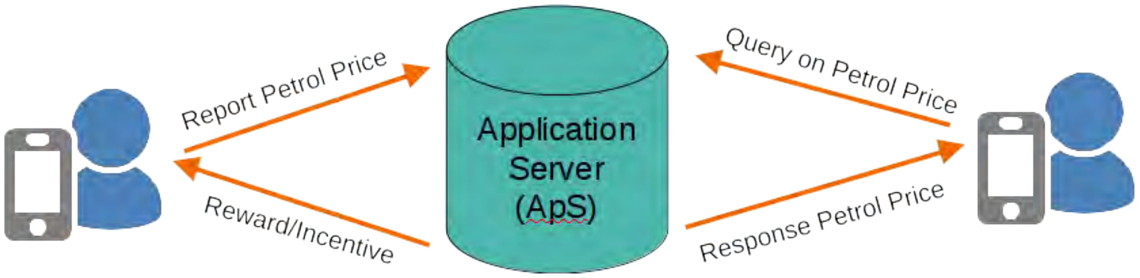


Figure 2.1: System Model of PSS (PetrolWatch).

2.2 Anonymization Model

In PSS, participants share their private information, e.g., location data to Application Server. To protect the privacy of the user, several privacy-protecting schemes have been proposed. Among them, the most popular scheme is anonymization. In this approach, an Anonymization Server (AS) is used in PSS. This server remains transparent from Application Server (ApS) which actually provides desired service to the community. ApS also computes incentive for participants and may monitor reputation of data source.

Thus, the whole cycle of participatory sensing in the perspective of user can be divided into two parts:

1. Anonymization Step: User sends actual observation to Anonymization Server (AS) that transforms and returns an anonymized report (AR). Since incentive or reputation schemes do not rely on this communication (between user and AS), user association with these reports need not be preserved. Therefore, hiding user identity is easier in this step and techniques from literature such as use of friend network can be adopted [1].

2. Reporting Step: User sends anonymized report to ApS along with his/her identity information. ApS has to de-anonymize these reports to decode the real value in order to offer desired service to the participants.

The observed report which is sent by the user to the AS for anonymization contains user preference of anonymity (denoted as k), OOI identification and the observed property/attribute. For example, a participant John observes the price of camera. He wants to report camera's price with 3-anonymity, i.e., $k = 3$. He is supposed to send the report $< Camera\{3\} >$: \$100 to the AS. Now, AS may anonymize his report as $< Camera, Phone, GPS >$: \$100 and returns this anonymized report (AR) to John. Then John sends this AR to the ApS with his identity. Hence, the information that John might have bought a camera can be protected.

As alluded in the previous chapter, users' association with multiple objects/events can also be protected simultaneously. Let John reports the price of camera on a particular location loc1 and wants to anonymize his report both in terms of location and product simultaneously. Hence, there are two types of OOIs for two dimensions. We use N and S to denote the total number of OOIs and the set of all OOIs, respectively in a single dimensional PSS. For d-dimensional PSS, the total number of OOIs for all dimensions are denoted as $N_1, N_2, \cdots, N_d$ and their sets as $S_1, S_2, \cdots, S_d$. Even the anonymity preference for all dimensions may be different. We use $k_1, k_2, \cdots, k_d$ to denote the anonymity preference. Suppose, John sends report to the AS as $< Camera\{2\}, loc1\{3\} >$: \$100 shown in Figure 2.2. This format denotes that $k_1 = 2$ and $k_2 = 3$ i.e. John wants to anonymize his report by adding two-anonymity on observed product (camera) and three-anonymity on his location (loc1). In this situation, a valid AR sent by AS might be $< \{Camera, Phone\}, \{loc1, loc2, loc3\} >$: \$100. This notion can be expressed in general term with the following definition.

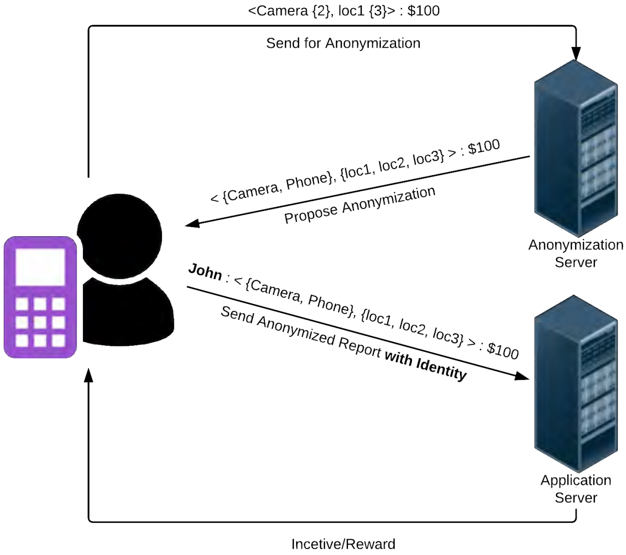

<Camera {2}, loc1 {3}> : $100
Send for Anonymization
< {Camera, Phone}, {loc1, loc2, loc3} > : $100
Propose Anonymization
Anonymization Server
John : < {Camera, Phone}, {loc1, loc2, loc3} > : $100
Send Anonymized Report with Identity
Application Server
Incetive/Reward

Figure 2.2: System model of a multi-dimensional anonymization process in PSS.

Definition 2. *(Anonymized Report): An Anonymized Report (AR) for an observation report* $< OOI_{i_1}\{k_1\}, OOI_{j_1}\{k_2\}, \cdots, OOI_{d_1}\{k_d\}, >$*: v is formed as* $< \{OOI_{i_1}\} \cup \{OOI_{i_2}\} \cup \cdots \cup \{OOI_{i_{k_i}}\}, \{OOI_{j_1}\} \cup \{OOI_{j_2} \cup \cdots \cup OOI_{j_{k_j}}\}, \cdots, \{OOI_{d_1}\} \cup \{OOI_{d_2} \cup \cdots \cup OOI_{d_{k_d}}\} >$*: a such that each* $OOI_{i_j} \in S_i$.

It is evident from the discussion above that the task of anonymization is basically to select some extra OOIs from the relevant available alternatives along with the real OOI according to the user's preference of anonymity. The quality of an anonymization algorithm is measured by the reduction of the number of required ARs to fully decode at the Application Server (ApS).

In our techniques, we assume that each OOI has a unique attribute; which may not be practical in some scenarios. However, it is established in [1] that the transformation of the non-unique scenario to the unique scenario can be accomplished by the AS. If AS receives an observation with an attribute that is duplicate, i.e., matches a different OOI's attribute, it adds a small insignificant value to the new attribute to make it unique.

2.3 Adversary model

The other remaining component of the system is the adversary model. The main objective of a malicious adversary is to reveal the location information as well as the sensitive information of the user. Moreover, it is quite natural that the adversary has close access to victim in real world. Therefore, the victim's user id might be known to adversary. We explain different types of adversary in Section 2.3.1 and discuss strategies taken against them in Section 2.3.2.

2.3.1 Different Types of Adversaries

In a single flow of reporting as discussed in Section 2.2, user requires to send information twice which cause potential privacy risks. We define the following terms to refer these communications:

- **RAS (Reports towards AS):** User sends report to AS without user identity. (i.e., RAS)
- **RApS (Reports towards ApS):** User sends anonymized report to ApS.

Malicious adversaries can eavesdrop messages in any part of these communications. Based on this eavesdrop messages the adversary can receive, adversary can mainly be divided into three types.

- **Type 1 Adversary (RAS):** This type of adversary may eavesdrop the reports which are sent to AS without user identity for anonymization purpose (shown in Figure 2.3). As it is pre-requisite for security surveillance that the communication can not be encrypted, we may assume that the service of AS is provided by some trusted third party e.g., a government agency.

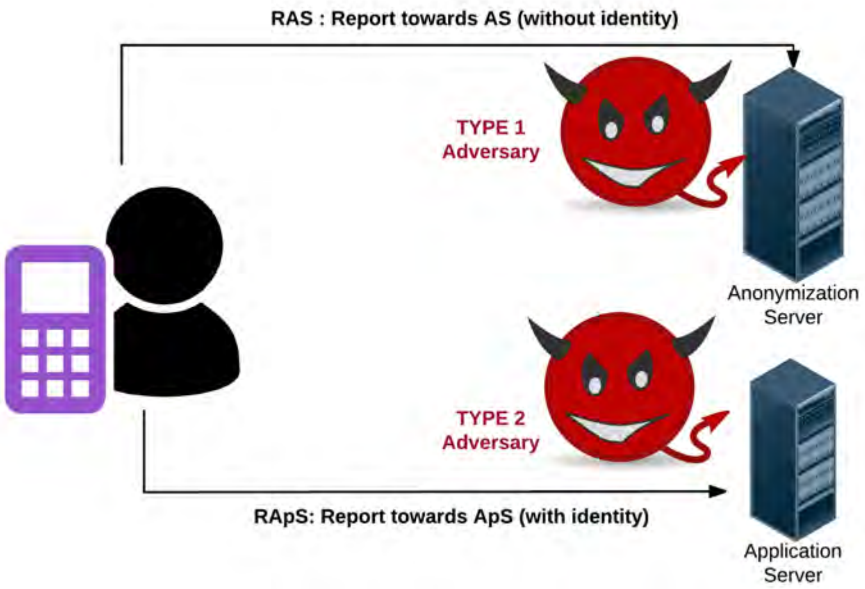

RAS : Report towards AS (without identity)
TYPE 1 Adversary
Anonymization Server
TYPE 2 Adversary
RApS: Report towards ApS (with identity)
Application Server

Figure 2.3: Different types of adversaries.

- **Type 2 Adversary (RApS):** Residing near ApS, adversary can receive quite a good number of RApS (shown in Figure 2.3). If the adversary is strong enough equipped with same decoding application as ApS, it is likely to infer the OOI value and thus reveal user-OOI association. This adversary is more stronger than others because of its decoding capability.

- **Type 3 Adversary (Compromised Friend Network + RApS):** Though RAS is not sent with user identity, it might be possible to trace back the originator's identity of the report. Hence, Mix network is used so that the reporter's identity cannot be traced back which is discussed in Section 3.2.1. Type 3 Adversary might reside in friend network. Thus it can receive certain amount of messages and learn some attributes of some OOI. This type of adversary becomes more malicious if it colludes with Type 2 adversary to improve decoding accuracy.

2.3.2 Strategies against Adversaries

In this section, we briefly explain different strategies to mitigate the risks against the three types of adversaries discussed above:

2.3.2.1 Strategies Against Type 1 Adversary

Type 1 adversary can eavesdrop the unanonymized reports (RAS) of users without identity which poses potential threat to privacy. This adversary might try to trace back the originator's identity from the eavesdropped messages. Using mix network, this privacy threat can be mitigated. However, from the RAS, adversary may learn the attributes of OOI directly which poses a significant privacy threat. The strategy taken against the Type 2 adversary can be applied here to mitigate the risk which is discussed in next Section.

2.3.2.2 Strategies Against Type 2 Adversary

Type 2 adversary is stronger enough because it has the decoding capability. Thus this adversary can learn attributes by decoding the eavesdropped RApS it received and thus reveals the OOI-user association. The strategy against this adversary is to divide the anonymization tasks among different anoymization servers instead of a single AS. The OOIs might be divided into two or more parts and assigned to different anonymization servers. In case of multi-dimensional scenario, any single dimension (i.e. location) can be divided among the AS while other dimensions remain same for all AS. As in real world, a single user is most likely to report information nearby his/her area, a user will be allowed to report from only one group with a single user id. Therefore, when a user is registered in PSS, it will be assigned to a group of OOI e.g. a single AS. If a user frequently travels to another area, he may register with another user id for that area. In this situation, the Type 2 adversary cannot distinguish the reports from different group of users as only ApS knows the user-AS association. So the attempt of decoding correct attributes of OOI of the adversary miserably fails. Moreover, in this strategy, users of different groups send RAS to their corresponding AS. Hence, Type 1 adversary can never get the complete information of the whole scenario by receiving good amount of RAS.

2.3.2.3 Strategies Against Type 3 Adversary

This type of adversary is a part of friend network and thus can learn attributes of some OOI. If it colludes with Type 2 adversary, it becomes more stronger. However, the strategy taken against the Type 2 Adversary significantly reduces the privacy threat. Because, though the adversary in friend network can receive good amount of RAS and thus learn some attributes, it cannot get the total information of the whole scenario because of different user groups. Moreover, as it cannot infer the user group of RApS, the previously learned attributes cannot help much in decoding improvement.

2.3.3 Summary

In summary, the stronger type of adversary (Type 2) assumes position near the ApS and tries to capture as many reports as possible and decodes them jointly using the same decoding algorithm used by the application server. These adversaries can be tackled by dividing OOIs in a region into several groups, each served by its dedicated AS, and allowing a user to either sense OOIs from only one group or use separate user id for each group. The other type of adversary (Type 1) that would try to trace an unanonymized report on its way to AS, was also befooled in [1] by employing random sequence of friend network.

2.4 Conclusion

In this chapter, we have demonstrated the basic system model of PSS. We have also defined some terms and concepts related to anonymization model of PSS which will be used frequently throughout this thesis. In the last section of this chapter, we have discussed about the Adversary model for PSS. The next chapter will briefly discuss on contemporary research works on privacy protection for PSS.

Chapter 3

Literature Review

Privacy protection is one of the most vital factors for the success of PSS. Hence, various types of approaches have been proposed in order to protect the privacy of the participants. In this chapter, the contemporary researches on the privacy-preserving techniques have been briefly explained.

The chapter is organized as follows. Section 3.1 categorizes the privacy concerns of participants in PSS. Section 3.2 briefly explains different techniques for preserving these privacy. This section also criticizes the existing techniques by referring to their limitation in real-world scenarios. Section 3.3discusses about subset-coding approach that provides state-of-the-art solution to the inferable privacy problem of PSS users. Finally Section 3.4 concludes the chapter.

3.1 Privacy in PSS

Assurance of privacy is the key factors for maintaining adequate participants in PSS [21]. Participants have to share their private information such as identity, location and other sensitive information which cause a great privacy threat. Privacy concern can be classified as follows with the help of Figure 3.1.

Ownership Privacy The identity of the participant is revealed to ApS while reporting data. There are many privacy schemes by which the ownership privacy can be achieved. However, if the ownership privacy is ensured, the ApS cannot give incentive/reward to the participant in exchange of his/her shared data. Hence, the currently proposed privacy schemes give more focus on Data privacy than the ownership privacy.

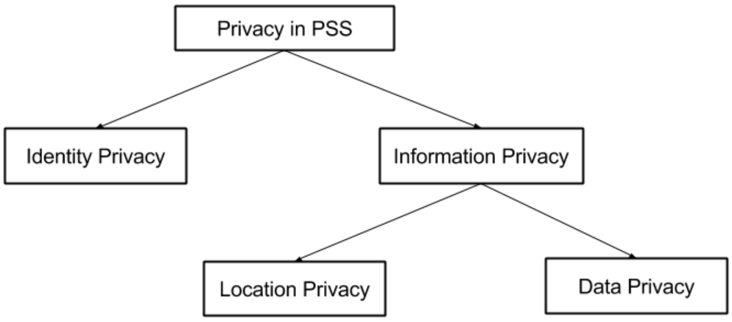


Figure 3.1: Types of privacy concerns in PSS.

Information Privacy Sensitive information such as location data of the observers is sent to ApS is also an important privacy concern. Even if the identity of the participant is hidden, background knowledge of participant may help the adversery to identify the participant. The data privacy can be divided into two categories:

- Location Information
- Sensitive Information (like weight of a person)

3.2 Privacy Preserving Approaches

Numerous research works have been conducted using quite a few techniques to protect the privacy (both identity and information) of participants in PSS. Some of these approaches depend on trusted third party while others use anonymization or friend network for privacy protection. The approaches can be classified into two types shown in Figure 3.2.

Unmodified Data In these approaches the ApS can retrieve the exact data from the recieved reports from participants. These approaches gurantee the complete data integrity.

Modified Data In these approaches, ApS cannot directly get the data from the reports. Here, ApS recieves noisy, anonymized or aggregated data from which ApS has to decode the real data or analyze patterns. Most of the time, these approaches degrade the data quality due to the additional noise added in the data.

Various types of privacy-preserving approaches have been proposed under these two categories to protect identity or location/information privacy. However, most of the approaches suffer from some challenging trade-off like anonymity vs. data integrity. These approaches are briefed shortly in this Section 3.2.

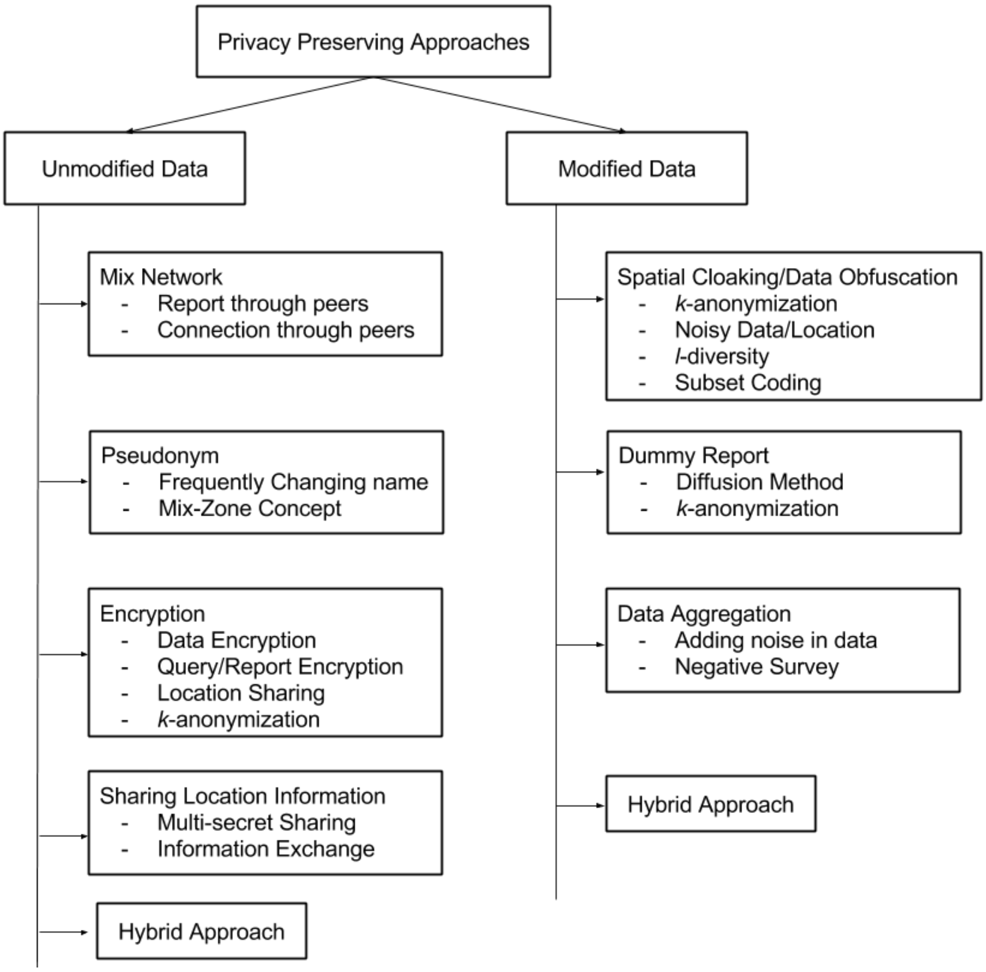


Figure 3.2: Categorization of popular privacy preserving approaches in PSS.

3.2.1 Mix Network

3.2.1.1 Reporting through peers

Onion Routing is a distributed overlay network where users can choose path to build a circuit where each node knows its predecessor and successor but no other nodes in the network. TOR [22], a second generation onion router, was designed to provide low-latency anonymity in TCP-based applications. This mix-network based concept have been used to protect location privacy in PSS. Hot-Potato-Privacy-Protection (HP^3) [20] scheme is designed based on this concept where user sends data to one of his/her friend and that friend chooses another friend to deliver the data to next hop. This process continues until the user-defined threshhold is reached. Then the last user sends the report to ApS. However, this approach has to bear higher latency to achieve strong anonymity. LAP [23] is proposed by Hsiao et. al which improves this latency and minimizes computational overhead. This scheme conceals end-hosts topological location to enhance anonymity against remote tracking. Each packet carries its forwarding state and only packet header is encrypted instead of whole packet. But this approach considers a weaker adversary model that attacks any peer except the reporting peer. So, this approach is appropriate where PSS demands more privacy from further located ISPs than the neighborhood peers.

3.2.1.2 Connection through peers

HP^3 [20], LAP [23] approaches transmit the actual payload through peers which increases the bandwidth and computing overhead. Wang and Ku [24] proposed a mix-network based scheme, where only the connection request is sent through peers while the actual payload is directly sent to ApS anonymously. This approach is called one-way anonymous protocol as ApS cannot send any acknowledgement to the participant in reply. Compared to other mix-network based solutions, this scheme consumes less bandwidth and computational power and simulteneously achieves secure data transmission and scalability.

3.2.1.3 Limitation

Mix-network based schemes suffer from delays due to slow network connection in volunteer peers. The bandwidth cost in these schemes are comparatively higher in these approaches. One-way mix-network based protocol like [24] improves the latency but cannot provide incentive/reward to the reported user and hence may suffer from low participation. Trustworthiness of peers is a big challenge in mix network. When there are limited number of hops, every peer becomes a suspect of malicious user. When the shared data is encrypted, the volunteer peers may transmit illegal data unknowingly.

3.2.2 Pseudonym

3.2.2.1 Frequent change of pseudonym

Pseudonym is the most common approach to protect identity privacy from ApS. However, long term pseudonym tends to be identified easily by adversary. Changing pseudonym is an ineadequate solution to this problem because linking old and new password is not difficult in high spatial and temporal resolution.

3.2.2.2 Mix-Zone

Beresford and Stajano [25] introduced mix-zone concept to protect users from this privacy attack. Mix Zone is defined as a group of users as a connected spatial region in which none of them have registered any application callback.In frequently chaning pseodonymous environment, adversary which see an user coming out from mix zone cannot distinguish from other users in mix zone. The size of the mix zone provides the anonymity strength of PSS. So, user might refuse to share data until the mix zone can provide minimum level of anonymity. MobiMix [26] approach was proposed also using this mix-zone concept. TrPF [27] has been proposed by Gao et al. which improves mix-zone concept for facilitating trajectory privacy preservation. This framework uses trajectory mix-zone model with low information loss to protect trajectory privacy. [28] presented a way to transfer user reputation information in a pseudonymous environment so that privacy and data trustworthiness can be simultaneously facilitated.

3.2.2.3 Limitation

Using historical data, strong adversary can detect the user's identity in pseodonymous environment. Mix Zone concept may suffer from low level of anonymity in high spatial-temporal resolution. Moreover, [29] the combined analysis of reported data with pattern may help to detect user.

3.2.3 Encryption

Encryption is a very common approach for protecting location and data privacy in PSS.

3.2.3.1 Location Sharing

Multi-secret sharing [30] concept is used in many privacy protecting schemes (detail in 3.2.4.1). In these schemes like [31], cryptography is used for sharing the locaiton data to buddies or location servers. In [32], users share their location to his/her friends with proximity preference using symmetric encryption techniques.

3.2.3.2 Data Encryption

For protecting the privacy of data reported by user, many cryptographic schemes have been proposed. For semi-trusted server, E. De Cristofaro and A. Durussel [33] presented an architecture of PSS where application server gets only encrypted data from users and blindly performs computation on the encrypted data. This approach is suitable for discovering common/shared interests among people or private scheduling withoug exposing any location information of user. However, encryption on small sensed data in energy-constrained devices is not feasible in real world. To report large multimedia data, an erasure coding based scheme named SLICER [34] has been proposed. In this scheme, the sensing record is sliced and each slice is transferred via other k participants or generator itself.

3.2.3.3 Query/Report Encryption

G. Ghinita, P. Kalnis [35] proposed a framework for protecting privacy of users in location-based queries based on Private Information Retrieval (PIR) theory. This framework achieves location privacy without any help of anonymizer or trusted third party. This scheme allows user to access the database of application server privately so that server does not know the location information of user. Using this approach, they also developed an algorithm for finding nearest neighbor. This approach solves the correlation attack for the first time. E. De Cristofaro; C. Soriente presented another approach named PEPSI [36] where queries or data reports are labled with specific keywords/id and users register to the system to get the authorized keyword/id corresponding to their data report/query. This framework uses Identity-Based Encryption (IBE) to enable non-interactivity in query protocol design.

3.2.3.4 k-anonymization

k-anonymization is very popular approach for protecting privacy in PSS. Cryptographic techniques have been applied in many proposed schemes to achieve k-anonymity. In [37], cryptographic scheme has been used to gurantee k-anonymity without any use of location broker, third party or trusted peers. In this approach, user learn whether k-anonimity is ensured within a query area with the help of location provider (e.g., cell phone operator). LotS [38] proposed by A. Michalas and N. Komninos also uses cryptographic techniques to maintain k-anonimity. In this scheme, user registers to the system and joins a group using public-key cryptography. Hence this ensures the unlinkability between the real identity of user and his/her reports and thus can support users' reputation by combining voting approach.

3.2.3.5 Limitation

Cryptographic encryption scheme is used to prevent external and internal attack which is a limitation for energy-constrained mobile devices. Specially, for reporting large data like multi-media data encryption schemes costs energy, network and computational overhead. In many location-sharing approaches like [31, 32, 39] assumes that location servers are always online. Also these approach requires overhead of secret-setup between user and his frinds or location servers. In [33,34], where computation is done blindly over encrypted data assumes that both server and the users are rational. Most of the above cryptographic operations are not optimized. And incentive mechanism cannot be applied here as the data is anonymously sent to Server.

3.2.4 Sharing Location Information

3.2.4.1 Multi Secret Sharing

Multi-secret sharing [30] is a concept where some arbitrarily related secrets are shared among a set of participants who are not trusted individually. This concept is used for protecting location privacy in [31] where cryptographic approach is used for secret sharing. S. Mascetti, D. Freni proposed two protocols named C/-Hide& Seek and C/-Hide& Hash in [32] using this multi-secret sharing concept. In this approach, users share their location to his/her friends with proximity preference and thus control their privacy with respect to his friends. PShare [39] is proposed by M. Wernke and F. Durr where private location information is split up into position shares of limited precisions and then distributed to different non-trusted location servers. Then user can combine several shares queried from different location servers to satisfy users' requirements of high precision. PShare [39] uses discretized square-based space representation and the precision change between the nearest privacy levels cannot be smaller than factor 4. This limits the users' control on privacy preference. Moreover, users require to know the sophisticated cryptogrpahic functions of PShare. In order to overcome the limitations of PShare, PShare has been extended in [40] for secure management of location data stored by location servers. In this approach, each position share recieved by a location server has only a position with limited precision and it can be used for calculation. Share Fusion algorithm has been proposed to combine multiple shares into positions of higher precision. In PShare [39], adverseries may attempt to increase the location precision from a sub-set of shares. This attack has been prevented in [41]. Here Share generation algorithm has been proposed which takes map knowledge into account. to prevent adversaries attempting to increase the location precision.

3.2.4.2 Information Exchange

In PSS, users often takes sensor data and report to the server and thus his/her trajectory gets revealed. To overcome this situation, exchanging report has been proposed in [42]. In this scheme, users exchange his/her collective sensor readings with another user when they physically meet. Thus the user's trajectory path gets jumbled. By repeatation of this exchange, user can adjust the level of privacy and latency. To identify malicious users in this scheme, TrustMeter [43] has been proposed which assesses the user contribution as well as trust levels. I. Boutsis and V. Kalogeraki propsed another low-cost information exchange strategy in [44] named LOCATE (LOCation-based middlewAre for TrajEctory databases). Here user data is distributed among multiple users in local user databases. This distribution of location makes impossible for an attacker to breach the privacy of user.

3.2.4.3 Limitation

Information exchange strategies [42,43] require high trust among the participants or gurantee of physical meeting of another participant which make these systems impractical. Moreover, LOCATE [44] assumes that user data is stored in mobile phones which is not applicable for energy and memory-constrained devices. However, these information exchange strategies costs high communication and sharing overhead causing high network consumption. Availability of location servers is another challenge in multi-secret sharing schemes [39–41].

3.2.5 Dummy Report

Introducing synthetic/dummy report is also another approach for achieving location privacy in PSS. The dummy location or reports can be generated by either the Application Server or the user himself.

3.2.5.1 Server Generated Dummy

G. Lee and W. Kim proposed location diffusion method [45] which scatters the location information. Additionally, in this approach, dummy messages are transmitted periodically by the base station or Application Server to confuse the attacker. This flooding of dummy messages hide the real traffic and makes difficult for the attacker to distinguish among the real and dummy traffic. As the dummy messages have no real meaning, these are discarded by the location server.

3.2.5.2 User Generated Dummy

Another way of confusing the attacker as well as protection of location privacy from application server is generating dummy by user himself. H. Kido and Y. Yanagisawa proposed a technique in [46] for location-based services, where user reports several false positions along with his/her true position to the service provider without any help from 3rd party. Then user simply extracts the necessary information from the replies from the server. They also proposed a cost reduction technique for communication between client and server. P. Shankar and V. Ganapathy proposed an autonomous k-anonymity based client-side tool in [47] where user can generate $k - 1$ sybil query along with his/her real query. Several extensions like andomizing path selection, handling active adversaries, endpoint caching, providing path continuity, and adding GPS sensor noise have been proposed in order to generate more realistic dummy query for protection against potential attacks.

3.2.5.3 Limitation

The main challenge of dummy reporting based schemes is to generate realistic dummy reports with less computational overhead. These schemes also have communication overhead for transmitting extra dummmy queries and their replies. In [47], there is an additional computational overhead for handling different databases involved in the different steps.

3.2.6 Data Aggregation

Data Aggregation is another popular technique for preserving data privacy in PSS. In real life, there are some PSS scenarios where Application Server is not interested on individual input data rather than interested to find aggregated result e.g. average or maximum value of temperature. PriSense [48] has been proposed which is based on data slicing and mixing and supports non-additive aggregation functions too like average, min/max, histogram etc. Negative Survey have been used in [49] for protecting privacy for multi-dimensional data aggregation.

3.2.7 Obfuscation

Obfuscation was first introduced in [50] as a new technique to safeguard location privacy which degrades quality of service. k-anonymity based location-privacy schemes have been proposed in [51,52]. In these approaches, the $k-1$ participants are selected through a third-party or other participants which may suffer from privacy attack of adversary participants or third-party. To address this challenge, a distributed k-anonymity based scheme has been proposed [53] where participants cloak their location data and then use private set interaction mechanism in order to match cloaked regions without disclosing their exact location to third party or other participants. But these approaches incorporate delay in real time op eration and not suitable where fine-grained information is required.

3.2.8 Hybrid

A hybrid privacy-preserving mechanism has been proposed in [54] that combines anonymization, data obfuscation, and encryption techniques to increase the privacy of the users while improving the quality of information and the energy consumption. Ensuring user privacy and data trustworthiness are two conflicting challenges in PSS [55]. ARTSense [56] designs trust assessment algorithms to compute the trust of sensing reports based on anonymous user reputation while maintaining privacy of the users. Wang et al. [57] proposed a framework to dynamically assess the trustworthiness of information as well as the participants. In [58] both privacy and incentive issues have been addressed using token-based authentication and blind signature. Here, task and credit are transferred through real id while the reports and credit reciept are transferred anonymously. Still, this approach can cause credit-based inferable attack. IncogniSense [59] scheme addresses this challenge by periodically changing pseudonym and dynamically cloaking the reputation score. But this approach is not feasible in real life for its additional management overhead and heavy communication cost.

3.3 Subset-Coding Based Techniques

All these privacy schemes described so far cannot achieve data quality, location and context-based privacy, latency and energy-constraint simultaneously. Subset-coding [1] has been proposed to solve this challenging work by introducing greedy k-anonymization with efficient joint de-anonymization. In this approach, Anonymization server (AS) generates anonymized report (AR) and Application server (ApS) decodes the real attributes of OOI from the recieved ARs from the users. Three types of approaches named as BGAS, DGAS and EGAS have been proposed in [1] as the anonymization technique used in subset-coding shown in Figure 3.3 (figure borrowed from [1]). DGAS is the best among the three and in our work we use the philosophy of this approach. Hence, it is discussed in detail in the subsequent section.

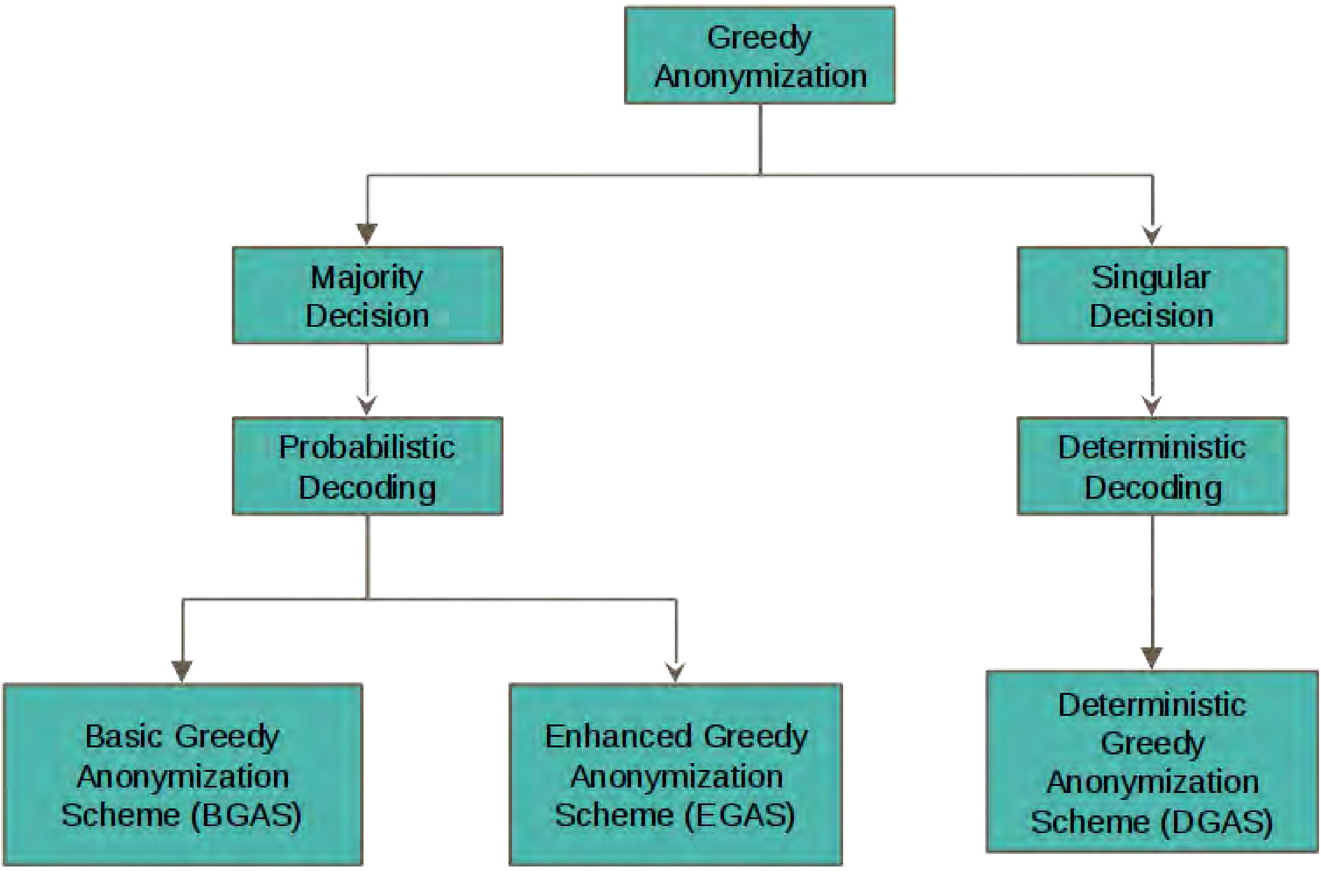

Greedy Anonymization
Majority Decision
Singular Decision
Probabilistic Decoding
Deterministic Decoding
Basic Greedy Anonymization Scheme (BGAS)
Enhanced Greedy Anonymization Scheme (EGAS)
Deterministic Greedy Anonymization Scheme (DGAS)

Figure 3.3: Different types of subset-coding approaches.

3.3.1 Concept of DGAS

Here we explain the basic concept of subset-coding (DGAS approach) with an example. Let us consider an PetrolWatch scenario with $N = 4$ where Application server wants to know the petrol price in 4 different petrol pumps with id $1, 2, 3$ and 4 and the desired anonymity $k = 3$. Both AS and ApS have prior knowledge of all the OOI. However, they do not know their corresponding attributes. First they assume the unique attributes for the OOI as v_1, v_2, v_3 and $, v_4$. However, the mapping between the OOI and the attributes are unknown to both AS and ApS. There might be $N! = 24$ possible mapping between OOIs

Serial	Report
1	$A : \$10$
2	$C : \$30$
3	$B : \$20$
4	$D : \$40$
5	$B : \$20$
6	$B : \$20$
7	$A : \$10$

Table 3.1: User reports according to their arrival order.

and attributes. Any of the mapping is the real mapping which have to be decoded by ApS. The user reports according to their arrival order are shown in Table 3.1.

Whenever the AS recieves an report according to the order shown in Table 3.1, the AS tries to anonymize such that the cardinality of its mapping C reduces as much as possible. In our example, the first report recieved by AS is A : \$10. After recieving this report, AS substitutes any of its assumed attributes (which has not been yet substituted) as \$10. Here we assume that AS substitutes v_1 as \$10. Now, as $k = 3$, there are three possilbe AR generation is possible i.e. $< A, B, C >$: \$10, $< A, B, D >$: \$10 and $< A, C, D >$: \$10. Among them, if AS chooses $< A, B, C >$: \$10, then ApS will be convinced that D OOI cannot have the attribute \$10. According to this logic, if AS chooses this AR, then the mappings where D's attribute is v_1 i.e. \$10 will be removed. There are 6 such mappings where $D = \$10$ shown in Table 3.2a. So, by choosing this AR, ApS will be able to reduce the cardinality by 6 from its possible mappings. With the same logic, if AS chooses $< A, B, D >$: \$10 or $< A, C, D >$: \$10, the cardinality reduction is 6 shown in Table 3.2a. As in all these cases, the cardinality reduction (CR) is 6, AS chooses any of them randomly. After receiving this AR, ApS will have total $24 - 6 = 18$ mappings.

The second report recieved is C : \$30. In this case, AS have to choose any of the three ARs among $< A, C, D >$: \$30, $< A, B, C >$: \$30 and $< B, C, D >$: \$30. Now, if AS chooses $< A, B, C >$: \$30, then the cardinality reduction is 6. Because there are 6 mappings in the current mapping sets where $D = \$30$. But if AS chooses the other two i.e. $< A, C, D >$: \$30 or $< B, C, D >$: \$30, then the cardinality reduction is 4 in each case explained in Table 3.2b. So, the AS chooses the first AR i.e. $< A, B, C >$: \$30 so that the maximum possible mapping can be removed and the real mapping can be found as soon as possible. So, whenever an user report is arrived, AS chooses the AR which gives the maximum possible cardinality reduction CR. The mapping and the possible CR count is explained for $3 - 7$ user reports

Table 3.2: Conforming tuples of gradually generated AR Set where dummy attributes are identified with leading v

(a) 1st Observation

Mappings	1^{st} Observation: A : \$10			selected ARs
	$< A, B, C >$: \$10	$< A, B, D >$: \$10	$< A, C, D >$: \$10	
(v_1, v_2, v_3, v_4)	$(\$10, v_2, v_3, v_4)$	$(\$10, v_2, v_3, v_4)$	$(\$10, v_2, v_3, v_4)$	
(v_1, v_2, v_4, v_3)	$(\$10, v_2, v_4, v_3)$	$(\$10, v_2, v_4, v_3)$	$(\$10, v_2, v_4, v_3)$	
(v_1, v_3, v_2, v_4)	$(\$10, v_3, v_2, v_4)$	$(\$10, v_3, v_2, v_4)$	$(\$10, v_3, v_2, v_4)$	
(v_1, v_3, v_4, v_2)	$(\$10, v_3, v_4, v_2)$	$(\$10, v_3, v_4, v_2)$	$(\$10, v_3, v_4, v_2)$	
(v_1, v_4, v_2, v_3)	$(\$10, v_4, v_2, v_3)$	$(\$10, v_4, v_2, v_3)$	$(\$10, v_4, v_2, v_3)$	
(v_1, v_4, v_3, v_2)	$(\$10, v_4, v_3, v_2)$	$(\$10, v_4, v_3, v_2)$	$(\$10, v_4, v_3, v_2)$	
(v_2, v_1, v_3, v_4)	$(v_2, \$10, v_3, v_4)$	$(v_2, \$10, v_3, v_4)$	~~$(v_2, \$10, v_3, v_4)$~~	
(v_2, v_1, v_4, v_3)	$(v_2, \$10, v_4, v_3)$	$(v_2, \$10, v_4, v_3)$	~~$(v_2, \$10, v_4, v_3)$~~	
(v_2, v_3, v_1, v_4)	$(v_2, v_3, \$10, v_4)$	~~$(v_2, v_3, \$10, v_4)$~~	$(v_2, v_3, \$10, v_4)$	
(v_2, v_3, v_4, v_1)	~~$(v_2, v_3, v_4, \$10)$~~	$(v_2, v_3, v_4, \$10)$	$(v_2, v_3, v_4, \$10)$	
(v_2, v_4, v_1, v_3)	$(v_2, v_4, \$10, v_3)$	~~$(v_2, v_4, \$10, v_3)$~~	$(v_2, v_4, \$10, v_3)$	
(v_2, v_4, v_3, v_1)	~~$(v_2, v_4, v_3, \$10)$~~	$(v_2, v_4, v_3, \$10)$	$(v_2, v_4, v_3, \$10)$	
(v_3, v_1, v_2, v_4)	$(v_3, \$10, v_2, v_4)$	$(v_3, \$10, v_2, v_4)$	~~$(v_3, \$10, v_2, v_4)$~~	
(v_3, v_1, v_4, v_2)	$(v_3, \$10, v_4, v_2)$	$(v_3, \$10, v_4, v_2)$	~~$(v_3, \$10, v_4, v_2)$~~	
(v_3, v_2, v_1, v_4)	$(v_3, v_2, \$10, v_4)$	~~$(v_3, v_2, \$10, v_4)$~~	$(v_3, v_2, \$10, v_4)$	
(v_3, v_2, v_4, v_1)	~~$(v_3, v_2, v_4, \$10)$~~	$(v_3, v_2, v_4, \$10)$	$(v_3, v_2, v_4, \$10)$	
(v_3, v_4, v_1, v_2)	$(v_3, v_4, \$10, v_2)$	~~$(v_3, v_4, \$10, v_2)$~~	$(v_3, v_4, \$10, v_2)$	
(v_3, v_4, v_2, v_1)	~~$(v_3, v_4, v_2, \$10)$~~	$(v_3, v_4, v_2, \$10)$	$(v_3, v_4, v_2, \$10)$	
(v_4, v_1, v_2, v_3)	$(v_4, \$10, v_2, v_3)$	$(v_4, \$10, v_2, v_3)$	~~$(v_4, \$10, v_2, v_3)$~~	
(v_4, v_1, v_3, v_2)	$(v_4, \$10, v_3, v_2)$	$(v_4, \$10, v_3, v_2)$	~~$(v_4, \$10, v_3, v_2)$~~	
(v_4, v_2, v_1, v_3)	$(v_4, v_2, \$10, v_3)$	~~$(v_4, v_2, \$10, v_3)$~~	$(v_4, v_2, \$10, v_3)$	
(v_4, v_2, v_3, v_1)	~~$(v_4, v_2, v_3, \$10)$~~	$(v_4, v_2, v_3, \$10)$	$(v_4, v_2, v_3, \$10)$	
(v_4, v_3, v_1, v_2)	$(v_4, v_3, \$10, v_2)$	~~$(v_4, v_3, \$10, v_2)$~~	$(v_4, v_3, \$10, v_2)$	
(v_4, v_3, v_2, v_1)	~~$(v_4, v_3, v_2, \$10)$~~	$(v_4, v_3, v_2, \$10)$	$(v_4, v_3, v_2, \$10)$	
CR	6	6	6	

Mappings	2^{nd} Observation: $C : \$30$			selected ARs
	$< A, B, C >: \$30$	$< A, C, D >: \$30$	$< B, C, D >: \$30$	
$(\$10, v_2, v_3, v_4)$	$(\$10, v_2, \$30, v_4)$	$(\$10, v_2, \$30, v_4)$	$(\$10, v_2, \$30, v_4)$	$< A, B, C >: \$10$
$(\$10, v_2, v_4, v_3)$	~~$(\$10, v_2, v_4, \$30)$~~	$(\$10, v_2, v_4, \$30)$	$(\$10, v_2, v_4, \$30)$	
$(\$10, v_3, v_2, v_4)$	$(\$10, \$30, v_2, v_4)$	~~$(\$10, \$30, v_2, v_4)$~~	$(\$10, \$30, v_2, v_4)$	
$(\$10, v_3, v_4, v_2)$	$(\$10, \$30, v_4, v_2)$	~~$(\$10, \$30, v_4, v_2)$~~	$(\$10, \$30, v_4, v_2)$	
$(\$10, v_4, v_2, v_3)$	~~$(\$10, v_4, v_2, \$30)$~~	$(\$10, v_4, v_2, \$30)$	$(\$10, v_4, v_2, \$30)$	
$(\$10, v_4, v_3, v_2)$	$(\$10, v_4, \$30, v_2)$	$(\$10, v_4, \$30, v_2)$	$(\$10, v_4, \$30, v_2)$	
$(v_2, \$10, v_3, v_4)$	$(v_2, \$10, \$30, v_4)$	$(v_2, \$10, \$30, v_4)$	$(v_2, \$10, \$30, v_4)$	
$(v_2, \$10, v_4, v_3)$	~~$(v_2, \$10, v_4, \$30)$~~	$(v_2, \$10, v_4, \$30)$	$(v_2, \$10, v_4, \$30)$	
$(v_2, v_3, \$10, v_4)$	$(v_2, \$30, \$10, v_4)$	~~$(v_2, \$30, \$10, v_4)$~~	$(v_2, \$30, \$10, v_4)$	
$(v_2, v_4, \$10, v_3)$	~~$(v_2, v_4, \$10, \$30)$~~	$(v_2, v_4, \$10, \$30)$	$(v_2, v_4, \$10, \$30)$	
$(v_3, \$10, v_2, v_4)$	$(\$30, \$10, v_2, v_4)$	$(\$30, \$10, v_2, v_4)$	~~$(\$30, \$10, v_2, v_4)$~~	
$(v_3, \$10, v_4, v_2)$	$(\$30, \$10, v_4, v_2)$	$(\$30, \$10, v_4, v_2)$	~~$(\$30, \$10, v_4, v_2)$~~	
$(v_3, v_2, \$10, v_4)$	$(\$30, v_2, \$10, v_4)$	$(\$30, v_2, \$10, v_4)$	~~$(\$30, v_2, \$10, v_4)$~~	
$(v_3, v_4, \$10, v_2)$	$(\$30, v_4, \$10, v_2)$	$(\$30, v_4, \$10, v_2)$	~~$(\$30, v_4, \$10, v_2)$~~	
$(v_4, \$10, v_2, v_3)$	~~$(v_4, \$10, v_2, \$30)$~~	$(v_4, \$10, v_2, \$30)$	$(v_4, \$10, v_2, \$30)$	
$(v_4, \$10, v_3, v_2)$	$(v_4, \$10, \$30, v_2)$	$(v_4, \$10, \$30, v_2)$	$(v_4, \$10, \$30, v_2)$	
$(v_4, v_2, \$10, v_3)$	~~$(v_4, v_2, \$10, \$30)$~~	$(v_4, v_2, \$10, \$30)$	$(v_4, v_2, \$10, \$30)$	
$(v_4, v_3, \$10, v_2)$	$(v_4, \$30, \$10, v_2)$	~~$(v_4, \$30, \$10, v_2)$~~	$(v_4, \$30, \$10, v_2)$	
CR	6	4	4	

(b) 2nd Observation

Mappings	3^{rd} Observation: $B : \$20$			selected ARs
	$< A, B, C >: \$20$	$< A, B, D >: \$20$	$< B, C, D >: \$20$	
$(\$10, v_2, \$30, v_4)$	$(\$10, \$20, \$30, v_4)$	$(\$10, \$20, \$30, v_4)$	$(\$10, \$20, \$30, v_4)$	$< A, B, C >: \$10$ $< A, B, C >: \$30$
$(\$10, \$30, v_2, v_4)$	$(\$10, \$30, \$20, v_4)$	~~$(\$10, \$30, \$20, v_4)$~~	$(\$10, \$30, \$20, v_4)$	
$(\$10, \$30, v_4, v_2)$	~~$(\$10, \$30, v_4, \$20)$~~	$(\$10, \$30, v_4, \$20)$	$(\$10, \$30, v_4, \$20)$	
$(\$10, v_4, \$30, v_2)$	~~$(\$10, v_4, \$30, \$20)$~~	$(\$10, v_4, \$30, \$20)$	$(\$10, v_4, \$30, \$20)$	
$(v_2, \$10, \$30, v_4)$	$(\$20, \$10, \$30, v_4)$	$(\$20, \$10, \$30, v_4)$	~~$(\$20, \$10, \$30, v_4)$~~	
$(v_2, \$30, \$10, v_4)$	$(\$20, \$30, \$10, v_4)$	$(\$20, \$30, \$10, v_4)$	~~$(\$20, \$30, \$10, v_4)$~~	
$(\$30, \$10, v_2, v_4)$	$(\$30, \$10, \$20, v_4)$	~~$(\$30, \$10, \$20, v_4)$~~	$(\$30, \$10, \$20, v_4)$	
$(\$30, \$10, v_4, v_2)$	~~$(\$30, \$10, v_4, \$20)$~~	$(\$30, \$10, v_4, \$20)$	$(\$30, \$10, v_4, \$20)$	
$(\$30, v_2, \$10, v_4)$	$(\$30, \$20, \$10, v_4)$	$(\$30, \$20, \$10, v_4)$	$(\$30, \$20, \$10, v_4)$	
$(\$30, v_4, \$10, v_2)$	~~$(\$30, v_4, \$10, \$20)$~~	$(\$30, v_4, \$10, \$20)$	$(\$30, v_4, \$10, \$20)$	
$(v_4, \$10, \$30, v_2)$	~~$(v_4, \$10, \$30, \$20)$~~	$(v_4, \$10, \$30, \$20)$	$(v_4, \$10, \$30, \$20)$	
$(v_4, \$30, \$10, v_2)$	~~$(v_4, \$30, \$10, \$20)$~~	$(v_4, \$30, \$10, \$20)$	$(v_4, \$30, \$10, \$20)$	
CR	6	2	2	

(c) 3rd Observation

Mappings	4^{th} Observation:D : \$40			selected ARs
	$< A, B, D >$: \$40	$< A, C, D >$: \$40	$< B, C, D >$: \$40	
$(\$10, \$20, \$30, v_4)$	$(\$10, \$20, \$30, \$40)$	$(\$10, \$20, \$30, \$40)$	$(\$10, \$20, \$30, \$40)$	$< A, B, C >$: \$10 $< A, B, C >$: \$20 $< A, B, C >$: \$30
$(\$10, \$30, \$20, v_4)$	$(\$10, \$30, \$20, \$40)$	$(\$10, \$30, \$20, \$40)$	$(\$10, \$30, \$20, \$40)$	
$(\$20, \$10, \$30, v_4)$	$(\$20, \$10, \$30, \$40)$	$(\$20, \$10, \$30, \$40)$	$(\$20, \$10, \$30, \$40)$	
$(\$20, \$30, \$10, v_4)$	$(\$20, \$30, \$10, \$40)$	$(\$20, \$30, \$10, \$40)$	$(\$20, \$30, \$10, \$40)$	
$(\$30, \$10, \$20, v_4)$	$(\$30, \$10, \$20, \$40)$	$(\$30, \$10, \$20, \$40)$	$(\$30, \$10, \$20, \$40)$	
$(\$30, \$20, \$10, v_4)$	$(\$30, \$20, \$10, \$40)$	$(\$30, \$20, \$10, \$40)$	$(\$30, \$20, \$10, \$40)$	
CR	0	0	0	

(d) 4th obsevation

Mappings	5^{th} Observation: B : \$20			selected ARs
	$< A, B, C >$: \$20	$< A, B, D >$: \$20	$< B, C, D >$: \$20	
$(\$10, \$20, \$30, \$40)$	$(\$10, \$20, \$30, \$40)$	$(\$10, \$20, \$30, \$40)$	$(\$10, \$20, \$30, \$40)$	$< A, B, C >$: \$10 $< A, B, C >$: \$20 $< A, B, C >$: \$30 $< A, B, D >$: \$40
$(\$10, \$30, \$20, \$40)$	$(\$10, \$30, \$20, \$40)$	~~$(\$10, \$30, \$20, \$40)$~~	$(\$10, \$30, \$20, \$40)$	
$(\$20, \$10, \$30, \$40)$	$(\$20, \$10, \$30, \$40)$	$(\$20, \$10, \$30, \$40)$	~~$(\$20, \$10, \$30, \$40)$~~	
$(\$20, \$30, \$10, \$40)$	$(\$20, \$30, \$10, \$40)$	$(\$20, \$30, \$10, \$40)$	~~$(\$20, \$30, \$10, \$40)$~~	
$(\$30, \$10, \$20, \$40)$	$(\$30, \$10, \$20, \$40)$	~~$(\$30, \$10, \$20, \$40)$~~	$(\$30, \$10, \$20, \$40)$	
$(\$30, \$20, \$10, \$40)$	$(\$30, \$20, \$10, \$40)$	$(\$30, \$20, \$10, \$40)$	$(\$30, \$20, \$10, \$40)$	
CR	0	2	2	

(e) 5th observation

Mappings	6^{th} Observation: B : \$20			selected ARs
	$< A, B, C >$: \$20	$< A, B, D >$: \$20	$< B, C, D >$: \$20	
$(\$10, \$20, \$30, \$40)$	$(\$10, \$20, \$30, \$40)$	$(\$10, \$20, \$30, \$40)$	$(\$10, \$20, \$30, \$40)$	$< A, B, C >$: \$10 $< A, B, C >$: \$20 $< A, B, C >$: \$30 $< A, B, D >$: \$40 $< A, B, D >$: \$20
$(\$20, \$10, \$30, \$40)$	$(\$20, \$10, \$30, \$40)$	$(\$20, \$10, \$30, \$40)$	~~$(\$20, \$10, \$30, \$40)$~~	
$(\$20, \$30, \$10, \$40)$	$(\$20, \$30, \$10, \$40)$	$(\$20, \$30, \$10, \$40)$	~~$(\$20, \$30, \$10, \$40)$~~	
$(\$30, \$20, \$10, \$40)$	$(\$30, \$20, \$10, \$40)$	$(\$30, \$20, \$10, \$40)$	$(\$30, \$20, \$10, \$40)$	
CR	0	0	2	

(f) 6th observation

Mappings	7^{th} Observation:A : $\$10$			selected ARs
	$< A, B, C >$: $\$10$	$< A, B, D >$: $\$10$	$< A, C, D >$: $\$10$	
$(\$10, \$20, \$30, \$40)$ $(\$30, \$20, \$10, \$40)$	$(\$10, \$20, \$30, \$40)$ $(\$30, \$20, \$10, \$40)$	$(\$10, \$20, \$30, \$40)$ ~~$(\$30, \$20, \$10, \$40)$~~	$(\$10, \$20, \$30, \$40)$ $(\$30, \$20, \$10, \$40)$	$< A, B, C >$: $\$10$ $< A, B, C >$: $\$20$ $< A, B, C >$: $\$30$ $< A, B, D >$: $\$40$ $< A, B, D >$: $\$20$ $< B, C, D >$: $\$20$
CR	0	1	0	

(g) 7th observation

Table 3.2c, 3.2d, 3.2e, 3.2f and 3.2g. After recieving the last report i.e. 7^{th} report, the mapping set size becomes one. So, after recieving this AR, ApS finds the real mappings i.e. all OOIs are decoded correctly by ApS.

3.3.2 Limitation

Subset-coding based approaches presented in [1] are not feasible in real world scenario for its high computational complexity i.e. $O(N!)$. Moreover, this approach considers the following assumptions which might be inappropriate in real-world PSS scenario:

- Application Server is aware of the list of all OOIs of PSS scenario.
- The attribute of an OOI varies for only in one dimension i.e it works for only single dimension.
- No faulty report is allowed to be recieved by both ApS and AS.

3.4 Conclusion

This chapter has presented the contemporary research works on privacy protection strategies for PSS and their limitations. The most relevant technique to our problem, Subset-Coding [1] has been explained with an example. However, none of the techniques explained above cannot guarantee data quality with acceptable computational complexity and all of these limit to protecting privacy in single dimension. In the next chapter, we design comparatively efficient schemes keeping these limitations in mind.

Chapter 4

Privacy Protection Scheme for Single-Dimensional Scenario

In Chapter 3, we have discussed the limitations of current privacy-preserving approaches in terms of data integrity, computational complexity and flexibility on user preferences. In this chapter, we are going to discuss our proposed algorithms in Single-Dimensional scenario that aims to provide solution to all these issues.

This chapter is organized as follows. Section 4.1 explains the basic concept of our proposed algorithm (both Anonymization and Decoding steps) along with some defined terms. Section 6.1.6 proves that our algorithm is more fault-tolerant than DGAS subset-coding approach with an example. Finally Section 4.5 concludes the chapter.

4.1 Basic Concept

The privacy-protective PSS scheme can be divided into two parts, i.e., Anonymization and Decoding. Here we discuss the concept of our proposed schemes. First, we describe a simple example of anonymization and decoding schemes for single dimension. Let us assume a PSS scenario with $N = |S| = 3$ where the AS receives three different products (OOIs) named A, B, and C which have prices (attributes) \$10, \$20, and \$30, respectively. As mentioned in Chapter 2, S denotes the set of OOIs and N is the number of OOIs in the system. For the sake of simplicity, we assume that the anonymity preference (k) is 2 for all users. For demonstrating this example, we have assumed a list of reports shown in Table 4.1 in order of appearance.

Incoming Order	Report
1	$< A\{2\} >$ \$10
2	$< B\{2\} >$ \$20
3	$< B\{2\} >$ \$20
4	$< C\{2\} >$ \$30

Table 4.1: List of reports observed in single-dimensional scenraio.

4.1.1 Anonymization Step

The AS maintains a data structure named Inverse Occurrence Checklist (IOC) for N OOIs. For example, IOC for OOI p, denoted as IOC_p, contains the absence-count of each other OOI $q|q \in S \cap q \neq p$. Also, we use the notion $IOC_p(q)$ to express how many times q has not been included in ARs of OOI_p. For example, if the IOC count for OOI B in IOC_A is 2 i.e. $IOC_A(B) = 2$, it denotes that B has not been included twice among all ARs for A processed so far. Initially, all IOC values for all OOIs are set to zero. Each time an observed report is received, AS anonymizes it and updates the corresponding IOC. The update of IOC values after each AR generations is shown in Figure 4.1a.

The AS can identify whether an OOI can be fully mapped to its actual value by ApS from a set of ARs. When an OOI is mapped uniquely to its attribute by decoding a set of ARs, the OOI is considered fully decoded. Formally, it is expressed as follows.

Rule 4.1. *The OOI p is fully decoded if for each OOI $q|q \in S \wedge q \neq p$, either $IOC_p(q) \neq 0$ or q has been decoded earlier.*

As discussed before, while choosing the extra OOIs for anonymizing the report of the observed OOI p, AS considers the following criteria to select each other OOI $q|q \in S \wedge q \neq p$.

Rule 4.2. *a) For each OOI $x|x \in S \wedge x \neq p \neq q, IOC_p(q) >= IOC_p(x)$*

b) q is already decoded according to Rule 4.1

After anonymizing the report, AS increases the count of those OOIs in S that have not been included in this AR.

In the example of Table 4.1, the first report carries the price of OOI A. To anonymize this report, AS checks only IOC_A. In IOC_A, there are two IOC counts for OOI B and C denoted as $IOC_A(B)$ and $IOC_A(C)$, respectively. Both the counts are zero and neither B nor C have been decoded yet. Consequently, AS randomly chooses any of them among B or C as extra OOI and form an AR such as $< A, B >$: \$10. As C has not been used in this AR, $IOC_A(C)$ is increased by one (Figure 4.1b).

	A	B	C
IOC_A	▨	0	0
IOC_B	0	▨	0
IOC_C	0	0	▨

(a) Initial

	A	B	C
IOC_A	▨	0	**1**
IOC_B	0	▨	0
IOC_C	0	0	▨

(b) 1 report anonymized

	A	B	C
IOC_A	▨	0	1
IOC_B	0	▨	**1**
IOC_C	0	0	▨

(c) 2 reports anonymized

	A	B	C
IOC_A	▨	0	1
IOC_B	**1**	▨	1
IOC_C	0	0	▨

(d) 3 reports anonymized

	A	B	C
IOC_A	▨	0	1
IOC_B	1	▨	1
IOC_C	0	**1**	▨

(e) 4 reports anonymized

Figure 4.1: Demonstration of anonymization process for single-dimensional scenario.

In the same manner, the second report is anonymized and $IOC_B(C)$ is updated (shown in Figure 4.1c). The third report comes for B again. AS checks IOC_B and finds that among the two IOC counts, $IOC_B(C) > IOC_B(A)$. Hence, AS chooses C as extra attribute (Rule 4.2). So, the AR is $< B, C >$: \$20 and $IOC_B(A)$ is increased by one. In this state of IOC_B, both $IOC_B(A)$ and $IOC_B(C)$ are non-zero. So, according to Rule 1, B can be fully decoded. Now $IOC_A(C) \neq 0$ and B has already been decoded. Therefore,according to Rule 4.1, A can also be decoded after receiving the 3rd AR. For the last report for OOI C, AS chooses extra OOI at random e.g. $< A, C >$: \$30 because both A and B have been decoded. According to Rule 4.2, C is now marked as decoded as both A and B are decoded and C has been reported at least once.

4.1.2 Decoding Step

With the same example used in anonymization step, we explain the decoding step. Like AS, the ApS also uses IOC count for each reported values, denoted as IOC_v to track the OOIs which were not selected till that point. An OOI is decoded i.e. the attribute v of an OOI is determined by ApS if the corresponding IOC for v meets the following criteria:

Rule 4.3. *a) Having a single OOI* $p|p \in S$ *and* $IOC_v(p) = 0$ *and* p *is not decoded for any other reported attribute.*

b) For any OOI $q|q \neq p$*, either* $IOC_v(q) \neq 0$ *or* q *is the decoded OOI for any other reported attribute.*

However, the ApS may not have prior knowledge of all OOIs. So, like AS, ApS cannot maintain fixed length *IOC* table. Instead, ApS keeps track the total count of reports denoted as T_v for reported attribute v and the occurance counts of those OOIs which has been reported to ApS. The occurance count for OOI p is denoted as $OC_v(p)$. Using this knowledge, ApS can easily calculate *IOC* count by following equality:

$$IOC_v(p) = T_v - OC_v(p) \tag{4.1}$$

According to the decoding Rule 4.3, the *IOC* count of one single OOI has to be zero among all reported OOIs in order to decode an attribute. Using Equation (4.1), the condition $IOC_v(p) = 0$ will be true if and only if $T_v = OC_v(p)$ is satisfied. Hence, the decoding rule can be re-written as below:

Rule 4.4. *a) Having a single OOI $p|p \in S$ such that $T_v = OC_v(p)$ and p is not decoded for any other reported attribute.*

b) For any OOI $q|q \neq p$, either $T_v(q) \neq OC_v(p)$ or q is the decoded OOI for any other reported attribute.

Thus the ApS keeps *OC* table instead of *IOC* table and performs decoding using the modified Rule 4.4. As illustrated in Section 4.1.1, the anonymized reports generated by the AS are shown in Table 4.2. Considering this example, ApS receives the 1st AR, $< A, B >$: \$10. Since \$10 has not been reported before, ApS creates an *OC* row for \$10 denoted as $OC_{\$10}$ and corresponding $T_{\$10}$ column for total report counts for \$10. As it is the first report of \$10, ApS sets $T_{\$10}$ to one. This report indicates that \$10 is the actual attribute of either A or B. So ApS creates two *OC* columns for A and B denoted as $OC_{\$10}(A)$ and $OC_{\$10}B$ and increases their *OC* count shown in Figure 4.2a. For the second AR, the same technique is applied (Figure 4.2b).

Incoming Order	Report	
	Real	Anonymized
1	$< A\{2\} >$ \$10	$< A, B >$ \$10
2	$< B\{2\} >$ \$20	$< A, B >$ \$20
3	$< B\{2\} >$ \$20	$< B, C >$ \$20
4	$< C\{2\} >$ \$30	$< B, C >$ \$30

Table 4.2: List of reports

	A	B	**T**
$OC_{\$10}$	**1**	**1**	**1**

(a) 1 report received

	A	B	**T**
$OC_{\$10}$	1	1	1
$OC_{\$20}$	**1**	**1**	**1**

(b) 2 reports received

	A	B	C	**T**
$OC_{\$10}$	1	1	N/A	1
$OC_{\$20}$	1	**2**	**1**	**2**

(c) 3 reports received

	A	B	C	**T**
$OC_{\$10}$	1	1	N/A	1
$OC_{\$20}$	1	2	1	2
$OC_{\$30}$	**1**	0	**1**	**1**

(d) 4 reports received

Figure 4.2: Demonstration of Decoding process for Single-dimensional scenario.

The 3rd anonymized report received by ApS is $< B, C >$: \$20. As \$20 has been reported before, ApS only updates the corresponding $OC_{\$20}$. However, the OOI C has been reported to ApS for the first time in this report. Hence, ApS creates additional column for OOI C and increases $T_{\$20}$, $OC_{\$20}(B)$ and $OC_{\$20}(C)$ by one (Fig. 4.2c). At this stage, the relations between OC counts and total counts for attribute \$20 are as follows:

$$OC_{\$20}(B) = T_{\$20} = 2 \tag{4.2}$$

$$OC_{\$20}(B) \neq T_{\$20} \text{ and } OC_{\$20}(B) \neq T_{\$20} \tag{4.3}$$

Thus, \$20 is decoded as the attribute of B by ApS according to Rule 4.4 since only $OC_B(B) = T_{\$20}$. As now B has been decoded, \$10 is also decoded for A because $OC_{\$10}(A) = T_{\$10} = 1$. Other OOI B and C cannot be the attribute of \$10 as B has been already decoded and C has not been occured in any report of \$10. Continuing in the same manner, for the last anonymized report $< A, C : \$30 >$, ApS creates $OC_{\$30}$, creates columns for total count and OC counts of reported OOIs and increases their count (Figure 4.2d). Since, both A and B are decoded for \$10 and \$20 respectively. Hence, according to the Rule 4.4, the attribute of C is identified as \$30.

4.2 Faulty Report Scenario

In real world, all the user reports received by ApS or AS are not guaranteed to be correct. A malicious user might report incorrect attribute or the data can be corrupted along the transmission path. Previous subset-coding approaches like DGAS fails to decode in such situation. But our proposed scheme can adopt to this environment and can decode successfully if a minimal percentage reports are faulty. This is a realistic assumption as number of malicious users should be nominal since this is a system for everyone's benefit. Here, we explain a scenario where DGAS fails to decode after receiving a faulty report, while our proposed scheme eventually can decode all OOIs.

Table 4.3: List of reports (with a faulty one).

Incoming Order	Real Report
1	$< A\{2\} > \$10$
2	$< B\{2\} > \$20$
3	$< B\{2\} > \$10$
4	$< C\{2\} > \$30$

Let us assume that the previously used example scenario in Section 4.1 had a faulty report shown in Table 4.3. In DGAS approach, if the reports are received by ApS in this order, then the mapping set after each anonymization will be as follows:

Table 4.4: Conforming tuples of gradually generated AR set for three observations where the third one is faulty

(a) 1st observation

Mappings	1^{st} Observation: A : \$10		selected ARs
	$< A, B >$: \$10	$< A, C >$: \$10	
(v_1, v_2, v_3)	$(\$10, v_2, v_3)$	$(\$10, v_2, v_3)$	
(v_1, v_3, v_2)	$(\$10, v_3, v_2)$	$(\$10, v_3, v_2)$	
(v_2, v_1, v_3)	$(v_2, \$10, v_3)$	~~$(v_2, \$10, v_3)$~~	
(v_2, v_3, v_1)	~~$(v_2, v_3, \$10)$~~	$(v_2, v_3, \$10)$	
(v_3, v_1, v_2)	$(v_3, \$10, v_2)$	~~$(v_3, \$10, v_2)$~~	
(v_3, v_2, v_1)	~~$(v_3, v_2, \$10)$~~	$(v_3, v_2, \$10)$	
CR	2	2	

Table 4.4c shows the mapping states of AS after receiving the faulty report i.e. $B : \$10$. Because of this wrong report, AS removes the real mapping $(\$10, \$20, \$30)$ as the AR $< B, C >$: \$10 maximizes the cardinality of mapping set. After this state, the mapping set size becomes one and hence AS considers it as the real decoded mapping. After that, even if all other reports received by AS are not faulty, AS cannot revert to its previous mapping set and hence can never achieve correct decoded mapping. This example clearly shows that

Mappings	2^{nd} Observation: B : \$20		selected ARs
	$< A, B >$: \$20	$< B, C >$: \$20	
$(\$10, v_2, v_3)$	$(\$10, \$20, v_3)$	$(\$10, \$20, v_3)$	$< A, B >$: \$10
$(\$10, v_3, v_2)$	~~$(\$10, v_3, \$20)$~~	$(\$10, v_3, \$20)$	
$(v_2, \$10, v_3)$	$(\$20, \$10, v_3)$	~~$(\$20, \$10, v_3)$~~	
$(v_3, \$10, v_2)$	~~$(v_3, \$10, \$20)$~~	$(v_3, \$10, \$20)$	
CR	2	1	

(b) 2nd observation

Mappings	3^{rd} Observation: B : \$10		selected ARs
	$< A, B >$: \$10	$< B, C >$: \$10	
$(\$10, \$20, v_3)$	$(\$10, \$20, v_3)$	~~$(\$10, \$20, v_3)$~~	$< A, B >$: \$10
$(\$20, \$10, v_3)$	$(\$20, \$10, v_3)$	$(\$20, \$10, v_3)$	$< A, B >$: \$20
CR	0	1	

(c) 3rd observation

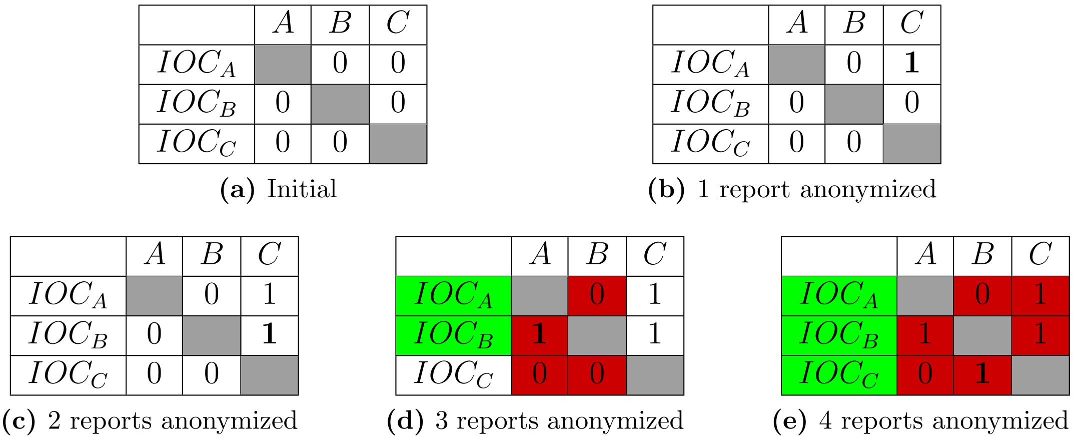
even a single faulty report is not allowed in DGAS approach which is very unlikely in real world.

On the other hand, our proposed approach can achieve 100% decodability even if some reports are faulty. With the same example used before, we explain here how AS can recover from the impact of the faulty report. As AS never considers the OOI's real attribute while choosing anonymized OOIs, the ARs generated with faulty report example shown in Table 4.5 will be same as the ARs generated in normal scenario. Figure 4.3b-4.3e show the ARs generated and the update of *IOC* tables after receiving all four reports respectively.

Table 4.5: List of anonymized reports

Incoming Order	Report	
	Real	Anonymized
1	$< A\{2\} >$ \$10	$< A, B >$ \$10
2	$< B\{2\} >$ \$20	$< A, B >$ \$20
3	$< B\{2\} >$ \$10	$< B, C >$ \$10
4	$< C\{2\} >$ \$30	$< B, C >$ \$30

Though the ARs generated in both correct and faulty report scenario are same, the ApS updates its corresponding *OC* tables differently due to the faulty attribute of third report. After receiving first two reports, the *OC* table contains two row for \$10 and \$20 shown in Figure 4.4b. After receiving the third report (which is faulty), ApS increases the *OC* count



	A	B	C
IOC_A		0	0
IOC_B	0		0
IOC_C	0	0	

(a) Initial

	A	B	C
IOC_A		0	**1**
IOC_B	0		0
IOC_C	0	0	

(b) 1 report anonymized

	A	B	C
IOC_A		0	1
IOC_B	0		**1**
IOC_C	0	0	

(c) 2 reports anonymized

	A	B	C
IOC_A		0	1
IOC_B	**1**		1
IOC_C	0	0	

(d) 3 reports anonymized

	A	B	C
IOC_A		0	1
IOC_B	1		1
IOC_C	0	**1**	

(e) 4 reports anonymized

Figure 4.3: Demonstration of anonymization process in single-dimensional scenario along with a faulty report.

for B and C in $OC_{\$10}$ instead of $OC_{\$20}$. After receiving the fourth report, ApS increases $OC_{\$10}(B)$ and $OC_{\$10}(C)$. However, because of the faulty report $< B, C >$: \$10, ApS can not decode any of the OOI as no attribute meets the decoding criteria i.e. $IOC_v = 0$ for a single OOI.

Table 4.6: List of newly arrived reports (all information is correct)

Incoming Order	Report	
	Real	Anonymized
1	$< A\{2\} >$ \$10	$< A, B >$ \$10
2	$< A\{2\} >$ \$10	$< A, C >$ \$10
3	$< A\{2\} >$ \$10	$< A, B >$ \$10

But eventually the ApS will able to decode the OOIs after receiving some non-faulty report. Let us assume that, the next new reports are all correct and the list is shown in Table 4.6 according to their order of arrival. When AS receives these reports, it follows the anonymization rule and generates AR. But AS does not know about the faulty report and hence it considered each OOI i.e. A, B and C decoded by ApS. Hence, for the next reports, AS chooses attributes randomly as all OOIs are decoded by AS according to Rule 4.2. The Table 4.6 shows the randomly generated ARs by AS for the newly received reports for anonymization. When these ARs are received by ApS, it updates $OC_{\$10}$ according to Rule 4.1. After decoding all new reports the OC table of ApS is shown in Figure 4.4g. At

	A	B	**T**
$OC_{\$10}$	**1**	**1**	**1**

(a) 1 report received

	A	B	**T**
$OC_{\$10}$	1	1	1
$OC_{\$20}$	**1**	**1**	**1**

(b) 2 reports received

	A	B	C	**T**
$OC_{\$10}$	1	**2**	**2**	**2**
$OC_{\$20}$	1	1	N/A	1

(c) 3 reports received (3$^{\text{rd}}$ one is faulty

	A	B	C	**T**
$OC_{\$10}$	1	2	2	2
$OC_{\$20}$	1	1	N/A	1
$OC_{\$30}$	**1**	0	**1**	**1**

(d) 4 reports received

	A	B	C	**T**
$OC_{\$10}$	**2**	**2**	2	**3**
$OC_{\$20}$	1	1	N/A	1
$OC_{\$30}$	**1**	0	**1**	**1**

(e) 5 reports received

	A	B	C	**T**
$OC_{\$10}$	**3**	2	**3**	**4**
$OC_{\$20}$	1	1	N/A	1
$OC_{\$30}$	**1**	0	**1**	**1**

(f) 6 reports received

	A	B	C	**T**
$OC_{\$10}$	**4**	**3**	3	**5**
$OC_{\$20}$	1	1	N/A	1
$OC_{\$30}$	**1**	0	**1**	**1**

(g) 7 reports received

Figure 4.4: Demonstration of decoding process for single-dimensional scenario along with a faulty report.

this stage, the IOC counts for \$10 is calculated as below:

$$IOC_{\$10}(A) = T_{\$10} - OC_{\$10}(A) = 5 - 4 = 1 \tag{4.4}$$

$$IOC_{\$10}(B) = T_{\$10} - OC_{\$10}(B) = 5 - 3 = 2 \tag{4.5}$$

$$IOC_{\$10}(C) = T_{\$10} - OC_{\$10}(C) = 5 - 3 = 2 \tag{4.6}$$

Though in $IOC_{\$10}$, there is a single OOI whose IOC value is almost zero (1) while others are greater than 1. Hence, A is the best candidate for \$10. Due to faulty reports, its IOC value could not be zero. When additional correct reports for \$10 will be received, the impact of this faulty report will disappear. That is, the $IOC_{\$10}(A)$ will be almost zero and the IOC count of other OOI i.e. $IOC_{\$10}(B)$ and $IOC_{\$10}(C)$ will differ by big margin from $T_{\$10}$. Thus the ApS can eventually achieve correct decodability.

4.3 Algorithms

This section describes the anonymization and decoding algorithms for single-dimensional scenario.

4.3.1 Anonymization Algorithm

Each time a report is received by AS for anonymization, it uses Algorithm 4.1 to anonymize it and update the IOC values. The input for this algorithm is a single report (p, k, v) where p is the OOI, k is anonymity preference and v is the observed value by user. First, the algorithm picks $(k-1)$ extra OOI from S by prioritizing the OOIs which have either higher IOC count in the corresponding IOC_p or is already decoded. The output of this algorithm is an AR in form of $< p_1, p_2, \cdots, p_k >$: v where $p_1 = p$ and $p_2, p_3, \cdots, p_k$ are the extra OOIs used in the anonymization. After anonymization, it increases the IOC count of the OOIs which have not been selected in this AR.

Algorithm 4.1 $\{S', v\}$: Anonymize SD (p, v, k)

```
Input:
    $p$ : Observed OOI
    $v$ : Actual attribute of OOI
    $k$ : Anonymity preference
Output:
    $S', v$ : An anonymized report where $S' \subset S \wedge |S'| = k$
Set $S' = \{p\}$
Select $(k-1)$ OOIs from $S$ applying Rule 4.2.
for all OOI $q|q \in (S - S')$ do
    Update $IOC_p$ by incrementing $IOC_p(q)$
end for
return $\{S', v\}$
```

4.3.2 Decoding Algorithm

Algorithm 4.2 is used by ApS for decoding the ARs. The input for this algorithm is an AR in form of $\{S', v\}$ where S' is the set of anonymized OOIs and v is the reported attribute. If v is reported to ApS for the first time, this algorithm creates new OC_v and T_v. Otherwise, it uses the OC_v and T_v from previous decoding processes.Then it increases T_v and the OC counts of reported OOIs. The updated OC_v and T_v is used in subsequent decoding processes. If only one single OOI p meets the decoding Rule 4.4 in OC_v, then the attribute v is considered as decoded for p.

Algorithm 4.2 Decode SD(S', v)

```
Input:
$S'$ : Set of OOIs in an AR
$v$ : Reported attribute
if $v$ is reported for first time then
    create $OC_v$ and $T_v$
else
    $OC_v, T_v$ : Updated $OC$ and $T$ for $v$ from previous decoding processes.
end if
$T_v = T_v + 1$
for all OOI $q \in S'$ do
    if $q$ not exists in $OC_v$ then
        $OC_v(q) = 0$
    end if
    $OC_v(q) = OC_v(q) + 1$
end for
$D$ : Set of previously decoded OOIs
if $OC_v(p) = T_v$ for a single $p \in OC_v$ and $p \notin D$ then
    $D = D \cup \{p\}$                    ▷ Attribute $v$ is decoded against OOI $p$
end if
```

4.4 Computational Complexity

In single-dimensional PSS, the order of anonymizing an user's report is N. For anonymizing a report, AS needs to choose $k-1$ extra OOIs according to higher IOC count and then update the IOC. Trivially, we may sort all IOC counts to get $k-1$ extra OOIs which takes $O(N \log N)$ operations. And k operations are required ($k \leq N-1$) to update the IOC values of the selected k OOIs in AR. Hence the total order is $O(N \log N) + O(N) = O(N \log N)$. However, if we use efficient data structure to keep the IOC counts, then the order reduces to $O(N)$. For instance, AS can maintain a hash table where the key is IOC count, i and the value is a collection of OOIs whose IOC value is equal to i. In this way if hash table or similar type data structure is used for keeping IOC counts, then AS picks OOIs from the hash table with the lowest key value, i.e., lowest IOC count. If the number of OOIs in the collection with lowest IOC count is less than $k-1$, AS picks the rest of them from the container with next lower key value. In this manner, AS can pick $k-1$ OOIs by performing only $k-1$ operation. The order of updating IOC count remains same as before. Because AS only moves those OOIs which are not present in AR from their current key to next increasing key. So, the total order of anonymizing a report is $O(N) + O(N) = O(N)$. For decoding an anonymized report, ApS only increases the IOC count of those OOIs which are not included in AR. It takes only $O(N)$ operations.

4.5 Conclusion

In this chapter, we have explained our proposed approach for single-dimensional scenario. We have also shown with an example that our proposed algorithm is capable of identifying correct values in scenarios where some reports carry false information. We have presented the pseudocode of our algorithm for both anonymization and decoding steps and discussed about its computational complexity. We extend our algorithm for multi-dimensional scenario which we present in the next chapter.

Chapter 5

Privacy Protection Scheme for Multi-Dimensional Scenario

We explained our proposed algorithms for Single-Dimensional scenario in Chapter 4. The extension of our algorithm for supporting multi-dimensional scenarios are discussed in this chapter. We explain the basic concept of the extended algorithm in Section 5.1. Section 5.2 presents the algorithms to implement the scheme. In Section 5.3, the computational complexity of the algorithm is discussed. Finally, 5.4 concludes the chapter.

5.1 Basic Concept

The rules for anonymization and decoding in multi-dimensional scenario are similar to that of single-dimensional one on principle. Here, we denote the total number of OOI for i^{th} dimension as N_i and the set of all OOIs in i^{th} dimension as S_i. The term OOI Combination is used to denote the combination of each OOI from each of the d dimensions for which an attribute is reported. Accordingly, the total number of OOI combinations is $\prod_{i=1}^{d} N_i$.

For the sake of simplicity, we discuss this process by restricting our example scenario in two dimensions i.e. $d = 2$. Accordingly, we assume a PSS system which deals with the price of 3 products e.g. $S_1 = A, B, C$ in three different locations e.g. $S_2 = X, Y, Z$. Hence, the total number of OOI combinations is $3 \times 3 = 9$ and their set is

$$R = (A, X), (A, Y), \cdots, (C, Y), (C, Z) \tag{5.1}$$

Let some observed attributes for each OOI combination be shown in Figure 5.1. Without any loss of generality, we assume that the anonymity preference for both dimensions (product

	X	Y	Z
A	\$10	\$11	\$12
B	\$20	\$21	\$22
C	\$30	\$31	\$32

Figure 5.1: Price of different OOI combinations.

Incoming Order	Report	
	Real	Anonymized
1	$< A\{2\}, X\{2\} >$: \$10	$< \{A, B\}, \{X, Y\} >$: \$10
2	$< A\{2\}, Y\{2\} >$: \$11	$< \{A, B\}, \{X, Y\} >$: \$11
3	$< A\{2\}, Z\{2\} >$: \$10	$< \{A, C\}, \{X, Z\} >$: \$10

Figure 5.2: List of reports observed in multi-dimensional scenraio.

and location) is $k_1 = k_2 = 2$ for all users. The observed reports are shown in Figure 5.2 in order of appearance along with a sample anonymized form. In the next section, we discuss how the anonymization was performed with a goal of achieving high data integrity at the ApS.

5.1.1 Anonymization Step

For each dimension i, AS chooses $k_i - 1$ OOIs along with the real OOI where k_i is user's preference of anonymity for i^{th} dimension. AS maintains d IOCs for each OOI combination in R. In our example, the first report $< A\{2\}, X\{2\} >$: \$10 refers to the price of A from location X. To anonymize these two OOIs, i.e., product and location AS randomly chooses additional OOIs B (for product) and Y (for location), respectively at the initial step. After producing this AR, i.e.,$< \{A, B\}, \{X, Y\}$: \$10 $>$, AS increases the count of $IOC^1_{A,X}(C)$ and $IOC^2_{A,X}(Z)$ as C and Z are not included in this AR.

	IOC^1			IOC^2		
	A	B	C	X	Y	Z
A, X	▨	0	1	▨	0	1
A, Y	▨	0	1	0	▨	1

(a) 2 reports anonymized

	IOC^1			IOC^2		
	A	B	C	X	Y	Z
A, X	▨	1	1	▨	1	1
A, Y	▨	0	1	0	▨	1

(b) 3 reports anonymized

Figure 5.3: Demonstration of anonymization process in multi-dimensional scenario.

For the second report, the same technique is applicable. When the 3rd report $< A\{2\}, X\{2\} >$: \$10 is received, AS chooses C and Z for anonymization according to 4.3.Since

$IOC^1_{A,X}(C)$ is higher than other IOC count in first dimension and $IOC^2_{A,X}(Z)$ is higher than other OOIs in second dimension, the AR is formed as $< \{A, C\}, \{X, Z\} >$: \$10. And the counts of $IOC^1_{A,X}(C)$ and $IOC^2_{A,X}(Z)$ are increased. For OOI combination $r = (A, X)$, the corresponding IOC^1 and IOC^2 satisfy the Rule 4.3. This is because in IOC^1, only $IOC^1_{A,X}(A) = 0$ and in IOC^2, only $IOC^2_{A,X}(X) = 0$. Therefore, the value of (A, X) can be decoded after receiving the 3rd AR.

5.1.2 Decoding Step

In multi-dimensional PSS, the ApS also uses OC counts and total counts for each reported attribute like single dimension as ApS does not have any prior knowledge of available OOI. We use the notation $OC^i_v(p)$ to denote the OC of i$^{\text{th}}$ dimension for a reported attribute v and $OC^i_v(p)$ to denote the OC count for OOI p in that corresponding OC. Initilly the OC table is empty.

	OC^1		OC^2		T
	A	B	X	Y	
\$10	**1**	**1**	**1**	**1**	**1**
\$11	**1**	**1**	**1**	**1**	**1**

(a) 2 reports received

	OC^1			OC^2			T
	A	B	C	X	Y	Z	
\$10	2	1	1	2	1	1	2
\$11	1	1	N/A	1	1	N/A	1

(b) 3 reports received

Figure 5.4: Demonstration of decoding process in multi-dimensional scenario.

In our example, after receiving the first AR, $< \{A, B\}, \{X, Y\} >$: \$10, the ApS creates one row for keeping the information of 10. In one column, ApS keeps the total report count $T_{\$10}$ and two other column to keep the OC counts for two dimensions i.e.$OC^1_{\$10}$ and $OC^2_{\$10}$. As it is the first report for 10, $T_{\$10}$ is set to one. In this report, ApS gets to know about A, B as the OOI for 1st dimension and X and Y for 2nd dimension. So, it creates columns for the OOI in respective dimensional OC and increases $OC^1_v(A)$, $OC^1_v(B)$, $OC^2_v(X)$ and $OC^2_v(Y)$ by one. In the same manner, when ApS receives the second AR $< \{A, B\}, \{X, Y\} >$: \$11 , it creates new row for \$11 and corresponding OC list for each dimensions. Then it increases the count of $T_{\$11}$, $OC^1_{\$11}(A)$, $OC^1_{\$11}(B)$, $OC^2_{\$11}(X)$ and $OC^2_{\$11}(Y)$ shown in Figure 5.4a.

The third AR $< \{A, C\}, \{X, Z\} >$: \$10 comes for \$10 again. It is the second report for \$10. So, ApS increases $T_{\$10}$ by one. In this report, ApS also knows about two new OOI called C and Z. Hence, ApS adds columns for the newly reported OOIs in their corresponding dimension to keep their occurance counts. Then ApS increases the occurance counts for

$OC^1_{\$10}(A)$, $OC^1_{\$10}(C)$, $OC^1_{\$10}(X)$ and $OC^2_{\$10}(Z)$ (Figure 5.3b). At this stage, for \$10, in the first dimension, the relations between OC counts and T_v are as follows:

$$T_{\$10} = OC^1_{\$10}(A) = 2 \tag{5.2}$$

$$T_{\$10} \neq OC^1_{\$10}(B) \text{ and } T_{\$10} \neq OC^1_{\$10}(C) \tag{5.3}$$

So in first dimension, only A's OC count is equal to T_v. In the same manner, if we observe the OC counts for second dimension, we find that only $OC^2_{\$10}(X) = T_{\$10}$. Therefore, following the Rule 4.4 for each dimension, the ApS successfully associated the attribute \$10 with product A and location X.

5.2 Algorithms

The algorithms for multi-dimensional PSS are philosophically similar to that of single dimensional one. Naturally, the number of OOIs increases and hence the number of IOCs maintained by both AS and ApS becomes larger.

5.2.1 Anonymization Algorithm

Algorithm 5.1 is used by AS to anonymize a multi-dimensional observation. It takes a set of user preferences $(k_1, k_2, k_3, \cdots k_d)$ for each dimension and the OOI combination $r = (p_1, p_2, \cdots, p_d)$ as input. The algorithm uses corresponding IOC i.e. IOC_r to anonymize this report. For each dimension i, the algorithm chooses $(k_i - 1)$ extra OOIs from the set of OOIs in that dimension, i.e. S_i by preferring the OOIs with highest IOC count and the decoded ones following Rule 4.2. These selected additional OOIs along with the observed OOI are put into the set S'_i. After preparing the set S'_i, the algorithm updates IOC_r by incrementing the IOC count for each OOI $q | q \in S_i \wedge q \notin S'_i$. Thus, the returning set is formulated, i.e., $S' = \{S'_1, S'_2, \cdots S'_d\}$ where S'_i denotes the anonymized OOIs set for i^{th} dimension.

Algorithm 5.1 $\{S', v\}$: Anonymize MD ($\{p_1, \cdots, p_d\}, v, \{k_1, \cdots, k_d\}$)

```
Input:
    $p_1, p_2, \cdots, p_d$ : Reported OOI combination
    $v$ : Reported attribute
    $k_1, k_2, \cdots, k_d$ : Anonymity preference for d dimensions
Output:
    $S', v$ : An anonymized report
Set $S' = \{\}$
for all dimension $i \in \{1, 2, \cdots, d\}$ do
    $S'_i = \{p_i\}$
    From all OOI $q | q \in S_i \wedge q \neq p_i$ , add($k_i$−1)OOIs to $S'_i$ by preferring the ones with higher $IOC_v(q)$ and which are already decoded
    $S' = S'S'_i$
    Update $IOC_v$ by incrementing each $IOC_v(q) | q \notin S'_i$    ▷ The IOC refers to the updated one from previous anonymization processes
end for
return $\{S', v\}$
```

5.2.2 Decoding Algorithm

Algorithm 5.2 is used by ApS for decoding the ARs. Here, input S'_i and v denote the set of anonymized OOIs in i$^{\text{th}}$ dimension and the reported attribute respectively. The ApS first checks whether the corresponding OC_v and T_v exists. If not, ApS creates corresponding OC_v and T_v. Then for each dimension i, the algorithm increases the OC count for all OOI q i.e. $OC^i_v(q) | q \in S'_i$. To check whether a reported value v has been fully decoded, ApS checks Rule 4.3 for each dimension i. If all dimensions' observed OOIs are decoded by ApS following Rule 4.3, then the actual OOI combination for v is known by ApS.

Algorithm 5.2 Decode MD $(S'_1, S'_2, \cdots, S'_d, v)$

```
Input:
    $S'_1, S'_2, \cdots, S'_d$ : Set of OOIs in $d$ dimensions of AR
    $v$ : Reported attribute
if $v$ is reported for first time then
    create $OC_v$ and $T_v$
else
    $OC_v, T_v$ : Updated $OC$ and $T$ for $v$ from previous decoding processes.
end if
$T_v = T_v + 1$
for all dimension $i \in \{1, 2, \cdots, d\}$ do
    for all OOI $q \in S'_i$ do
        if $q$ not exists in $OC^i_v$ then
            $OC^i_v(q) = 0$
        end if
        $OC^i_v(q) = OC^i_v(q) + 1$
    end for
end for
for all dimension $i \in \{1, 2, \cdots, d\}$ do
    if $OC^i_v(p_i) = T_v$ for a single $p_i \in OC^i_v$ then
        $i^{\text{th}}$ dimension is decoded against OOI $p_i$
    end if
end for
if all dimension is decoded then
    Attribute $v$ is fully decoded against $p_1, p_2, \cdots p_d$
end if
```

5.3 Computational Complexity

In single dimension, choosing extra OOIs requires $O(N)$ operations if efficient data structure is used. In multi-dimensional scenario, this operation for choosing extra OOIs is done for each dimension. Therefore, the order for anonymizing an user's report in multi-dimensional scenario is $O(N_1 + N_2 + \cdots + N_d) = O(d \times N)$. The order of decoding a report in single dimension is $O(N)$. Hence, in d-dimensional PSS system, the order is $O(d \times N)$.

5.4 Conclusion

In real world, most of the PSS scenarios intend to discover the attributes for complex OOIs which cause privacy threat in multiple dimensions. Hence, we extend our algorithm for multiple dimensions which has been presented in this chapter. We have also demonstrated the pseudocode for the proposed algorithm for multi-dimensional scenario and deduced their computational complexity. In the next chapter, we present our simulation results based on our implementation.

Chapter 6

Experimental Results and Discussion

We have explained our proposed algorithms both for single and multi-dimensional scenario in last two chapters. To establish the applicability and assess the performance of our proposed schemes, we implement and evaluate our algorithms with both comprehensive simulation and android-based real world prototype. The analysis shows the impact of the number of OOIs and user preferences on the decodability rate.

This chapter is organized as follows. In Section 6.1, we present the results of our simulation on synthesis data with different types of setup. The next Section 6.2 present the performance of our android application as a prototype of our implemented algorithm. Section 6.3 concludes the chapter.

6.1 Simulation Setup

We have conducted several types of simulation of our proposed scheme using a custom simulator. We have implemented this simulator using python language. To generate graphs from the result of our simulation we have used matplotlib library (python 2D plotting library). To generate observed reports randomly for simulation, we have used uniform distribution. By varying the number of OOIs (N) and the anonymity preference (k) of users, we have analyzed the performance of the algorithm for both single and multi-dimensional scenarios. As we are mostly interested in evaluating performance of our system with a high degree of data quality we have investigated how many observations are required to achieve different extent of decodadiblity. We use a term called 'Decodability Rate' to represent our analysis graph. Decodability Rate of T observations is defined as the proportion of OOIs decoded among N number of OOIs on average. All the results presented here are obtained by averaging 1000 simulation runs.

6.1.1 Results for Single-Dimensional PSS

The anonymization schemes that have been proposed so far [1] is not feasible in real-life for reporting large number of OOIs. But our proposed scheme is scalable while preserving data integrity in real-life. Figure 6.1 shows the decodability rate in our simulation for fixed $N = 15$ by varying anonymity preference k from 8 to 14 in single dimension. Naturally, high anonymity preference requires more observations to achieve full decodability. For example, around 100 reports are needed to achieve full decodability for $k = 8$ while little more than 200 reports are needed for $k = 13$. However, the highest possible anonymity preference e.g. $k = 14$ requires considerably higher number of observations, i.e., 375. This result indicates that based on the observation frequency of a particular PSS application, a feasible k should be recommended. Figure 6.2 and Figure 6.3 show similar trend with PSS having different Ns and ks.

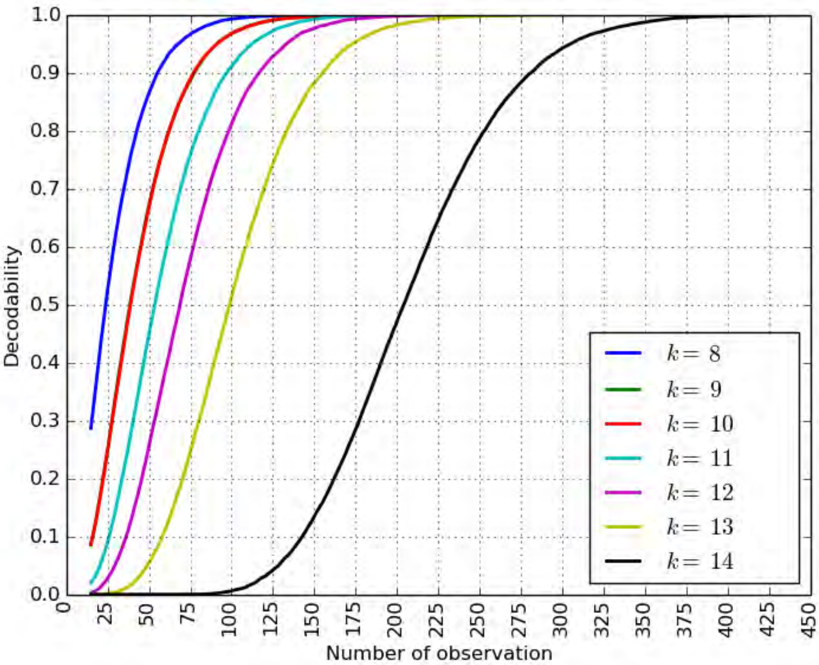


Figure 6.1: Decodability rate in single-dimensional scenario for $N(= 15)$ and varying k.

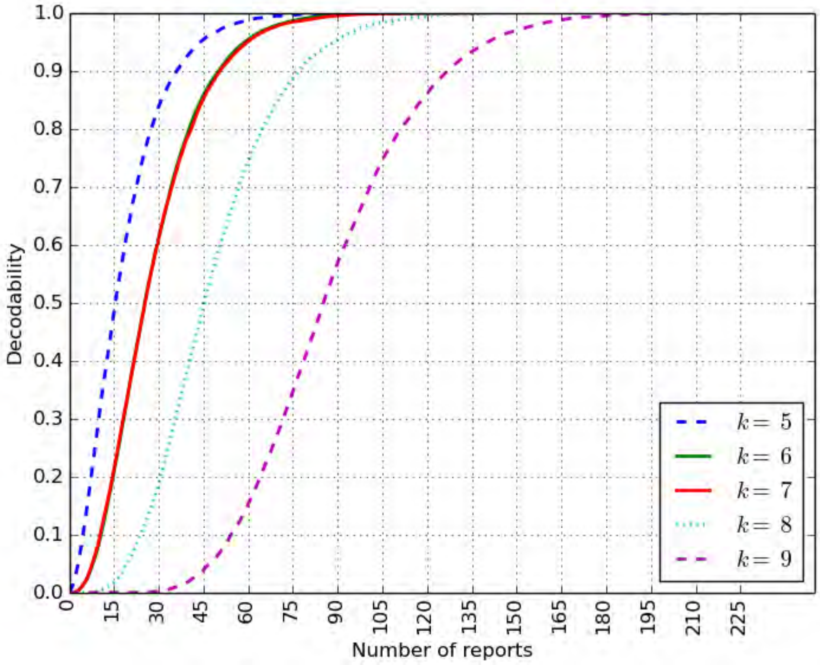


Figure 6.2: Decodability rate in single-dimensional scenario for $N(=10)$ and varying k.

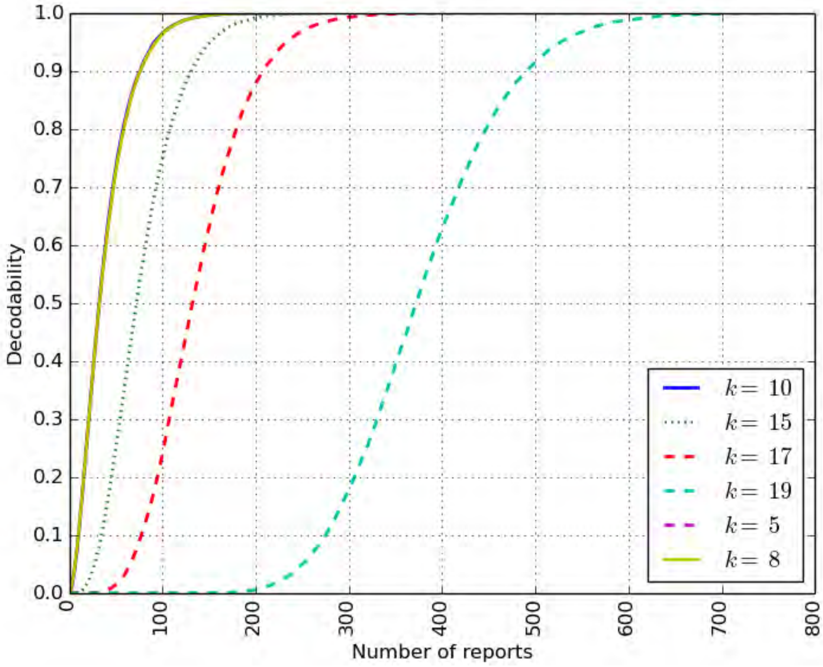


Figure 6.3: Decodability rate in single-dimensional scenario for $N(=20)$ and varying k.

Figure 6.4 presents the decodability rate for different N ($11 \sim 15$) with maximum anonymity, i.e. $k = N - 1$. Here we see that, for $N = 11$ the number of reports required to achieve full decodability is around 200, which is almost half to that for $N = 15$. Hence, we may conclude that even in spatio-temporal scenarios with fewer number of observations, the proposed technique may accommodate high anonymity preference if the number of OOIs is restricted. This can be controlled by reducing the operating area of a single PSS unit. Figure 6.5 and Figure 6.6 show similar results for $N = 5 \sim 20$ with highest anonymity preference. These graphs depict the robustness of our proposed system from smaller N to larger N.

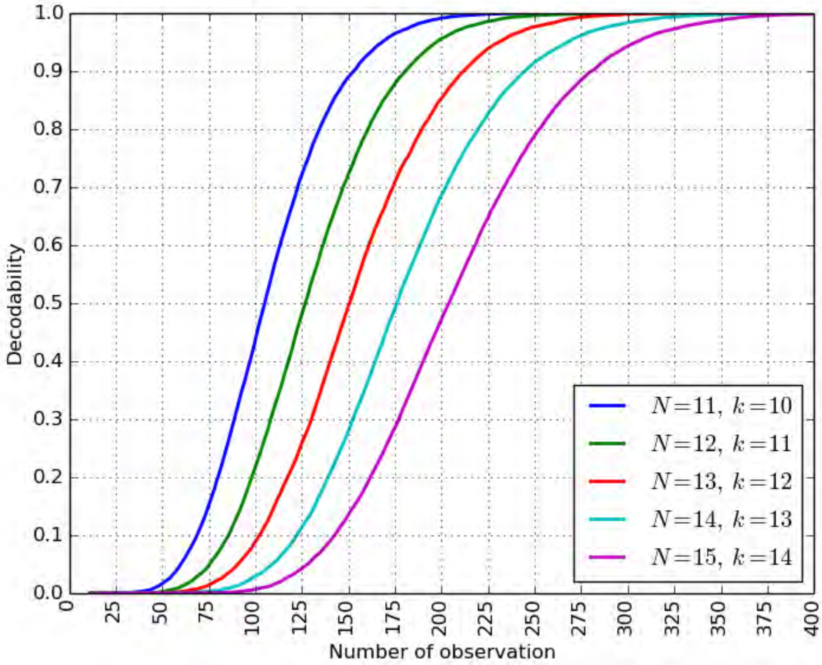


Figure 6.4: Decodability rate with highest possible $k(= N - 1)$ and $N = 11 \sim 15$.

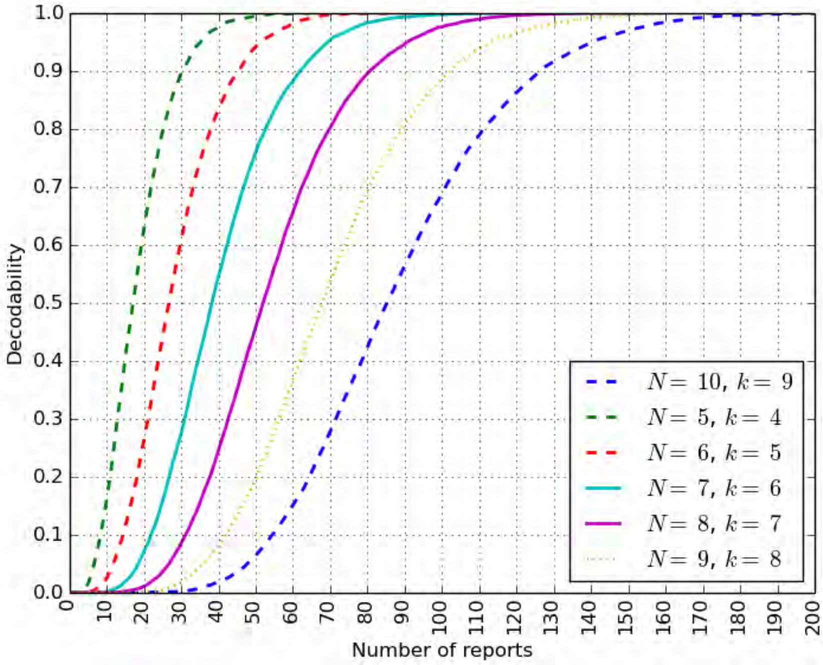


Figure 6.5: Decodability rate with highest possible $k(= N - 1)$ and $N = 5 \sim 10$.

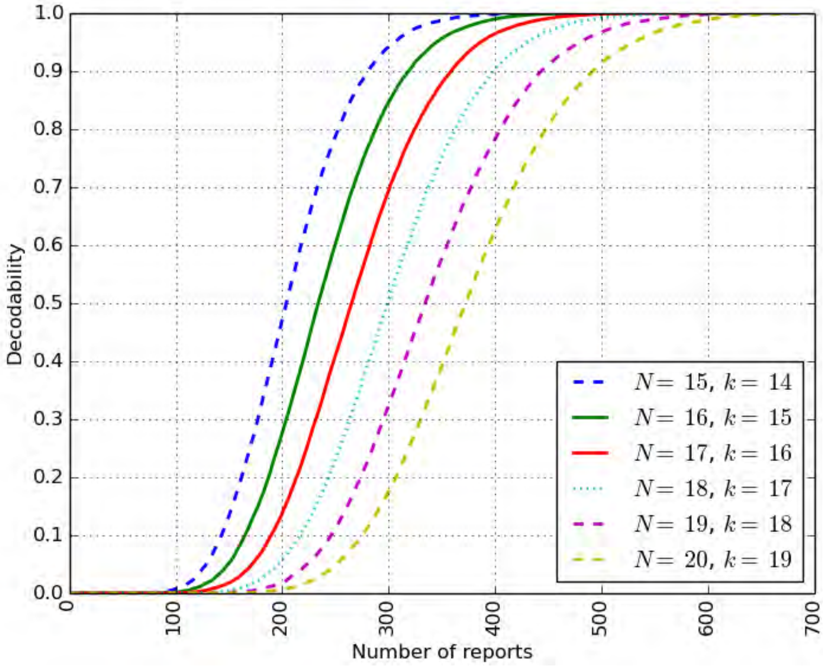


Figure 6.6: Decodability rate with highest possible $k(= N - 1)$. and $N = 15 \sim 20$.

6.1.2 Results for Multi-Dimensional PSS

Allowing anonymity in multiple dimensions and at once satisfying different anonymity preference for each dimension is the most desired performance for an anonymization scheme that we achieve without sacrificing the quality (data integrity) of PSS service. Figure 6.7 depicts the simulation result for two-dimensional anonymization which anonymizes both location and product. Here, the number of locations, $N_l = 8$ and the number of products $N_p = 4$ with variable anonymity preference k_l and for a fixed k_p. Naturally, number of reports required to achieve full decodability increases with the increase in k_l. Figure 6.8 shows the same scenario where k_p is fixed and k_l is varied. This also shows that the decodability rate decreases with the increase of k_l.

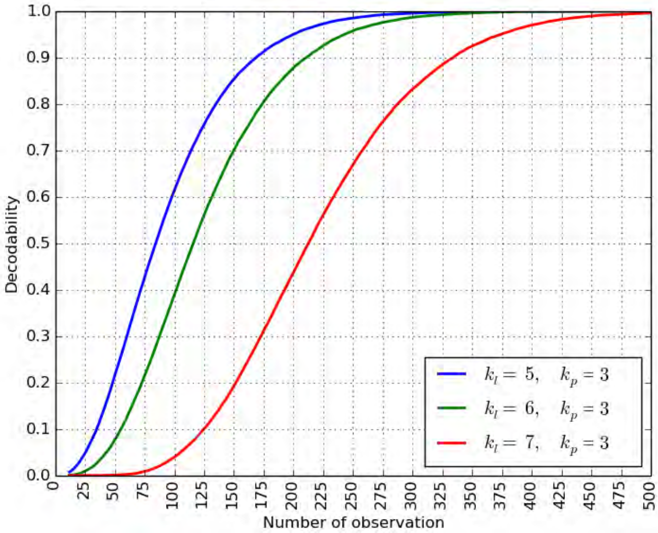


Figure 6.7: Decodability rate in multi-dimensional scenario for $N_l = 8$, $N_p = 4$ and varying k_l.

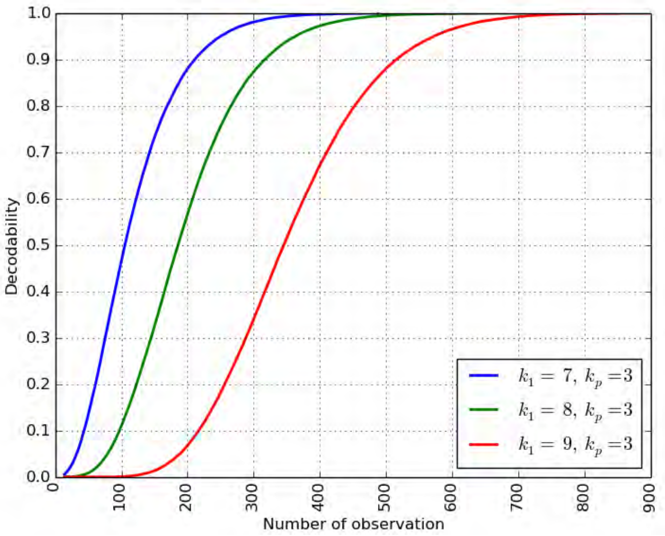


Figure 6.8: Decodability rate in multi-dimensional scenario for $N_l = 10$, $N_p = 4$ and varying k_l.

Figure 6.9 shows the decodability rate for different N in multiple dimensions with highest anonymity in each cases. We see that for quite large number of OOIs in both dimensions, i.e., $N_1 \in \{13, 14\}$ and $N_2 \in \{6, 7\}$ the required number of observations are in the range of $1800 - 2200$ to achieve full decodability. This is possible in a scenario with very high frequency of observations. This result also shows that increasing the number of OOIs in one dimension and decreasing that in another dimension does not change the decodability rate at considerable extent. Hence, in case of very large number of OOIs in multi-dimensional scenario, PSS can balance by varying the number of OOIs in different dimensions in order to achieve decodability with a finite number of observations.

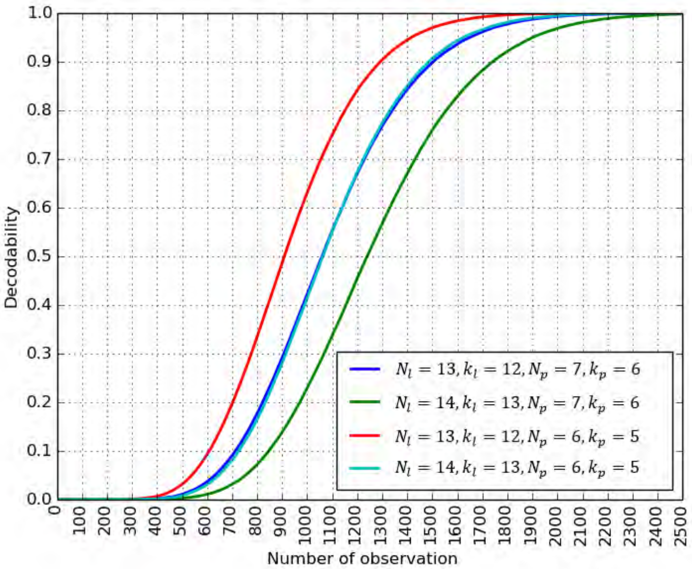


Figure 6.9: Decodability rate in multi-dimensional scenario for $k = N - 1$ and varying N_l, N_p.

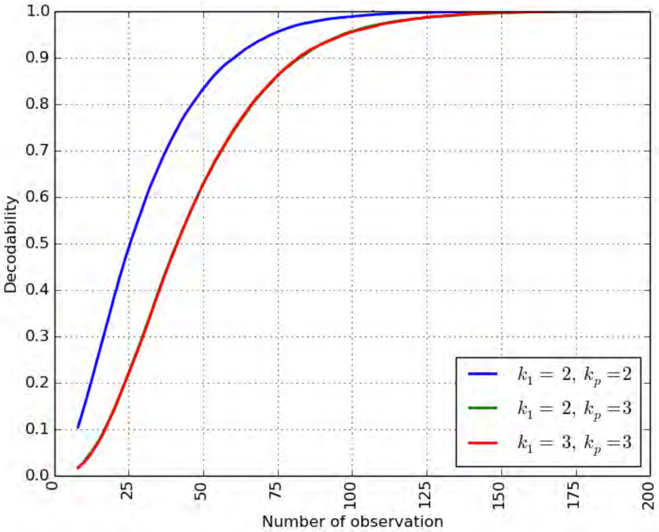


Figure 6.10: Decodability rate in multi-dimensional scenario for $N_l = N_p = 4$ and varying k_l, k_p.

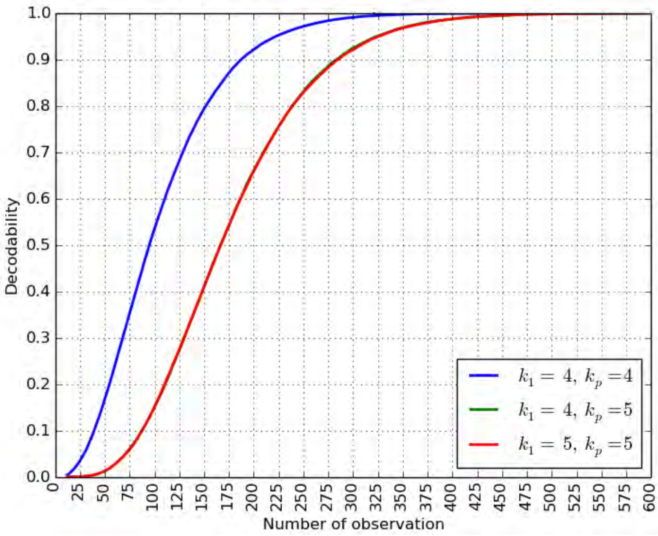


Figure 6.11: Decodability rate in multi-dimensional scenario for $N_l = N_p = 6$ and varying k_l, k_p.

Another interesting pattern for multi-dimensional scenario is shown in Figure 6.10 and Figure 6.11. Here N_l and N_p are fixed but the k_l and k_p are different. These two figures depict that if the the cardinality of both dimensions are same i.e. $N_l = N_p$, then the decodability rate mainly increases for the highest anonymity among k_l and k_p. Hence, decreasing the anonymity preference for only one dimension does not effect much on decodability rate if both dimensions' cardinality is same.

6.1.3 Comparison with Subset-Coding (DGAS)

Our proposed approach is robust compared to previous subset-coding approach (DGAS) [1]. We have shown in Section 6.1.1 that our algorithms can perform in manageable time even for larger number of OOIs e.g. $N = 10$. For the high time complexity, DGAS approach is not feasible for larger N. However, we have compared the decodability rate of DGAS with our results for smaller N in Fig. 6.12. This result shows that our decodability performance does not differ much compared to DGAS.

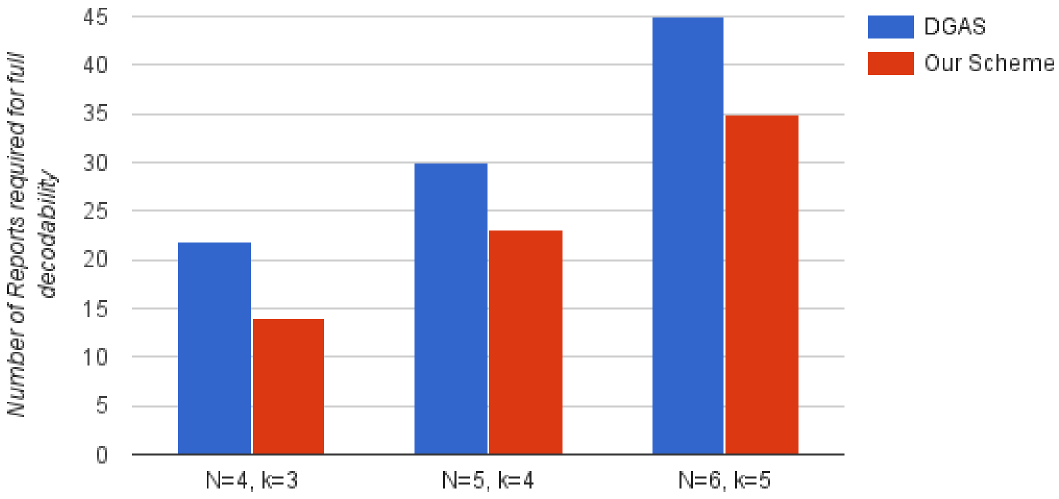


Figure 6.12: Comparison of decodability rate between DGAS and our approach for smaller $N = 4 \sim 6$ and highest anonymity $k = N - 1$.

6.1.4 Results for Variable User Preference

In real world, individual's privacy concern varies with many parameters such as culture of the society and family, job position, age, etc. Therefore, choosing a universal anonymity preference (k) for all users is sometimes impractical. Moreover, incentive schemes may reward lower anonymity preference more if it is found better to gain decodability. From this consideration, we would like to show the response of our proposed schemes against variable anonymity preference. Without loss of generality, we show result for three different configurations in Figure 6.13. First, we consider a fixed $k = 12$ for $N = 15$. Then, we compare it with $k = 10$ and 14 in equal proportion. Finally, we like to distribute user preferences in three equal portions for $k = 10$, 12, and 14, respectively. We find that there is not significant change in decodability for variable anonymity preference. Same pattern is observed in Figure 6.14. Thus, our algorithms offer a flexibility to satisfy users with diverse anonymity preference.

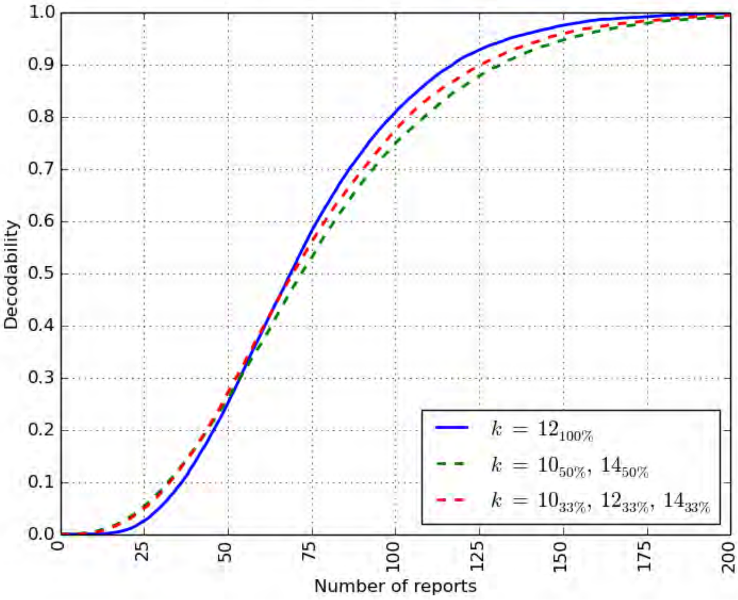


Figure 6.13: Decodability rate with different proportion of k between $12 \sim 14$.

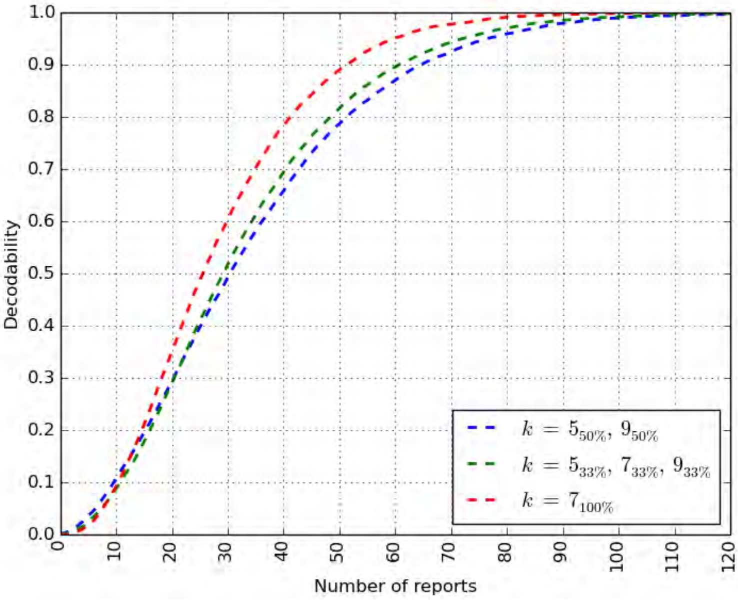


Figure 6.14: Decodability rate with different proportion of k between $5 \sim 9$.

6.1.5 Results with Missing Reports

All the reports anonymized by AS may not be received by ApS. It can happen if the participant does not send the anonymized report to ApS or packet is lost during transmission. Our proposed scheme is tolerant to this type of scenario. Figure 6.15 shows the decodability rate with respect to different percentage of missing reports denoted as μ for $N = 8$ and $k = N - 1$. The figure clearly shows that the decodability rate in ideal situation (no missing report) and in 10% missing rate does not differ significantly. The same scenario is shown in Figure 6.16 and Figure 6.17 for different N. Hence, in real world our scheme can decode at satisfactory rate even if some reports are lost. The decodability rate in presence of packet loss with respect to anonymity preference is shown in Figure 6.18 and Figure 6.19. The graphs show the decodability rate with fixed $N = 10$ with $k = 7$ and $k = 8$ respectively. Here, the decodability rate increases for increasing the anonimity preference. Hence, even with missing report rate, the decodability rate increases with anonymity for fixed N. This pattern shows that the decodability rate with packet loss does not have direct relationship with anonymity preference and simply slowers the decodability rate for missing reports.

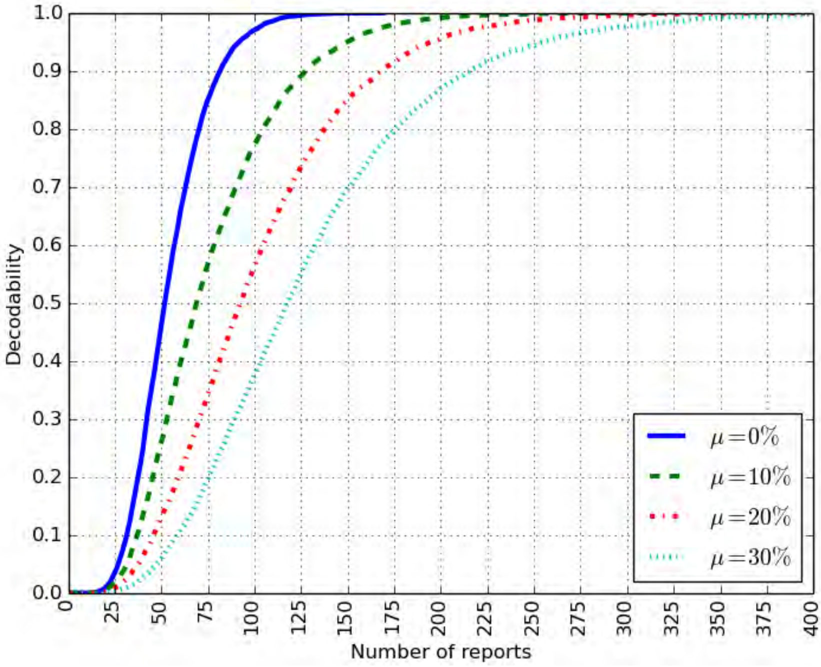


Figure 6.15: Decodability rate with varying missing report rate, μ for $N = 8$.

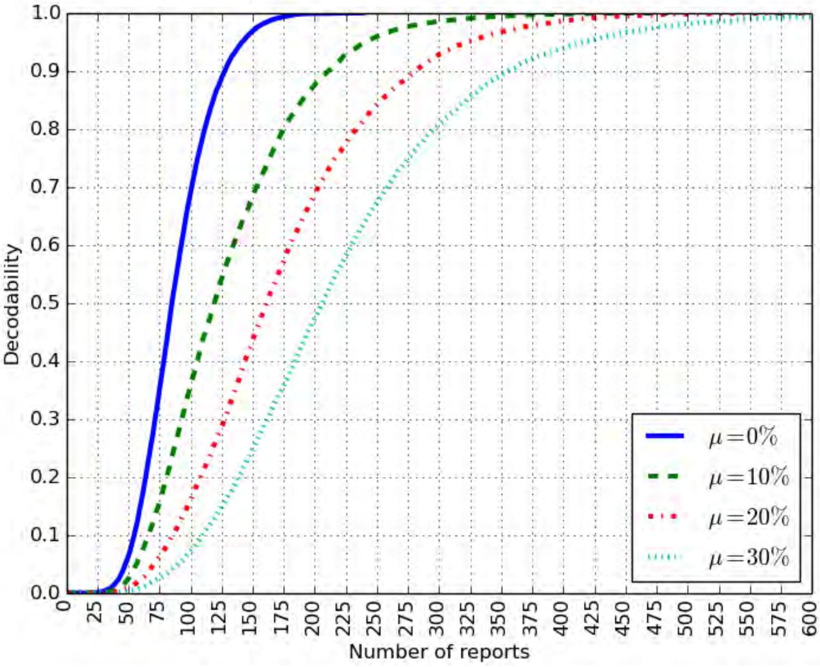


Figure 6.16: Decodability rate with varying missing report rate, μ for $N = 10$.

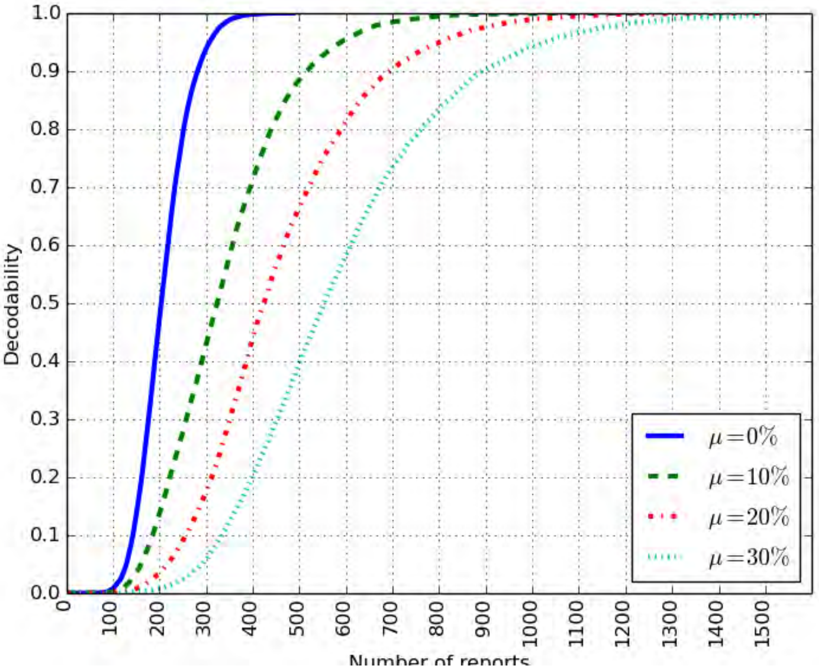


Figure 6.17: Decodability rate with varying missing report rate, μ for $N = 15$.

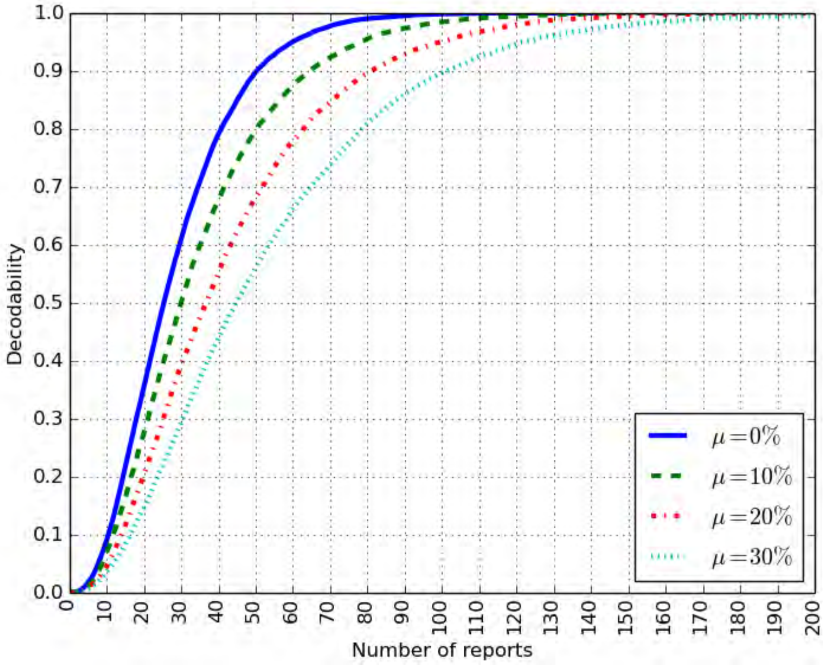


Figure 6.18: Decodability rate with varying missing report rate, μ for $N = 10, k = 7$.

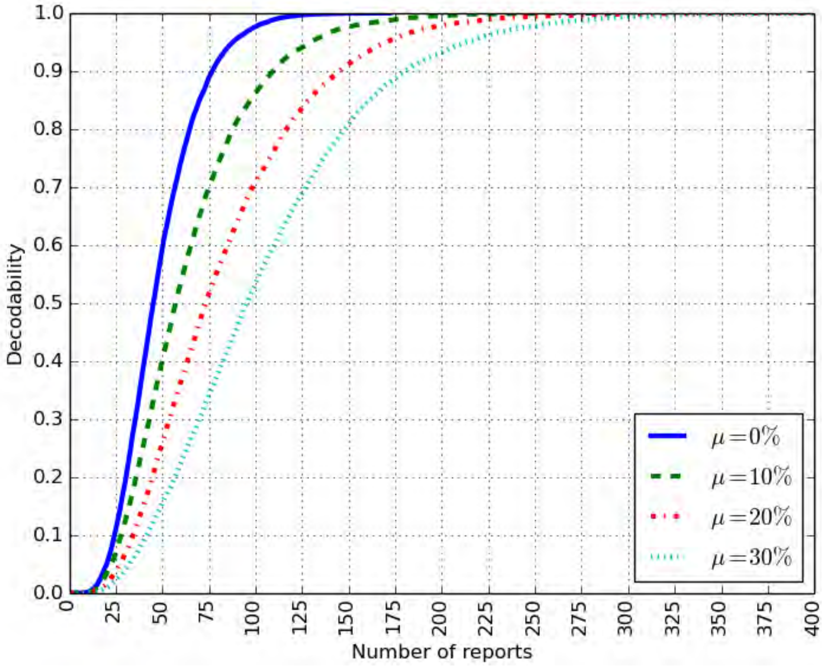


Figure 6.19: Decodability rate with varying missing report rate, μ for $N = 10, k = 8$.

6.1.6 Results with Faulty Reports

In real world, PSS may not always receive correct attributes of OOIs. An user might send false attribute intentionally or unintentionally. As discussed in Section 6.1.6, our scheme is capable of adapt to this situation and can decode the correct attributes eventually after getting enough correct reports. Figure 6.20 shows the decodability rate for different faulty report rate i.e. ϵ for multi-dimensional scenario for $N_l = 3$, $N_p = 3$ with highest anonymity $k = N - 1$. From the figure, it is apparent that our scheme can achieve full decodability at 15% false report rate after receiving 416 reports on average which is quite satisfactory.

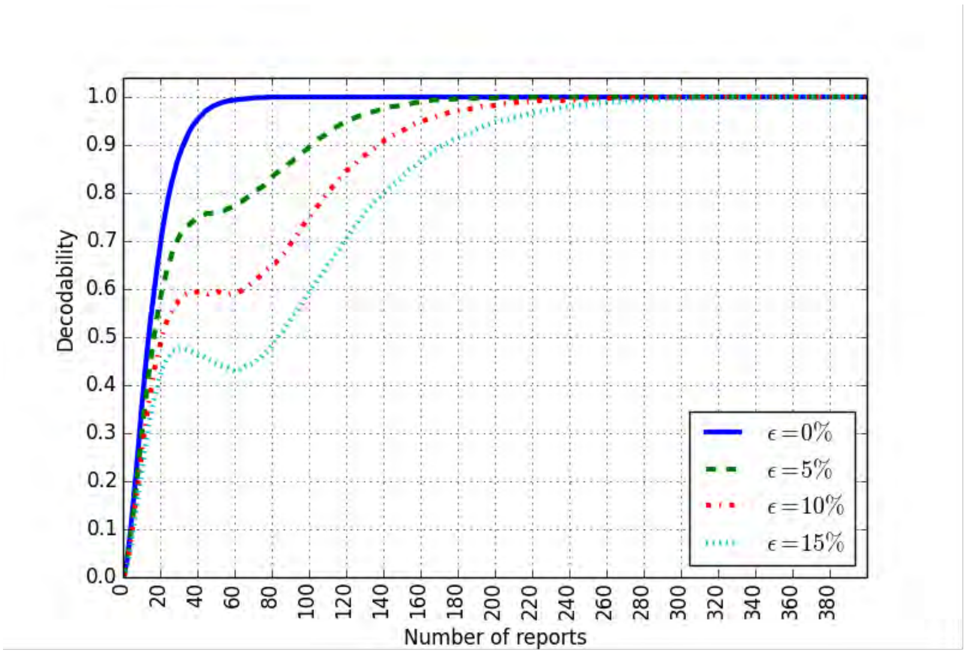


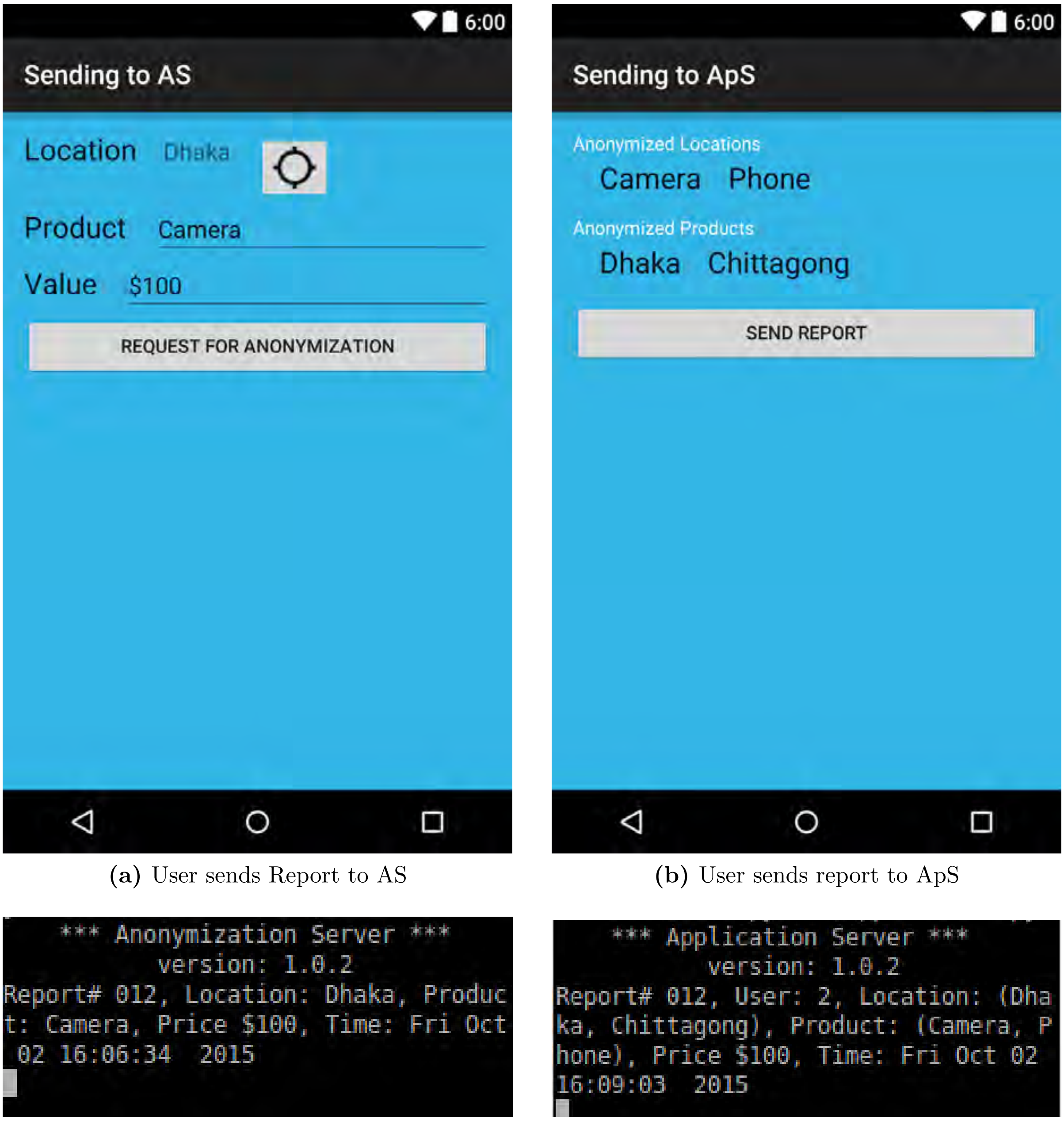
Figure 6.20: Decodability rate with varying faulty report rate, ϵ for $N_l = 3, N_p = 3$ and highest anonymity $(k = N - 1)$.

6.2 Android Prototype Based Experiment

We have developed Android based software prototypes as a proof of concept of our proposed scheme which can be applied in real-world scenario. Specifically, it has modules for the users to send actual report to the AS, receive ARs from AS and forward this with user id to the ApS. Figure 6.21a and Figure 6.21b show the user-interfaces for sending report to AS and to ApS. And Figure 6.21c and Figure 6.21d show the servers' responses.

With the help of this application using Android Smart-phones (connected to Internet and equipped with GPS) and two separate servers dedicated as AS and ApS built with

Python Tornado Web Framework, we test our anonymization and decoding algorithms that run in AS and ApS, respectively. Here the user's current location is obtained from device's GPS and other information like product and actual attribute is taken as user input. After receiving the anonymized report from AS, user can directly send the report to ApS as shown in Figure 6.21b. This user's report is received by ApS shown in Figure 6.21d. ApS can decode all the reported products successfully from the received ARs. We have simulated our application for 10 different PSS scenario where in each scenario, $N_l = 3$, $N_p = 3$ and $k_l = k_p = 2$. Figure 6.22 shows the result of our experiment where 100% decodability has been achieved after recieving 59 reports on average. We have calculated this decodability rate by averaging the results by running the application for 10 times.



Sending to AS

Location Dhaka

Product Camera

Value $100

(a) User sends Report to AS

Sending to ApS

Anonymized Locations

Camera Phone

Anonymized Products

Dhaka Chittagong

(b) User sends report to ApS

```
    *** Anonymization Server ***
          version: 1.0.2
Report# 012, Location: Dhaka, Produc
t: Camera, Price $100, Time: Fri Oct
 02 16:06:34  2015
```

(c) AS receives user's report without identity

```
    *** Application Server ***
          version: 1.0.2
Report# 012, User: 2, Location: (Dha
ka, Chittagong), Product: (Camera, P
hone), Price $100, Time: Fri Oct 02
16:09:03  2015
```

(d) ApS receives AR with user identity

Figure 6.21: Android prototype and server responses.

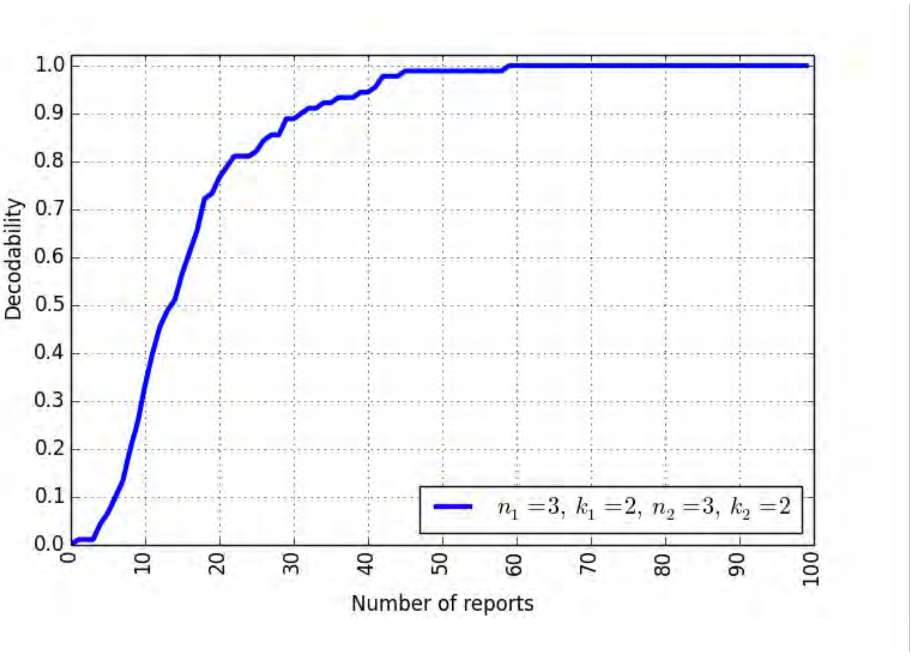


Figure 6.22: Android real world experimental result.

6.3 Conclusion

In this chapter, we have presented the results obtained from the simulator of PSS implemented according to our proposed approaches. Using different values for N and k we have demonstrated the impact of these parameters on the decodability rate. This analysis will help to decide suitable values of these parameters in different application scenarios without compromising the performance of PSS according to the current demand. Besides using synthesis data, we have also presented the results obtained from the Android test bed. In the next chapter, we are going to summarize our contributions on privacy-protection for PSS along with future direction.

Chapter 7

Conclusions and Future Works

Protecting privacy and rewarding the participators along with keeping data quality at the receivers' end is a challenging problem in PSS. This project aims in solving this problem with feasible computational complexity. It has also developed technique that preserves privacy in multiple directions. In pursuing this aim, the key achievements of our thesis and their significance are summarized as follows:

- The main innovation of this work is to provide privacy with anonymization model while maintaining data quality supporting incentive facility with feasible time complexity.
- In our proposed technique, we have provided the flexibility of choosing privacy preference for the participants. Hence, in the same PSS, user can demand higher incentive in return of his/her sacrifice on privacy by choosing anonymity preference. This flexibility also makes a balance in decodability rate by allowing different types of participants with different privacy preferences.
- In reality, reporting the attribute of an OOI may cause privacy threat in multiple dimensions. Hence, we have extended our algorithm for multi-dimensional scenario which is the first solution to this problem to the best of our knowledge. This multi-dimensional algorithm also provides the flexibility to the user for having different user preference for different dimensions.
- We have simulated a PSS for both single and multi-dimensional scenarios and investigated the performance of proposed anonymization and decoding schemes varying number of OOIs, anonymity preference and other relevant parameters. We have also implemented an Android prototype of our simulator to test the real world applicability of our proposed schemes.

Our simulation and experimental results show that we can achieve sufficient data integrity at the target end from a feasible number of user reports. This approach is likely to contribute in making participatory sensing a popular technology to the community ensuring privacy of participants without com-promising accuracy of data and also facilitate flexible incentive scheme for users with different anonymity preference. Our research findings presented in this report can be extended to the following areas:

- In our algorithm and simulation, we have assumed that each OOI has unique attribute. But in real time, there might be several OOIs which have the same attribute. Our algorithm can be extended by using efficient data structure to support this scenario.

- Most of the PSS aims at providing direct service to the users like finding cheapest petrol pump around an area. In this context, the attributes of OOIs keep changing in regular interval or frequently. Our algorithm can be extended to adapt to this dynamic environment to provide quality service to users.

- Our proposed algorithm requires a trusted Anonymization Server, AS. If the anonymization can be done without the help of AS, it will mitigate the communication cost and reduce the privacy risk. Achieving data integrity along with privacy protection and the facility of incentive scheme without Anonymizer is an open challenge in PSS.

Bibliography

[1] Manzur Murshed, Anindya Iqbal, Tishna Sabrina, and Kh Mahmudul Alam. A subset coding based k-anonymization technique to trade-off location privacy and data integrity in participatory sensing systems. In *Network Computing and Applications (NCA), 2011 10th IEEE International Symposium on*. IEEE, 2011.

[2] Yi F Dong, S Kanhere, Chun Tung Chou, and Nirupama Bulusu. Automatic collection of fuel prices from a network of mobile cameras. In *Distributed computing in sensor systems*. Springer, 2008.

[3] Shitiz Sehgal, Salil S Kanhere, and Chun Tung Chou. Mobishop: Using mobile phones for sharing consumer pricing information. In *Demo Session of the Intl. Conference on Distributed Computing in Sensor Systems*. Citeseer, 2008.

[4] Linda Deng and Landon P Cox. Livecompare: grocery bargain hunting through participatory sensing. In *Proceedings of the 10th workshop on Mobile Computing Systems and Applications*. ACM, 2009.

[5] Nirupama Bulusu, Chun Tung Chou, Salil Kanhere, Yifei Dong, Shitiz Sehgal, David Sullivan, and Lupco Blazeski. Participatory sensing in commerce: Using mobile camera phones to track market price dispersion.

[6] Joaquin Ballesteros, Mosaddequr Rahman, Bogdan Carbunar, and Naphtali Rishe. Safe cities. a participatory sensing approach. In *Local Computer Networks (LCN), 2012 IEEE 37th Conference on*. IEEE, 2012.

[7] S. B. Eisenman, E. Miluzzo, N. D. Lane, R. A. Peterson, G-S. Ahn, and A. T. Campbell. The bikenet mobile sensing system for cyclist experience mapping. In *Proceedings of the 5th International Conference on Embedded Networked Sensor Systems*. ACM, 2007.

[8] Bret Hull, Vladimir Bychkovsky, Yang Zhang, Kevin Chen, Michel Goraczko, Allen Miu, Eugene Shih, Hari Balakrishnan, and Samuel Madden. Cartel: A distributed mobile sensor computing system. In *Proceedings of the 4th International Conference on Embedded Networked Sensor Systems*. ACM, 2006.

[9] Raghu K. Ganti, Nam Pham, Hossein Ahmadi, Saurabh Nangia, and Tarek F. Abdelzaher. Greengps: A participatory sensing fuel-efficient maps application. In *Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services*. ACM, 2010.

[10] Prashanth Mohan, Venkata N. Padmanabhan, and Ramachandran Ramjee. Nericell: Rich monitoring of road and traffic conditions using mobile smartphones. In *Proceedings of the 6th ACM Conference on Embedded Network Sensor Systems*, 2008.

[11] Sasank Reddy, Andrew Parker, Josh Hyman, Jeff Burke, Deborah Estrin, and Mark Hansen. Image browsing, processing, and clustering for participatory sensing: Lessons from a dietsense prototype. In *Proceedings of the 4th Workshop on Embedded Networked Sensors*. ACM, 2007.

[12] Min Mun, Sasank Reddy, Katie Shilton, Nathan Yau, Jeff Burke, Deborah Estrin, Mark Hansen, Eric Howard, Ruth West, and Péter Boda. Peir, the personal environmental impact report, as a platform for participatory sensing systems research. In *Proceedings of the 7th International Conference on Mobile Systems, Applications, and Services*, 2009.

[13] Andrew T. Campbell, Shane B. Eisenman, Nicholas D. Lane, Emiliano Miluzzo, and Ronald A. Peterson. People-centric urban sensing. In *Proceedings of the 2Nd Annual International Workshop on Wireless Internet*, WICON '06, 2006.

[14] Hong Lu, Wei Pan, Nicholas D. Lane, Tanzeem Choudhury, and Andrew T. Campbell. Soundsense: Scalable sound sensing for people-centric applications on mobile phones. In *Proceedings of the 7th International Conference on Mobile Systems, Applications, and Services*, 2009.

[15] Elizabeth S. Cochran. The Quake-Catcher Network, 2010.

[16] Suhas Mathur, Tong Jin, Nikhil Kasturirangan, Janani Chandrasekaran, Wenzhi Xue, Marco Gruteser, and Wade Trappe. Parknet: Drive-by sensing of road-side parking

statistics. In *Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services*, 2010.

[17] Farrokh Jazizadeh and Burcin Becerik-Gerber. Toward adaptive comfort management in office buildings using participatory sensing for end user driven control. In *Proceedings of the Fourth ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*. ACM, 2012.

[18] Yuanqing Zheng, Pengfei Zhou, and Mo Li. How long to wait? predicting bus arrival time with mobile phone based participatory sensing. *IEEE Transactions on Mobile Computing*, 2014.

[19] Kuan Lun Huang, Salil S Kanhere, and Wen Hu. Preserving privacy in participatory sensing systems. *Computer Communications*, 2010.

[20] Ling Hu and Cyrus Shahabi. Privacy assurance in mobile sensing networks: go beyond trusted servers. In *Pervasive Computing and Communications Workshops (PERCOM Workshops)*. IEEE, 2010.

[21] Juong-Sik Lee and Baik Hoh. Sell your experiences: a market mechanism based incentive for participatory sensing. In *Pervasive Computing and Communications (PerCom), 2010 IEEE International Conference on*. IEEE, 2010.

[22] Roger Dingledine, Nick Mathewson, and Paul Syverson. Tor: The second-generation onion router. In *Proceedings of the 13th Conference on USENIX Security Symposium - Volume 13*, SSYM'04, pages 21–21, Berkeley, CA, USA, 2004. USENIX Association.

[23] Hsu-Chun Hsiao, TH-J Kim, Adrian Perrig, Akimasa Yamada, Samuel C Nelson, Marco Gruteser, and Wei Meng. Lap: Lightweight anonymity and privacy. In *Security and Privacy (SP), 2012 IEEE Symposium on*. IEEE, 2012.

[24] Chih-Jye Wang and Wei-Shinn Ku. Anonymous sensory data collection approach for mobile participatory sensing. In *Data Engineering Workshops (ICDEW), 2012 IEEE 28th International Conference on*. IEEE, 2012.

[25] Alastair R Beresford and Frank Stajano. Location privacy in pervasive computing. *IEEE Pervasive computing*, 2003.

[26] Balaji Palanisamy and Ling Liu. Mobimix: Protecting location privacy with mix-zones over road networks. In *Data Engineering (ICDE), 2011 IEEE 27th International Conference on*. IEEE, 2011.

[27] Sheng Gao, Jianfeng Ma, Weisong Shi, Guoxing Zhan, and Cong Sun. Trpf: A trajectory privacy-preserving framework for participatory sensing. *Information Forensics and Security, IEEE Transactions on*, 2013.

[28] Kuan Lun Huang, Salil S Kanhere, and Wen Hu. A privacy-preserving reputation system for participatory sensing. In *Local Computer Networks (LCN), 2012 IEEE 37th Conference on*. IEEE, 2012.

[29] Zhengli Huang, Wenliang Du, and Biao Chen. Deriving private information from randomized data. In *Proceedings of the 2005 ACM SIGMOD international conference on Management of data*. ACM, 2005.

[30] Carlo Blundo, Alfredo Santis, Giovanni Crescenzo, Antonio Giorgio Gaggia, and Ugo Vaccaro. *Advances in Cryptology — CRYPTO '94: 14th Annual International Cryptology Conference Santa Barbara, California, USA August 21–25, 1994 Proceedings*, chapter Multi-Secret Sharing Schemes, pages 150–163. Springer Berlin Heidelberg, Berlin, Heidelberg, 1994.

[31] G. F. Marias, C. Delakouridis, L. Kazatzopoulos, and P. Georgiadis. Location privacy through secret sharing techniques. In *World of Wireless Mobile and Multimedia Networks, 2005. WoWMoM 2005. Sixth IEEE International Symposium on a*, pages 614–620, June 2005.

[32] Sergio Mascetti, Dario Freni, Claudio Bettini, X. Sean Wang, and Sushil Jajodia. Privacy in geo-social networks: proximity notification with untrusted service providers and curious buddies. *The VLDB Journal*, 20(4):541–566, 2010.

[33] E. De Cristofaro, A. Durussel, and I. Aad. Reclaiming privacy for smartphone applications. In *Pervasive Computing and Communications (PerCom), 2011 IEEE International Conference on*, pages 84–92, March 2011.

[34] Fudong Qiu, Fan Wu, and Guihai Chen. Slicer: A slicing-based k-anonymous privacy preserving scheme for participatory sensing. In *Mobile Ad-Hoc and Sensor Systems (MASS), 2013 IEEE 10th International Conference on*, pages 113–121, Oct 2013.

[35] Gabriel Ghinita, Panos Kalnis, Ali Khoshgozaran, Cyrus Shahabi, and Kian-Lee Tan. Private queries in location based services: Anonymizers are not necessary. In *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*, SIGMOD '08, pages 121–132, New York, NY, USA, 2008. ACM.

[36] Emiliano De Cristofaro and Claudio Soriente. Short paper: Pepsi—privacy-enhanced participatory sensing infrastructure. In *Proceedings of the fourth ACM conference on Wireless network security*. ACM, 2011.

[37] H. Takabi, J. B. D. Joshi, and H. A. Karimi. A collaborative k-anonymity approach for location privacy in location-based services. In *Collaborative Computing: Networking, Applications and Worksharing, 2009. CollaborateCom 2009. 5th International Conference on*, pages 1–9, Nov 2009.

[38] Antonis Michalas and Nikos Komninos. The lord of the sense: A privacy preserving reputation system for participatory sensing applications. In *Computers and Communication (ISCC), 2014 IEEE Symposium on*. IEEE, 2014.

[39] M. Wernke, F. Drr, and K. Rothermel. Pshare: Position sharing for location privacy based on multi-secret sharing. In *Pervasive Computing and Communications (PerCom), 2012 IEEE International Conference on*, pages 153–161, March 2012.

[40] Pavel Skvortsov. *Position sharing for location privacy in non-trusted systems.* PhD thesis, Stuttgart, Universität Stuttgart, Diss., 2015, 2015.

[41] Pavel Skvortsov, Frank Dürr, and Kurt Rothermel. *Pervasive Computing: 10th International Conference, Pervasive 2012, Newcastle, UK, June 18-22, 2012. Proceedings*, chapter Map-Aware Position Sharing for Location Privacy in Non-trusted Systems, pages 388–405. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[42] D. Christin, J. Guillemet, A. Reinhardt, M. Hollick, and S. S. Kanhere. Privacy-preserving collaborative path hiding for participatory sensing applications. In *Mobile Adhoc and Sensor Systems (MASS), 2011 IEEE 8th International Conference on*, pages 341–350, Oct 2011.

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

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

[45] Gunhee Lee, Wonil Kim, and Dong-kyoo Kim. An effective method for location privacy in ubiquitous computing. In *Proceedings of the 2005 International Conference on Embedded and Ubiquitous Computing*, EUC'05, pages 1006–1015, Berlin, Heidelberg, 2005. Springer-Verlag.

[46] H. Kido, Y. Yanagisawa, and T. Satoh. An anonymous communication technique using dummies for location-based services. In *Pervasive Services, 2005. ICPS '05. Proceedings. International Conference on*, pages 88–97, July 2005.

[47] Pravin Shankar, Vinod Ganapathy, and Liviu Iftode. Privately querying location-based services with sybilquery. In *Proceedings of the 11th International Conference on Ubiquitous Computing*, UbiComp '09, pages 31–40, New York, NY, USA, 2009. ACM.

[48] J. Shi, R. Zhang, Y. Liu, and Y. Zhang. Prisense: Privacy-preserving data aggregation in people-centric urban sensing systems. In *INFOCOM, 2010 Proceedings IEEE*, pages 1–9, March 2010.

[49] Michael M Groat, Ben Edwards, James Horey, Wenbo He, and Stephen Forrest. Enhancing privacy in participatory sensing applications with multidimensional data. In *Pervasive Computing and Communications (PerCom), 2012 IEEE International Conference on*. IEEE, 2012.

[50] Matt Duckham and Lars Kulik. Simulation of obfuscation and negotiation for location privacy. In *Spatial Information Theory*. Springer, 2005.

[51] Buğra Gedik and Ling Liu. Location privacy in mobile systems: A personalized anonymization model. In *Distributed Computing Systems, 2005. ICDCS 2005. Proceedings. 25th IEEE International Conference on*. IEEE, 2005.

[52] Khuong Vu, Rong Zheng, and Lie Gao. Efficient algorithms for k-anonymous location privacy in participatory sensing. In *INFOCOM, 2012 Proceedings IEEE*. IEEE, 2012.

[53] D. Christin, D.M. Bub, A. Moerov, and S. Kasem-Madani. A distributed privacy-preserving mechanism for mobile urban sensing applications. In *Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), 2015 IEEE Tenth International Conference on*, pages 1–6, April 2015.

[54] Idalides J Vergara-Laurens, Diego Mendez, and Miguel A Labrador. Privacy, quality of information, and energy consumption in participatory sensing systems. In *Pervasive Computing and Communications (PerCom), 2014 IEEE International Conference on*, pages 199–207. IEEE, 2014.

[55] Daojing He, S. Chan, and M. Guizani. User privacy and data trustworthiness in mobile crowd sensing. *Wireless Communications, IEEE*, 22(1):28–34, February 2015.

[56] Xinlei Wang, Wei Cheng, Prasant Mohapatra, and Tarek Abdelzaher. Enabling reputation and trust in privacy-preserving mobile sensing. *Mobile Computing, IEEE Transactions on*, 2014.

[57] Xinlei Wang, K. Govindan, and P. Mohapatra. Collusion-resilient quality of information evaluation based on information provenance. In *Sensor, Mesh and Ad Hoc Communications and Networks (SECON), 2011 8th Annual IEEE Communications Society Conference on*, pages 395–403, June 2011.

[58] Qinghua Li and Guohong Cao. Privacy-preserving participatory sensing. *Communications Magazine, IEEE*, 53(8):68–74, August 2015.

[59] D. Christin, C. Rosskopf, M. Hollick, L.A. Martucci, and S.S. Kanhere. Incognisense: An anonymity-preserving reputation framework for participatory sensing applications. In *Pervasive Computing and Communications (PerCom), 2012 IEEE International Conference on*, pages 135–143, March 2012.