Int. J. Communications, Network and System Sciences, 2020, 13, 73-103
https://www.scirp.org/journal/ijcns
ISSN Online: 1913-3723
ISSN Print: 1913-3715

Multi-Dimensional Anonymization for Participatory Sensing Systems

Nafeez Abrar[1], Shaolin Zaman[1], Anindya Iqbal[2], Manzur Murshed[3]

[1]Bangladesh University of Engineering and Technology, Dhaka, Bangladesh
[2]Department of CSE, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh
[3]IEEE, Federation University, Melbourne, Australia
Email: nafeezabrar@gmail.com, shaolinkhusbu@gmail.com, anindya@cse.buet.ac.bd, manzur.murshed@federation.edu.au

How to cite this paper: Abrar, N., Zaman, S., Iqbal, A. and Murshed, M. (2020) Multi-Dimensional Anonymization for Participatory Sensing Systems. *Int. J. Communications, Network and System Sciences*, **13**, 73-103.
https://doi.org/10.4236/ijcns.2020.136006

Received: May 24, 2020
Accepted: June 27, 2020
Published: June 30, 2020

Abstract

Participatory sensing systems are designed to enable community people to collect, analyze, and share information for their mutual benefit in a cost-effective way. The apparently insensitive information transmitted in plaintext through the inexpensive infrastructure can be used by an eavesdropper to infer some sensitive information and threaten the privacy of the participating users. Participation of users cannot be ensured without assuring the privacy of the participants. Existing techniques add some uncertainty to the actual observation to achieve anonymity which, however, diminishes data quality/utility to an unacceptable extent. The subset-coding based anonymization technique, DGAS [LCN 16] provides the desired level of privacy. In this research, our objective is to overcome this limitation and design a scheme with broader applicability. We have developed a computationally efficient subset-coding scheme and also present a multi-dimensional anonymization technique that anonymizes multiple properties of user observation, e.g. both location and product association of an observer in the context of consumer price sharing application. To the best of our knowledge, it is the first work which supports multi-dimensional anonymization in PSS. This paper also presents an in-depth analysis of adversary threats considering collusion of adversaries and different report interception patterns. Theoretical analysis, comprehensive simulation, and Android prototype based experiments are carried out to establish the applicability of the proposed scheme. Also, the adversary capability is simulated to prove our scheme's effectiveness against privacy risk.

Keywords

Anonymization, Privacy, Location Privacy, Participatory Sensing

1. Introduction

Participatory Sensing System (PSS) is a framework facilitating community members sense, collect, analyze, and share information obtained from the surroundings for mutual benefit. This has evolved as a cost-effective alternative for reliable and impartial data collection, processing and dissemination. Smartphones equipped with high-precision localization capability and camera or other ad-hoc sensing devices mounted on vehicles may be used to record objects/events of interest by the participants. These captured data are then reported to the *Application Server* (ApS) using existing lightweight wireless communication networks. ApS is expected to extract valuable information from a collection of reports to repay the participants by responding to specific queries on-demand. Consumer price information sharing applications [1] [2], safety and monitoring [3] and query based user-services [1] [4] are some of the widely used PSS applications.

Privacy preservation of the participants is a pre-requisite for the success of PSS. An eavesdropper may infer sensitive information about an observer by intercepting some reports. Hiding the data ownership is not an option in this context as it infringes the reputation schemes needed by ApS to assess the reliability and trustworthiness of data and also for developing incentive mechanisms [5]. Hence, any mix-network based scheme that creates hard-to-trace communications by using a chain of proxies e.g. hot-potato-privacy-protection (HP^3) [6], cannot be used. Encrypting the reports before transmitting is also not a viable option. As the reported data is usually drawn from a small range, a smart adversary may encrypt some predicted messages with the public key of the application server and then match these against the received message. Moreover, PSS is expected to rely on public networks where encryption may unnecessarily raise concerns within the law enforcement agencies.

The existing privacy protection mechanisms, where information is transmitted with some anonymity or by adding Gaussian noise or at reduced precision, cannot be used here as the destination expects complete data recoverability at the individual level. For example, PetrolWatch [1] assists drivers to find the cheapest fuel station in the neighborhood. If it recommends a station with a higher price for a user looking for the cheapest one, its reputation would be destroyed. Another technique called Spatial cloaking [3] also fails to achieve data quality as they refer to a geographic region in reply to the query for a specific location. If more than one candidate objects exist in that region, it becomes confusing. Some data aggregation techniques [7] have been proposed where aggregate information (such as the average weight of a community) is required at service provider end but these approaches are limited to applications. 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. The aim of this paper is to provide a solution to this significant challenge.

Our previous *k*-anonymization techniques proposed in [8] [9] aimed to ensure data recoverability at the individual level while preserving the privacy of PSS participants. In this context, an observation is *k*-anonymous if the observed object/event is indistinguishable from $k-1$ other objects and data recoverability implies the ability to map an object to its attribute (e.g. map a petrol pump with its petrol price) by de-anonymizing a collection of anonymized reports. However, the Probabilistic Greedy Anonymization Scheme (PGAS) [8] was limited in application as it suffers from very high computational complexity, *i.e.* $O(N!)$ where N implies the number of objects of interest. In an attempt to overcome this limitation, a deterministic approach was adopted in [9] named Deterministic Greedy Anonymization Scheme (DGAS). This approach is a subtraction-based (Set difference) technique which iteratively rules out impossible mapping combinations from all possible set. The advantage of this approach is that it can achieve global optimization as it considers all possible combinations. However, the cardinality of all possible combinations of mappings is exponential number which leads to exponential order of anonymization. This limitation makes this approach infeasible in practical scenarios. Complexity of DGAS has been improved in Fast Deterministic Greedy Anonymization Scheme (FDGAS). However, it could work with only a fixed degree of anonymity (*i.e.* $k = N-1$) and thus sacrificed the flexibility of user preference in desired anonymity. In this work, we have developed a technique that allows the flexibility of any user preference of anonymity without increasing complexity than FDGAS [9]. Our proposed approach uses attribute-centric scheme which keeps track of possible objects/events per reported attribute instead of computing over all possible combinations. Thus the order of anonymization is reduced to polynomial time, *i.e.* $O(N)$ effectively. This achievement allows us to design the first anonymization scheme for the multi-dimensional scenario in PSS and overcomes the restriction over anonymity, *k*.

To the best of our knowledge, this is the first work that attempts simultaneously anonymizing multiple attributes of an observation. The proposed *OC*-based scheme is developed first for single-dimension anonymization and then extended to the multi-dimensional scenarios. The paper also presents comprehensive analysis on privacy risk by the adversaries. By considering colluding adversary models and different message interception patterns of the adversaries, the analysis confirms the robustness of the proposed *k*-anonymization scheme against a wide range of malicious attacks. The specific contributions of this paper are listed below:

- Developing the first Multi-dimensional Anonymization Scheme (MDEAS) of PSS that provides anonymity in multiple dimensions.
- Designing efficient anonymization and de-anonymization algorithms that preserve high data recoverability at the desired end and exploiting some optimization issues.
- Designing *k*-anonymization technique that works with variable *k*, *i.e.* different user preference of anonymity. This is useful to design incentive schemes

considering users' choice of privacy, *i.e.* offering more incentive to a user who demands less anonymity.

- Presenting theoretical analysis on desirable properties of MDEAS and validating those with extensive simulation.
- Comprehensively analyzing privacy risk with simulation in the presence of different types of adversaries.
- Conducting real-world experiments with Android-based prototype.

The organization of the rest of the paper is as follows. Section 2 describes the system model of PSS, its related terminologies, and the adversary model. Section 3 explains our proposed scheme with a detailed example. Section 4 presents the optimizations that can be applied in MDEAS with an example. We also present a theoretical analysis on required number of reports to achieve full de-anonymization in Section 5. We present the Algorithm and its computational complexity in Section 6. In Section 7, we present our simulation setup and the results of experiments. In Section 8, we discuss the previously proposed privacy schemes on PSS along with their limitations for real world applications. Finally, Section 9 concludes the paper.

2. System Model

2.1. PSS Architecture

In PSS, users (mobile nodes) report their observations about some objects or events to an Application Server (ApS). The ApS wants to collect information about some particular objects/events (e.g. price of fuel). We denote this domain as PSS scenario. We use the term Objects of Interest to represent the objects/events that a PSS scenario is interested about.

Definition 1 (of Interest). *An Object of Interest* (*OOI*) *is an object/event whose attribute/property is observed and reported by the participants of a PSS application*.

Here we discuss the major entities of a PSS scenario in brief. The users are independent and they do not collectively send reports and there is no apparent communication between them. To protect the privacy of a user, several anonymization schemes are applied in PSS that use an Anonymization Server (AnS) [5]. This server remains transparent from Application Server (ApS) which actually provides desired service to the community. The service of AS is provided by some trusted third party e.g. a government agency. The whole process of privacy preserving participatory sensing in the perspective of a user can be divided into two parts:

1) Anonymization Step: User reports the observed data to Anonymization Server (AnS) that transforms and returns an anonymized report (AR) to her. Since incentive or reputation schemes do not rely on this communication, user association with these reports need not be preserved. This reporting is done through mix-network based communication [6].

2) Reporting Step: User sends AR to ApS along with his/her identity infor-

mation. ApS de-anonymizes these reports to map an attribute to an *OOI* for offering the desired service to the participants. This reporting is done through conventional, inexpensive and un-encrypted communication network.

Note that the primary purpose of using AnS is to reduce the number of required observations to map all the *OOIs* to their corresponding attributes by the ApS. The report sent to the AnS for anonymization contains user preference of anonymity (denoted as *k*), *OOI* identification, and the observed value/attribute. For example, a participant John observes the price of a camera. He wants to report camera's price with 3-anonymity *i.e.* $k=3$. So he sends the report $\langle \text{Camera}\{3\}\rangle : \$100$ to the AnS. Now, AnS may anonymize his report as $\langle \text{Camera}, \text{Phone}, \text{GPS}\rangle : \$100$ and returns this AR to John. Next, John sends this AR to the ApS with his identity.

As alluded in the previous section, users' association with multiple objects/events can also be protected simultaneously in our proposed scheme. Let John reports the price of a camera on a particular location *loc*1 and wants to anonymize his report both in terms of location and product simultaneously. In many cases, this is essential as the price of the product varies with location. Let us 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 scenario, let the total number of *OOIs* for all dimensions be denoted as $N_1, N_2, \cdots, N_d$ and their respective sets of *OOIs* as $S_1, S_2, \cdots, S_d$. Even the anonymity preference for each dimension can be different. We use $k_1, k_2, \cdots, k_d$ to denote the anonymity preference. The term *OOI* Combination is used to denote the collection of *d OOIs* from *d* dimensions for which an attribute is reported. Accordingly, the total number of *OOI* combinations is $X = \prod_{i=1}^{d} N_i$. Suppose, John reports his observed data to the AnS as $\langle \text{Camera}\{2\}, loc1\{3\}\rangle : \$100$ as shown in **Figure 1**. Accordingly, here $k_1 = 2$ and $k_2 = 3$ which implies that John wants to anonymize his report with two-anonymity on the observed product (camera) and three-anonymity on his location of purchase (loc1). In this situation, a valid AR sent by AnS might be $\langle \{\text{Camera}, \text{Phone}\}, \{loc1, loc2, loc3\}\rangle : \$100$. This notion can be expressed in general term with the following definition.

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

Hence, 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. A good anonymization algorithm should select the extra *OOIs* in such a way that ApS can de-anonymize them with few ARs.

Data quality is achieved when the ARs of a PSS scenario are fully de-anonymized. Here, we define the term Full De-anonymization as follows:

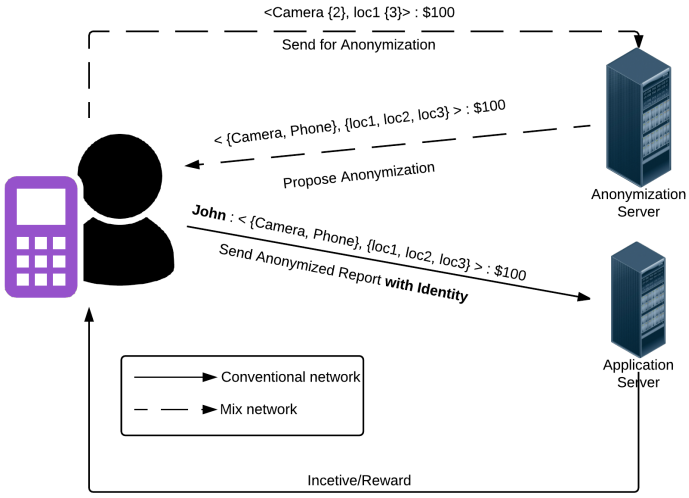


Figure 1. System model of a multi-dimensional PSS system.

Definition 3 (De-anonymization). *The outcome of an anonymization technique achieves full de-anonymization iff N OOIs* (*single-dimensional scenario*) *or X OOI combinations* (*multi-dimensional scenario*) *can be associated with their correct attributes.*

Another desired property of an anonymization technique in our context is to achieve full de-anonymization from a feasibly low number of anonymized reports. To measure this property, we define the term NRRFD as follows:

Definition 4 (NRRFD). *NRRFD* (*Number of Reports Required for Full De-anonymization*) *refers to the total number of ARs required to achieve full de-anonymization in a particular PSS scenario.*

The NRRFD depends on the order of appearance of ARs in PSS. However, anonymization techniques anonymize intelligently to keep the NRRFD minimal.

In our model, we assume that each *OOI* has a unique attribute; which may not be practical in some scenarios. However, the transformation of the non-unique scenario to the unique scenario can be accomplished by the AnS which can make the attribute unique by adding a small value below the level of significance when it receives the same attribute for different *OOIs*.

2.2. Adversary Model

The adversaries of PSS are assumed to be rational, *i.e.* does not attack the operation of the system. Rather they try to eavesdrop messages and reveal users' private information. As the AR sent to ApS includes participator's identity, the adversary residing near ApS is the strongest one (see **Figure 2**). This adversary tries to de-anonymize the eavesdropped ARs and thus reveal the *OOI*-user association.

The primary strategy against adversary is to divide the anonymization tasks of a PSS among different AnSs as presented in [8]. In this strategy, the *OOIs* under the PSS is divided into two or more groups and assigned to different AnSs. A user will be allowed to report from only one group with a single user id. The ad-

versaries are also assumed to be registered users of PSS and each adversary (one id) will be member of one group. A grouping example is presented in Figure 3 where multiple petrol stations with same names are assigned to different groups. If any user e.g. John registers in Group 1, he can report only the price of petrol pumps belonging to Group 1. However, as adversaries do not know the group id of John, by intercepting the AR of John, they will not be able to distinguish between the three petrol pumps with the same name *A* (see Figure 3). For example, *OOI A* of Group 1 from where John reported is not distinguishable from *OOI A* of Group 2 and thus creates confusion to the adversary.

Note that the user registration for a particular group of *OOIs* is done once in a while. Hence, it can be done via a secure website (e.g. HTTPS). Thus group

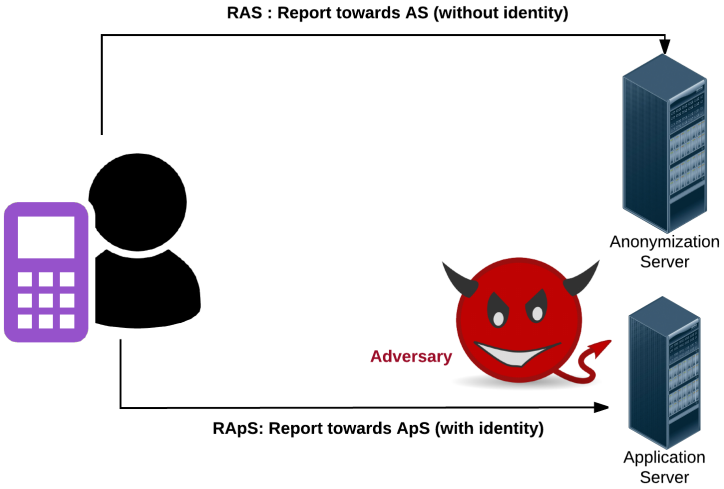


Figure 2. Adversary model of PSS.

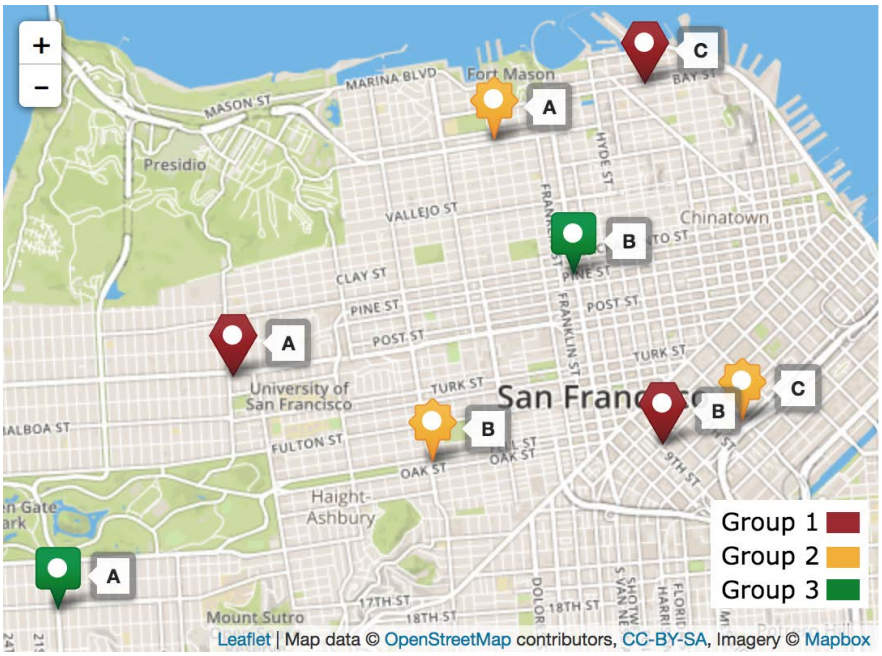


Figure 3. PSS grouping as the first line of defense against adversary.

association of user is not revealed to the adversary but known to ApS which can de-anonymize ARs according to their groups. However, our previous works [8] did not analyze the outcome of this strategy, *i.e.* to what extent adversary threat is mitigated. In this paper, we make a comprehensive analysis considering different possible attacks as discussed below.

2.2.1. Adversary Interception Pattern

We divide adversaries into three types depending on how they intercept the ARs of a PSS scenario as shown in Table 1. The Type 1 adversary is considered for theoretical interest only because it is practically impossible to intercept all reports for an adversary. The Type 2 adversary intercepts all reports for a limited period of time which is also less probable in practical scenarios. Type 3 is the most common type of adversary in practical PSS scenarios and the privacy risk against this type of adversary is investigated in detail.

Besides this interception pattern, we also consider some enhanced capabilities of adversaries as follows.

2.2.2. Collusion

Each adversary intercepts certain ARs and tries to de-anonymize those. However, the threat becomes stronger if multiple adversaries share information among them. As Type 2 and Type 3 adversaries cannot intercept all ARs, they can share their intercepted ARs among themselves and become stronger. Moreover, all types of adversaries can also share their own observed *OOI*-attribute mapping among themselves and de-anonymize attributes with the combined information.

2.2.3. Prediction

As adversaries cannot distinguish reports of different PSS groups, they cannot reveal the real *OOI* from their de-anonymization result. Even if we assume that the adversaries know the group mapping by registering in all groups or by collusion with members of other groups, it is still not possible to reveal the real *OOI* from the reported *OOI* name as adversaries do not know the users' group id. This argument is validated with simulation results.

Let adversaries make some prediction based on distance estimation to predict the real location (*OOI*) from the *OOIs* of different groups with same local ID, an equidistant location may be considered as a representative point of all these *OOIs*. The *OOI* nearest from this representative point can be considered as real *OOI* by adversary. We shall also empirically investigate if this strategy enables the adversary to cause more risk.

Table 1. Types of adversaries according to interception pattern.

Type of Adversary	Interception Pattern
Type 1	All ARs for whole period
Type 2	All ARs for a limited period
Type 3	Random proportion of ARs for whole period

3. Conceptual Framework of Proposed Scheme

We now discuss the conceptual framework with an illustration of examples of our proposed scheme, MDEAS (Multi-dimensional Effective Anonymization Scheme). MDEAS works efficiently and supports variable-length anonymity. Instead of keeping all possible combinations of attribute-*OOI* mapping, MDEAS keeps track of occurrence counts or absence counts for each reported attribute. When a user sends an actual report to the AnS, it tries to anonymize each observed report in such a way that ApS can de-anonymize maximum attributes. The whole scheme can be divided into two parts, *i.e.* Anonymization and De-anonymization. For the convenience of the readers, we first discuss the simple boundary case of the technique considering single dimension and then explain its expansion to multiple dimensions.

3.1. Single-Dimensional Scenario

In order to explain the concept, we assume a PSS scenario of consumer price sharing with $N = |S| = 4$ where it collects the price of four different products named *A*, *B*, *C*, and *D* that have prices \$10, \$20, \$30, and \$40, respectively. We have assumed these values of the parameters by analyzing real world application scenarios. For the sake of simplicity, we assume unique attributes. We assume the order of appearance of observations as shown in **Figure 4(a)**. However, the algorithm is applicable for any arbitrary order of appearance.

3.1.1. Anonymization Process

The goal of anonymization process is to generate AR by selecting $k-1$ additional *OOIs* along with the observed *OOI* in such a way that the joint de-

Incoming Order	Report	
	Real	Anonymized
1	$\langle A\{3\}\rangle$: \$10	$\langle A,B,C\rangle$: \$10
2	$\langle A\{3\}\rangle$: \$10	$\langle A,B,D\rangle$: \$10
3	$\langle A\{3\}\rangle$: \$10	$\langle A,C,D\rangle$: \$10
4	$\langle B\{2\}\rangle$: \$20	$\langle B,C\rangle$: \$20
5	$\langle B\{2\}\rangle$: \$20	$\langle B,D\rangle$: \$20
6	$\langle C\{2\}\rangle$: \$30	$\langle A,C\rangle$: \$30
7	$\langle C\{2\}\rangle$: \$30	$\langle B,C\rangle$: \$30
8	$\langle D\{2\}\rangle$: \$40	$\langle A,D\rangle$: \$30
9	$\langle D\{3\}\rangle$: \$40	$\langle C,D\rangle$: \$40

(a)

	A	*B*	*C*	*D*
IOC_A		0	0	0
IOC_B	0		0	0
IOC_C	0	0		0
IOC_D	0	0	0	

(b)

	A	*B*	*C*	*D*
IOC_A		0	0	**1**

(c)

	A	*B*	*C*	*D*
IOC_A		0	**1**	**1**

(d)

	A	*B*	*C*	*D*
IOC_A		**1**	**1**	**1**

(e)

	A	*B*	*C*	*D*
IOC_B	**1**		0	**1**

(f)

	A	*B*	*C*	*D*
IOC_B	**2**		**1**	**1**

(g)

	A	*B*	*C*	*D*
IOC_A		**1**	**1**	**1**
IOC_B	**1**		**1**	**2**
IOC_C	**2**	**1**		**1**
IOC_D	**1**	**2**	**1**	

(h)

Figure 4. Demonstration of the anonymization process in a single-dimensional scenario. (a) List of reports; (b) Initial; (c) First anonymization step; (d) Second anonymization step; (e) Third anonymization step; (f) Fourth anonymization step; (g) Fifth anonymization step; (h) Completion of anonymization.

anonymization can be done with a feasibly low number of ARs. Without any control over the distribution or order of observations, it cannot be done optimally. However, the AnS uses some heuristics such as maximizing the diversity of an AR with respect to previously generated ARs for same *OOI*. The AnS maintains a data structure named Inverse Occurrence Checklist (*IOC*) for N *OOIs*. The *IOC* for an *OOI p*, denoted as IOC_p, contains the absence count of each other *OOI* $q \mid q \in S \wedge q \neq p$ where S is the set of *OOIs*. We use the notion $IOC_p(q)$ to express how many times q could be, but has not been included in ARs of *OOI p*.

The rule for identifying whether an *OOI* might be de-anonymized is as follows:

Rule 1. (*Rule of being de-anonymizable*) *An OOI p is de-anonymizable if*

$$\forall_{q \in S \wedge q \neq p} \; IOC_p(q) > 0 \tag{1}$$

The rule for selecting $k-1$ *OOIs* is as follows:

Rule 2. (*Rule for selection of OOIs in an AR*) *To anonymize a report containing the attribute of OOI p, the set of selected OOIs,* S' *is formed as:*

$$S' = p \bigcup (k-1) \text{ largest IOC-valued OOIs in } (S \setminus p) \tag{2}$$

After anonymizing the report, AnS increases the count of those *OOIs* in S that have not been included in this AR. This rule of updating *IOC* values can be formalized as:

Rule 3. (*Rule of updating IOC*) *After producing an AR for OOI p, the* IOC_p *is updated as*

$$\forall_{q \in S \wedge q \neq p \wedge q \notin S'} \; IOC_p(q) = IOC_p(q) + 1 \tag{3}$$

All *IOC* values are initialized with zero (**Figure 4(b)**). Let the first three reports carry the price of *OOI A* (**Figure 4(a)**). To anonymize the first report, AnS chooses $k-1$ *OOIs* randomly. Hence, the AR might be any of $\langle A, B, C \rangle : \$10$ or $\langle A, B, D \rangle : \$10$ or $\langle A, C, D \rangle : \$10$. Without any loss of generality, we assume that the selected AR is $\langle A, B, C \rangle : \$10$. According to Rule 3, $IOC_A(D)$ is increased by one (**Figure 4(c)**) as D has not been used in this AR. When \$10 is reported again in the second report, AS chooses D as one of the anonymized *OOI* because $IOC_A(D)$ is greater than other *IOC* values. The other *OOI* is chosen randomly from B and C as both have same *IOC* value. Let, AnS chooses B as the other anonymized *OOI*. Hence, this time the AR is $\langle A, B, D \rangle : \$10$. According to Rule 3, now AS updates $IOC_A(C)$ because C does not exist in this AR (**Figure 4(d)**). The third report also contains the attribute of A. Now, AS chooses C and D as anonymized *OOIs* because $IOC_A(C) = IOC_A(D) = 1 > IOC_A(B)$. Hence, the AR is now $\langle A, C, D \rangle : \$10$. As B is not present, AnS increases $IOC_A(B)$ after producing this AR.

In this state, for *OOI* A, all the *IOC* values, *i.e.* $IOC_A(B), IOC_A(C), IOC_A(D) > 0$. Hence, according to Rule 1, A is de-anonymizable. In the same way, subsequent observations are de-anonymized. After the arrival of the ninth report, all the *OOIs* are de-anonymizable and *IOC*

table reaches the state shown in Figure 4(h) traversing through previous states shown in Figures 4(c)-(g).

3.1.2. De-Anonymization Process

This process runs in ApS. As ApS may not have prior knowledge of all *OOIs*, it cannot construct fixed length *IOC* table. Therefore, ApS maintains another data structure named *OC* (Occurrence Checklist) for each reported attribute. *OC* for a reported attribute *v*, denoted as OC_v, tracks the occurrence count of the candidate *OOIs* for *v*. We also use the notation $OC_v(p)$ to denote how many times *p* has been reported as candidate *OOIs* in all reports of attribute *v*. Besides OC_v, ApS tracks the total number of reports for each reported attribute *v* denoted as T_v. When an AR is received by the ApS, it follows the steps below:

1) Creates OC_v if *v* is reported for the first time to ApS. (all values initialized to zero).

2) Sets $T_v = T_v + 1$.

3) Updates $OC_v(p)$ as

$$\forall_{p \in \text{OOIs of AR}} \quad OC_v(p) = OC_v(p) + 1$$

An attribute *v* is mapped to an *OOI p i.e. p* is de-anonymized by ApS, if the following rule is satisfied for OC_v.

Rule 4. (*Rule of being de-anonymized*) *The attribute v is de-anonymized for OOI p if*

$$OC_v(p) = T_v \wedge \forall_{q \in S \wedge q \neq p \wedge q \notin S_D} OC_p(q) < T_v \tag{4}$$

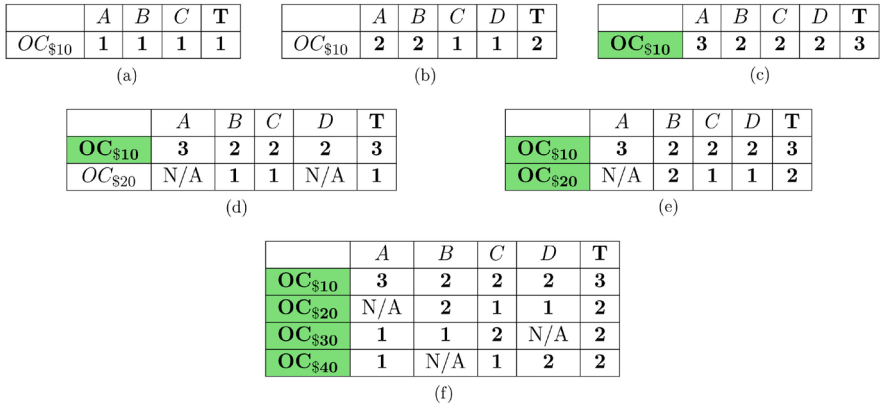
Considering our example, ApS receives the AR, $\langle A, B, C \rangle$: \$10 first. Since \$10 has not been reported before, ApS creates an *OC* for \$10 denoted as $OC_{\$10}$. As it is the first report of \$10, ApS sets $T_{\$10}$ to one. This report indicates that \$10 is a possible attribute of either *A*, *B* or *C*. Hence, ApS creates three *OC* columns for *A*, *B* and *C* denoted as $OC_{\$10}(A)$, $OC_{\$10}(B)$ and $OC_{\$10}(C)$ respectively and increases their *OC* values as shown in Figure 5(a).

The second AR received by ApS is $\langle A, B, D \rangle$: \$10. As \$10 has been reported before, ApS does not need to create $OC_{\$10}$ again. However, the *OOI D* has been reported to ApS for the first time in this report. Hence, ApS creates an additional column for *OOI D*. Next, ApS increases the *OC* values of the candidate *OOIs* of this report, *i.e.* $OC_{\$10}(A)$, $OC_{\$10}(B)$ and $OC_{\$10}(D)$ by one (Figure 5(b)). ApS also increases the total count *i.e.* $T_{\$10}$ by one. In the same manner, when the third AR is received by ApS which also contains the attribute \$10, ApS simply updates its corresponding $OC_{\$10}$. At this stage, the current *OC* counts and total counts for attribute \$10 is:

$$OC_{\$10}(A) = T_{\$10} = 3 \text{ and all other } OC\text{s, } i.e.,$$
$$OC_{\$10}(B) = OC_{\$10}(C) = OC_{\$10}(D) < T_{\$10}$$

Therefore, according to Rule 4, ApS can de-anonymize the attribute \$10 as only the OC_v value of *OOI A* is equal to $T_{\$10}$.

In the same process, *B*, *C* and *D* are de-anonymized by ApS after receiving fourth to ninth ARs (Figure 5(f)).



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

(a)

	A	B	C	D	**T**
$OC_{\$10}$	**2**	**2**	**1**	**1**	**2**

(b)

	A	B	C	D	**T**
$OC_{\$10}$	**3**	**2**	**2**	**2**	**3**

(c)

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

(d)

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

(e)

	A	B	C	D	**T**
$OC_{\$10}$	**3**	**2**	**2**	**2**	**3**
$OC_{\$20}$	N/A	**2**	**1**	**1**	**2**
$OC_{\$30}$	**1**	**1**	**2**	N/A	**2**
$OC_{\$40}$	**1**	N/A	**1**	**2**	**2**

(f)

Figure 5. Demonstration of the de-anonymization process in a single-dimensional scenario. (a) 1 report received; (b) 2 report received; (c) 3 report received; (d) 4 report received; (e) 5 report received; (f) all reports received.

3.2. Multi-Dimensional Scenario

For multi-dimensional PSS scenario, we apply the same rules of anonymization and de-anonymization explained in the previous section for each dimension.

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 application 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 $R=\{(A,X),(A,Y),\cdots,(C,Y),(C,Z)\}$. The observed attributes for each *OOI* combination are shown in **Figure 6(a)**. Without any loss of generality, we assume that the anonymity preference for both dimensions (product and location) is $k_1=k_2=2$ for all users and the observations are shown in **Figure 6(b)** in order of appearance.

3.2.1. Anonymization Process

For each dimension *i*, AnS chooses k_i-1 *OOIs* along with the real *OOI* where k_i is user's preference of anonymity for i^{th} dimension. AnS maintains *d* different *IOCs* for each *OOI* combination $r\in R$. In our example, the first report $\langle A\{2\},X\{2\}\rangle$:\$11 refers to the price of *A* from location *X*. To anonymize these two *OOIs*, *i.e.* product and location, AnS randomly chooses additional *OOIs B* (for product) and *Y* (for location), respectively at the initial step. After producing this AR, *i.e.* $\langle\{A,B\},\{X,Y\}\rangle$:\$11, AnS 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. Here, IOC^1 and IOC^2 refers to the respective dimensions.

Following the strategy as shown in **Figure 6(c)** and **Figure 6(d)**, (A,X) can be de-anonymized after receiving the third AR.

3.2.2. De-Anonymization Process

In multi-dimensional scenario, we use the notation $OC^i_v(p)$ to denote the *OC* of i^{th} dimension for a reported attribute *v* and $OC^i_v(p)$ to denote the *OC* value for *OOI p* in that corresponding *OC*.

	X	Y	Z
A	\$11	\$12	\$13
B	\$21	\$22	\$23
C	\$31	\$32	\$33

(a)

Incoming Order	Report	
	Real	Anonymized
1	$\langle A\{2\}, X\{2\}\rangle : \$11$	$\langle \{A,B\}, \{X,Y\}\rangle : \$11$
2	$\langle A\{2\}, Y\{2\}\rangle : \$12$	$\langle \{A,B\}, \{X,Y\}\rangle : \$12$
3	$\langle A\{2\}, X\{2\}\rangle : \$11$	$\langle \{A,C\}, \{X,Z\}\rangle : \$11$

(b)

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

(c)

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

(d)

Figure 6. Demonstration of the anonymization process in a multi-dimensional scenario. (a) Price of different *OOIs*; (b) List of reports; (c) Second anonymization step; (d) Third anonymization step.

In our example, after receiving the first AR $\langle \{A,B\}, \{X,Y\} \rangle : \$11$, the ApS creates one row for keeping the information of \$11. In one column, ApS keeps the total report count $T_{\$11}$ and two other columns to keep the *OC* values for two dimensions, *i.e.* $OC^1_{\$11}$ and $OC^2_{\$11}$. As it is the first report for 11, $T_{\$11}$ is set to one. From this report, the ApS comes to know about *A*, *B* as the *OOIs* for first dimension and *X*, *Y* for the second dimension. It creates columns for the *OOIs* in respective dimensions of *OC* and increases $OC^1_v(A)$, $OC^1_v(B)$, $OC^2_v(X)$ and $OC^2_v(Y)$ by one. The remaining de-anonymization process continues in the same manner (Figure 7).

4. Optimization Strategies to Reduce NRRFD in MDEAS

MDEAS can boost up its performance by adopting some optimization techniques. In anonymization process, the *IOC* counts refer to the *OOIs* which are ruled out from the possible mappings at the end of ApS. Hence, while choosing *OOIs* for anonymization, we prefer the *OOIs* with highest *IOC* values. However, the *OOIs* which are already de-anonymized, are more preferable candidates for being selected in ARs because they are already ruled out by ApS. We can redefine the Rule 1 and Rule 2 as follows where S_D denotes the set of de-anonymizable *OOIs* in current anonymization process.

Rule 5. (*Rule of being de-anonymizable*) *An OOI p is de-anonymizable if*

$$\forall_{q \in S \wedge q \neq p} \quad IOC_p(q) > 0 \vee q \in S_D \tag{5}$$

The rule of anonymizing an observed report by AnS is as follows:

Rule 6. (*Rule for selection of OOIs in an AR*) *To anonymize a report containing the attribute of OOI p, the set of selected OOIs, S' is formed as:*

$$S' = \begin{cases} p \bigcup \text{any } k-1 \text{ OOIs in } S_D, & \text{if } |S_D| \geq k-1 \\ p \bigcup S_D \bigcup \left(k-1-|S_D|\right) & \text{largest} \\ \text{IOC-valued OOIs in } (S \setminus p) \setminus S_D & |S_D| \leq k-1 \end{cases} \tag{6}$$

Similarly, while de-anonymizing reports, the *OOIs* which are already de-anonymized for other attributes are automatically ruled out from possible candidate

	OC^1		OC^2		T
	A	B	X	Y	
\$11	**1**	**1**	**1**	**1**	**1**
\$12	**1**	**1**	**1**	**1**	**1**

(a)

	OC^1			OC^2			T
	A	B	C	X	Y	Z	
\$11	2	1	1	2	1	1	2
\$12	1	1	N/A	1	1	N/A	1

(b)

Figure 7. Demonstration of the de-anonymization process in a multi-dimensional scenario. (a) 2 reports received; (b) 3 reports received.

lists and their *OC* values are not considered while de-anonymizing. The Rule 4 is modified as follows.

Rule 7. (*Rule of being de-anonymized*) *The attribute v is de-anonymized for OOI p if*

$$OC_v(p) = T_v \wedge \forall_{q \in S \wedge q \neq p} OC_p(q) < T_v \tag{7}$$

Here, we present a simple example referring to our example in Section 1. In our example, B is de-anonymized after receiving fifth report. However, if this optimization is applied, B would be de-anonymized at fourth report. According to Rule 2, AR would choose $\langle A, B \rangle : \$20$ instead of $\langle B, D \rangle : \$20$. Here, A is chosen instead of D because A is already de-anonymized. In this case, the IOC_B will look like the following:

	A	B	C	D
IOC_B			**1**	**1**

In this state, B will be de-anonymized according to the Rule 5 as $IOC_B(C), IOC_B(D) > 0$ and A is already de-anonymized for other attribute.

5. Theoretical Analysis on NRRFD for MDEAS

In this section, we present a theoretical analysis on NRRFD as explained earlier. As the order of appearance of reports (ARs) is probabilistic, we derive the expected NRRFD using probability theory. Consider a d-dimensional PSS scenario where the i-th dimension has N_i *OOIs* that are reported with k_i anonymity for all $1 \le i \le d$. Overall, there are $X = \prod_{i=1}^{d} N_i$ distinct *OOI* combinations that need to be reported with as many unique attributes. Our de-anonymization scheme carries out "attribute-centric" independent de-anonymization process for each of the X unique attributes.

For each AR received by ApS for a particular attribute, it eliminates $N_i - k_i$ *OOIs* from the potential list of *OOIs* in the i-th dimension on the basis that their *OC* values are less than the number of ARs received so far. As the anonymization process selects unobserved *OOIs* in the order of their *IOC*, the de-anonymization process of ApS is able to continually eliminate $N_i - k_i$ *OOIs* from the potential list of observed *OOIs* in the i-th dimension for each received report. To isolate the actual attribute, ApS needs to eliminate $N_i - 1$ other *OOIs*. Therefore, the de-anonymization process in the i-th dimension requires $\left\lceil \frac{N_i - 1}{N_i - k_i} \right\rceil$ reports to

isolate observed *OOI* in i^{th} dimension from the candidate list altogether. As the de-anonymization process in each dimension is independent of other dimensions and they are carried out in parallel, the de-anonymization process of a particular attribute is completed by isolating the *OOIs* in all d dimensions after receiving $Y = \max_{1 \le i \le d} \left\lceil \frac{N_i - 1}{N_i - k_i} \right\rceil$ reports.

Ideally, the anonymization framework needs no more than $n_{ideal} = XY$ reports to de-anonymize the attributes of all X *OOI* combinations. This lower-bound, however, can only be met if and only if each unique attribute is observed exactly Y times, which is an unrealistic assumption. The probability of a particular attribute being observed is $p = \frac{1}{X}$. After n observations, the number of times each unique attribute v is reported, n_v, can be assumed normally distributed with mean, $\mu = np = \frac{n}{X}$ and variance, $\sigma^2 = np(1-p) = \frac{n}{X}\left(1 - \frac{1}{X}\right) = n\frac{X-1}{X^2}$ according to the Central Limit Theorem, *i.e.* $n_v \sim N\left(\mu, \sigma^2\right)$. The minimum of X number of n_v's follows the Gumbel distribution ([10], §10.5), one of the Generalized Extreme Value (GEV) distributions, with mean $\mu - z\sigma = \frac{n}{X} - \sqrt{n}\frac{z\sqrt{X-1}}{X}$ where

$$z = \Phi^{-1}\left(1 - \frac{1}{X}\right) + \gamma\left(\Phi^{-1}\left(1 - \frac{1}{X_e}\right) - \Phi^{-1}\left(1 - \frac{1}{X}\right)\right) \tag{8}$$

and Φ^{-1} is the inverse CDF (cumulative distribution function) of the standard normal distribution $N(0,1)$, and $\gamma = 0.5772$ is the Euler-Mascheroni constant [11].

We may now find the expected NRRFD, $\bar{n}$ needed to de-anonymize the values of all X *OOI* combinations by finding the root of the following quadratic equations.

$$\frac{\bar{n}}{X} - \sqrt{\bar{n}}\frac{z\sqrt{X-1}}{X} = Y \tag{9}$$

Simplifying the equation above, we find

$$\sqrt{n} = \frac{1}{2}\left(z\sqrt{X-1} + \sqrt{z^2(X-1) + 4XY}\right) \tag{10}$$

$$\bar{n} = \frac{1}{4}\left(2z^2(X-1) + 4XY + 2\sqrt{z^2(X-1)\left(z^2(X-1) + 4XY\right)}\right) \tag{11}$$

Finally, by simplifying more the above equation, we get

$$\cong n_{ideal} = \left(1 + \frac{z^2}{2Y} + \frac{z}{\sqrt{Y}}\right) \tag{12}$$

Figure 8 shows the NRRFD by varying N_2 with fixed N_1 (e.g. 5, 10, 20, 60). When N_2 is smaller than N_1, the n_{ideal} increases linearly with X. And when N_2

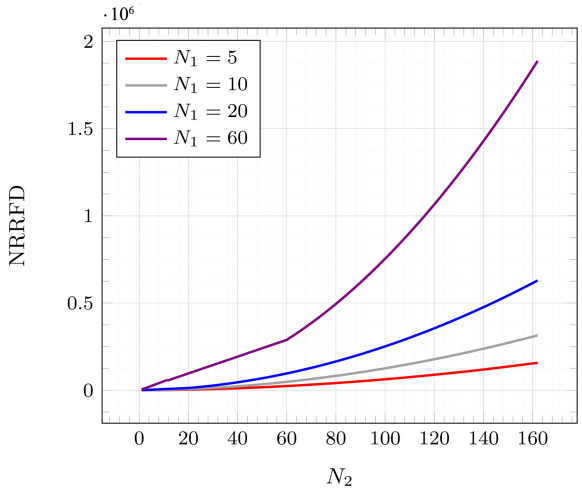


Figure 8. NRRFD on different N and k in multi-dimensional PSS scenario.

is larger than N_1, n_{ideal} increases quadratically with both X and Y. This graph also depicts that the configurations with similar n_{ideal} values require similar NRRFD.

The value of NRRFD obtained from this mathematical analysis conforms to the results obtained from our simulaitons. We present both theoretical and simulation results in Section 7.2.

6. Algorithms in MDEAS

In this section, we present the algorithms to be used for anonymizing observations and de-anonymizing them at ApS. These can be applied in any-dimensional PSS scenario.

6.1. Anonymization Algorithm

Algorithm 1 is used by the AnS to anonymize a d-dimensional observation. It takes a set of user preferences $(k_1, k_2, k_3, \cdots, k_d)$ for d dimensions and the corresponding *OOI* combination $r = (p_1, p_2, \cdots, p_d)$ as input. To remind the readers, here $p_1, p_2, \cdots, p_d$ are *OOIs* of different dimensions such as location, product etc. The Algorithm uses corresponding *IOC*, *i.e.* IOC_r to anonymize this report. For each dimension i, $(k_i - 1)$ extra *OOIs* are selected from the set of *OOIs* in that dimension, *i.e.* S_i by preferring the *OOIs* with highest *IOC* value and the de-anonymized ones following Rule 2. These selected *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* $q \mid q \in S_i \wedge q \notin S_i'$. Thus, the returning set is formulated, *i.e.* $S' = \{S_1', S_2', \cdots, S_d'\}$ where $S_{i'}$ denotes the anonymized *OOI* set for i^{th} dimension.

6.2. De-Anonymization Algorithm

Algorithm 2 is used by the ApS for de-anonymizing the ARs. Here, input S_i' and v denote the set of anonymized *OOIs* in the i^{th} dimension and the reported

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
1: Set $S' = \emptyset$
2: **for each** dimension $i \in \{1, 2, \cdots, d\}$ **do**
3: $S'_i = \{p_i\}$
4: Add $(k_i - 1)$ OOIs from $(S_i - \{p_i\})$ to S'_i by preferring the already de-anonymized ones and (if needed) others in decreasing order of IOC_v.
5: $S' = S' \cup S'_i$
6: Update IOC_v by incrementing each $IOC_v(q) | q \notin S'_i$ ▷ *IOC_v refers to updated one from previous anonymization processes*
7: **end for**
8: **return** $\{S', v\}$

Algorithm 1. $\{S', v\}$: Anonymize $(\{p_1, \cdots, p_d\}, v, \{k_1, \cdots, k_d\})$.

Input:
$S'_1, S'_2, \cdots, S'_d$: Set of OOIs in d dimensions of AR
v: Reported attribute
Output:
P: De-anonymized OOI combination for attribute v.
1: Set $P = \emptyset$
2: **if** v is reported for the first time **then**
3: Create OC_v and T_v ▷ *otherwise, OC and T for v from previous de-anonymization processes will be considered.*
4: **end if**
5: $T_v = T_v + 1$
6: **for all** dimensions $i \in \{1, 2, \cdots, d\}$ **do**
7: Update $OC^i_v(q)$ for all $q \in S'_i$
8: **end for**
9: **for all** dimension $i \in \{1, 2, \cdots, d\}$ **do**
10: **if** $OC^i_v(p_i) = T_v$ for a unique $p_i \in OC^i_v$ **then**
11: $P = P \cup \{p_i\}$
12: **end if**
13: **end for**
14: **if** $|P| = d$ **then**
15: **return** P
16: **end if**
17: **return** NULL

Algorithm 2. P: De-anonymize $(S_{1'}, S_{2'}, \cdots, S_{d'}, v)$.

attribute, respectively. For each dimension *i*, the Algorithm increases the *OC* value for all *OOIs q*, *i.e.* $OC^i_v(q) \mid q \in S'_i$. To check whether a reported attribute, *v* has been de-anonymized, ApS checks Rule 1 for each dimension *i*. If all dimensions' observed *OOIs* are de-anonymized following Rule 1, then the actual *OOI* combination for *v* is known by the ApS.

Note that, as discussed in Section 3, the anonymization **Algorithm 1** can be optimized for single dimension to achieve faster de-anonymization. In order to do this, we need to keep track of the already de-anonymized *OOIs* and prioritize those to add in anonymized set in Line 4 of **Algorithm 1**. However, this optimization is not applicable in multi-dimensional scenario as the attribute of *OOI* depends on multiple dimensions and the anonymization is done separately for each dimension.

7. Results and Discussions

To establish the applicability and assess the performance of our proposed schemes, we have experimented with both comprehensive simulation and android-based

real world prototype. Adversary capabilities are modeled considering realistic approach to encourage the replication; we have shared our implementations of both simulation and android-based prototype[1].

7.1. Simulation

7.1.1. Simulation Setup

We have conducted a simulation of our proposed schemes using custom simulator. User Observations are generated randomly with uniform distribution. By varying the number of *OOIs* and the anonymity preference of users, we have analyzed the performance of the algorithms for both single and multi-dimensional scenarios. As we are mostly interested in evaluating the performance of proposed schemes in terms of data quality, we investigated how many observations are required to achieve different extent of de-anonymization. We use a term called "De-anonymization rate" to present our results. De-anonymization rate of T observations is defined as the proportion of *OOIs* de-anonymized among the N *OOIs*. We shall also analyze the impact of anonymity preference on de-anonymization rate. All the results presented here are obtained by averaging 1000 simulation runs.

7.1.2. Results for Single-Dimensional PSS

In this section, we have presented the results of simulation for single-dimensional PSS scenario by applying the simple optimization explained in Section 4. As alluded earlier, our proposed scheme is scalable in the number of *OOIs*. Hence, we could experiment with PSS scenarios with a reasonably large number of *OOIs*. Figure 9 shows the de-anonymization rate for fixed $N=15$ by varying anonymity preference k from 8 to 14. Naturally, high anonymity preference requires more observations to de-anonymize all *OOIs*. For example, around 100 reports are needed to de-anonymize all *OOIs* for $k=8$ while little

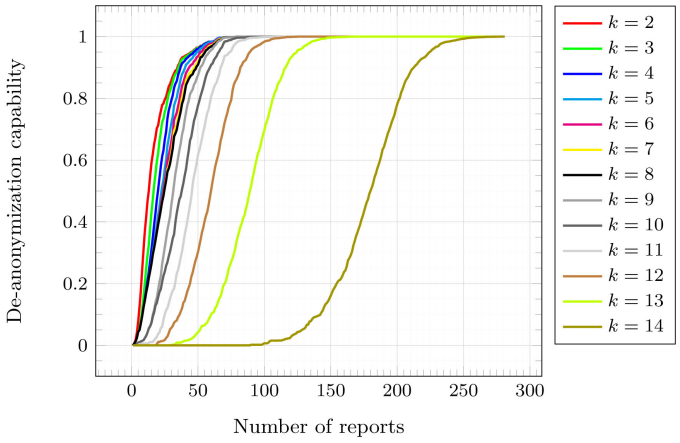


Figure 9. De-anonymization rate of MDEAS in single-dimensional PSS scenario for $N=15$ and varying k.

[1]Simulator: https://bitbucket.org/shaolinkhusbu/pss-simulator android prototype: https://bitbucket.org/nafeezabrar/pss-server-front-end.

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 in the *OOIs*, a feasible k should be selected.

7.1.3. Comparison with Existing Subset-Coding Techniques

As we have discussed in Section 1 that we are interested in scenarios where individual data need to be retrieved at the destination, existing techniques such as spatial cloaking, obfuscation are not applicable in our context. Hence, k-Anonymization Techniques, e.g. PGAS [8], DGAS/FDGAS [9] are the only comparable techniques with our schemes. Although we have outperformed them completely in terms of computational overhead, we may compare in terms of their required number of observations to achieve certain data quality. Since the requirement of observations to achieve full de-anonymization is almost similar in all 3 cases, we consider it sufficient to compare only DGAS with our approach. We have simulated both approaches with the same set of observations with different N and k to maintain fairness. Since the computational overhead of DGAS is much higher than MDEAS, we have simulated it with smaller N. From **Figure 10**, we see that we require a higher NRRFD compared to DGAS but almost similar NRRFD in optimized MDEAS explained in Section 4.

7.1.4. 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. We achieved this without sacrificing recoverability of data. **Figure 11** depicts the result for two-dimensional anonymization which anonymizes both location and product with different anonymity preference. We see that for quite a large number of *OOIs* in both dimensions, *i.e.* $N_1 \in 8,10,15$ and $N_2 \in 4,6,8$, the required number of observations are in the range of 500 -

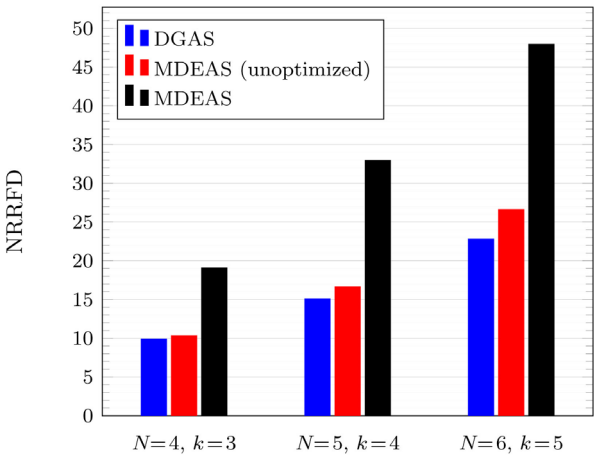


Figure 10. Comparison of de-anonymization rate between DGAS and MDEAS (both unoptimized and optimized) for $N=4$ to 6 and highest anonymity preference, $k=N-1$ in single-dimensional PSS scenario.

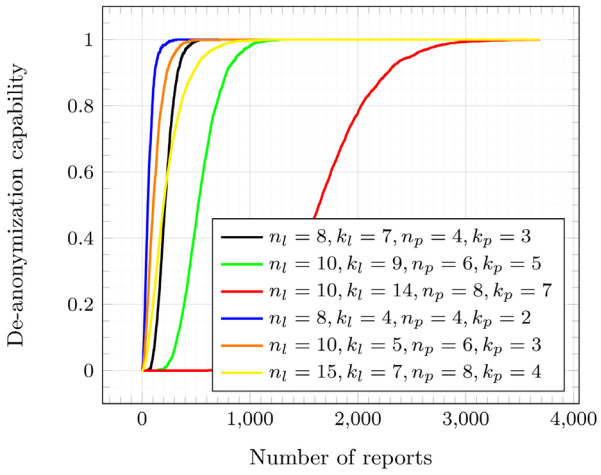


Figure 11. Impact of location anonymity k_l on the de-anonymization rate of MDEAS in multi-dimensional PSS where, $N_l = 8$, $N_p = 4$ and $k_p = 3$.

3000 to de-anonymize all *OOIs* with highest anonymity $k_i = N_i - 1$. However, when the anonymity preference is reduced to half, *i.e.* $k_i = N_i/2$, the required number of observations also declines significantly. For example, for $N_l = 8$ and $N_p = 4$, if k_l and k_p is reduced from $k_i = N_i - 1$ to $k_i = N_i/2$, the required number of observations decreases by 49% which is approximately half compared to the highest anonymity. Hence, in the case of a very large number of *OOIs* in multi-dimensional scenraio, PSS can vary the anonymity preference in different dimensions in order to achieve good de-anonymization with a finite number of observations.

7.2. Comparison with Theoretical Result

Figure 12 depicts the result for two-dimensional anonymization which anonymizes both location and product *i.e.* for $N_l \in \{13,14\}$ and $N_p \in \{6,7\}$. Our simulation result matches with the theoretical results found explained in Section 5. The red dot shown in this figure denotes the theoretical mean while the diamond-shaped black dot denotes the mean achieved by 1000 simulation run. The mean NRFFD achieved from our simulation differs by only 4% from the theoretical mean which is inacceptable range and validates each other Moreover, an interesting pattern is observed in this figure. The total reports required for full de-anonymization is 2864, 2812 and 2679 for $N_l = 14, N_p = 8$, $N_l = 15, N_p = 7$ and $N_l = 16, N_p = 6$ which are very close. Similarly the reports required are similar for $N_l = 13, N_p = 8$, $N_l = 14, N_p = 7$ and $N_l = 15, N_p = 6$ also. This similar pattern is observed in both our simulation and theoretical result. Both results depicts that the NRRFD depends mostly on n_{ideal}. Therefore, similar configurations with closer $n_{ideal} = XY$ requires similar NRRFD.

Results for Variation in User Preference

In the real world, individual's privacy concern varies with many parameters such

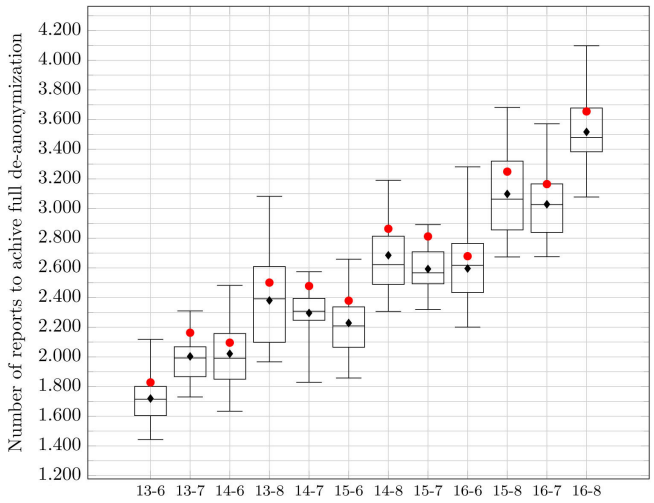


Figure 12. Theoretical and experimental results different N_l and N_p.

as the 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 helps to gain better de-anonymization. From this consideration, we would like to evaluate the response of our proposed scheme against variable anonymity preference. Without loss of generality, we show result for three different configurations in **Figure 13**. First, we consider a fixed $k=12$ for $N=15$. Then, we compare it with $k=10$ for half the observations and $k=14$ for the other half. Finally, we like to distribute user preferences in three equal proportions for $k=10$, 12, and 14, respectively. We find that there is not a significant change in de-anonymization rate for these variable anonymity preferences. Thus our algorithms offer a flexibility to satisfy users with diverse anonymity preference without compromising de-anonymization performance.

7.3. Privacy Risk Analysis with Adversary

Adversary residing near ApS can eavesdrop the ARs sent by the participants and thus reveal actual attribute of *OOIs* and find the users' association with the *OOI*. We have discussed the adversary model in details in Section 2.2 and a grouping strategy is proposed to mitigate the adversary risk. We also discussed some additional adversary capabilities. In our simulation, we have evaluated MDEAS's performance under the presence of adversaries with additional capabilities where grouping strategy is applied. We use τ and T_A to define the colluding group size and adversary type, respectively.

7.3.1. Simulation Setup

We simulated all 3 types of adversaries as defined in **Table 1** (Section 2.2) who collude among themselves and try to de-anonymize by sharing observed *OOIs* among the colluding team members. We have also analyzed the de-anonymization capability of adversaries by varying their pattern of intercepting ARs, T_A and

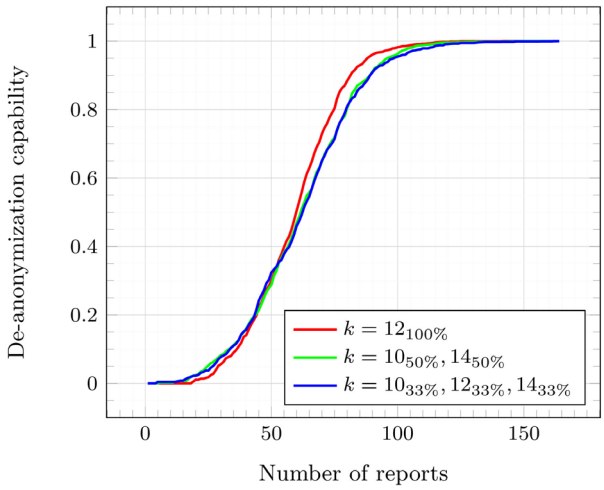


Figure 13. Impact of variation of anonymity preference, k on the de-anonymization rate of MDEAS in single-dimensional PSS where $N = 15$.

collusion *i.e.* the size of their colluding team, τ. In this simulation, we have assumed an arbitrary PSS scenario with two dimensions (location and product) where $N_l = 18$ and $N_p = 12$ and $3 \le k_l, k_p \le 5$. We have applied grouping as discussed in Section 2.2 and varied the total number of groups, G by 4, 6 and 9. The generation of observed reports and grouping of *OOIs* are done randomly with uniform distribution. Adversaries are also selected from the registered users randomly (uniform distribution) according to the colluding team size. All the simulation results are prepared by taking the average of 100 runs. For the sake of completeness, we also simulated the behavior of adversary that might use prediction as discussed in 1 for de-anonymization.

7.3.2. Results of the De-Anonymization Capability of Adversary

We have compared the de-anonymization capability of different types of adversaries in **Figure 14**. In this experiment, only the AR interception pattern differs while the colluding team size, τ and the number of groups of PSS, G are fixed. Here, it is clear that the de-anonymization capability of all three types of adversaries does not differ much. As grouping is applied in PSS, no matter how many ARs they intercept, they only know the *OOI*-attribute mapping of their own observed reports which is only 1.4% of total attributes of PSS.

The colluding adversaries incur more privacy risk on PSS as more collusion means more shared information and revelation of ARs. **Figure 15** depicts the impact of colluding team size, τ on the de-anonymization capability of adversaries by keeping N, k, G, and T_A fixed. It is clear from this result that the increase of colluding team size increases the privacy risk of users. Still, a practically reasonable size of colluding groups (e.g. 5, 10, 15 when the total number of users is 500) can only reveal 11%, 21% and 30% attributes of PSS respectively that are actually revealed from the observed reports of adversaries themselves.

Figure 16 shows the de-anonymization rate of adversary by varying the total number of groups, G in PSS while other parameters are fixed. This analysis clearly

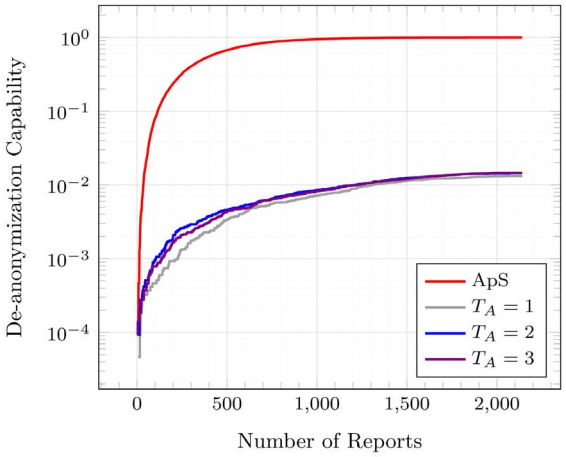


Figure 14. De-anonymization rate of different adversary types, T_A in multi-dimensional PSS with $N_l = 18, N_p = 12$ where adversary colluding team size, $\tau = 1$, the total number of groups, $G = 6$ and anonymity, $k_l = 3 = k_p = 3$.

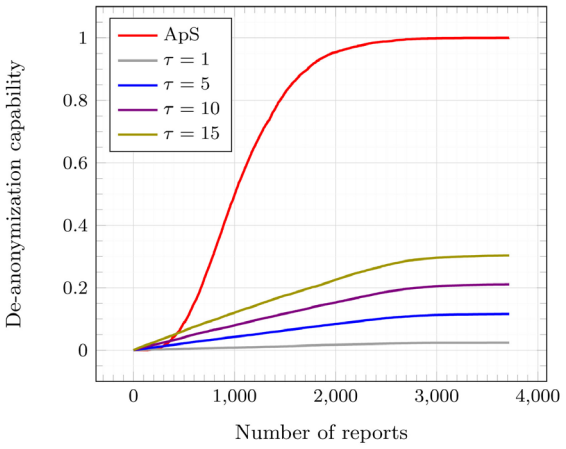


Figure 15. De-anonymization rate of the adversary for different colluding team size, τ where $T_A = 2, G = 4, k_l = k_p = 5$.

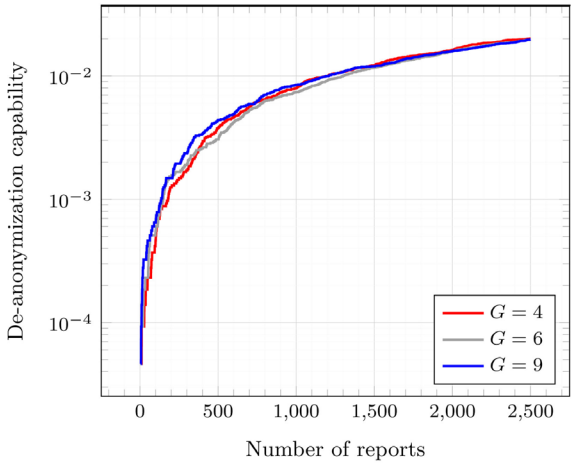


Figure 16. De-anonymization rate of the adversary for multi-dimensional PSS by varying the total number of groups, G where $T_A = 2, \tau = 1, k_l = k_p = 3$.

depicts that G does not have much impact on the de-anonymization capability of the adversary.

From Figures 14-16, we may conclude that the interception of ARs by adversary does not help them in their de-anonymization. However, more collusion means more revelation of attributes and consequently the risk increases. Still, this privacy risk is not significant as such large portion of a colluding group of users practically does not exist.

As discussed in Section 2.2, the adversary might make some prediction on their set of most probable de-anonymized *OOIs*. As an example, we explained how the adversary can predict location *OOIs* using distance estimation. We have simulated such adversary with location prediction capability and shown the result in Figure 17. Here, we have shown the de-anonymization capability of different possible types of adversaries with respect to colluding team size τ. Even though the adversary has assumed distance based prediction, adversaries with colluding team size 10 can only de-anonymize about 50% *OOIs* which is equivalent to a random prediction.

7.4. Android Prototype Based Experiment

We have developed an Android-based software prototype 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 the actual report to the AS, receive ARs from AS and forward this with user id to the ApS. Figure 18(a) and Figure 18(b) show the user interface for sending a report to AnS and ApS respectively. Figure 18(c) and Figure 18(d) show the servers' responses.

With the help of this application using Android Smart-phones (connected to the Internet and equipped with GPS) and two separate servers dedicated as AnS and ApS built with Python Tornado Web Framework, we test our anonymization and de-anonymization algorithms. Here the user's current location is obtained from device's GPS and other information like the product and its actual

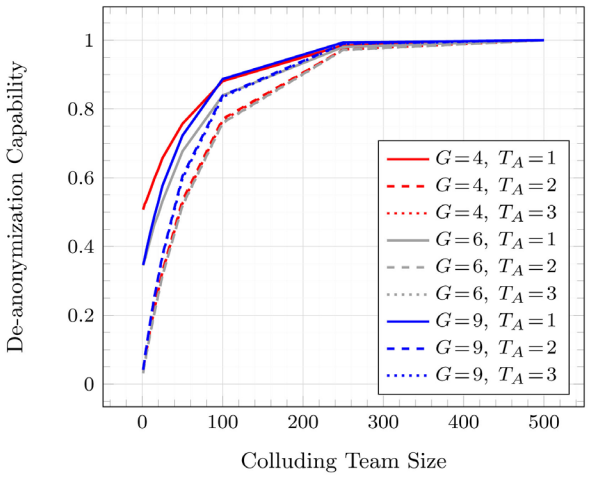
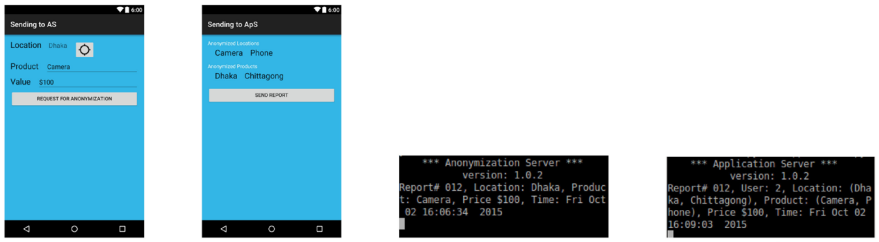


Figure 17. De-anonymization rate of adversary with location-estimation capability by varying colluding team size, τ and number of groups, G where $T_A = 2, G = 4, k_l = k_p = 5$.



(a) (b) (c) (d)

Figure 18. Android prototype of sending a report to AnS and ApS and the corresponding server response on receiving the reports. (a) User sends Report to AnS; (b) User sends report to ApS; (c) AS receives user's report without identity; (d) ApS receives AR with user identity.

attribute is taken as user input. After receiving the anonymized report from AnS, a user can directly send the report to ApS shown in **Figure 18(b)**. This user's report is received by ApS shown in **Figure 18(d)**. ApS can de-anonymize all the reported products successfully from the received ARs. We have simulated our application for 2-dimensional PSS scenario where $N_l = 3$, $N_p = 3$ and $k_l = k_p = 2$. **Figure 19** shows our experimental result where 90% de-anonymization has been achieved after receiving 59 reports on average were required to achieve full de-anonymization. We have calculated this de-anonymization rate by averaging the results by running the application for 10 times.

8. Related Works

Assurance of privacy in accordance with users' contribution is the key factor for maintaining adequate participants in PSS system [12] and hence numerous research works using a few techniques have been conducted to protect the privacy of the users. These approaches are briefly discussed below.

8.1. Mix Network

Hot-Potato-Privacy-Protection (HP^3) [6] scheme is designed based on mix network concept where a user sends a report to one of his/her friend and that friend chooses another friend to deliver the report to next hop. This process continues until a threshold is reached and then the last recipient sends the report to ApS. LAP [13] reduces the latency and minimizes computational overhead of this mix-network scheme. Wang and Ku proposed another variant of mix-network approach in [14] where only the connection request is transferred through the peers and thus it consumes less bandwidth and computational power. Recently, using mix network a collaborative data exchange method is proposed in [15] where participants exchange data before submission and submit mixed data for privacy protection. Another decentralized peer-to-peer exchange platform named Privacy Aware Incentivization (PAI) has been proposed [16] to provide anonymous, untraceable and secured data submission alongwith adaptive, adjustable and incentive-compatible reward computation. However, all these mix-network based schemes suffer from delays due to slow network connection

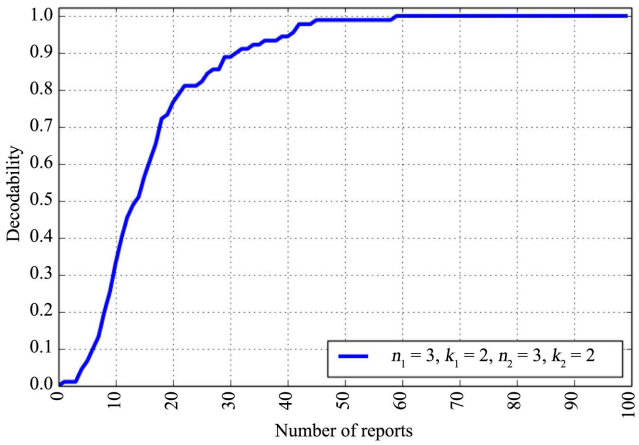


Figure 19. De-anonymization rate of ApS achieved in the android experiment.

in volunteer peers. Ensuring the trustworthiness of the peers is a big challenge of these schemes.

8.2. Pseudonym

The pseudonym-based approaches are also common for protecting identity privacy from ApS. But long-term pseudonym tends to be identified easily by adversary. Mix-zone concept is proposed in [17] where users register in a connected spatial region so that adversary cannot distinguish a user coming out from a mix-zone. Mobimix [18], TrPf [19] are proposed using this mix-zone concept. However, these approaches may suffer from low level of anonymity in high spatial-temporal resolution.

8.3. Encryption

Encryption is one of the most common approaches for protecting privacy in PSS. E. De Cristofaro *et al.* presented an approach where server gets encrypted data and blindly performs computation on the encrypted data. LotS [20] maintains *k*-anonymity with the use of cryptographic techniques and combines voting approaches to support users' reputation. To report large multimedia data, an erasure coding based scheme named SLICER [21] has been proposed where the sensing record is sliced and each slice is transferred via other participants or generator itself using cryptographic encryption scheme. However, these encryption-based approaches are often prohibited by government as illegal data might be transferred.

Multi Secret Sharing and Information Exchange

Multi-secret sharing [22] is a concept where some arbitrarily related secrets are shared among a set of participants who are not trusted individually. This approach is improved in PShare [23] [24]. For protecting trajectory privacy, exchanging report has been proposed in [25]. In this scheme, users exchange his/her collective sensor readings with another user when they physically meet. To identify malicious users in this scheme, TrustMeter [26] has been proposed

which assesses the user contribution as well as trust levels. I. Boutsis and V. Kalogeraki proposed another low-cost information exchange strategy in [27] 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.

8.4. Techniques for Aggregated Data

Different techniques for privacy protection have been proposed in the PSS scenarios where the ApS is only interested in aggregated result. PriSense [28] supports non-additive aggregate functions like average, min/max, histogram etc. Negative surveys are used in [29] to facilitate the complemented sensory data as an input and get aggregated result of the actual ones without revealing actual individual data.

8.5. Spatial Cloaking

Obfuscation is first introduced in [30] as a new technique to safeguard location privacy which degrades quality of service. *k*-anonymity based location-privacy schemes have been proposed in [31] [32] where $k-1$ 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 [33] where participants cloak their location data without disclosing their exact location to third party or other participants. However, these approaches incorporate delay in real time operation and not suitable where fine-grained information is required.

8.6. Differential Privacy

Differential privacy (DP) protection is a new paradigm based on the notion that some aggregate property of a large data-set remains unchanged even if individual data are tweaked with controlled random noise. Many researchers have utilized this differential privacy technique for providing privacy protection in mobile crowd-sourcing (task assignments), aggregation-based queries and location-based services. For example, DP-MDBScan schema proposed in [34] focuses on clustering analysis of network user data. Another differential Privacy protection approach is proposed in [35] which is applicable to arbitrary aggregate query function by avoiding too much noise. In [36] [37] controlled noise is added to protect workers' from revealing exact locations. Another framework has been proposed in [38], where cellular service providers (CSP) release workers' locations to third party application servers in noisy form applying Differential privacy based technique. However, none of these approaches aim for 100% accuracy on query response or task acceptance ratio. As we aim for achieving 100% de-anonymization, differential privacy cannot be applied in our technique of data collection.

8.7. Combination of Different Techniques

Many privacy-preserving mechanisms [39] have been proposed 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 [40]. ARTSense [41] 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.* [42] proposed a framework to dynamically assess the trustworthiness of information as well as the participants. In [43] 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 receipt are transferred anonymously. Still, this approach can cause credit-based inferable attack. IncogniSense [44] 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.

9. Conclusion

In this paper, we have presented efficient algorithms for anonymization and de-anonymization of user observations in the context of PSS. To the best of our knowledge, this is the first work that presented anonymization technique in multiple dimensions with flexible anonymity preference. Theoretical analysis and simulation results show that our proposed scheme achieves sufficient data recoverability at the target end from a feasible number of user reports. We have also implemented an Android prototype and conducted experiments in real-world. Our proposed approach is likely to contribute to making participatory sensing a popular technology to the community ensuring privacy of participants without compromising the quality of data.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Dong, Y.F., Kanhere, S., Chou, C.T. and Bulusu, N. (2008) Automatic Collection of Fuel Prices from a Network of Mobile Cameras. In: *Distributed Computing in Sensor Systems*, Springer, Berlin, 140-156.

[2] Deng, L. and Cox, L.P. (2009) LiveCompare: Grocery Bargain Hunting through Participatory Sensing. In: *Proc. 10th Workshop on Mobile Computing Systems and Applications*, ACM, New York. https://doi.org/10.1145/1514411.1514415

[3] Ballesteros, J., Rahman, M., Carbunar, B. and Rishe, N. (2012) Safe Cities. A Participatory Sensing Approach. *IEEE 37th Conference Proc. Local Computer Networks*, Clearwater, 22-25 October 2012, 626-634. https://doi.org/10.1109/LCN.2012.6423684

[4] Zheng, Y.Q., Zhou, P.F. and Li, M. (2014) How Long to Wait? Predicting Bus Arrival Time with Mobile Phone Based Participatory Sensing. *IEEE Transactions on Mobile Computing*, **13**, 1228-1241.

[5] Huang, K.L., Kanhere, S.S. and Hu, W. (2010) Preserving Privacy in Participatory Sensing Systems. *Computer Communications*, **33**, 1266-1280. https://doi.org/10.1016/j.comcom.2009.08.012

[6] Hu, L. and Shahabi, C. (2010) Privacy Assurance in Mobile Sensing Networks: Go beyond Trusted Servers. 2010 8*th IEEE International Conference on Pervasive Computing and Communications Workshops*, Mannheim, May 2010, 613-619. https://doi.org/10.1109/PERCOMW.2010.5470509

[7] Ganti, R.K., Pham, N., en Tsai, Y. and Abdelzaher, T.F. (2008) PoolView: Stream Privacy for Grassroots Participatory Sensing. *Proceedings of the* 6*th International Conference on Embedded Networked Sensor Systems*, Raleigh, 5-7 November 2008, 281-294. https://doi.org/10.1145/1460412.1460440

[8] Murshed, M., Iqbal, A., Sabrina, T. and Alam, K.M. (2011) A Subset Coding Based k-Anonymization Technique to Trade-Off Location Privacy and Data Integrity in Participatory Sensing Systems. 2011 *IEEE* 10*th International Symposium on Network Computing and Applications*, Cambridge, 25-27 August 2011, 107-114. https://doi.org/10.1109/NCA.2011.22

[9] Sabrina, M.M.T. and Iqbal, A. (2016) Anonymization Techniques for Preserving Data Quality in Participatory Sensing. 2016 *IEEE* 41*st Conference on Local Computer Networks* (*LCN*), Dubai, 7-10 November 2016, 607-610.

[10] David, H.A. and Nagaraja, H.N. (2004) Order Statistics. Wiley Series in Probability and Statistics. Wiley, Hoboken. https://doi.org/10.1002/0471722162

[11] Euler-Mascheroni Constant. https://mathworld.wolfram.com/Euler-MascheroniConstant.html

[12] Lee, J.-S. and Hoh, B. (2010) Sell Your Experiences: A Market Mechanism Based Incentive for Participatory Sensing. *IEEE International Conference on Pervasive Computing and Communications*, Mannheim, 29 March-2 April 2010, 60-68.

[13] Hsiao, H.-C., Kim, T.H.-J., Perrig, A., Yamada, A., Nelson, S.C., Gruteser, M. and Meng, W. (2012) Lap: Lightweight Anonymity and Privacy. 2012 *IEEE Symposium on Security and Privacy*, San Francisco, CA, May 2012, 506-520. https://doi.org/10.1109/SP.2012.37

[14] Wang, C.-J. and Ku, W.-S. (2012) Anonymous Sensory Data Collection Approach for Mobile Participatory Sensing. *IEEE* 28*th International Conference on Data Engineering Workshops*, Arlington, 1-5 April 2012, 220-227. https://doi.org/10.1109/ICDEW.2012.78

[15] Zhang, T.Q., Zhang, R. and Wang, J. (2020) Privacy Preservation with Unequal Data Exchange Strategy in Participatory Sensing. *Journal of Physics*: *Conference Series*, **1486**, Article ID: 052004. https://doi.org/10.1088/1742-6596/1486/5/052004

[16] Connolly, M., Dusparic, I., Iosifidis, G. and Bouroche, M. (2019) Privacy Aware Incentivization for Participatory Sensing. *Sensors* (*Basel*, *Switzerland*), **19**, 4049. https://doi.org/10.3390/s19184049

[17] Beresford, A.R. and Stajano, F. (2003) Location Privacy in Pervasive Computing. *IEEE Pervasive Computing*, **2**, 46-55. https://doi.org/10.1109/MPRV.2003.1186725

[18] Palanisamy, B. and Liu, L. (2011) MobiMix: Protecting Location Privacy with Mix-Zones over Road Networks. 27*th International Conference on Data Engineering*, Hannover, 11-16 April 2011, 494-505.

https://doi.org/10.1109/ICDE.2011.5767898

[19] Gao, S., Ma, J.F., Shi, W.S., Zhan, G.X. and Sun, C. (2013) Trpf: A Trajectory Privacy Preserving Framework for Participatory Sensing. *Information Forensics and Security*, *IEEE Transactions*, **8**, 874-887. https://doi.org/10.1109/TIFS.2013.2252618

[20] Michalas, A. and Komninos, N. (2014) The Lord of the Sense: A Privacy Preserving Reputation System for Participatory Sensing Applications. *IEEE Symposium on Computers and Communications* (*ISCC*), Funchal, 23-26 June 2014, 1-6. https://doi.org/10.1109/ISCC.2014.6912480

[21] Qiu, F.D., Wu, F. and Chen, G.H. (2013) SLICER: A Slicing-Based k-Anonymous Privacy Preserving Scheme for Participatory Sensing. *IEEE* 10*th International Conference on Mobile Ad-Hoc and Sensor Systems*, Hangzhou, 14-16 October 2013, 113-121. https://doi.org/10.1109/MASS.2013.33

[22] Blundo, C., Santis, A., Crescenzo, G., Gaggia, A.G. and Vaccaro, U. (1994) Advances in Cryptology. *CRYPTO* '94: 14*th Annual International Cryptology Conference*, Santa Barbara, 21-25 August 1994. https://doi.org/10.1007/3-540-48658-5_17

[23] Wernke, M., Drr, F. and Rothermel, K. (2012) Pshare: Position Sharing for Location Privacy Based on Multisecret Sharing. 2012 *IEEE International Conference on Pervasive Computing and Communications* (*PerCom*), Lugano, 19-23 March 2012, 153-161. https://doi.org/10.1109/PerCom.2012.6199862

[24] Skvortsov, P. (2015) Position Sharing for Location Privacy in Non-Trusted Systems. PhD Thesis, Universitat Stuttgart, Stuttgart.

[25] Christin, D., Guillemet, J., Reinhardt, A., Hollick, M. and Kanhere, S.S. (2011) Privacy-Preserving Collaborative Path Hiding for Participatory Sensing Applications. 2011 *IEEE* 8*th International Conference on Mobile Adhoc and Sensor Systems*, Valencia, 17-21 October 2011, 341-350. https://doi.org/10.1109/MASS.2011.41

[26] Christin, D., Rodriguez Pons-Sorolla, D., Hollick, M. and Kanhere, S.S. (2014) Trustmeter: A Trust Assessment Scheme for Collaborative Privacy Mechanisms in Participatory Sensing Applications. 2014 *IEEE Ninth International Conference on Intelligent Sensors*, *Sensor Networks and Information Processing*, Singapore, 21-24 April 2014, 1-6. https://doi.org/10.1109/ISSNIP.2014.6827614

[27] Boutsis, I. and Kalogeraki, V. (2013) Privacy Preservation for Participatory Sensing Data. 2013 *IEEE International Conference on Pervasive Computing and Communications* (*PerCom*), San Diego, 18-22 March 2013, 103-113. https://doi.org/10.1109/PerCom.2013.6526720

[28] Shi, J., Zhang, R., Liu, Y. and Zhang, Y. (2010) Prisense: Privacy-Preserving Data Aggregation in People-Centric Urban Sensing Systems. 2010 *Proceedings IEEE INFOCOM*, San Diego, CA, March 2010, 1-9. https://doi.org/10.1109/INFCOM.2010.5462147

[29] Groat, M.M., Edwards, B., Horey, J., He, W.B. and Forrest, S. (2012) Enhancing Privacy in Participatory Sensing Applications with Multidimensional Data. 2012 *IEEE International Conference on Pervasive Computing and Communications*, Lugano, 19-23 March 2012, 144-152. https://doi.org/10.1109/PerCom.2012.6199861

[30] Duckham, M. and Kulik, L. (2005) Simulation of Obfuscation and Negotiation for Location Privacy. In: *Spatial Information Theory*, Springer, Berlin, 31-48. https://doi.org/10.1007/11556114_3

[31] Gedik, B. and Liu, L. (2005) Location Privacy in Mobile Systems: A Personalized Anonymization Model. 25*th IEEE International Conference on Distributed Computing Systems*, Columbus, 6-10 June 2005, 1-18.

[32] Vu, K., Zheng, R. and Gao, L. (2012) Efficient Algorithms for k-Anonymous Loca-

tion Privacy in Participatory Sensing. *Proceedings IEEE INFOCOM*, Orlando, 25-30 March 2012, 2399-2407. https://doi.org/10.1109/INFCOM.2012.6195629

[33] Christin, D., Bub, D.M., Moerov, A. and Kasem-Madani, S. (2015) A Distributed Privacy-Preserving Mechanism for Mobile Urban Sensing Applications. *IEEE Tenth International Conference on Intelligent Sensors*, *Sensor Networks and Information Processing* (*ISSNIP*), Singapore, 7-9 April 2015, 1-6. https://doi.org/10.1109/ISSNIP.2015.7106932

[34] Ni, L., Li, C., Wang, X., Jiang, H. and Yu, J. (2018) DP-MCDBSCAN: Differential Privacy Preserving Multi-Core DBSCAN Clustering for Network User Data. *IEEE Access*, **6**, 21053-21063. https://doi.org/10.1109/ACCESS.2018.2824798

[35] Chen, J., Ma, H., Zhao, D. and Liu, L. (2017) Correlated Differential Privacy Protection for Mobile Crowdsensing. *IEEE Transactions on Big Data*, 1. https://doi.org/10.1109/TBDATA.2017.2777862

[36] Xiong, P., Zhang, L.F. and Zhu, T.Q. (2017) Reward-Based Spatial Crowdsourcing with Differential Privacy Preservation. *Enterprise Information Systems*, **11**, 1500-1517. https://doi.org/10.1080/17517575.2016.1253874

[37] Andrés, M.E., Bordenabe, N.E., Chatzikokolakis, K. and Palamidessi, C. (2013) Geo-Indistinguishability: Differential Privacy for Location-Based Systems. *Proceedings of the* 2013 *ACM SIGSAC Conference on Computer & Communications Security*, November 2013, 901-914. https://doi.org/10.1145/2508859.2516735

[38] To, H., Ghinita, G. and Shahabi, C. (2014) A Framework for Protecting Worker Location Privacy in Spatial Crowdsourcing. *Proceedings of the VLDB Endowment*, **7**, 919. https://doi.org/10.14778/2732951.2732966

[39] Vergara-Laurens, I.J., Mendez, D. and Labrador, M.A. (2014) Privacy, Quality of Information, and Energy Consumption in Participatory Sensing Systems. 2014 *IEEE International Conference on Pervasive Computing and Communications* (*PerCom*), Budapest, 24-28 March 2014, 199-207. https://doi.org/10.1109/PerCom.2014.6813961

[40] He, D.J., Chan, S. and Guizani, M. (2015) User Privacy and Data Trustworthiness in Mobile Crowd Sensing. *IEEE Wireless Communications*, **22**, 28-34.

[41] Wang, X.L., Cheng, W., Mohapatra, P. and Abdelzaher, T. (2014) Enabling Reputation and Trust in Privacy-Preserving Mobile Sensing. *IEEE Transactions on Mobile Computing*, **13**, 2777-2790. https://doi.org/10.1109/TMC.2013.150

[42] Wang, X.L., Govindan, K. and Mohapatra, P. (2011) Collusion-Resilient Quality of Information Evaluation Based on Information Provenance. *Proceedings of the* 8*th Annual IEEE Communications Society Conference on Sensor*, *Mesh and Ad Hoc Communications and Networks*, Salt Lake City, 27-30 June 2011, 395-403. https://doi.org/10.1109/SAHCN.2011.5984923

[43] Li, Q.H. and Cao, G.H. (2015) Privacy-Preserving Participatory Sensing. *IEEE Communications Magazine*, **53**, 68-74.

[44] Christin, D., Rosskopf, C., Hollick, M., Martucci, L.A. and Kanhere, S.S. (2012) IncogniSense: An Anonymitypreserving Reputation Framework for Participatory Sensing Applications. 2012 *IEEE International Conference on Pervasive Computing and Communications*, Lugano, 19-23 March 2012, 135-143. https://doi.org/10.1109/PerCom.2012.6199860