Identifying Anomalies while Preserving Privacy

Hafiz Asif, Jaideep Vaidya, *Fellow, IEEE*, and Periklis A. Papakonstantinou

Abstract—Identifying anomalies in data is vital in many domains, including medicine, finance, and national security. However, privacy concerns pose a significant roadblock to carrying out such an analysis. Since existing privacy definitions do not allow good accuracy when doing outlier analysis, the notion of sensitive privacy has been recently proposed to deal with this problem. Sensitive privacy makes it possible to analyze data for anomalies with practically meaningful accuracy while providing a strong guarantee similar to differential privacy, which is the prevalent privacy standard today. In this work, we relate sensitive privacy to other important notions of data privacy so that one can port the technical developments and private mechanism constructions from these related concepts to sensitive privacy. Sensitive privacy critically depends on the underlying anomaly model. We develop a novel n-step lookahead mechanism to efficiently answer arbitrary outlier queries, which provably guarantees sensitive privacy if we restrict our attention to common a class of anomaly models. We also provide general constructions to give sensitively private mechanisms for identifying anomalies and show the conditions under which the constructions would be optimal.

Index Terms—sensitive privacy, differential privacy, outlier detection, anomaly identification, outlier analysis

1 Introduction

Outlier analysis is a fundamental data analysis task and is extremely useful in practice. It is used to discover complex non-conforming patterns in the data and can generate actionable insights. The ability to identify outliers, i.e. anomalies, is an essential prerequisite to numerous applications in various domains [1], [2], [3], [4], [5]. For example, to treat cancer, we must tell if a tumor is malignant; to counter email scams, we must filter spam; and to stop bank fraud, we must flag the suspicious transactions — and the *de facto* approach to solve these problems is outlier analysis. However, as it is common for data analyses, outlier analysis is a double-edged sword. While it helps us to solve challenging problems [6], [7], [8], it also poses a risk to our privacy.

The fundamental question that we face today for outlier analysis is: *Can we lift privacy restrictions, i.e. accurately analyze data looking for outliers without hurting the privacy of those who contribute their data?* In this work, we answer this question for the most practically relevant case, when outliers are defined in a data-dependent way (i.e. a record is anomalous only if it is ”different” from other collected data). “Data-dependent” means a query for anomalous records in a database, models anomalies by comparing each database record with nearby records, aiming to infer significant dissimilarities. We should note that if one defines an anomalous record independent to other records then conceptually much simpler solutions can be given – in particular, those that go through the typical differential privacy literature.

An immediate thought is to use differential privacy (DP) [9], [10] to protect privacy in outlier analysis in the general data-dependent case. However, DP is inherently unable to identify or find outlier records accurately [11], [12], [13], [14], [15]. Under DP, the privacy is controlled through a parameter $\varepsilon > 0$. This parameter also controls the accuracy/utility: the higher the ε, the lower the privacy and the higher the utility. However, there is no non-trivial ε-controlled-tradeoff between privacy and utility for outlier detection or identification.

Consider outlier detection; it outputs all the records in the database, which are outliers. For this task, any ε-DP mechanism (i.e., algorithm) is bound to have a very poor utility [12] even for not very small values of ε (e.g., $\varepsilon \approx 1$). This is because a DP mechanism's output must remain almost the same under addition or removal of any record including the outlier records. Outliers are, however, typically in sparse regions of (data)-space and may have no other identical records in the database. In this case, removing an outlier record would typically eliminate the outlier entirely whereas adding a new record in any sparse region would create an outlier, in both cases significantly reducing the utility. Therefore, a DP mechanism performs very poorly for this task. Although one can achieve a desired level of utility by setting ε high enough, this effectively provides no privacy.

One way to obviate the low-utility problem is to vary the privacy guarantee based on the outlyingness. However this is not possible under DP, wherein every record is guaranteed the same privacy. There are some variants of differential privacy [13], [14], [15] that relax DP guarantee and are relevant to outlier analysis. They, however, are either limited in their application or are unable to deal with the data-dependent notions of outliers — the focus of this work.

The modern data-dependent methods of anomaly identification label a record as anomalous based on its degree of dissimilarity from other existing records. Consequently, the labeling of a record as anomalous is specific to a dataset, and knowing that a record is anomalous can leak a significant amount of information about the other existing records. This is the key challenge that any privacy-preserving anomaly identification method must overcome.

To solve this problem, we recently introduced the notion

- *The authors are with Rutgers University, New Jersey, USA E-mails: hafiz.asif@rutgers.edu, periklis.research@gmail.com, jsvaidya@business.rutgers.edu*

Manuscript received April 19, 2005; revised August 26, 2015.

of *sensitive privacy* (SP) [11], which conceptualizes what it means to protect privacy in outlier analysis. SP enables one to accurately identify if a record is anomalous while simultaneously affording a strong differential-privacy- like guarantee to *most of the records*. SP is both computationally realizable and amenable to analysis.

The key contributions of this article are to:

- Present the notion of sensitive privacy — which, compared to DP, provides an additional knob to tune the privacy-utility trade-off in outlier analysis — as well as discuss its limitations.
- Relate sensitive privacy to other privacy definitions.
- Develop a novel n-step lookahead mechanism to efficiently and privately answer arbitrary outlier queries for a restricted class of anomaly models and prove why it cannot be used in a general sense.
- Propose constructions to develop sensitively private mechanisms for identifying anomalies, prove their privacy guarantees, and characterize the conditions under which they yield optimal mechanisms.
- Establish the effectiveness of sensitively private mechanism (developed via our constructions) by empirical evaluation of over real-world datasets.

The current article is a significantly extended version of our prior work [11], further developing the concept of Sensitive Privacy and improving its theoretical understanding.

2 Related Work

Protecting privacy in data analysis aims to limit the information disclosure about individuals from the result of a data analysis. Differential privacy [9], [16] was the first mathematical notion to define privacy by quantifying this information disclosure — and it has shaped the field of private data analysis. It guarantees that no attacker can use a differentially private result (of data analysis) to find out with certainty if a particular person's data were used in the analysis. Thus, it affords "plausible deniability" to people, that is, any person can claim that her data was not used in the data analysis even if her data were used.

Differential privacy is an algorithmic definition of privacy, which requires that the probability for any output of a privacy-protecting mechanism (i.e. an algorithm that takes a database as input) "should not change much" by adding or removing any one record in the input database. Two databases that differ by one record are called *neighbors*. That is, we think of the huge space where each point in this space is a whole database and a neighboring relation between any two databases (in this simple case neighboring databases are those that differ by one record). Here, "should not change much" means that the probabilities (for any output) corresponding to any two neighboring databases should be within a multiplicative factor of e^{ε} — this is referred to as the *privacy constraint*. The $\varepsilon > 0$ is the privacy parameter. The smaller the value of value of ε, the higher the privacy. Typically, to achieve differential privacy, a mechanism probabilistically perturbs the correct answer using noise sampled from a carefully calibrated distribution. Differential privacy works well for many classes of aggregate and statistical data analysis tasks. However, there are various data analysis problems, anomaly identification being one of them, where the privacy-utility relationship essentially makes it impossible to achieve both practically meaningful privacy and utility [11], [12], [17], [18], [19]. In particular, these are data analysis problems where the output of the database query is too sensitive to replacing the input database by one of its neighbors. A good example is when the output is a yes/no result (binary function). That is, there are databases and records in these databases where a simple local modification flips the output from yes to no. Compare this with the typical application of differential privacy in reporting averages or variances etc, where small changes in a database do not have a drastic effect on the output. This has led to the development of variants of differential privacy to address important practical challenges in data analysis. Many of these variants either generalize the notion of neighboring databases or redefine what is meant by "the output should not change much" for neighboring databases (see [20] for a survey for different generalizations and variants of differential privacy).

Below, we review some of the important variants of differential privacy and identify and discuss the gaps in the context of private outlier analysis that still exist.

Pufferfish [18] and Blowfish [19] are two frameworks to give generalized versions of differential privacy. Both of them provide a way to redefine neighboring databases based on what secret (i.e. the kind of information disclosure) we want to protect (i.e. limit the information disclosure about individuals). These frameworks add to our theoretical understanding of private data analysis and are useful for applications the secret that needs to be protected can be clearly demarcated. However, these frameworks do not provide any method or direction to deal with outlier detection or identification, especially, when the outliers are defined in a data-dependent fashion. We solve this problem by conceptualizing the notion of sensitive record and sensitive neighborhood graph, both of which take into account the data-dependent nature of the problem.

Protected (differential) privacy [13], which was proposed for analyzing networks, divides the set of all possible records into two categories: one is protected while the other is not. Although we can use protected privacy (instead of differential privacy) to boost the accuracy for some types of outlier analysis, this does not work in many cases, and especially, for the case when outliers are defined in a data-dependent way. This shortcoming of protected privacy is due to the fixed and data-independent categorization of records into protected group and unprotected group (let's say the outliers), which is not possible when outliers are defined in data-dependent way. This is particularly fragile since without seeing the database it is not possible to characterize a record as an outlier or not, and additionally, changing (or adding/removing) a few records in the given data may also affect the outlying status of a record as per the specification of the chosen anomaly model. Thus, the privacy guarantee cannot be quantified in the order specified in protected privacy. This is one of the main problems that we tackle when defining sensitive privacy.

One-sided (differential) privacy [14] uses a similar approach as in protected privacy. Similar to protected privacy, it is useful for the cases where outliers are defined independent of the data as it also defines the records to be protected independent of the database. Additionally, it further relaxes the definition by only considering a subset of pair of neighbors that must satisfy the privacy constraint. This leaves one-sided privacy open to attacks that can infer if a particular ”protected” record is present in the data or not. Note that sensitive privacy is immune to such an attack.

Another way to generalize differential privacy is to have different levels of privacy (i.e. the value of ε) for different records, which Personalized (differential) privacy adopts [21]. Personalized privacy requires that the level of privacy be pre-specified for each record. For examples, when sharing their data, people can specify the level of privacy they want. However, when the outliers are defined in a data-dependent fashion, and we want to provide privacy as per the degree of outlyingness of each record (which is required to make the analysis useful), this notion of privacy (for similar reasons discussed above) is also not applicable.

As opposed to Personalized privacy, Tailored (differential) privacy quantifies the level of privacy for each record as a function of the record and the database [22]. Thus it allows one to tailor the privacy guarantee across all records. Outlier privacy, an instantiation of tailored privacy, defines privacy in the presence of outliers, however, the problem that [22] focuses on is different than ours as it aims to protect outliers with higher privacy guarantee compared to rest of the records in the data. Below, we discuss some limiting features of outlier privacy to highlight the problems it presents in carrying out an accurate private outlier analysis.

Outlier privacy affords a stronger privacy guarantee to outliers (depending upon their degree of outlyingness) compared to the other records in the data. For the problem of computing private histograms, the notion of outlier that [22] uses is equivalent to that of $(\beta, 0)$-anomaly (defined in Section 3). However, this notion of outlier is too simple to work in practice for many tasks, and the mechanisms introduced in [22] cannot address the problem of identifying anomalies with practically meaningful accuracy. In most practical cases, the outlyingness of a record i also depends upon the other records in the data, this nature of data-dependence must also be taken into account. Furthermore, when we provide more privacy to outliers, the utility of the outlier analysis degrades, even more than when we use plain differential privacy. Thus, this work [22] does not apply to the problem of identifying anomalies accurately.

We propose the notion of sensitive privacy to address all of the above mentioned shortcomings. Additionally, we consider outlier models that are more general, data-dependent, and provide constructions of privacy mechanisms to identify outliers in data — these are constructions of mechanisms that preserve utility and protect privacy.

A final related work is that of anomaly-restricted (differential) privacy [15] which does take into account the data-dependent nature of anomalies (i.e. outlier). However, it does so in a very restricted setting: in [15] the input databases are guaranteed to have only one outlier, a structure not present in typically available databases, which is in addition to other restrictions on the input database. Although it has theoretical value, it is inapplicable for most practical settings for outlier analysis. Sensitive privacy does not make such restricting assumptions, has immediate interpretation, is amenable to analysis, and efficiently realizable in practice.

In general, when mechanisms are developed to meet privacy as per the models stated above, the assumption is that a trusted curator has access to the data and is able to employ the mechanism to answer a given query. However, when no such trusted curator exists, secure multiparty computation [23], [24] can be used to simulate such a trusted curator in a distributed setting. This has also been done for outlier analysis [25], [26], [27], but it is not the focus of this work.

3 Preliminaries

Databases

In this work a database is a histogram. The set of all possible databases is denoted by $\mathcal{D} = \{y \in \mathbb{N}^{\mathcal{X}} : ||y||_1 < \infty\}$, where $\mathbb{N} = \{0, 1, 2, \dots\}$, $\mathcal{X}$ is an arbitrary finite set of possible values of records, and $|| \cdot ||_1$ is the ℓ_1 norm. Thus, for any database x and $i \in \mathcal{X}$, x_i is the number of records[1] in x that are identical to i , $i \in x$ is the binary predicate which is true if and only if (iff) $x_i \geq 1$. That is, the notations $i \in \mathcal{X}$ and $i \in x$ have completely different interpretation. We assume that each record in the database is associated with a single person.

Furthermore, for any $i \in \mathcal{X}$, we use $\mathbf{e}^i$ to denote the database consisting only of one record of value i, that is, $\mathbf{e}^i_i = 1$ and for all $j \neq i$, $\mathbf{e}^i_j = 0$. Finally, we use ξ to denote the function that makes the negative coordinates of any given $w \in \mathbb{R}^{\mathcal{X}}$, zero, that is, for every i, $\xi(w)_i = w_i$ if $w_i \geq 0$ and $\xi(w)_i = 0$ otherwise.

Anomalies (i.e. Outliers)

To characterize the outlyingness of a record, we will use the concept of normality property instead of an anomaly model. We first define an anomaly model and then use it to define the normality property.

An *anomaly model* is a predicate F over the domain $\mathcal{X} \times \mathcal{D}$ such that for any $i \in \mathcal{X}$ and $x \in \mathcal{D}$, $F(i, x) = 1$ if a record of value i is anomalous with respect to the database x, otherwise $F(i, x) = 0$. We emphasize that for the predicate to be true, a record of value i need not present in x, namely, $F(i, x) = 1$ does not imply $i \in x$.

Now, for a fixed anomaly model F, a *normality property* is a predicate $p : \mathcal{X} \times \mathcal{D} \to \{0, 1\}$ such that for every $i \in \mathcal{X}$ and every $x \in \mathcal{D}$, $p(i, x) = 1$ iff i is present in x (i.e. $i \in x$) *and* i is non-anomalous with respect to x (i.e. $F(i, x) = 0$). Note that the normality property is not simply the negation of anomaly model since the normality value is also predicated on the presence of the record in the given database, which is not the case for the anomaly model.

Henceforth, when we fix a normality property, we assume an underlying fixed anomaly model. Thus, we shall omit mentioning the anomaly model when it is clear from the context.

We use $\mathfrak{P} = \{\text{property} : \mathcal{X} \times \mathcal{D} \to \{0, 1\}\}$ to denote the set of all properties, wherein the set of all normality properties makes a subset of $\mathfrak{P}$.

[1] It is common in the literature to call $i \in \mathcal{X}$ as “type”, whereas for added clarity in the context of anomaly identification we will be referring to i as “record”.

Anomaly identification function :
We now introduce the important notion of *anomaly identification function*. It tells us if a record is present in the database as an anomalous record. For a fixed anomaly model, we say a boolean function $g : \mathcal{X} \times \mathcal{D} \rightarrow \{0,1\}$ is an anomaly identification function if for every $i \in \mathcal{X}$ and $x \in \mathcal{D}$, $g(i,x) = 1 \iff i \in x$ and i is an anomalous record with respect to x (note that no change is made to x).
We will mainly focus on the above formulation of identifying anomalies because it represents a fundamental and one of the most conceptually difficult to deal with cases of privacy-preserving outlier analysis, especially, for differential privacy such as definitions of data privacy [11], [12]. Alternatively, we could have defined g without predicating on the existence of i in x. But when we drop the predicate on the existence of i, we in effect blur the distinction between the notion of a void spot (that in a different database could have been occupied by a record) in the database and the notion of an anomaly. We, however, note that the above given formulation is extensible to the case where the database, over which anomaly identification is performed, is considered to include the record for which anomaly identification is desired. Here, for example, the anomaly identification for a record i over a data x can be computed over the database that consists of all the records in x as well as the record i.

The main anomaly model we are considering :
Although, some of our developments are for general anomalies/outliers, we will be focusing on (β, r)-anomaly as the notion of outlier. We choose the (β, r)-anomaly model since it is quite prevalent in practice, it generalizes many statistical anomaly models, and has many well-known variants and extensions [4], [28], [29] that our work naturally extends to them. Under (β, r)-anomaly model, *a record is considered anomalous (i.e. outlier) if there are no more than β records similar to it*. The parameters β and r are given by the domain experts [29] or found through exploratory analysis.
We use the following notation to give the definition of (β, r)-anomaly. For any $x \in \mathcal{D}$, $i \in \mathcal{X}$, $r \geq 0$, and distance function $d_{\mathcal{X}} : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}_{\geq 0}$, $B_x(i,r) = \sum\limits_{j \in \mathcal{X} : d_{\mathcal{X}}(i,j) \leq r} x_j$.

Definition 1 ((β, r)-anomaly [29])**.** *An* anomaly *is defined for a database $x \in \mathbb{N}^{\mathcal{X}}$ and a record $i \in \mathcal{X}$ as follows. We say that i is a (β, r)*-anomaly *in the database x if $i \in x$ (i.e. i is present in x), and $B_x(i,r) \leq \beta$ (i.e. there are at most β records in x that are within distance r from i).*

Whenever we refer to a (β, r)-anomaly, we assume that there is any fixed distance function $d_{\mathcal{X}}$. That is, $(\mathcal{X}, d_{\mathcal{X}})$ is a metric space.

Privacy

A remark on terminology: The term "query" is overloaded. In general, a "query" is a function. A "query instance" is a function together with the inputs to the function. We think of queries as the problem we have to solve, whereas a mechanism is an algorithmic solution to this problem. In the differential privacy literature mechanisms are randomized algorithms whose only input is a database (and nothing else). However, our intuitive real-world problem regards "querying" a specified database *and* a specified record; i.e. we ask whether the specified record is an anomaly in the database. To deal with this notational issue we consider a family of algorithms $\{M_{\text{record}}\}_{\text{record} \in \mathcal{X}}$ one for each record. In other words, a "query" resolved by a privacy mechanism regards a fixed record, whereas an intuitive notion of a query gets as input both the database and the record.

Differential privacy :

Definition 2 (differential privacy [9], [16])**.** *For a given $\varepsilon > 0$, we say a mechanism M is ε-differentially private if for every $x, y \in \mathcal{D}$ such that $||x - y||_1 = 1$, and every measurable $R \subseteq Range(M)$,*

$$\mathrm{P}\left(M(x) \in R\right) \leq e^{\varepsilon} \mathrm{P}\left(M(y) \in R\right).$$

Private anomaly identification query & mechanism :
The query associated with an anomaly identification function, g, is called *anomaly identification query* (AIQ). Since, the input to the private mechanism is only the database, for AIQ, we think of each AIQ for each fixed record. Thus, we specify an AIQ by the pair (i, g), and we also write g_i, where i is a record and g is an anomaly identification function. Now, for a fixed AIQ, g_i, a private anomaly identification mechanism, $M_i : \mathcal{D} \rightarrow \{0,1\}$ is identified by its distribution, where for every x, $\mathrm{P}\big(M_i(x) = g_i(x)\big)$ is the probability that M_i outputs correctly, and $\mathrm{P}\big(M_i(x) \neq g_i(x)\big)$ is the probability that M_i errs on x. When there is no confusion we write M instead of M_i.

Privacy induced graphs :
We rely on the following graphs to define privacy for outlier analysis. These graphs are used to appropriately generalize differential privacy and play a central role in our analysis, for more details see [18], [19], [30]. We consider simple graphs over the databases and call them *neighborhood graphs*.

Definition 3 (neighborhood graph)**.** *A simple graph $G = (\mathcal{D}, E)$, where $\mathcal{D}$ is the set of all nodes, is called a* neighborhood graph *if the set of edges, E, is such that*

$$E \subseteq \big\{\{x,y\} : \ x, y \in \mathcal{D} \text{ and } ||x - y||_1 = 1\big\}.$$

Note that a neighborhood graph is typically infinite. For any given neighborhood graph, G, $\mathcal{E}(G)$ denotes the set of edges in G. For any two neighborhood graphs G and G' over the same set of vertices, we say G' is subgraph of G if $\mathcal{E}(G') \subseteq \mathcal{E}(G)$. As an example, let us look at the neighborhood graph, $\mathbb{G}$, associated with differential privacy. $\mathbb{G}$ is such that for every $x, y \in \mathcal{D}$, $\{x,y\} \in \mathcal{E}(\mathbb{G}) \iff ||x-y||_1 = 1$. We call $\mathbb{G}$ the DP neighborhood graph. Note that every neighborhood graph is a subgraph of $\mathbb{G}$.
For any given neighborhood graph, G, we define the shortest path metric over each of the connected components, G' of G, i.e. $d_{G'}$, which gives the shortest path length between any two nodes of the connected component G', where the path length corresponding to any two nodes directly connected by an edge is 1. We refer to this metric as the *shortest path metric*. For simplicity, we abuse the notation, and write d_G to denote the collection of all metrics, each for a connected component of the neighborhood graph G. We stress out that the d_G is only defined for the databases , i.e. nodes, that are connected in G. Any two databases x and

y that are connected by an edge in the neighborhood graph G (i.e. $d_G(x, y) = 1$) are called neighbors.

Another important concept in this context is that of *Lipschitz continuity*, a property of a function f with respect to the neighborhood graph G (f is defined over the entire G — we do not have a different f for each connected component of G). We use this notion to define a necessary constraint for privacy-protecting mechanisms to identify outliers. In our exposition we will consider f from $\mathcal{X} \times \mathcal{D}$ to $\mathbb{R}_{\geq 0} \cup \{\perp\}$. So we will extend the standard notion of Lipschitz continuity, considered in privacy literature [30], to cover the non-real part of the function (i.e. $\perp$) as well.

Definition 4 (Lipschitz continuity)**.** *For any given neighborhood graph, G, $\alpha > 0$, we say a function $f : \mathcal{X} \times \mathcal{D} \to \mathbb{R}_{\geq 0} \cup \{\perp\}$ is α-*Lipschitz continuous *if for every $i \in \mathcal{X}$ and neighboring x and y in G, the following holds:*

if $f(i, x) = \perp$ *or* $f(i, y) = \perp$ *then* $f(i, x) = f(i, y) = \perp$
otherwise $|f(i, x) - f(i, y)| \leq \alpha$

Privacy setting :

We consider the trusted curator setting for privacy. The trusted curator is a fully trusted third party, who has access to the database. It receives the query (for example, an AIQ), uses a mechanism to compute the query, and sends the result back. Now, if the curator uses a private mechanism, then we are guaranteed that the query is answered in a privacy-preserving fashion.

4 Sensitive Privacy (SP)

We now define the notion of *sensitive privacy*. *Sensitive privacy* requires privacy protection of every record that may be normal under a small change in the database. By "privacy of a record $i \in \mathcal{X}$" we mean the "privacy of the mechanism M_i". To achieve this, we define sensitive privacy using a metric space over the databases and require a private mechanism to statistically blur the distinction between databases that are close in the metric space.

While differential privacy uses the ℓ_1-metric, we utilize a different metric over databases, which we define using the notion of sensitive record. Informally, we say a record is *sensitive* with respect to a database if it is normal or becomes normal under a small change in the database. Although, sensitive records are technically important just because they help us appropriately generalize differential privacy, as a bonus we also get that the definition of a sensitive record it happens to be a natural one and is inspired from the existing literature on outlier analysis [31], [32].

Recall that, by definition, an anomalous record significantly diverges from other records in the database. [31], [32]; That is, if we make a small change (e.g. add or remove a few records) to the database, the label (provided by the anomaly model) of an anomalous record should not change. So, the records whose labels do change are not anomalous, and all of these are protected[2] under sensitive privacy.

Now, given the notion of sensitive record, we define the metric over the databases by considering a graph over the databases. In this graph, called *sensitive neighborhood graph*, every two nodes (i.e. databases) that differ by a sensitive record are connected by an edge. The metric distance over the databases is now given as the shortest path length between the databases in the same connected component of the graph. This metric space has the property that databases differing by a sensitive record are closer compared to the databases differing in a non-sensitive record. As a result this metric space enables us to fine-tune the trade-off between accuracy and privacy in outlier analysis. A conceptual contribution of this work is to express this trade-off only via two parameters (i.e., ε and k, see below).

Formalization

Let us start by formalizing the concepts of sensitive record and sensitive neighborhood graph.

We use the notion of *normality property* p to identify the normal records that exist in the database. We formalize the notion of "small change" in the database as the addition or removal of k records from the database. We consider this change to be typical and want to protect the privacy of every record that may become normal under this small change. Next, we use this notion of small change in the database to define the key concept of sensitive record. Informally, for a fixed normality property, all the records whose privacy must be protected are termed as *sensitive records*.

Definition 5 (sensitive record)**.** *Arbitrarily fix $k \geq 1$ and a normality property p. For any given database $x \in \mathcal{D}$, we say an $i \in \mathcal{X}$ is* k-sensitive with respect to x *if there is a database $y \in \mathcal{D}$ such that $||x - y||_1 \leq k$ and $p(i, y) = 1$.*

As an example of sensitive record, consider $(\beta = 3, r = 0)$-anomaly and $k = 2$. In this case, for any given x, an i will be 2-sensitive with respect to x if $x_i \geq 2$. This is because if $x_i \leq 1$, then adding (and/or removing) any two records from x can never make i a non-outlier in x as x_i will remain at max 3. However, when $x_i \geq 2$, adding 2 records of value i to x ensures that i will be a non-outlier.

Next, we give the important notion of *k-sensitive neighborhood graph*. It generalizes the DP neighborhood graph through the concept of sensitive records. Throughout this work, we use $\mathbb{G}$ to depict the DP neighborhood graph (see Preliminaries for the definition).

Definition 6 (sensitive neighborhood graph)**.** *For arbitrarily fixed $k \geq 1$ and normality property p, a neighborhood graph G_S is called k-*sensitive neighborhood graph *for p if for every $\{x, y\} \in \mathcal{E}(\mathbb{G})$, $\{x, y\} \in \mathcal{E}(G_S) \iff$ there is an $i \in \mathcal{X}$ such that:* 1) *$|x_i - y_i| = 1$ and* 2) *i is k-sensitive with respect to x or y.*

Note that although sensitivity of a record is determined by the addition and/or removal of k records, the neighboring databases in k-sensitive neighborhood graph (G_S) differ by exactly one record. For instance, if we are given that i is k-sensitive with respect to x, the neighbors of x that differ in i can be at most two: one is obtained by adding i to x and the other is obtained by removing i from x.

Furthermore, G_S (as opposed to $\mathbb{G}$) is tied to the normality property, and hence, the anomaly definition. Distances (i.e. shortest path lengths) is these graphs allow us to generalize in a clean way the differential privacy as well as other privacy definitions (see Section 7).

[2] A mechanism M_i for a record i is where we quantify privacy. We say that a "record i is protected" meaning that there is a mechanism M_i (which should be clear from the context) that has sufficiently good privacy guarantees.

Definition 7 (sensitive privacy). *Arbitrarily fix $\varepsilon > 0$, $k \geq 1$, and a normality property p, and let G_S be the k-sensitive neighborhood graph for p. Then we say a mechanism M is (ε, k)-*sensitively private *if for every two neighbors x and y in G_S, and every measurable $R \subseteq Range(M)$,*

$$\mathrm{P}\left(M(x) \in R\right) \leq e^{\varepsilon}\mathrm{P}\left(M(y) \in R\right).$$

Sensitive privacy requires that for every two neighbors, any statistical test R one may be concerned about should occur with "almost the same probability". Namely, the presence or the absence of a k-sensitive record should not affect the likelihood of occurrence of any event. Here, "almost the same probability" means that the above probabilities are within a multiplicative factor e^{ε}, which for reasonably small $0 < \varepsilon < 1$ we have $e^{\varepsilon} \approx (1+\varepsilon)$.

When ε, k and p are fixed, for any input database x, we are guaranteed that the output of the sensitively private mechanism remains the "same" under addition or removal of any one record from x that is k-sensitive for x or any of its neighbors. Thus, from a sensitively private output, an attacker cannot infer if a sensitive record was present in the database. This always holds for all the normal records since every normal record is k-sensitive for all values of k.

When p is fixed, ε and k relate to the SP guarantee as follows: the lower the value of ε, the higher the privacy guarantee of sensitive privacy (which is similar to DP); in contrast to ε, the higher the value of k, the higher the privacy guarantee.

Furthermore, sensitive privacy also has the properties of composition and post-processing [11]. However, unlike differential privacy (where the composition is with respect to ε), the composition for sensitive privacy is with respect to ε and a fixed k-sensitive neighborhood graph.

For outlier analysis, we will show that sensitive privacy admits mechanisms that can accurately identify whether a record is anomalous while simultaneously guaranteeing strong privacy by making it statistically impossible to infer if a non-anomalous record was included in the database.

5 Understanding Sensitive Privacy

In this section, we first discuss how the parameters ε and k relate to the SP-guarantee, especially, in terms of the sensitive neighborhood graph. Since the analysis for the third parameter (the normality property) is non-trivial and complicated, it is separately detailed in Section 6. Second, we elaborate on the distinguishing characteristics of sensitive privacy for anomaly identification, and how it compares to differential privacy through indicative experimental results over data generated from 2D Gaussian distribution.

SP: the role of ε and k

The privacy parameter ε plays the same role in sensitive privacy as it does in differential privacy: the smaller its value, the higher the privacy. For neighboring databases in a sensitive neighborhood graph (G_S), the guarantee of sensitive privacy is exactly the same as that of differential privacy. However, if the two databases x and y, differ by one record that is non-sensitive (for both the databases), then they are not neighbors in G_S, and the SP-guarantee for such a pair is weaker than of differential privacy, nevertheless, it has the same form.

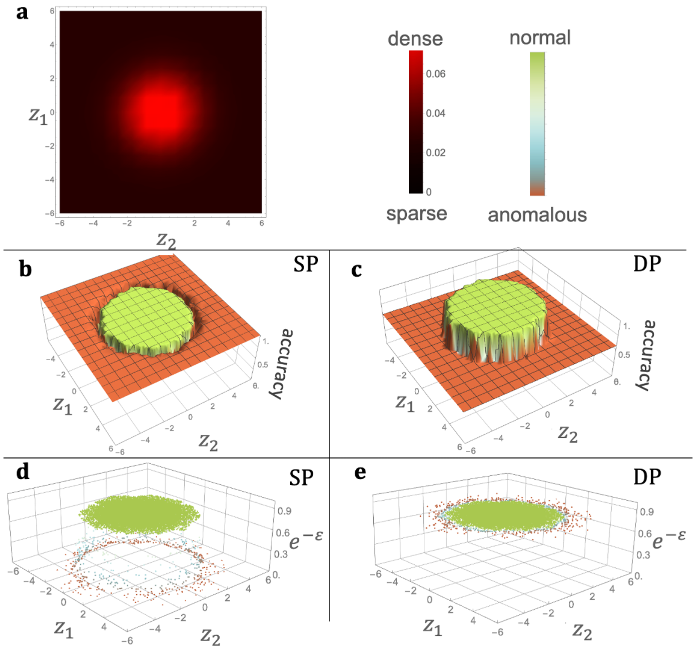


Fig. 1: **(a)** gives the density plot of the distribution of the example data. z_1 and z_2 axes give the coordinate of a point (record). **(b)** and **(c)** respectively show the accuracy (on vertical axis) for (β, r)-AIQ's (for each record, i.e., data point) via $(0.25, 1)$-SP and 0.25-DP mechanisms (in Section 8.4). The plots give the interpolated results(footnote# 5) to clarify the relationship of outlyingness and accuracy. **(d)** and **(e)** give the privacy (on vertical axis for $\varepsilon = \ln(A_i)$) for each i in the data for private (β, r)-AIQ. All the green (normal) points in **(d)** are at the same level as all the points in **(e)**.

Next, we discuss the parameter k. It quantifies the level of sensitivity of records: (for a fixed ε and p) the higher the value of k, the stronger the SP-guarantee. A higher value of k results in more records being considered sensitive. This is because a record that is k-sensitive (with respect to a database x) is also $(k+1)$-sensitive (with respect to x). In a typical setting[3], if G_S and G'_S are two sensitive neighborhood graphs for k and k' such that $k \geq k'$, then, compared to G'_S, more databases are neighbors in G_S ($\mathcal{E}(G_S) \supseteq \mathcal{E}(G'_S)$). Thus, if k is large enough, the SP-guarantee is the same as the DP-guarantee. Hence, k provides a way to trade-off privacy and utility in anomaly identification.

Indicative Experimental Results comparing SP and DP

Here, we wish to use Figure 1 and 2 (showing empirical results[4]) as a means to understand what distinguishes sen-

[3] In a typical setting, every record i can be an anomalous or normal depending upon the database, i.e., for each i, there are two databases: in one, i is anomalous, and in the other i is normal.

[4] The results are obtained over database of size $n = 20,000$, sampled from the 2D normal distribution, $N(\mu, \Sigma)$, for $\mu = (0,0)$ and $\Sigma = \left[\begin{smallmatrix} 2 & 0 \\ 0 & 2 \end{smallmatrix}\right]$. The β and r are set as follows: $\beta = 1.2 \times 10^{-3} n$ and $r = 0.13\sqrt{\sigma_1^2 + \sigma_2^2}$. This relationship is adapted from [29], which established this relationship for one-dimensional (1D) data (i.e. not the same as in the example of Figure 1). For (the 1D case) any record i in a sampled database, with empirical mean μ and empirical standard deviation σ, we say i is anomalous if $|i - \mu| \geq 3\sigma$. This fundamental notion of anomaly is equivalent to $(\beta = 1.2 \times 10^{-3} n,\ r = 0.13\sigma)$-anomaly [29], where n is the size of the database.

sitive privacy from differential privacy for anomaly identification.These plots are based on our SP and DP mechanisms for (β, r)-AIQ (from Section 8.4), which we used to compute and plot the accuracy[5] and privacy for each record.

We know that under SP – unlike DP – all the records do not have the same privacy guarantee; and additionally, SP-guarantee for every record i varies across databases (due to the data-dependent nature of sensitivity, i.e., sensitive record). So, for the purpose of comparing SP and DP, we quantify the privacy guarantee of a mechanism – whether SP or DP – for any i in the context of the given database (this approach is similar to that of tailored DP [22], discussed in Section 7). For our mechanisms (both SP and DP) and a given database x, the privacy guarantee for any record i can be given as[6] $\varepsilon_i = \ln(A_i)$ (which measures the max divergence in the probability of M_i's outputs for x and its neighbors in i – note that i is not necessarily sensitive):

$$A_i = \max_{w \in \{y,z\}} \max_{b \in \{0,1\}} \left\{ \frac{P(M_i(x) = b)}{P(M_i(w) = b)}, \frac{P(M_i(w) = b)}{P(M_i(x) = b)} \right\}$$

where $M_i : \mathcal{D} \to \{0, 1\}$ is the privacy mechanism (either SP or DP) for (β, r)-AIQ for i, $y = \xi(x + \mathbf{e}^i)$ and $z = \xi(x - \mathbf{e}^i)$.

Let us now discuss the distinguishing features of sensitive privacy and how they compare with differential privacy.

- Under sensitive privacy, every k-sensitive i has $\varepsilon_i = \varepsilon$. For instance, in Figure 1d, all the green points are non-anomalous and have the same privacy ε. This privacy guarantee is the same as provided by DP (green points in Figure 1d are at the same level as all the points in Figure 1e). This is because most of the records in real-world databases are not anomalous, and hence, are sensitive.
- In practice, for anomaly identification, sensitive privacy guarantee strong and similar privacy guarantee for most of the records in the data. In Figure 1d, the vast majority of points is green (or greenish), i.e. non-anomalous, and for each such i, $\varepsilon_i \approx \varepsilon$.
- Sensitive privacy provides a novel way to define outlyingness, and relates it to the privacy-accuracy trade-off of the mechanism. For a given database x and record i, outlyingness of i is measured by the smallest distance from x at which there is a database y in G_S (the sensitive neighborhood graph) such that i is k-sensitive for y — the greater the distance, the stronger the outlyingness. Under this notion of outlyingness, the weaker the outlyingness of a record, the higher privacy it gets with sensitive records getting the highest privacy, i.e. ε (see the level of green, blue, and brown points in Figure 1d). In contrast, under DP all record have the same privacy ε.
- Furthermore, the higher the outlyingness of a record, the higher the accuracy the SP mechanism is allowed to achieve for AIQ the record, which is in contrast to DP (compare Figures 1b and 1c). Similarly, the higher the non-outlyingness of a record (i.e., the record is 'super' normal), the higher the accuracy the SP mechanism can achieve, and this is akin to DP.
- The parameter k defines the boundary between the sensitive and non-sensitive records: the higher the value of k, the more records are considered sensitive, and therefore, must be protected with higher privacy guarantee. See Figure 2, where we show how increasing k, increases the number of sensitive records and the privacy guarantee.

6 SP-Guarantee Analysis w.r.t. Normality Property

The normality property plays a central role in defining the sensitive privacy guarantee. For instance, assume a constant normality property p, which always outputs 0 (i.e. $p(i, x) = 0$ for every i and x), and a mechanism (to compute a function f) that always outputs the correct answer. This mechanism is sensitively private (for p), but it is not differentially private.

Now, on the other hand, consider a normality property p that outputs 1 if the database's size is odd and 0 otherwise (i.e. $p(i, x) = 1 \iff |x|$ is odd, for every i and x).

For such a property, every record is sensitive with respect to every database. Therefore, in this case, a mechanism is sensitively private if and only if it is differentially private (see Section 7 for details on how DP is related to SP).

Between the aforementioned two extremes lies a set of practically-meaningful properties ($P \subsetneq \mathfrak{P}$) that we call *regular* (defined shortly). Regular properties not only provide a meaningful SP guarantee for anomaly identification but also corresponds to many of the anomaly models used in practice. Informally, a property $p \in \mathfrak{P}$ is called regular if every two nodes (i.e. databases), each of which has at least one sensitive record, are connected in the sensitive neighborhood graph for p.

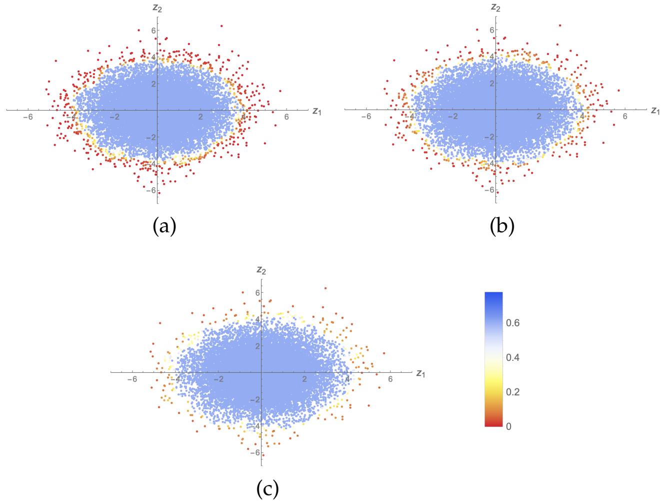
To define the notion of regular property, we need to clarify some notation and definitions. For arbitrarily fixed $k \geq 1$ and a property $p \in \mathfrak{P}$, let D^p be a subset of $\mathcal{D}$ such that $D^p = \{x \in \mathcal{D} : \exists i \in \mathcal{X}$ s.t. i is k-sensitive w.r.t. $x\}$. Now for a given D^p and a k-sensitive neighborhood graph, G_S, corresponding to p, let $G_S(D^p) = (D^p, E)$ be a subgraph of G_S such that for every $x, y \in D^p$, $\{x, y\} \in E$ if an only if $\{x, y\} \in \mathcal{E}(G_S)$. The definition of regular property follows.

Definition 8 (regular property). *For any given property $p \in \mathfrak{P}$, we say p is* regular *if for every $k \geq 1$, $G_S(D^p)$ is connected, where G_S is the k-sensitive neighborhood graph for p.*

Thus, for any regular property p, all the databases that have at least one sensitive record are in one connected component, $G_S(D^p)$, of G_S, and the databases that are not connected with $G_S(D^p)$ do not have any sensitive record. However, such databases do not correspond to the databases for outlier analysis that we encounter in practice or for which the anomaly models are conceived. Since by definition, outliers make a minority of the records in a database [4], [5], hence, regular properties are representative of the practical settings for outlier analysis.

[5] To make the visualization clear, we interpolate the results from 30 sampled databases by using one-degree polynomial in the two coordinates (Figure 1b-c). For this, we used "ListPlot3D" function of Mathematica 12 with "InterpolationOrder" as 1.

[6] Under our mechanisms, for any i, it suffices to measure the privacy loss under M_i (as opposed to M_j for any other $j \neq i$) by contrasting x with databases y and z. This is because, for (β, r)-AIQ for the record i, the center of the ball is at i and thus it is exactly for M_i and databases y and z where the privacy loss is maximized. This can be confirmed by looking at the mdd-function for (β, r)-anomaly identification function (under DP) and λ_k for SP, both given in Section 8.4.



(a) (b) (c)

Fig. 2: **(a)-(c)**, the plot is for the same data and SP mechanism as in Figure 1 but for varying k. The two axes give the coordinates of a point. The color gives the level of privacy in terms of the value $e^{-\varepsilon}$ (for each i, $\varepsilon = \ln(A_i)$), under $(0.25, k)$-SP AIQ. **(a)**, $k = 1$. **(b)**, $k = 7$. **(c)**, $k = 14$.

On the other hand, if a property p is not regular, then the databases in $G_S(D^p)$ belong to at least two different connected components. Now for an (ε, k)-SP mechanism M, it becomes possible that for two databases $x, y \in G_S(D^p)$ and an $R \subseteq Range(M)$, $P\left(M(x) \in R\right) > 0$ while $P\left(M(y) \in R\right) = 0$. However, this situation does not occur for regular properties (Lemma 2).

In the next section, we discuss and analyze normality properties associated with (β, r)-anomaly model.

Regular vs. non-regular (β, r)-normality properties

We use $p_{[\beta,r]}$ to depict an arbitrary (β, r)-normality property, i.e. the normality property corresponding to (β, r)-anomaly. Different values of β and r and the distance function over the records correspond to different settings. It turns out that for most practical settings, $p_{[\beta,r]}$ is indeed regular, but not for all.

Since the notion of regular property is defined via sensitive records, let us first discuss what makes a record k-sensitive under $p_{[\beta,r]}$. For a given $k, p_{[\beta,r]}$ and x, a record i is k-sensitive for x if there are $(\beta+1-k)$-many records in x that are within distance r of i (i.e., $B_x(i,r) \geq \beta+1-k$; Lemma 1). For instance, consider the case for $k = 2$ and $B_x(i,r) = \beta-1$ (for $\beta \geq 2$). Here, i is k-sensitive for x. This is because by adding two records – one of value i and the other of any value j such that $d_{\mathcal{X}}(i,j) \leq r$ – to x, i will become a non-outlier. There is, however, a subtlety when it comes to considering the neighbors in sensitive records. Take a y such that $B_y(i,r) = \beta - 2$. Clearly, i is not k-sensitive for y; because by adding or removing any two records from y, i cannot become a non-outlier. Yet x is a neighbor of y. This is because i is k-sensitive for x, and thus, $x + \mathbf{e}^i$ and $x - \mathbf{e}^i = y$ are neighbors of x.

Lemma 1 ([11]). *For every $k \geq 1$, $p_{[\beta,r]}$,$i \in \mathcal{X}$, and $x \in \mathcal{D}$, i is k-sensitive for $x \iff B_x(i,r) \geq \beta + 1 - k$.*

For instance, for $k < \beta$, $(\beta, 0)$-normality property is not regular (Claim 1). The non-regularity of $p_{[\beta,0]}$ causes anomalous records to have no privacy, however, the privacy for the sensitive records remains the same. At the same time the non-regularity of $p_{[\beta,0]}$ also allows us to develop a very simple and efficient mechanism to achieve sensitive privacy to compute any given query (see Section 8.1).

Note that when $k \geq \beta$, the k-sensitive neighborhood graph, G_S, for $p_{[\beta,0]}$ is the same as $\mathbb{G}$. This holds because for every $i \in \mathcal{X}$ and $x \in \mathcal{D}$, there is a neighbor y of x (in G_S) such that i is k-sensitive for x or y. Any property $p \in \mathfrak{P}$ that yields a k-sensitive neighborhood graph equal to $\mathbb{G}$ is regular. However in such a case, the SP guarantee is identical to the DP guarantee.

Claim 1. *For every $1 \leq k < \beta$ and $\beta > 1$, $p_{[\beta,0]}$ is not regular.*

To confirm the above claim, arbitrarily fix $\beta > 1$ and $k < \beta$. Let x and y be two databases such that for some $i \neq j$, $x_i = \beta+1-k$ and $x_j = 0$, and $y_i = 0$ and $y_j = \beta+1-k$. Note that i is k-sensitive for x but not y, and j is k-sensitive for y but not x (follows from Lemma 1). Although there is one sensitive record for x and one for y, they are not connected in the k-sensitive neighborhood graph. This holds because for every database z that differs from x by one record, j is not k-sensitive for z (as $z_j + k \leq \beta$); thus, $z_j = x_j = 0$. Hence, for every neighbor z of x, $z_j = x_j = 0$ as well. Furthermore, a simple inductive argument (on the databases at distance $\ell \in \mathbb{R}$ from x) shows that any database, w, that is connected to x has $w_j = 0$. Thus, we conclude that y is not connected to x, and the $(\beta, 0)$-normality property (for $k < \beta$) is not regular.

We now characterize the condition that makes a $p_{[\beta,r]}$ regular. For this, we use the parameter r. When r is *non-trivial*, $p_{[\beta,r]}$ is regular (Claim 2). To define what makes r non-trivial, we first clarify some notation. We use $\mathcal{S} : \mathcal{X} \times \mathbb{N} \to 2^{\mathcal{X}}$ (where $2^{\mathcal{X}}$ is the power set of $\mathcal{X}$) to define the sets of records that are reachable from any record i. For a given r, every $\ell \in \mathbb{N}$ and $i \in \mathcal{X}$, $\mathcal{S}(i,\ell) = \bigcup_{j \in \mathcal{S}(i,\ell-1)} X(j,r)$, where $\mathcal{S}(i,0) = \{i\}$ and $X(i,r) = \{j' \in \mathcal{X} : d_{\mathcal{X}}(i,j') \leq r\}$ (recall that (β, r)-anomaly comes with a distance function $d_{\mathcal{X}}$, see Section 3).

For (β, r)-anomaly and $d_{\mathcal{X}}$, we say the parameter $r \geq 0$ is *non-trivial* if there exists $s \in \mathbb{N}$ such that for every $i \in \mathcal{X}$, $\mathcal{S}(i,s) = \mathcal{X}$. Claim 2 follows.

Claim 2. *For any given distance function $d_{\mathcal{X}}$, if the parameter r is non-trivial, then $p_{[\beta,r]}$ is regular.*

Proof. Arbitrarily fix a set $\mathcal{X}$ of order m and a (β, r)-normality property, $p_{[\beta,r]}$, for arbitrarily fixed $\beta \geq 1$, $r \geq 0$ and $d_{\mathcal{X}}$ such that r is non-trivial for the distance function $d_{\mathcal{X}}$. Next, fix an arbitrary value of $k \geq 1$, and let G_S be the k-sensitive neighborhood graph for $p_{[\beta,r]}$.

We now recall that $D^{p_{[\beta,r]}}$ is the maximal set such that $D^{p_{[\beta,r]}} \subseteq \mathcal{D}$ and for every $x \in \mathcal{D}$, $x \in D^{p_{[\beta,r]}} \iff$ there exists $i \in \mathcal{X}$ that is sensitive with respect to (w.r.t.) x.

We prove the claim in two steps through a reachability argument, where we show every database (node) in $G_S(D^{p_{[\beta,r]}})$ is connected to a fixed database that has β mass in each coordinate. For the proof, we arbitrarily fix an $x \in D^{p_{[\beta,r]}}$

and an $i \in \mathcal{X}$ such that i is k-sensitive w.r.t x, and hence, $B_x(i,r) \geq \beta + 1 - k$ (follows from Lemma 1).

Next, we define a function, ω_x, to give databases that are same as x except they differ from x in the coordinate i. For every $a \in \mathbb{N}$, we define $\omega_x(a) = x + a \cdot \mathsf{sgn}(\beta - x_i)\mathbf{e}^i$, where sgn is the standard *signum* function that outputs 0 for the input 0. Thus, $\omega_x(a)$ is the same as x except for ith coordinate where its mass is more (less) by a than that of x if $x_i < \beta$ (respectively if $x_i > \beta$).

Part (a): Here we show that x is connected to a database in $G_S(D^{p_{[\beta,r]}})$ that is same as x except for the coordinate i, where it has mass β (i.e. the database $\omega_x(|\beta - x_i|)$).

When $|\beta - x_i| = 0$, the claim holds trivially (as x is reachable from itself). So, we consider the case for $|\beta - x_i| > 0$. Here, for $a = 1, \ldots, |\beta - x_i|$, $\omega_x(a-1)$ is a neighbor of $\omega_x(a)$ in G_S, and both $\omega_x(a-1)$ and $\omega_x(a)$ belong to $D^{p_{[\beta,r]}}$ because for every a (as given above), i is sensitive w.r.t. $\omega_x(a)$ due to $B_{\omega_x(a)}(i,r) \geq \beta + 1 - k$.

From the above it follows that x is connected to the database $\omega_x(|\beta - x_i|) \in D^{p_{[\beta,r]}}$ through the path given by $< x, \omega_x(1), \ldots, \omega_x(|\beta - x_j|) >$.

Part (b): Here, we will use an inductive argument to show that the database, which has β mass in each coordinate, is reachable from x.

For our fixed database x and any $J \subseteq \mathcal{X}$, let y^J be a database such that for every $j \in J$, $y^J_j = \beta$, and for every $j \in \mathcal{X} \setminus J$, $y^J_j = x_j$. And let s be the smallest integer such that $\mathcal{S}(i,s) = \mathcal{X}$ — this holds because r is non-trivial. We want to show that $y^{\mathcal{S}(i,s)}$ reachable from x.

Firstly, note that $y^{\mathcal{S}(i,0)}$ can be reached from x (Part (a)) — this proves the base case. Next, assume that for some ℓ such that $0 \leq \ell < s$, $y^{\mathcal{S}(i,\ell)}$ is reachable from x (inductive hypothesis). We show that $y^{\mathcal{S}(i,\ell+1)}$ is reachable from x.

Let $J = \mathcal{S}(i,\ell) \cup (\mathcal{S}(i,\ell+1) \setminus \mathcal{S}(i,\ell)) = \mathcal{S}(i,\ell) \cup \{i_1, i_2, \ldots, i_n\}$ for some n, and for every t in $\{0,1,2,\ldots,n\}$, $J_t = \mathcal{S}(i,\ell) \cup \{i_1, i_2, \ldots, i_t\}$, where $J_0 = \mathcal{S}(i,\ell)$.

Note that every $j \in J$ is k-sensitive with respect to $y^{\mathcal{S}(i,\ell)}$. To confirm this, arbitrarily fix a $j \in J$. If $j \in \mathcal{S}(i,\ell)$ then j is sensitive with respect to $y^{\mathcal{S}(i,\ell)}$ (follows from Lemma 1 as $y^{\mathcal{S}(i,\ell)}_j = \beta$). If, however, $j \in J \setminus \mathcal{S}(i,\ell)$, then there exists a $j' \in \mathcal{S}(i,\ell)$ such that $d_{\mathcal{X}}(j,j') \leq r$ (follows from the definition of $\mathcal{S}$), and for $z = y^{\mathcal{S}(i,\ell)}$, $B_z(j,r) \geq \beta$; thus, j is k-sensitive w.r.t. $z = y^{\mathcal{S}(i,\ell)}$ (Lemma 1).

Now, from Part (a), it follows that, y^{J_1} is reachable from y^{J_0}, and y^{J_2} is reachable from y^{J_1}, and so on. Thus, y^{J_n} ($= y^J$) is reachable from y^{J_0}, that is, the inductive hypothesis implies that $y^{\mathcal{S}(i,\ell+1)}$ is reachable from x.

Thus, we conclude that $y^{\mathcal{S}(i,s)}$ is reachable from x. Since k and $x \in D^{p_{[\beta,r]}}$ were fixed arbitrarily, the claim holds true for every $k \geq 1$ and $x \in D^{p_{[\beta,r]}}$. This completes the proof. □

7 SP in Relation to Other Definitions

In this section, we show how sensitive privacy relates to the other related data privacy concepts in the literature.

Differential privacy: We begin by restating the definition of differential privacy in terms of the neighborhood graph.

Definition 9. *For any given $\varepsilon > 0$, we say a mechanism M is ε-differentially private if for every two neighbors x and y in $\mathbb{G}$ and every $R \subseteq Range(M)$,*

$$\mathrm{P}\left(M(x) \in R\right) \leq e^{\varepsilon} \mathrm{P}\left(M(y) \in R\right).$$

The restatement of differential privacy makes it easy to see that if a k-sensitive neighborhood graph, G_S, is the same as $\mathbb{G}$ (i.e. $G_S = \mathbb{G}$), then a mechanism is (ε, k)-sensitively private if and only if it is ε-differentially private. One can easily confirm this by: (1) comparing Definition 7 with Definition 9, and (2) using the fact that $G_S = \mathbb{G}$.

Furthermore, if a mechanism is ε-differentially private then it is also (ε, k)-sensitively private for all $k \geq 1$ and normality properties. This follows from the fact that every k-sensitive neighborhood graph is a subgraph of DP neighborhood graph.

Protected differential privacy (PDP): In [13], the authors present a definition of privacy and private algorithms for a targeted search in social networks, which can be used to search for anomalies (e.g. terrorists) as well. Their algorithms are private under protected differential privacy, their proposed notion.

However, the definition of anomaly (or the normality property in our context) that can be considered under PDP cannot be data-dependent in a non-trivial manner. Namely, for a pre-specified $X \subseteq \mathcal{X}$ (the set of the protected population), we can define the corresponding normality property, p, where p is such that for every $i \in \mathcal{X}$ and $x \in \mathcal{D}$, $p(i,x) = 1 \iff i \in X$. Thus, protected differential privacy deals with a subclass of definitions from sensitive privacy. Furthermore, in this context, solving the problem for the data-dependent definition of anomalies is an open question [13] that sensitive privacy answers.

Tailored differential privacy: *Tailored differential privacy* [22] generalizes differential privacy, wherein the privacy parameter, ε, is a function of a record and a database (i.e. $\epsilon : \mathcal{X} \times \mathcal{D} \to \mathbb{R}_{\geq 0} \cup \{\infty\}$). Hence, it allows for having different levels of privacy (i.e. the value of ε) for different records.

Definition (TDP). For any given $\epsilon : \mathcal{X} \times \mathcal{D} \to \mathbb{R}_{\geq 0} \cup \{\infty\}$, we say a mechanism M is $\epsilon(\cdot)$-*tailored differentially private* if for every$R$ and $x \in \mathcal{D}$ and every $i \in \mathcal{X}$ such that $x_i \geq 1$,

$$\Pr[M(x) \in R] \leq e^{\epsilon(i,x)} \Pr[M(x - \mathbf{e}^i) \in R]$$

and

$$\Pr[M(x - \mathbf{e}^i) \in R] \leq e^{\epsilon(i,x)} \Pr[M(x) \in R]$$

Sensitive privacy deals with a subclass of mechanisms that are private under tailored differential privacy. If a mechanism is sensitively private then there exists a function ϵ such that the mechanism is also $\epsilon(\cdot)$-tailored differentially private. Alternatively, we can say that for a specific function ϵ, a mechanism is sensitively private if and only if it is tailored differentially private (Claim 3).

Next, we define the *privacy-functions*, ϵ_α, that we need to formally state Claim 3. For any given k-sensitive neighborhood graph G_S (for an arbitrary $p \in \mathfrak{P}$) and a fixed $\alpha > 0$, we say $\epsilon_\alpha : \mathcal{X} \times \mathcal{D} \to \mathbb{R}_{\geq 0} \cup \{\infty\}$ is a *privacy-function* if for every $i \in \mathcal{X}$ and every $x \in \mathcal{D}$, $\epsilon_\alpha(i,x) = \alpha \cdot d_{G_S}(x, \xi(x - \mathbf{e}^i))$ if x and $\xi(x - \mathbf{e}^i)$ are connected in G_S otherwise $\epsilon_\alpha(i,x) = \infty$.

Recall that ξ, for a given input, replaces the negative value of each coordinate with zero.

Claim 3. *Arbitrarily fix $k \geq 1$, $\alpha > 0$, and a normality property p, and let G_S be the k-sensitive neighborhood graph for p. If ϵ_α is the privacy-function for G_S (as given above) then for every mechanism M, M is (α, k)-sensitively private $\iff$ it is $\epsilon_\alpha(\cdot)$-tailored differentially private.*

Proof. Arbitrarily fix $k \geq 1$, a normality property along with the corresponding k-sensitive neighborhood graph, G_S, and $\alpha > 0$. Let ϵ_α be the privacy-function for G_S, and M be a mechanism with domain $\mathcal{D}$.

We first show the $\implies$ direction. Let M be (α, k)-sensitively private (SP). Arbitrarily pick $x \in \mathcal{D}$ and $i \in \mathcal{X}$ such that $x_i \geq 1$. Let $y = x - \mathbf{e}^i$. Thus, by definition of G_S, y is a node in G_S. Now, if x and y are connected in G_S, then for an arbitrarily fixed $R \subseteq Range(M)$, it follows that

$$\begin{aligned} \mathrm{P}\big(M(x) \in R\big) &\leq e^{\alpha d_{G_S}(x,y)} \mathrm{P}\big(M(y) \in R\big) \\ \mathrm{P}\big(M(x) \in R\big) &\leq e^{\epsilon_\alpha(i,x)} \mathrm{P}\big(M(y) \in R\big) \end{aligned}$$

The first inequality follows from the definition of sensitive privacy, whereas the second one follows from the definition of ϵ_α. Thus one of the constraints of tailored differential privacy (TDP) holds for x and i. Similarly, from symmetry of d_{G_S} and the second constraint imposed by SP, the other privacy constraint of TDP follows.

On the other hand, if x and y are not connected in G_S, the probabilities corresponding to x and y are allowed to arbitrarily diverge from each other. The same is the case for TDP under our ϵ_α, which is equal to ∞ here. Thus, in this case, the claim holds as well. Since x, i, and R were picked arbitrarily, the constraints hold for every x and i (such that $x_i \geq 1$) and R. Hence by the definition of TDP, M is $\epsilon_\alpha(\cdot)$-TDP.

Next, we prove the $\impliedby$ direction. Let M be $\epsilon_\alpha(\cdot)$-TDP. Arbitrary pick neighboring database x and y in G_S, and $R \subseteq Range(M)$. Since x and y are neighbors, there exists $i \in \mathcal{X}$ such that $||x - y||_1 = |x_i - y_i| = 1$; let us fix this i. Now, if $x_i > y_i$, then it follows that

$$\begin{aligned} \mathrm{P}(M(x) \in R) &\leq e^{\epsilon_\alpha(i,x)} \mathrm{P}(M(y) \in R) \\ \mathrm{P}(M(x) \in R) &\leq e^{\alpha d_{G_S}(x,y)} \mathrm{P}(M(y) \in R) \\ \mathrm{P}(M(x) \in R) &\leq e^{\alpha} \mathrm{P}(M(y) \in R) \end{aligned}$$

The first inequality holds because M is $\epsilon_\alpha(\cdot)$-TDP; the second one follows by the definition of ϵ_α, and the third one holds because $d_{G_S}(x, y) = 1$ for neighbors. Similarly, the second privacy constraint for SP follows from the other privacy constraint of TDP. In the case, $x_i < y_i$, we get $\mathrm{P}(M(x) \in R) \leq e^{\epsilon_\alpha(i,y)} \mathrm{P}(M(y) \in R)$; here again we can show, in a similar fashion as above, that both the privacy constraints for SP hold. Since we picked the neighboring databases and R arbitrarily, the privacy constraints for privacy parameter α and k-sensitive neighborhood graph G_S hold for all the neighboring databases and R. Hence, M is (α, k)-SP. This completes the proof. □

8 Sensitive Privacy and Mechanism Construction

How can we achieve sensitive privacy for identifying anomalies? In this section, we answer this question. We will first look at the simple case, where we look for $(\beta, r = 0)$-anomalies with the guarantee of sensitive privacy. To do this, we develop a sensitively private mechanism, named n-step lookahead mechanism, which works well for this case. However, when r is non-trivial (defined in Section 6), we show that n-step lookahead mechanism cannot achieve sensitive privacy.

We will then introduce a construction to give SP mechanism for AIQ. We will also show, how this construction can be used to give optimal mechanisms for AIQ, and develop SP as well as DP mechanisms to identify (β, r)-anomalies.

8.1 n-Step lookahead mechanism

Here, we give a sensitively private mechanism, named *n-step lookahead mechanism*, to compute a given query, f, for outlier analysis. For a fixed normality property p and $n \in \mathbb{N}$, n-step lookahead mechanism, $M_{p,n}$ adds noise to each coordinate i of the given database x if i is n-sensitive with respect to x, that is, if in n-steps (i.e. adding or removing n records) from x gives a database y such that $p(i, y) = 1$. It then computes f on the perturbed database.

n-Step lookahead mechanism, $M_{i,p,n}$:

```
1) Input: database $x$
2) For the fixed $i \in \mathcal{X}$:
3)     If $\texttt{isSensitive}(i, x, p, n) = True$:
4)         set $x = x + \texttt{Lap}(1/\varepsilon) \times \mathbf{e}^i$
5) Return $f(\xi(x))$.
```

In n-step lookahead mechanism, $f : \mathcal{D} \to \mathcal{R}$ denotes the query function for analyzing data for anomalies, e.g. an anomaly identification function — we assume f is computable. $\texttt{Lap}(1/\varepsilon)$ denotes an independent sample from Laplace distribution of mean 0 and scale $1/\varepsilon$. For given n and p (normality property) the mechanism $M_{i,p,n}$ also uses `isSensitive` function, which for given record i and database x returns $True$ if there is a y such that $||x - y||_1 \leq n$ and $p(i, y) = 1$, and $False$ otherwise.

For any $p_{[\beta,0]}$, the n-step lookahead mechanism $M_{i,p,n}$ is (ε, k)-SP if $n = k + 1$ (Claim 4). For this, `isSensitive` is simple and easily computable. For any $i, x, p_{[\beta,0]}$, and n, $\texttt{isSensitive}(i, x, p_{[\beta,0]}, n) = (x_i \geq \beta + 1 - n)$.

n-Step lookahead mechanism works well for $(\beta, 0)$-normality properties. For the practical settings (i.e. $\beta \geq k + 1$), $M_{i,p_{[\beta,0]},n}$ only has an additional linear (in size of the input database) computational overhead. Because, for any given database x, we only need to perturb the coordinates that have non-zero mass (i.e. $i \in \mathcal{X}$ such that $x_i \geq 1$) as others will not be k-sensitive.

Claim 4. *Arbitrarily fix an $\varepsilon > 0$, a query f, and let $\texttt{isSensitive}(i, x, p_{[\beta,0]}, n)$ be as given above. Then, for every $k \geq 1$ and $p_{[\beta,0]}$, $(k + 1)$-step lookahead mechanism is (ε, k)-sensitively private.*

Proof. Arbitrary fix $\mathcal{X}$ (a finite set), ε, and f as given above. Next, fix arbitrary values of $\beta \geq 1$ and $k \geq 1$. And let

$\texttt{isSensitive}(i, x, \mathfrak{p}_{[\beta,0]}, n) = (x_i \geq \beta + 1 - n)$ for every $i \in \mathcal{X}$ and $x \in \mathcal{D}$, and let $M_{\mathfrak{p}_{[\beta,0]},n}$ be the n-step lookahead mechanism for $n = k + 1$.

For the proof, we note that the mechanism $(M_{i,\mathfrak{p}_{[\beta,0]},n})$ perturbs the coordinates of the input database using Laplace distribution of mean zero and scale $1/\varepsilon$. Thus, the perturbed database it generates will be guaranteed to be sensitively private if, for every two neighbors x and y, (1) the mechanism perturbs the same set of coordinates $J \subseteq \mathcal{X}$ in x and in y, and (2) J is a super set of all coordinates that are either sensitive in one neighbor or the other.

We first prove (1). Arbitrarily fix an $i \in \mathcal{X}$ and two neighbors x and y in G_S such that $||x - y||_1 = |x_i - y_i| = 1$. Thus, all the coordinates of x and y except for i are the same, and $M_{i,\mathfrak{p}_{[\beta,0]},n}$ will perturb the same coordinates in x and y except for i. So, we next show that $M_{i,\mathfrak{p}_{[\beta,0]},n}$ perturbs the coordinate i for both x and y.

When i is k-sensitive w.r.t. to both the databases, the mechanism perturbs the i coordinate of both x and y. Because when i is k-sensitive w.r.t. to $z \in \mathcal{D}$, $B_z(i, r = 0) = z_i \geq \beta + 1 - k \geq \beta + 1 - n$ (from Lemma 1 and $n = k + 1$).

Thus, without loss of generality, let i be k-sensitive w.r.t x, but not w.r.t y. Since i is k-sensitive w.r.t x, $x_i \geq \beta + 1 - k$ and $y_i \geq \beta - k$ (as $|x_i - y_i| = 1$). In this case, $M_{i,\mathfrak{p}_{[\beta,0]},n}$ will perturb the coordinate i in both x and y.

Lastly, note that, for every database z and every coordinate j, $M_{i,\mathfrak{p}_{[\beta,0]},n}$ perturbs the coordinate j if $z_j \geq \beta - k$. Hence, the set of coordinate that $M_{i,\mathfrak{p}_{[\beta,0]},n}$ perturbs for both x and y is a super set of the coordinates that are sensitive either with respect to x or y. Since, the x, y, i were chosen arbitrarily, the claim holds for all the neighbors in G_S.

Now, given that the perturbed database is guaranteed to be sensitively private, from post-processing property, we conclude the claim holds. This completes the proof. □

Limitations of the look-ahead mechanism

We show that the n-step lookahead mechanism does not work for (β, r)-normality properties in general; especially, when the normality property is regular. That is we show that for a regular (β, r)-normality properties, it is impossible for n-step lookahead mechanism to achieve (ε, k)-SP when $n, k \leq \beta$ (Theorem 1).

In Theorem 1, we only consider $n, k \leq \beta$. Since for every $k > \beta$, every $i \in \mathcal{X}$ is k-sensitive with respect to every database $x \in \mathcal{D}$. Thus, in such a case, the k-sensitive neighborhood graph (G_S) is the same as DP neighborhood graph (i.e. $G_S = \mathbb{G}$). And using sensitive privacy instead of differential privacy will not result in any gain in utility.

Theorem 1. *Arbitrarily fix $\varepsilon > 0$, finite set $\mathcal{X}$, distance function $d_{\mathcal{X}}$, regular $\mathfrak{p}_{[\beta,r]}$, and let g be the (β, r)-anomaly identification function. If there exist $i, j \in \mathcal{X}$ such that $d_{\mathcal{X}}(i, j) > 2r$ then for every k and n such that $1 \leq k, n \leq \beta$, n-step lookahead mechanism for (β, r)-AIQ, (j, g), is* not *(ε, k)-sensitively private.*

Proof. Arbitrarily fix $\mathcal{X}$ (a finite set), $d_{\mathcal{X}}$, and $\mathfrak{p}_{[\beta,r]}$ such that (1) $\mathfrak{p}_{[\beta,r]}$ is regular and (2) $i, j \in \mathcal{X}$ such that $d_{\mathcal{X}}(i, j) > 2r$. Choose an arbitrary value of $\varepsilon > 0$, and let g as given above. Next, fix an arbitrary value of k such that $1 \leq k \leq \beta$, and let G_S be the k-sensitive neighborhood graph for $\mathfrak{p}_{[\beta,r]}$.

To prove the claim, we consider two databases that are connected in G_S. Note that there exist x and y in $\mathcal{D}$ such that $x_i = x_j = \beta$, $||x||_1 = 2\beta$, and $y_i = \beta$, $||y||_1 = \beta$. Thus, i and j both are k-sensitive w.r.t. x, while i is k-sensitive w.r.t. y but j is not (Lemma 1). Since $\mathfrak{p}_{[\beta,r]}$ is regular, $x, y \in D^{\mathfrak{p}_{[\beta,r]}}$ are connected (by Definition 8).

Now, let (j, g) be the (β, r)-AIQ, and let $f = g_j$. Note that $f(x) = 1$ and $f(y) = 0$. Next, fix any n such that $1 \leq n \leq \beta$, and let $M_{\mathfrak{p}_{[\beta,r]},n}$ be the n-step lookahead mechanism for f.

Here, j is n-sensitive w.r.t. to x but not w.r.t. y because $B_y(j, r) + n \leq \beta$ (Lemma 1). So, $M_{\mathfrak{p}_{[\beta,r]},n}$ will perturb x_j but not y_j, and hence, $\mathrm{P}\big(M_{\mathfrak{p}_{[\beta,r]},n}(x) \in \{1\}\big) > 0$ but $\mathrm{P}\big(M_{\mathfrak{p}_{[\beta,r]},n}(y) \in \{1\}\big) = 0$. Now, from Lemma 2, we conclude that $M_{\mathfrak{p}_{[\beta,r]},n}$ is not (ε, k)-SP. Since, n and k were fixed arbitrarily the claim follows. □

Lemma 2. *Arbitrarily fix $\varepsilon > 0$, $k \geq 1$, and $p \in \mathfrak{P}$, and let G_S be the k-sensitive neighborhood graph for p. Now, for any mechanism $M : \mathcal{D} \to \mathcal{R}$ that is (ε, k)-SP, it holds that for every $x, y \in \mathcal{D}$ that are connected in G_S and every $R \subseteq \mathcal{R}$,*

$$P\big(M(x) \in R\big) = 0 \iff P\big(M(y) \in R\big) = 0$$

Proof sketch. We prove the claim by contradiction. Assume $M : \mathcal{D} \to \mathcal{R}$ is an (ε, k)-SP mechanism. Fix any $x, y \in \mathcal{D}$ and $R \subseteq \mathcal{R}$ such that for some $\ell \in \mathbb{R}$ such that $\ell > 0$, $d_{G_S}(x, y) = \ell$, and $\mathrm{P}\big(M(y) \in R\big) > \mathrm{P}\big(M(x) \in R\big) = 0$. Since M is (ε, k)-SP, $\mathrm{P}\big(M(y) \in R\big) \leq e^{\varepsilon\ell}\mathrm{P}\big(M(x) \in R\big)$ must hold. But there are no $\ell, \varepsilon \in \mathbb{R}_{>0}$ that satisfy the above constraint. This implies M is not sensitively private, which contradicts our assumption. One can prove the other direction using the other privacy constraint in a similar way. Thus, the lemma follows. □

8.2 SP Mechanism Construction for AIQ

We now present our construction (Construction 1) to develop sensitively private mechanism for AIQ. We will also show how to use Construction 1 to develop an optimal sensitively private mechanism as well as a differentially private mechanism. Finally, we will instantiate Construction 1 for (β, r)-anomaly [29] (a prevalent anomaly model). Although we instantiate the construction for (β, r)-anomaly, it is not tied to any specific anomaly definition or model, and hence, is generally applicable for other anomaly models.

We define the notion of *minimum discrepant distance* (mdd) over a sensitive neighborhood graph (G_S), which plays the central role in our construction. Roughly speaking, for given G_S and a function f, mdd of a database x is the distance to the closest point y where f changes the value it has on x, i.e. $f(x) \neq f(y)$. To measure mdd for anomaly identification, we define the mdd-function (Definition 10).

Construction 1 uses mdd-function to give an SP mechanism for an arbitrary AIQ, which has very high accuracy in practice. For instance, the SP mechanism for (β, r)-AIQ errs with exponentially small probability on most of the typical inputs (Theorem 5).

Definition 10 (mdd-function)**.** *Let G_S be a sensitive neighborhood graph (for an arbitrary $k \geq 1$ and a normality property). Then, for any anomaly identification function g, we say a function, $\Delta_{G_S} : \mathcal{X} \times \mathcal{D} \to \mathbb{N} \cup \{\bot\}$, is the* minimum discrepant

distance function *for g, if for every $i \in \mathcal{X}$ and database $x \in \mathcal{D}$, the following holds:*
If there is $z \in \mathcal{D}$ such that $z \sim x$ (i.e. connected to x in G_S) and $g_i(z) \neq g_i(x)$ then

$$\Delta_{G_S}(i,x) = \min_{\substack{z \in \mathcal{D}:\ z \sim x \text{ and} \\ g_i(z) \neq g_i(x)}} d_{G_S}(x,z)$$

otherwise, $\Delta_{G_S}(i,x) = \perp$.

A simple and efficient mechanism for anomaly identification — which is both accurate and sensitively private — can be given if g and the corresponding mdd-function, Δ_{G_S}, can be computed efficiently.

However, for an arbitrary normality property, computing the mdd-function efficiently is a conceptually non-trivial task, one that we conjecture it cannot be done efficiently. This is because the metric, d_{G_S}, which gives rise to the metric-based property captured by the mdd-function, is induced by (a) the definition of anomaly (e.g. specific values of β and r as we saw in Section 6) and (b) the distance function over the records. Thus, making it exceedingly difficult to analyze it in general.

Therefore, in the next section, we present our constructions that instead uses a lower bound on the mdd-function to give sensitively private mechanism.

8.3 SP-mechanism via lower bounding mdd

Construction 1 uses a lower bound, λ (a function over $\mathcal{X} \times \mathcal{D}$), for the minimum discrepant distance (mdd). We parameterize Construction 1 by λ, which is associated with a sensitive neighborhood graph. Since the sensitive neighborhood graph is defined for an anomaly definition, hence, the graph becomes concrete for each concrete anomaly definition (e.g. see Section 8.4).

Now, for any fixed AIQ, (i, g), and given λ, Construction 1 provably gives an SP mechanism as long as λ is an *acceptable lower bound* on the mdd-function (Theorem 2).

Below, we define the notion of acceptable lower bound on an mdd-function. Arbitrarily fix a neighborhood graph G and an anomaly identification function g (for the definitions given below). We say $\lambda : \mathcal{X} \times \mathcal{D} \to \mathbb{R}_{\geq 0} \cup \{\perp\}$ is a ***lower bound*** on the mdd-function Δ_G for g if for every $i \in \mathcal{X}$ and $x \in \mathcal{D}$, the following holds: if $\Delta_G(i,x) = \perp$, then $\lambda(i,x) = \perp$ or $\lambda(i,x) \in \mathbb{R}_{\geq 0}$, otherwise $\lambda(i,x) \leq \Delta_G(i,x)$. For any given $\alpha > 0$, we say λ is α-***acceptable*** if: (1) for every i and x, if $\lambda(i,x) \in \mathbb{R}_{\geq 0}$, then $\lambda(i,x) \geq 1$, and (2) λ is α-Lipschitz continuous over G (defined in Section 3). Finally, for any given mdd-function, Δ_G, and $\alpha \geq 1$, we say λ is an α-***acceptable lower bound*** on Δ_G, if it is α-acceptable and a lower bound on Δ_G.

Remark: Although at first it appears that the *Lipschitz continuity condition* is some side technicality, in fact bounding its value constitute the main part of our argument for privacy of our mechanisms. Thus giving an SP mechanism for (i, g) via Construction 1 reduces to giving a Lipschitz continuous lower bound for the mdd-function corresponding to g.

Construction 1. U_λ

```
1) Input $x \in \mathcal{D}$
2) If $\lambda(i,x) = \perp$, set $t = 0$
3) Else, set $t = e^{-\varepsilon(\lambda(i,x)-1)}/(1+e^{\varepsilon})$
4) Sample $b$ from $\{0,1\}$ such that $\mathrm{P}(b \neq g(i,x)) = t$
5) Return $b$
```

Note that as it is common in the privacy literature, the notation for sampling with probability $P(b \neq g(i,x)) = t$, in practice means to sample with probability exponentially close to t (within error $1/2^n$ for input length parameter n). Also, note that the above is a family of constructions parameterized by λ, i.e. one construction, U_λ, for each λ. This construction is very efficiently realizable as long as we can efficiently compute g and λ. Furthermore, the error of the mechanism, yielded by the construction, for any input is exponentially small in λ (Theorem 2).

Theorem 2 (U_λ is SP)**.** *Arbitrarily fix $\varepsilon > 0$, $k, \alpha \geq 1$, and a normality property p. Let G_S be the k-sensitive neighborhood graph for p, and (i, g) be an arbitrary AIQ, where g and p are for the same anomaly definition.*
For every $\lambda : \mathcal{X} \times \mathcal{D} \to \mathbb{R}_{\geq 0} \cup \{\perp\}$, if λ is an α-acceptable lower bound on Δ_{G_S} for g, then Construction 1 yields an $(\varepsilon\alpha, k)$-SP mechanism, U_λ, such that for every $x \in \mathcal{D}$ and $\lambda(i,x) \in \mathbb{R}_{\geq 0}$,

$$P(U_\lambda(x) \neq g(i,x)) = e^{-\varepsilon(\lambda(i,x)-1)}/(1+e^{\varepsilon})$$

Proof of Theorem 2. Arbitrarily fix $\varepsilon > 0$, $k \geq 1$, $\alpha \geq 1$, and a normality property p. Let G_S be the k-sensitive neighborhood graph for p, and g be the anomaly identification function for the anomaly definition for p.

Fix λ to be an α-acceptable lower bound on the mdd-function, Δ_{G_S}, for g. Let U_λ be as given by Construction 1. Next, arbitrarily fix an AIQ, (i, g), and $x, y \in \mathcal{D}$ that are neighbors in G_S (i.e. $d_{G_S}(x,y) = 1$).

We will show that the privacy constraint for both the outputs 0 and 1 are satisfied by U_λ.

If $\lambda(i,x) = \perp$ then $\lambda(i,y) = \perp$ (and vice versa) because λ is Lipschitz continuous (follows from λ being α-acceptable). Since λ is a lower bound on Δ_{G_S}, $\lambda(i,x) = \perp$ implies that either (a) there is no z in G_S such that $g(i,z) \neq g(i,x)$, or (b) every z' that is connected with x is such that $g(i,z') = g(i,x)$. In both cases we get that $g(i,x) = g(i,y)$ as there is no z connected to x, and hence to y, such that $g(i,z) \neq g(i,x)$. And in this case, the privacy constraint hold trivially. Therefore we will assume that $\lambda(i,x), \lambda(i,y) \neq \perp$.

Firstly, consider the case, when $g(i,x) = g(i,y) = b$ for some $b \in \{0,1\}$. Here, from the α-Lipschitz continuity the following holds.

$$\frac{\mathrm{P}(U_\lambda(x) = 1-b)}{\mathrm{P}(U_\lambda(y) = 1-b)} = e^{\varepsilon(-\lambda(i,x)+\lambda(i,y))} \leq e^{\alpha\varepsilon}$$

Now for the other constraint, we have the following:

$$\begin{aligned}\frac{\mathrm{P}(U_\lambda(x) = b)}{\mathrm{P}(U_\lambda(y) = b)} &= \frac{1-\mathrm{P}(U_\lambda(x) = 1-b)}{1-\mathrm{P}(U_\lambda(y) = 1-b)} \\ &= \frac{1+e^{\varepsilon}-e^{-\varepsilon(\lambda(i,x)-1)}}{1+e^{\varepsilon}-e^{-\varepsilon(\lambda(i,y)-1)}} \qquad (1)\end{aligned}$$

Next, we show that $e^{\varepsilon\alpha}$ is indeed an upper bound for the expression given in (1). Since $\varepsilon > 0$ and $\lambda(i,y)$ and α are at

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2021.3129633, IEEE Transactions on Knowledge and Data Engineering

least 1, we get the following:

$$
\begin{aligned}
& e^{\varepsilon}(e^{\varepsilon\alpha}+1) \leq e^{\varepsilon(\alpha+\lambda(i,y))}(1+e^{\varepsilon}) \\
\iff & e^{\varepsilon}(e^{2\varepsilon\alpha}-1) \leq e^{\varepsilon(\alpha+\lambda(i,y))}(1+e^{\varepsilon})(e^{\varepsilon\alpha}-1) \\
\iff & e^{\varepsilon}(e^{\varepsilon\alpha}-e^{-\varepsilon\alpha}) \leq e^{\varepsilon\lambda(i,y)}(1+e^{\varepsilon})(e^{\varepsilon\alpha}-1) \\
\iff & e^{\varepsilon\lambda(i,y)}(1+e^{\varepsilon})-e^{\varepsilon}e^{-\varepsilon\alpha} \leq e^{\varepsilon\alpha}\left[e^{\varepsilon\lambda(i,y)}(1+e^{\varepsilon})-e^{\varepsilon}\right] \\
\iff & \frac{e^{\varepsilon\lambda(i,y)}(1+e^{\varepsilon})-e^{\varepsilon}e^{-\varepsilon\alpha}}{e^{\varepsilon\lambda(i,y)}(1+e^{\varepsilon})-e^{\varepsilon}} \leq e^{\varepsilon\alpha} \\
& \text{because } -\alpha \leq \lambda(i,y)-\lambda(i,x)\text{, the following holds} \\
& \frac{e^{\varepsilon\lambda(i,y)}(1+e^{\varepsilon})-e^{\varepsilon}e^{\varepsilon(\lambda(i,y)-\lambda(i,x))}}{e^{\varepsilon\lambda(i,y)}(1+e^{\varepsilon})-e^{\varepsilon}} \leq e^{\varepsilon\alpha} \\
\iff & \frac{1+e^{\varepsilon}-e^{-\varepsilon(\lambda(i,x)-1)}}{1+e^{\varepsilon}-e^{-\varepsilon(\lambda(i,y)-1)}} \leq e^{\varepsilon\alpha}
\end{aligned}
$$

Now consider the case of $g(i,x) \neq g(i,y)$. Here, $\lambda(i,x) = \lambda(i,y) = 1$. To confirm this, recall that $d_{G_S}(x,y) = 1$, and $\Delta_{G_S}(j,z) \geq \lambda(j,z) \geq 1$ for every $j \in \mathcal{X}$ and $z \in \mathcal{D}$. Thus, $\Delta_{G_S}(i,x) = \Delta_{G_S}(i,y) = 1$ implies $\lambda(i,x) = \lambda(i,y) = 1$. Therefore, the privacy constraints hold trivially for this case. Since, x and y were picked arbitrarily, the above shows that all the privacy constraints hold for all the neighbors. This concludes the formal argument. □

Optimal SP mechanism via Construction 1

For $\lambda = \Delta_{G_S}$, Construction 1 gives a pareto optimal sensitively private mechanism (Theorem 3).

For arbitrarily fixed $\varepsilon > 0$, $k \geq 1$, and a normality property p, we say a mechanism U is ***pareto optimal*** (ε,k)-sensitively private if (1) it is (ε,k)-SP and (2) for every (ε,k)-SP mechanism $M : \mathcal{D} \to \{0,1\}$ and every database $x \in \mathcal{D}$, $P(U(x) = g_i(x)) \geq P(M(x) = g_i(x))$. Particularly, this implies that of all the SP mechanisms yielded by Construction 1, each corresponding to a different λ, *the "best" mechanism is for* $\lambda = \Delta_{G_S}$.

Theorem 3. *Arbitrarily fix $\varepsilon > 0$, $k \geq 1$, and a normality property p. Let G_S be the k-sensitive neighborhood graph for p, and (i,g) be an arbitrary AIQ, where g and p are for the same anomaly definition.*

If Δ_{G_S} is the mdd-function for g, then $U_{\Delta_{G_S}}$ (Construction 1) is pareto optimal (ε,k)-sensitively private.

Proof. Let ε, k, p, g, and G_S be as given above. Arbitrarily fix $i \in \mathcal{X}$, and let Δ_{G_S} be the mdd-function for g, and $U_{\Delta_{G_S}}$ be as given by Construction 1.

Firstly, note that $U_{\lambda=\Delta_{G_S}}$ is (ε,k)-SP. This follows from Theorem 2 and the fact that, for $\lambda = \Delta_{G_S}$, λ is 1-acceptable lower bound on Δ_{G_S} (Lemma 3 since G_S is a neighborhood graph).

Next, we prove the optimality claim by contradiction. Assume that $U_{\Delta_{G_S}}$ is not pareto optimal. That is, there exits an (ε,k)-SP mechanism M (for p) such that

- for every x, $P\big(M(x) = g_i(x)\big) \geq P\big(U_{\Delta_{G_S}}(x) = g_i(x)\big)$ and
- for a database y, $P\big(M(y) = g_i(y)\big) > P\big(U_{\Delta_{G_S}}(y) = g_i(y)\big)$

Let us fix the y given above. Using y and our assumption about it, we show that there is an input database z, where M does worse than $U_{\Delta_{G_S}}$.

Let z be a database such that $d_{G_S}(y,z) = \Delta_{G_S}(i,y)$ and $g_i(z) \neq g_i(y)$. Note that if there is no such z, then $\Delta_{G_S}(i,y) = \perp$. In this case, our assumption about the database y, and hence, M cannot hold because, in this case, $P\big(U_{\Delta_{G_S}}(y) = g_i(y)\big) = 1$. Furthermore, similar would be the case if z is not connected to y (i.e. $\Delta_{G_S}(i,y) = \perp$). Thus, if the assumptions about M holds, then $d_{G_S}(y,z) \in \mathbb{N}$ (i.e. y and z are connected).

If we let w be a neighbor of z on the shortest path from y to z such that $g_i(w) \neq g_i(z)$, then $g_i(w) = g_i(y)$ and $d_{G_S}(y,w) = \Delta_{G_S}(i,y) - 1$. Since M (ε,k)-SP, for $b = g_i(w)$, it follows that

$$
\begin{aligned}
P\big(M(w) \neq b\big) \leq & e^{\varepsilon d_{G_S}(y,w)} P\big(M(y) \neq b\big) \\
= & e^{\varepsilon(\Delta_{G_S}(i,y)-1)} P\big(M(y) \neq b\big) \\
< & e^{\varepsilon(\Delta_{G_S}(i,y)-1)} P\big(U_{\Delta_{G_S}}(y) \neq b\big) \\
= & 1/(1+e^{\varepsilon})
\end{aligned}
\tag{2}
$$

The first inequality is due to the SP constrains on M. The second inequality is due to the fact that M is strictly better than $U_{\Delta_{G_S}}$ on y. The last equality holds because $P\big(U_{\Delta_{G_S}}(y) \neq g_i(y)\big) = e^{-\varepsilon(\Delta_{G_S}(i,y)-1)}/(1+e^{\varepsilon})$ (follows from Construction 1).

Since M is assumed to (ε,k)-SP, we get the following:

$$
\begin{aligned}
P\big(M(z) \neq 1-b\big) \geq & e^{-\varepsilon} P\big(M(w) \neq 1-b\big) \\
= & e^{-\varepsilon} P\big(M(w) = b\big) \\
= & e^{-\varepsilon}\big(1 - P\big(M(w) \neq b\big)\big) \\
> & 1/(1+e^{\varepsilon})
\end{aligned}
\tag{3}
$$

The first inequality is due to M being SP. The first equality is due to the fact that there are only two possible outputs. The last inequality holds because $P\big(M(w) \neq b\big) < 1/(1+e^{\varepsilon})$ (which follows from the inequality given by (2)).

Finally, for $b = g_i(w)$, it follows that

$$
\begin{aligned}
P\big(M(z) = g_i(z)\big) = & P\big(M(z) = 1-b\big) \\
= & 1 - P\big(M(z) \neq 1-b\big) \\
< & e^{\varepsilon}/(1+e^{\varepsilon}) \\
= & P\big(U_{\Delta_{G_S}}(z) = g_i(z)\big)
\end{aligned}
$$

The first equality is due to the fact that $b = g_i(w) \neq g_i(z)$. The first inequality is due to the inequality given by (3). The last equality holds due to the following facts: (1) $P\big(U_{\Delta_{G_S}}(z) \neq g_i(z)\big) = e^{-\varepsilon(\Delta_{G_S}(i,z)-1)}/(1+e^{\varepsilon})$, and (2) $\Delta_{G_S}(i,z) = 1$ because $d_{G_S}(z,w) = 1$ and $g_i(z) \neq g_i(w)$.

From the above, we reach a conclusion that contradicts our assumption that M is strictly "better" than $U_{\Delta_{G_S}}$. Thus, we conclude the $U_{\Delta_{G_S}}$ is pareto optimal. □

Lemma 3. *Arbitrarily fix a neighborhood graph G and an anomaly identification function g, and let Δ_G be the mdd-function for g. Then for $\lambda = \Delta_G$, λ is 1-acceptable lower bound on Δ_G.*

Proof. Let G, g, Δ_G, and λ be as given above. We show that λ is 1-acceptable and is a lower bound on Δ_G.

To show that λ is 1-acceptable, we first prove that Δ_G is 1-Lipschitz continuous. For this, arbitrarily fix an $i \in \mathcal{X}$ and two neighbors x and y in G, i.e. $d_G(x,y) = 1$.

Firstly, consider the case when $\Delta_G(i,x) = \perp$ (this is without loss of generality as x and y were picked arbitrarily).

$\Delta_G(i,x) = \perp$ implies that either (a) there is no z in G such that $g(i,z) \neq g(i,x)$, or (b) every z' that is connected with x is such that $g(i,z') = g(i,x)$. In both the scenario, we get that $g(i,x) = g(i,y)$ as there is no z connected to x, and hence to y, such that $g(i,z) \neq g(i,x)$; thus, $\Delta_G(i,y) = \perp$. Hence, for this case, the Lipschitz continuity constraints hold.

Next, we consider the case where $\Delta_G(i,x), \Delta_G(i,y) \in \mathbb{N}$ (i.e. there is a z connected to x, and hence to y, such that $g(i,z) \neq g(i,x)$). Let $\mathcal{D}^x \subseteq \mathcal{D}$ be such that every $z \in \mathcal{D}$ is connected to x in G_S.
Now, note that by triangular inequality we get that for every database $z \in \mathcal{D}^x$, $d_G(x,z) \leq d_G(x,y) + d_G(y,z) = 1 + d_G(y,z)$. Thus, $\Delta_G(i,x) = \min_{z \in \mathcal{D}^x : g_i(z) \neq g_i(x)} d_G(x,z) \leq 1 + \Delta_G(i,y)$. Since x and y were chosen arbitrarily, swapping x and y gives $\Delta_G(i,y) \leq 1 + \Delta_G(i,x)$. Thus, 1-Lipschitz continuity constrains holds for x and y. Because we arbitrarily picked, i, and neighbors x, and y, the claim holds for all the neighbors and every $i \in \mathcal{X}$. Hence, Δ_G is 1-Lipschitz continuous.

Next, we show that if, for any i and x, $\Delta_G(i,x) \in \mathbb{N}$, then $\Delta_G(i,x) \geq 1$. For this, arbitrarily fix $i \in \mathcal{X}$ and x such that $\Delta_G(i,x) \in \mathbb{N}$. This implies that there is a database z at distance $d_G(x,y)$ from x such that $g(i,z) \neq g(i,x)$. For this to hold, $z \neq x$, and hence, $d_G(z,x) \geq 1$. Thus, by definition of mdd-function, $\Delta_G(i,x) \geq 1$. Since i and x were picked arbitrarily, the claim holds. Thus, the above shows that Δ_G is 1-acceptable.

Since for every i and x, $\lambda(i,x) = \Delta_G(i,x)$, λ is indeed a lower bound on Δ_G. This completes the proof. □

DP mechanism via Construction 1

For Construction 1, if we use a λ that is an α-acceptable lower bound on the mdd-function of the DP neighborhood graph, $\Delta_{\mathbb{G}}$, then the construction yields a differentially private mechanism for AIQ (Corollary 1).

Corollary 1. *Arbitrarily fix $\varepsilon > 0$, $\alpha \geq 1$, and an AIQ (i,g). For every λ such that it is α-acceptable lower bound on $\Delta_{\mathbb{G}}$ for g, U_λ (given by Construction 1) is an $\varepsilon\alpha$-differentially private mechanism.*

To confirm above claim, note that from Definition 7 (of sensitive privacy) and Definition 9, it follows that differential privacy is a special case of sensitive privacy, when the k-sensitive neighborhood graphs, G_S, is the same as neighborhood graph, $\mathbb{G}$, i.e., $G_S = \mathbb{G}$ (for more details see Section 7). Thus, for $G_S = \mathbb{G}$, a mechanism is ε-differentially private if and only if it is ε-sensitively private. Hence, Corollary 1 follows from Theorem 2.

8.4 (β, r)-anomaly identification via Construction 1

Here, we first use Construction 1 to give an optimal differentially private (DP) mechanism for (β,r)-AIQ (Theorem 4). We do this for two reasons. First, we will use this optimal DP mechanism as a baseline to compare our SP mechanisms' performance. Second, and importantly, we need $\Delta_{\mathbb{G}}$ (the mdd-function for $\mathbb{G}$ and (β,r)-anomaly identification function), which we use to give the optimal DP mechanism, to define a lower bound for the mdd-function for k-sensitive neighborhood graph (for $\mathscr{p}_{[\beta,r]}$) and (β,r)-anomaly identification function. We use this lower bound to develop SP mechanism for (β,r)-AIQ.

Optimal ε-DP mechanism for (β, r)-AIQ

Construction 1, for $\lambda = \Delta_{\mathbb{G}}$, gives a pareto optimal DP mechanism for (β,r)-AIQ, where $\Delta_{\mathbb{G}}$ (mdd-function) is given by (4) (Theorem 4).
$\Delta_{\mathbb{G}}$, provided below, is for arbitrary (β,r)-anomaly identification function with arbitrary values for $\beta \geq 1$, $r \geq 0$, $i \in \mathcal{X}$, and $x \in \mathcal{D}$ — which is indeed the mdd-function for (β,r)-AIQ [11]. Recall that $B_x(i,r) = \sum_{j \in \mathcal{X} : d_{\mathcal{X}}(i,j) \leq r} x_j$.

$$\Delta_{\mathbb{G}}(i,x) = \begin{cases} 1 & x_i = 0 \wedge B_x(i,r) < \beta \\ 2 + B_x(i,r) - \beta & x_i = 0 \wedge B_x(i,r) \geq \beta \\ \min\left(x_i, \beta + 1 - B_x(i,r)\right) & x_i \geq 1 \wedge B_x(i,r) \leq \beta \\ B_x(i,r) - \beta & x_i \geq 1 \wedge B_x(i,r) > \beta \end{cases} \tag{4}$$

Theorem 4 ($U_{\Delta_{\mathbb{G}}}$ is optimal and DP)**.** *Arbitrarily fix $\varepsilon > 0$ and (β,r)-anomaly identification function g, and let $\Delta_{\mathbb{G}}$ be as given by (4) for g. Then, for any fixed $i \in \mathcal{X}$, $U_{\Delta_{\mathbb{G}}}$ (Construction 1) for (β,r)-AIQ, (i,g), is pareto optimal ε-DP mechanism.*

Since $\Delta_{\mathbb{G}}$ is the mdd-function for (β,r)-anomaly identification function, the claim that the $U_{\Delta_{\mathbb{G}}}$ is differentially private follows from Lemma 3 and Corollary 1. And for $G_S = \mathbb{G}$, Theorem 3 establishes the optimality claim of $U_{\Delta_{\mathbb{G}}}$.

(ε, k)-SP mechanism for (β, r)-AIQ

Below, we have provided λ_k, a $(1,1)$-acceptable lower bound on the mdd-function for the k-sensitive neighborhood graph for (β,r)-normality property [11]. For this λ_k, Construction 1 yields (ε,k)-SP mechanism, U_{λ_k}, for (β,r)-AIQ such that, for non-sensitive records, U_{λ_k} can have exponentially small error in β (Theorem 5). Below, we give λ_k for arbitrary $k, \beta \geq 1, r \geq 0, i \in \mathcal{X}$, and $x \in \mathcal{D}$.

$$\lambda_k(i,x) = \begin{cases} \Delta_{\mathbb{G}}(i,x) & B_x(i,r) \geq \beta + 1 - k \\ \begin{array}{l} \beta + 1 - B_x(i,r) \\ + \min(0, x_i - k) \end{array} & B_x(i,r) < \beta + 1 - k \end{cases} \tag{5}$$

It is clear from the definition of λ_k (given by (5)) that when a record, i, is k-sensitive with respect to x, $\lambda_k(i,x) = \Delta_{\mathbb{G}}(i,x)$, which implies that there is no gain in utility (i.e. accuracy) compared to the pareto optimal DP mechanism. However, when a record is not sensitive, $\lambda(i,x) > \Delta_{\mathbb{G}}(i,x)$, our SP mechanism achieves much higher utility compared to the optimal DP mechanism, which is especially true for the records that are (β,r)-anomalous with a higher degree of outlyingness, for example, the records that lie in a very sparse region.

Theorem 5 (accuracy and privacy of U_{λ_k})**.** *Arbitrarily fix $\varepsilon > 0$, $k \geq 1$, and a (β,r)-AIQ, (i,g). Let λ_k be as given by (5) and G_S be the k-sensitive neighborhood graph for (β,r)-normality property. Then, the mechanism, U_{λ_k} (Construction 1)*

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2021.3129633, IEEE Transactions on Knowledge and Data Engineering

is (ε, k)-SP such that for every $i \in \mathcal{X}$ and $x \in \mathcal{D}$ if i not k-sensitive for x, then

$$P(U_{\lambda_k}(x) \neq g(i,x)) \leq e^{-\varepsilon|\beta+1-k-B_x(i,r)|}$$

For the λ_k, the privacy claim follows from Theorem 2, while the error bound follows from Theorem 2 and the definition of λ_k — note that $B_x(i,r) < \beta+1-k$ implies that i is not sensitive for x (Lemma 1).

The theorem suggests that the error of the SP mechanism, U_{λ_k}, will be exponentially small with respect to $\varepsilon|\beta-k|$ in the typical settings, where an outlier has a very small number of records nearby, i.e., $B_x(i,r)$ is very small. This will be further confirmed in the empirical evaluation in the next section. We give an example to show that U_{λ_k} achieves high accuracy in typical settings. Fix $k \leq \beta/10$. Now for any record i in a database x, satisfying $B_x(i,r) \leq \beta/2$ is an outlier for which U_{λ_k} *will err with probability less that* $e^{-2\varepsilon\beta/5}$.

9 Empirical Evaluation

We empirically evaluated the performance of the SP and DP mechanisms for (β, r)-AIQ (given in Section 8.4) over synthetic as well as real-world datasets from diverse domains: Credit Fraud [33] (available at Kaggle [8]), Mammography and Thyroid (available at Outlier Detection DataSets Library [34]), and APS Trucks (APS Failure at Scania Trucks, available at UCI machine learning repository [35]). We also compared the performance of our SP mechanism with that of the pareto optimal DP mechanism for (β, r)-AIQ. Table 1 provides the datasets specifications.

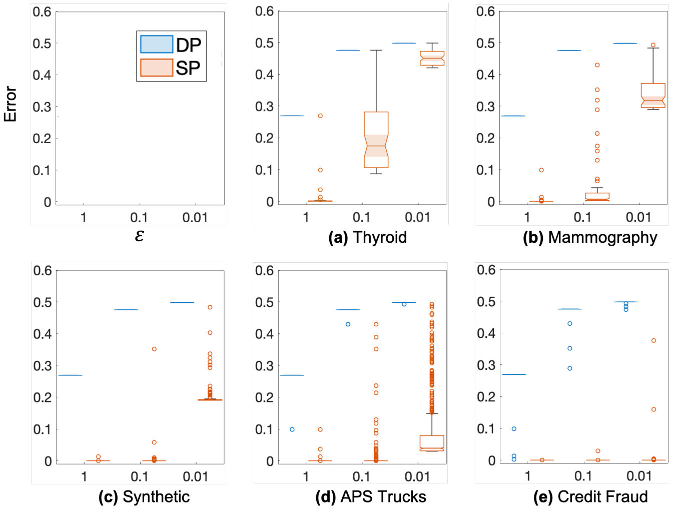
We used synthetic data to evaluate the performance in the high-dimensional case. To generate the synthetic data with outliers we used a mixed and 200-dimensional Gaussian distribution. For this, we followed the strategy of Dong et al. [36], which is the standard practice in the literature and is described in [11].

We emphasize that the aim of this work is to study the effect of achieving privacy in identifying anomalies. So we focus on evaluating the proposed approach for achieving privacy for this problem, and how it compares to differential privacy in real world settings.

For the higher dimensional datasets, we first performed dimension reduction using PCA (i.e., principal component analysis, which is the standard practice for identifying outliers in high-dimensional data [31], [37]). From the output of PCA, we chose, top $6, 9$, and 12 features for the Credit Fraud, Synthetic, and APS Trucks datasets respectively. Next, we obtain the values of β and r by using the protocol outlined in [11] (i.e., typically these values are provided by the domain experts [29]). Table 1 gives the values of β and r, along with the number of true (β, r)-anomalies (true anomalies identifiable by (β, r)-anomaly method for the given parameter values).

Dataset	**size**	**dim**	$(\boldsymbol{\beta}, \boldsymbol{r})$	**true** (β, r)-**anomalies**
Credit Fraud	$284,807$	28	$(1022, 6.7)$	103
APS Trucks	$60,000$	170	$(282, 16.2)$	677
Synthetic	$20,000$	200	$(97, 3.8)$	201
Mammography	$11,183$	6	$(55, 1.7)$	75
Thyroid	$3,772$	6	$(18, 0.1)$	61

TABLE 1: **Dataset specifications and parameter values.**



(a) Thyroid

(b) Mammography

(c) Synthetic

(d) APS Trucks

(e) Credit Fraud

Fig. 3: Box plots of the errors of the SP and DP mechanisms for (β, r)-AIQ over all the true (β, r)-anomalies in each dataset for $\varepsilon = \{1, 0.1, 0.01\}$. The sub-figures (a)-(e), follow the legends and axis labels given in the top-left sub-figure.

Results

Our first evaluation metric is the *error*. The error of a private mechanism (i.e., a randomized algorithm) is its probability of a wrong answer — recall that in the case of AIQ, there are only two possible answers, i.e. 0 and 1. For each AIQ for a *fixed record*, we estimate the error as the fraction of wrong answers over the m computations of the AIQ. For our experiments we choose $m = 10000$.

In the first set of experiments, we evaluated the errors of SP and DP mechanisms and how they vary for different levels of privacy, i.e., $\varepsilon = 1,\ 0.1,\ 0.01$. For this, we computed the error of each mechanism for all true (β, r)-anomaly in each dataset, which are given by the box plots in Figure 3.

The error of the SP-mechanism is overwhelmingly concentrated about zero (Figure 3), which is also true for the smaller values of ε. This is in direct contrast with the error of DP-mechanism. In many cases, the error of the SP mechanism is so small (e.g. of the order 10^{-15} or even smaller for larger values of ε) that it can be considered zero for all practical purposes.

Furthermore, we see that as the dataset's size increases, the error of SP-mechanism as well as its variance both decrease. This is due to the fact that, for typical (β, r)-outliers, the error of the SP-mechanism is exponentially small in $\varepsilon|\beta-k|$ (as discussed earlier), and β is directly proportional to the size of the dataset. Additionally, compared to the optimal DP-mechanism – which performs very poorly on almost all the outliers – the error for the SP-mechanism grows at much smaller rate (Figure 3). In fact, the errors of DP-mechanism are concentrated about $1/(1+e^{\varepsilon})$ (Figures 3 and 4). This is in accordance with our theoretical results and the assumption that the databases are typically sparse. For stronger privacy guarantee (i.e. the smaller values of ε), the error of DP-mechanism is consistently close to that of random coin flip

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2021.3129633, IEEE Transactions on Knowledge and Data Engineering

(i.e. selecting 0 or 1 with probability close to 1/2) except for a few cases.

One interesting point to note in Figure 3 is that for the two high dimensional datasets (APS and Synthetic), while the majority of the outlier records have very low (and concentrated error), relative to the other datasets, they do have a larger number of boxplot-outliers (i.e., outlier records which have a much higher error, in some cases close to the error of the DP-mechanism.) We believe that this may be due to problems faced by the underlying (β, r)-outlier model in high dimensional data. Specifically, (β, r)-outlier model works well, if $\beta - B_x(i, r)$ is large (e.g., greater than 0.5β) if i is an outlier. It turns out that this does not hold for the outlier records corresponding to the boxplot-outliers, which is why the error is high in these cases.

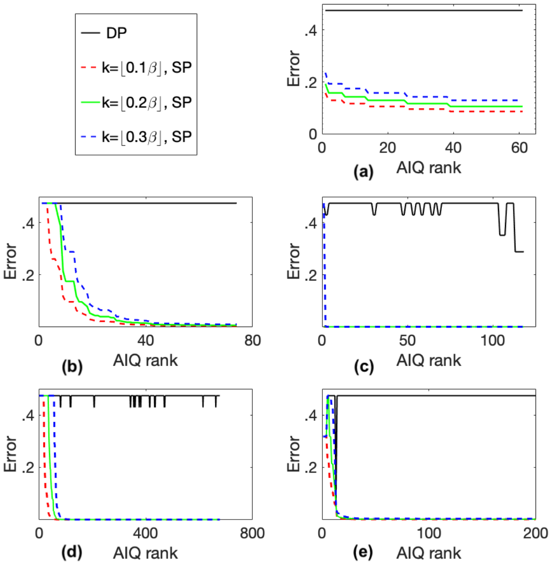


Fig. 4: **(a)-(e)**, give the errors of SP and DP mechanisms. AIQ rank is given by the error of SP-mechanism for each anomaly: the higher the rank, the lower the error. In all the figures, $\varepsilon = 0.1$. **(a)**, Thyroid, **(b)**, Mammography, **(c)**, Credit Fraud, **(d)**, APS Trucks, **(e)**, Synthetic data.

Next, we evaluate how the error of the SP-mechanism changes by varying k. We choose k as 10%, 20%, and 30% of β. Recall that for $k \geq 1$, most of the records in the typical real-world databases are protected with the privacy guarantee of ε (see Section 5 for details). Further, recall that increasing k increases the number of records that would be considered sensitive, regardless of whether they are present in the database. Infact, when $k \geq \beta$, all possible records would be considered sensitive, and the sensitive graph is identical to the DP-neighborhood graph, further implying that the error of the SP mechanism will be *exactly the same* as the DP mechanism. Figure 4 plots the errors of (ε, k)-SP mechanism for (β, r)-AIQ for varying k[7]. Here, we also plot the error for the optimal DP mechanism as a reference for comparison.

The results in Figure 4 show that even for the higher values of k SP-mechanism performs reasonably well. Here again, if the size of the dataset is large enough, the loss in accuracy for most of the records is negligible.

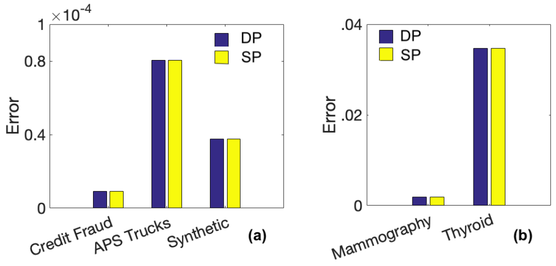


Fig. 5: **(a)** and **(b)**, give the average errors of SP and DP mechanism for AIQ over all the normal records from each data set; $\varepsilon = 0.1$.

Dataset	mean error		mean error (anomalies)
	SP	DP	SP
Credit Fraud	1.1127E−21	0.4750	1.1127E−21
APS Trucks	2.9719E−13	0.4750	2.9719E−13
Synthetic	3.2173E−5	0.4750	3.2173E−5
Mammography	0.0022	0.4749	0.0021
Thyroid	0.0870	0.4750	0.0867

TABLE 2: "mean error" is over the randomly picked n records from the possible values of the records for each dataset for SP and DP mechanisms for (β, r)-AIQ. "mean error (anomalies)" is only over the anomalous records in the n picked records. Here, n is 20% of the size of the dataset, $\varepsilon = 0.1$.

Next, we evaluated the performance over the normal records. Here, both the SP and the DP mechanisms perform equally (Figure 5). This is because for a fixed value of ε, all the sensitive records — which include all the normal records — have the same privacy under (ε, k)-SP as all the records in under ε-DP. Again the pattern continues, the datasets with larger sizes exhibit very small error.

To evaluate the performance over future queries, we picked n records uniformly at random from the space of possible (values of) records for each dataset, where n was set to be 20% of the size of the dataset. Here too the SP-mechanism outperforms the DP-mechanism significantly (Table 2). This is because most of the randomly picked records are anomalous as per the (β, r)-anomaly, which is due to the sparsity of the databases. This fact becomes very clear when we compare the mean error over the random records to the mean error over the anomalous records in the randomly picked records (see the second and the last column of Table 2). Since the probability of observing a mistake is extremely small (e.g., 1 in 10^{10} trials) , in Table 2, the mean is computed over the actual probability of error of the mechanism instead of the estimated error.

Finally, to evaluate the overall performance of our SP-mechanism for (β, r)-AIQ, we computed precision, recall, and F_1-score [31]. We also provide a comparison with two different baseline mechanisms, B_1, B_2 in addition to pareto optimal DP mechanism (see Table 3).

B_1 and B_2 are the *best* performing mechanisms (i.e., with the highest F_1-score) from two families of mechanisms. This mechanisms serve as the naive base lines. Each mechanism in each of the family is identified by a threshold t, where

[7] Recall that k quantifies the typicality of the 'small change' in the database (as discussed in Sections 4 and 5) via the notion of sensitive record.

Dataset	Precision				Recall				F_1-score			
	B_1	B_2	DP	SP	B_1	B_2	DP	SP	B_1	B_2	DP	SP
Credit Fraud	0.0101	0.0230	0.9930	0.9963	1.0000	0.0498	0.5250	0.9968	0.0199	0.0315	0.6868	0.9966
APS Trucks	0.0115	0.0165	0.9870	0.9931	1.0000	0.0753	0.5263	0.9954	0.0227	0.0271	0.6865	0.9943
Synthetic	0.0101	0.0114	0.9930	0.9963	1.0000	0.1189	0.5250	0.9968	0.0199	0.0208	0.6868	0.9966
Mammography	0.0070	0.0081	0.0211	0.2004	0.8244	0.1000	0.5250	0.9977	0.0138	0.0149	0.0435	0.3337
Thyroid	0.0174	0.0191	0.1427	0.3100	0.6656	0.2918	0.5250	0.8993	0.0339	0.0358	0.2244	0.4610

TABLE 3: B_1 and B_2 are the best mechanisms from two families of mechanism. DP and SP are the mechanisms from Section 8.4 and Section 8.4 respectively. Going from red to blue the value decreases. For our SP and DP mechanisms, $\varepsilon = 0.1$

$0 \leq t \leq 1$. Below, we describe the mechanisms from both the families for fixed ε, threshold t, record $i \in \mathcal{X}$, and database $x \in \mathcal{D}$.

The mechanism in the first family is given as follows. $B^i_{1,t}(x) = 1$ if and only if $\mathcal{O}(x) + \texttt{Lap}(1/\varepsilon) > t \times (||x||_1 + \texttt{Lap}(1/\varepsilon))$; here $\mathcal{O}(x)$ gives the number of anomalies in x and $\texttt{Lap}(1/\varepsilon)$ is independent noise from Laplace distribution of mean zero and scale $1/\varepsilon$. The mechanism in the second family is given as follows. $B^i_{2,t}(x) = 1$ if and only if $\mathcal{O}(x) + \texttt{Lap}(\beta/\varepsilon) > t \times (||x||_1 + \texttt{Lap}(1/\varepsilon))$.

Note that, the mechanism from the first family are ε_1-DP, where $\varepsilon_1 = (\beta + 1)\varepsilon$. This is due to composition of DP and the fact that $\max_{x,y \in \mathcal{D}:||x-y||_1=1} |\mathcal{O}(x) - \mathcal{O}(y)| = \beta$ and $\max_{x,y \in \mathcal{D}:||x-y||_1=1} \Big| ||x||_1 - ||y||_1 \Big|$ [16]. However the mechanism from the second family are ε_2-DP, where $\varepsilon_2 = 2\varepsilon$.

Thus, our evaluation over a range of real-word datastes show that *we can have higher privacy guarantee for sensitive records, while still being able to accurately identify anomalies*. And the fact that the error for SP mechanism for (β, r)-AIQ becomes smaller with the increase of the size of the dataset indicates that our approach is even more appropriate for big data settings.

10 Conclusion

In this article, we have considered the problem of anomaly identification, which has numerous applications, while taking privacy into consideration. While differential privacy is the state of the art model for privacy, it is inherently incapable of allowing good utility if reasonable privacy is to be met. Therefore, we develop the novel notion of sensitive privacy and relate sensitive privacy to other important notions of data privacy so that one can port the technical developments and private mechanism constructions from these related concepts to sensitive privacy. We have developed a novel n-step lookahead mechanism to efficiently and privately answer arbitrary outlier queries for a restricted class of anomaly models. We also provide general constructions to give sensitively private mechanisms for identifying anomalies and show the conditions under which the constructions would be optimal. In the future we plan to examine how effective mechanisms can be developed for other anomaly models, and examine how the sensitive privacy model can be used for other types of data analysis problems.

Acknowledgments

Research reported in this publication was supported by the National Science Foundation under awards CNS-1564034, CNS-2027789, and CNS-1747728, and by the National Institutes of Health under award R35GM134927. The content is solely the responsibility of the authors and does not necessarily represent the official views of the agencies funding the research.

References

[1] Seppo Karrila, Julian Hock Ean Lee, and Greg Tucker-Kellogg. A comparison of methods for data-driven cancer outlier discovery, and an application scheme to semisupervised predictive biomarker discovery. *Cancer informatics*, 10:CIN–S6868, 2011.

[2] Soumi Ray, Dustin S McEvoy, Skye Aaron, Thu-Trang Hickman, and Adam Wright. Using statistical anomaly detection models to find clinical decision support malfunctions. *Journal of the American Medical Informatics Association*, 2018.

[3] Gordon D Schiff, Lynn A Volk, Mayya Volodarskaya, Deborah H Williams, Lake Walsh, Sara G Myers, David W Bates, and Ronen Rozenblum. Screening for medication errors using an outlier detection system. *Journal of the American Medical Informatics Association*, 24(2):281–287, 2017.

[4] Charu C. Aggarwal. *Outlier Analysis*. Springer, 2013.

[5] Vic Barnett and Toby Lewis. *Outliers in Statistical Data*. John Wiley and Sons, 3rd edition, 1994.

[6] Alison M Darcy, Alan K Louie, and Laura Weiss Roberts. Machine learning and the profession of medicine. *Jama*, 315(6):551–552, 2016.

[7] Ziad Obermeyer and Ezekiel J Emanuel. Predicting the future—big data, machine learning, and clinical medicine. *The New England journal of medicine*, 375(13):1216, 2016.

[8] Machine Learning Group. Credit card fraud detection, Mar 2018.

[9] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *TCC*, pages 265–284. Springer, 2006.

[10] Cynthia Dwork, Aaron Roth, et al. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407, 2014.

[11] Hafiz Asif, Periklis A. Papakonstantinou, and Jaideep Vaidya. How to accurately and privately identify anomalies. In *SIGSAC CCS*. ACM, 2019.

[12] Hafiz Asif, Periklis A Papakonstantinou, and Jaideep Vaidya. A guide for private outlier analysis. *IEEE Letters of the Computer Society*, 3(1):29–33, 2020.

[13] Michael Kearns, Aaron Roth, Zhiwei Steven Wu, and Grigory Yaroslavtsev. Private algorithms for the protected in social network search. *Proceedings of the National Academy of Sciences*, 113(4):913–918, 2016.

[14] Stelios Doudalis, Ios Kotsogiannis, Samuel Haney, Ashwin Machanavajjhala, and Sharad Mehrotra. One-sided differential privacy. *arXiv preprint arXiv:1712.05888*, 2017.

[15] Daniel M Bittner, Anand D Sarwate, and Rebecca N Wright. Using noisy binary search for differentially private anomaly detection. In *International Symposium on Cyber Security Cryptography and Machine Learning*, pages 20–37. Springer, 2018.

[16] Cynthia Dwork. Differential privacy. In Michele Bugliesi, Bart Preneel, Vladimiro Sassone, and Ingo Wegener, editors, *Automata, Languages and Programming*, pages 1–12, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.

[17] Daniel Kifer and Ashwin Machanavajjhala. No free lunch in data privacy. In *Proceedings of the 2011 ACM SIGMOD*, pages 193–204. ACM, 2011.

[18] Daniel Kifer and Ashwin Machanavajjhala. Pufferfish: A framework for mathematical privacy definitions. *ACM Transactions on Database Systems (TODS)*, 39(1):3, 2014.

[19] Xi He, Ashwin Machanavajjhala, and Bolin Ding. Blowfish privacy: Tuning privacy-utility trade-offs using policies. In *Proceedings of the 2014 ACM SIGMOD*, pages 1447–1458. ACM, 2014.

[20] Damien Desfontaines and Balázs Pejó. Sok: Differential privacies. *Proceedings on Privacy Enhancing Technologies*, 2020(2):288–313, 2020.

[21] Zach Jorgensen, Ting Yu, and Graham Cormode. Conservative or liberal? personalized differential privacy. In *2015 IEEE 31st International Conference on Data Engineering (ICDE)*, pages 1023–1034. IEEE, 2015.

[22] Edward Lui and Rafael Pass. Outlier privacy. In *TCC*, pages 277–305. Springer, 2015.

[23] Andrew C. Yao. Protocols for secure computation. In *Proceedings of the 23rd Annual IEEE Symposium on Foundations of Computer Science*, pages 160–164, 1982.

[24] Oded Goldreich, Silvio Micali, and Avi Wigderson. How to play any mental game - a completeness theorem for protocols with honest majority. In *Proc. of the Symp. on the Theory of Computing*, 1987.

[25] Jaideep Vaidya and Chris Clifton. Privacy-preserving outlier detection. In *Data Mining, 2004. ICDM'04. Fourth IEEE International Conference on*, pages 233–240. IEEE, 2004.

[26] Hafiz Asif, Tanay Talukdar, Jaideep Vaidya, Basit Shafiq, and Nabil Adam. Collaborative differentially private outlier detection for categorical data. In *IEEE CIC*, pages 92–101. IEEE, 2016.

[27] Hafiz Asif, Tanay Talukdar, Jaideep Vaidya, Basit Shafiq, and Nabil Adam. Differentially private outlier detection in a collaborative environment. *IJCIS*, 27(03):1850005, 2018.

[28] Edwin M. Knorr, Raymond T. Ng, and Vladimir Tucakov. Distance-based outliers: algorithms and applications. *The VLDB Journal*, 8(3-4):237–253, 2000.

[29] Edwin M Knorr and Raymond T Ng. Algorithms for mining distancebased outliers in large datasets. In *Proceedings of the 1998 VLDB*, pages 392–403. Citeseer, 1998.

[30] Konstantinos Chatzikokolakis, Miguel E Andrés, Nicolás Emilio Bordenabe, and Catuscia Palamidessi. Broadening the scope of differential privacy using metrics. In *International Symposium on Privacy Enhancing Technologies Symposium*, pages 82–102. Springer, 2013.

[31] Charu C Aggarwal. Outlier analysis. In *Data mining*, pages 237–263. Springer, 2015.

[32] Vic Barnett and Toby Lewis. *Outliers in statistical data*. Wiley, 2000.

[33] Andrea Dal Pozzolo, Olivier Caelen, Reid A Johnson, and Gianluca Bontempi. Calibrating probability with undersampling for unbalanced classification. In *Computational Intelligence, 2015 IEEE Symposium Series on*, pages 159–166. IEEE, 2015.

[34] Shebuti Rayana. ODDS library, 2016. Available at http://odds.cs.stonybrook.edu.

[35] Dheeru Dua and Casey Graff. UCI machine learning repository, 2017.

[36] Yihe Dong, Samuel B Hopkins, and Jerry Li. Quantum entropy scoring for fast robust mean estimation and improved outlier detection. *arXiv preprint arXiv:1906.11366*, 2019.

[37] Ian Jolliffe. Principal component analysis. In *International encyclopedia of statistical science*, pages 1094–1096. Springer, 2011.

Hafiz Asif is a Postdoctoral Associate at Rutgers. His research interests are in the areas of privacy, security, machine learning, and data analytics.

Jaideep Vaidya is a Professor of Computer Information Systems with Rutgers University and is the Director of the Rutgers Institute for Data Science, Learning, and Applications. He has published over 190 papers in international conferences and journals. His research interests are in privacy, security, and data management. He is an IEEE Fellow, an ACM Distinguished Scientist, and is the Editor in Chief of IEEE TDSC.

Periklis A. Papakonstantinou is an Associate Professor in the Management Science and Information Systems Department at Rutgers University. His research interests include theory of computing, cryptography, and privacy.