Differentially Private Outlier Detection in Correlated Data

Kwassi H. Degue, Karthik Gopalakrishnan, Max Z. Li, Hamsa Balakrishnan

***Abstract*— The detection of outliers has become increasingly important for the control and monitoring of large-scale networked systems such as transportation and smart grids. Data from these systems, such as location traces or power consumption, are collected from individual agents, and are often privacy-sensitive. Furthermore, the networked nature of these systems results in the data of different individuals being correlated with each other. In this paper, we use the concept of differential privacy to design a privacy-preserving algorithm for outlier detection in correlated data. We determine analytic formulas to evaluate the performance of the proposed differentially private algorithm, and we analyze the trade-off between privacy level and detection accuracy. We illustrate our methodology using an example based on outlier detection in household electricity usage data.**

I. Introduction

The identification of distribution changes in a stream of data plays a key role in numerous practical settings, such as fault detection, syndromic surveillance, signal detection, finance, and security systems. To perform this task, a metric computed from a sequence of observations must be consistent with the hypothesis that the data are realizations of a given distribution. For example, smart home Internet of Things (IoT) devices may wish to detect outliers in activity within a home; Public Health Services may wish to detect a disease outbreak by using individuals' medical records; congestion-aware routing applications may want to detect traffic congestion on roads by relying on location data provided by smartphones and connected vehicles [1]. In these scenarios, the data often contains highly privacy-sensitive information. By using off-the-shelf statistical techniques, smart meter data can be used to deduce if certain individuals were at home, or even deviated from their routines [2]. Furthermore, it has been shown that even aggregate location data can lead to the reconstruction of individual trajectories [3], [4]. The above observations motivate the development of privacy-preserving mechanisms for outlier detection.

Anonymization techniques such as k-anonymity [5] or removing explicitly identifying information from personal data are known to be inefficient in preserving privacy [6], [7]. The notion of *differential privacy* [8]–[10] provides a much stronger privacy guarantee to individuals' data, and lends itself to several applications [11]–[16]. It provides each individual agent with the guarantee that the output of the considered query will not be significantly altered by whether or not they contribute their data, or what value they contribute. Differential privacy can be achieved through *input perturbation*, *output perturbation* [11], [17] or the *sparse vector technique* [18], [19]. In this paper, we consider input perturbation, which has the advantage that each individual agent can perturb its data before sending it to a data aggregator, thereby eliminating the need to trust the aggregator.

Classical hypothesis testing under differential privacy constraints has been previously considered in several contexts. Differentially private algorithms for categorical data that followed a multinomial distribution were considered by [20] and [21]. In addition, differentially private outlier detection using Monte Carlo (MC) approaches were proposed in [20], [21], as well as using machine learning-based techniques in [22]. Furthermore, differentially private statistical tests were studied for the data of individual agents under a Gaussian assumption in [23] and [24], in order to decide whether or not the mean of a sequence of independent and identically distributed (i.i.d.) scalar Gaussian random variables differed from a given value. However, data from different agents were assumed to be uncorrelated in these prior works, which may be an unrealistic assumption in some systems [12], [16], [25].

Example (Privacy-aware outlier detection in correlated electricity usage data). Suppose that the agents consist of individual households in a neighborhood, and that the data signal contributed by each agent is the per-day electricity consumption collected via a smart meter. The electricity consumption of households may be correlated (e.g., because of similar weather or the similar age of homes); indeed, such correlations can be observed in Fig. 1.

A data aggregator (e.g., the power company) is interested in monitoring the smart meter data to detect unusual patterns (*outliers*) in observed usage, to determine what, if any, control interventions are needed. However, potentially private and identifying information can be discerned from

The NASA University Leadership Initiative (grant #80NSSC20M0163) provided funds to assist the authors with their research, but this article solely reflects the opinions and conclusions of its authors and not any NASA entity. The authors were also supported in part by NSF CPS Award No. 1739505. M. Z. Li was supported in part by a NSF fellowship.

The authors are with the Department of Aeronautics and Astronautics at the Massachusetts Institute of Technology, Cambridge, MA. {kdegue, karthikg, maxli, hamsa}@mit.edu

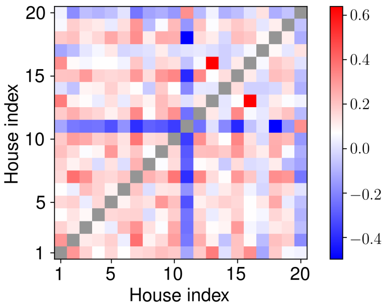


Fig. 1. Correlation coefficients for the daily electricity consumption of 20 individual households (REFIT data [26]).

such data [2]. We would like to design a differentially private mechanism for detecting outliers in correlated data. Such a privacy guarantee will encourage more households to participate and contribute their smart meter data, which can help unlock the full potential of smart grids [27].

Differentially private outlier detection algorithms considering the Euclidean distance have been proposed in [28]. However, this metric does not account for the correlation between variables; in fact, it has been shown that correlated data is excessively weighted when the Euclidean distance is used for outlier detection [29], [30]. We propose to overcome this limitation by using the *Mahalanobis distance* [31], which adapts the Euclidean distance to account for correlations between variables.

In this paper, we design and analyze a differentially private algorithm to detect outliers in multivariate Gaussian signals using the squared Mahalanobis distance as a metric. In contrast to [23], [24], [32] where i.i.d. scalar quantities are considered, we consider multivariate signals provided by individual agents who may be correlated. Unlike [19] where the sparse vector technique is used, we consider scenarios in which individual agents do not necessarily trust the data aggregator, and therefore we design an input perturbation architecture to guarantee differential privacy for the agents' data. Using the squared Mahalanobis distance, we derive analytical formulas for the trade-offs between the accuracy of detection and privacy level.

Section II presents the problem statement and provides us with some background on differential privacy and on outlier detection using the squared Mahalanobis distance. These results are applied to design a differentially private outlier detection algorithm in Section III. Section IV provides us with analytic formulas for probabilities of detection of the outlier detection algorithm and for thresholds as functions of user-defined target probabilities of false-alarm. Finally, in Section IV, we present an application to outlier detection in a smart metering scenario.

II. Problem Statement

A. Notation

We fix a generic probability triple $(\Omega, \mathcal{A}, \mathbb{P})$, with $\mathcal{A}$ a σ-algebra on the sample space Ω, and $\mathbb{P}$ a probability measure defined on $\mathcal{A}$. We denote the ℓ_p-norm of a vector $\boldsymbol{x} = (x_1, \dots, x_n)^\intercal \in \mathbb{R}^{n\times 1}$ by $|\mathbf{x}|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}$, for $p \in [1, \infty]$; $|\cdot|$ without subscripts is used for absolute values. We denote the group of orthogonal $n \times n$ matrices by $\mathrm{O}\left(n\right)$. Boldface variables denote vectors, and non-boldface, capital letters denote matrices. Special matrices such as covariance matrices are denoted by Σ. We index observations by superscripts enclosed in parentheses, and index components of vectors in subscripts.

B. Outlier Detection in Distribution Problem

Consider a sequence of m observations $\mathcal{O}_m := \left\{\mathbf{x}^{(k)}\right\}_{k=1}^{k=m}$, with $\mathbf{x}^{(k)} = \left[x_i^{(k)}\right] \in \mathbb{R}^{n\times 1}$. The vector $\mathbf{x}^{(k)}$ contains data from multiple agents at time k, where $x_i^{(k)}$ denotes the data for an individual agent i at time k. We observe the data $\mathcal{O}_m$ sequentially. The signal vectors $\mathbf{x}^{(k)}$ are assumed to be realizations of $\mathbf{X} \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\boldsymbol{\mu}, \Sigma\right)$, i.e., the samples across k are i.i.d. However, the data values of different individuals (i.e., $x_i^{(k)}$ and $x_j^{(k)}$ for $i \neq j$) may be correlated.

The first goal is to design a statistical test to check whether or not an observation $\mathbf{x}^{(k)}$ deviates significantly from the known historical distribution. We use the following quantity as a metric for our task of detecting such outlying observations:

Definition 1. *The squared Mahalanobis distance [31] of a signal* $\mathbf{y}$ *drawn from some distribution* $\mathcal{D}\left(\boldsymbol{\mu}, \Sigma\right)$ *with finite mean and variance is*

$$\left(d_M(\mathbf{y})\right)^2 = (\mathbf{y} - \boldsymbol{\mu})^\intercal \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu}). \tag{1}$$

The squared Mahalanobis distance can be viewed as a measure of the distance between a signal and the underlying data distribution. It scales each individual agent's contribution with respect to the variability of its data. Notice that $\left(d_M\left(\mathbf{x}^{(k)}\right)\right)^2$ increases with the distance between $\mathbf{x}^{(k)}$ and the mean $\boldsymbol{\mu}$. In our setup, we consider $\mathbf{x}^{(k)}$ to be drawn from $\mathcal{N}\left(\boldsymbol{\mu}, \Sigma\right)$, then $\mathbf{x}^{(k)} - \boldsymbol{\mu}$ is a zero-mean Gaussian variable, and we deduce the following proposition by applying [33, Section 2.3].

Proposition 1. Let $\mathbf{x}^{(k)}$ be realizations of $\mathbf{X} \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\boldsymbol{\mu}, \Sigma\right)$. Then, the squared Mahalanobis distance $\left(d_M\left(\mathbf{x}^{(k)}\right)\right)^2$ defined in (1) is a random variable that satisfies

$$\left(d_M\left(\mathbf{x}^{(k)}\right)\right)^2 \sim \chi_n^2, \tag{2}$$

where χ_n^2 represents a chi-squared distribution with n degrees of freedom.

Accordingly, we formally define the term *outlier* in the context of this article:

Definition 2. *An observation $\mathbf{x}^{(k)}$ is labeled as an outlier if*

$$\left(d_M\left(\mathbf{x}^{(k)}\right)\right)^2 \geq h,$$

where h is a user-specified threshold. Note that a higher value of h makes the outlier detection more conservative (i.e., decreases the false alarm rate).

Consequently, the observations in the data set $\mathcal{O}_m$ are mapped to one of the two following hypotheses:

$$\begin{cases} \mathrm{H}_0: & \left(d_M\left(\mathbf{x}^{(k)}\right)\right)^2 < h : \mathbf{x}^{(k)} \text{ is not an outlier,} \\ \mathrm{H}_1: & \left(d_M\left(\mathbf{x}^{(k)}\right)\right)^2 \geq h : \mathbf{x}^{(k)} \text{ is an outlier.} \end{cases}$$

We can then compute the following decision rule:

$$\operatorname{dec}\left(\mathbf{x}^{(k)}\right) = \begin{cases} 0, & \text{if } q\left(\mathbf{x}^{(k)}\right) < h : \mathrm{H}_0 \text{ is chosen,} \\ 1, & \text{if } q\left(\mathbf{x}^{(k)}\right) \geq h : \mathrm{H}_1 \text{ is chosen,} \end{cases} \tag{3}$$

where the *query* $q : \mathcal{O}_m \to \mathbb{R}_{\geq 0}$ is the squared Mahalanobis distance of $\mathbf{x}^{(k)}$:

$$q\left(\mathbf{x}^{(k)}\right) = \left(d_M\left(\mathbf{x}^{(k)}\right)\right)^2. \tag{4}$$

Rule (3) decides whether or not an observation $\mathbf{x}^{(k)}$, belonging to the data set $\mathcal{O}_m$, is an outlier. The threshold h, the mean $\boldsymbol{\mu}$, the matrix Σ and the outcome of the statistical test (3) are assumed to be publicly known information. In this article, we consider the case where the entries $x_i^{(k)}$ of the agents contributing to the dataset are privacy-sensitive, i.e., the agents want to ensure that sharing the outcome of the outlier-detection decision rule does not reveal any information about their contribution to $\mathbf{x}^{(k)}$. More formally, we would like to ensure that $\mathcal{O}_m$ privacy-sensitive data set with respect to the outlier detection rule. This would require modifying the outlier-detection decision rule (3) in order to satisfy this privacy requirement. In the next subsection, we present a brief introduction to differential privacy, a concept that formalizes such requirements [9].

C. Differential Privacy

Let $\mathcal{H}$ be a space of data sets. Throughout this article, the space that contains the observation sequence $\mathcal{O}_m$ is denoted by $\mathcal{H} \equiv \mathbb{R}^{n \times m}$. We define a *mechanism* M as a random map from $\mathcal{H}$ to some measurable output space. A differentially private mechanism aims to provide similarly-distributed outputs for inputs that need to be made indistinguishable [10].

A symmetric binary relation Adj on $\mathcal{H}$, called adjacency, is used to describe inputs that are considered "close". For example, two inputs can be *adjacent* if the inputs are the same for all but one individual agent, where that one agent's input differs, but the difference is bounded. More formally, throughout this article, two sequences of observations $\mathcal{O}_m := \left\{\mathbf{x}^{(k)}\right\}_{k=1}^{k=m}$ and $\widetilde{\mathcal{O}}_m := \left\{\widetilde{\mathbf{x}}^{(k)}\right\}_{k=1}^{k=m}$ are termed adjacent if, and only if:

$$\left|x_i^{(k)} - \widetilde{x}_i^{(k)}\right| \leq \rho^{(k)}, \quad \text{for some } 1 \leq k \leq m,\ 1 \leq i \leq n,$$

$$\text{and } x_j^{(\ell)} = \widetilde{x}_j^{(\ell)}, \text{ for all } \ell \neq k,\ j \neq i, \tag{5}$$

where the set of positive values $\left\{\rho^{(k)}\right\}_{k=1}^{k=m} \in \mathbb{R}_{>0}^m$ is given. In other words, we consider two sequences of observations to be adjacent if, and only if, they differ only by the value of a single element $x_i^{(k)}$ within a single vector $\mathbf{x}^{(k)}$, and this difference is bounded as well. We denote two sequences of adjacent observations $\mathcal{O}_m$ and $\widetilde{\mathcal{O}}_m$ by $\operatorname{Adj}(\mathcal{O}_m, \widetilde{\mathcal{O}}_m)$.

Next, we formally define differential privacy as established by [8], [9]. To do so, we define a new variable $\rho = \max\limits_{1 \leq k \leq m} \rho^{(k)}$, i.e., the maximum of the bounds on differences between adjacent data sets.

Definition 3. *Let $\mathcal{H}$ be a space provided with a symmetric binary relation denoted* Adj*, and consider $(\mathcal{P}, \mathcal{M})$ a measurable space, with $\mathcal{M}$ a given σ-algebra over $\mathcal{P}$. Let $\epsilon, \delta \geq 0$. A randomized mechanism M from $\mathcal{H}$ to $\mathcal{P}$ is (ϵ, δ)-differentially private (for* Adj*) if, for all $\mathcal{O}_m, \widetilde{\mathcal{O}}_m \in \mathcal{H}$ such that $\operatorname{Adj}\left(\mathcal{O}_m, \widetilde{\mathcal{O}}_m\right)$, and for all sets S in $\mathcal{M}$:*

$$\mathbb{P}\left(M\left(\mathcal{O}_m\right) \in S\right) \leq e^{\epsilon}\, \mathbb{P}\left(M\left(\widetilde{\mathcal{O}}_m\right) \in S\right) + \delta. \tag{6}$$

In other words, when $\mathcal{O}_m$ and $\widetilde{\mathcal{O}}_m$ are adjacent, (6) dictates that the distributions of the random variables $M\left(\mathcal{O}_m\right)$ and $M\left(\widetilde{\mathcal{O}}_m\right)$ are close. Next, we define the notion of *sensitivity*, which plays a major role in the design of differentially private mechanisms.

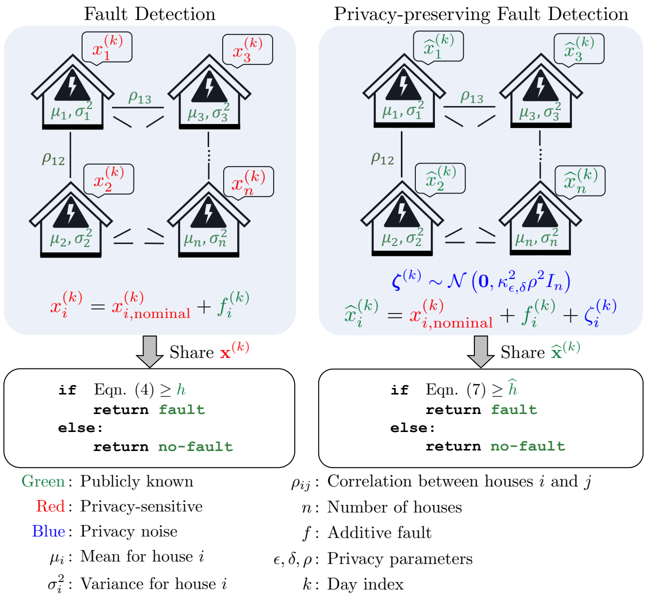
Definition 4. *Let $\mathcal{H}$ be a space of data sets with an adjacency relation* Adj*, and consider $\mathcal{P}$ a vector space with norm $\|\cdot\|_{\mathcal{P}}$. The sensitivity of a query $q : \mathcal{H} \to \mathcal{P}$ is defined as the quantity $\Delta_{\mathcal{P}q} := \sup_{\left\{\mathcal{O}_m, \widetilde{\mathcal{O}}_m \,:\, \operatorname{Adj}\left(\mathcal{O}_m, \widetilde{\mathcal{O}}_m\right)\right\}} \left\|q\left(\mathcal{O}_m\right) - q\left(\widetilde{\mathcal{O}}_m\right)\right\|_{\mathcal{P}}$. In particular, when $\mathcal{P} \triangleq \mathbb{R}^n$ (with $n = +\infty$ being a possibility), and given the p-norm for $p \in [1, \infty]$, $\Delta_{\mathcal{P}q}$ denotes the ℓ_p-sensitivity. For brevity, we simply write Δ_p instead of $\Delta_{\mathcal{P}q}$.*

The *Gaussian mechanism* [34] can be achieved by adding Gaussian noise proportional to the ℓ_2-sensitivity of a mapping to enforce (ϵ, δ)-differential privacy.

First, the Q-function [11] is given by $Q(x) := \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp(-\frac{u^2}{2})\, du$. Then, for $1 > \delta > 0$, define $\kappa_{\delta,\epsilon} := \frac{1}{2\epsilon}\left(Q^{-1}(\delta) + \sqrt{(Q^{-1}(\delta))^2 + 2\epsilon}\right)$.

Theorem 1 (from [11])**.** *Let $\epsilon > 0$, and $1 > \delta > 0$. Consider a system $\mathcal{G} : \mathbb{R}^{n_1} \to \mathbb{R}^{n_2}$. Then, the mechanism $M(\boldsymbol{y}) = \mathcal{G}(\boldsymbol{y}) + \boldsymbol{\nu}$, with $\boldsymbol{\nu}$ a white Gaussian noise (i.e., sequence of i.i.d zero-mean Gaussian vectors) $\boldsymbol{\nu} \sim \mathcal{N}\left(\mathbf{0}, \kappa_{\delta,\epsilon}^2 \left(\Delta_2(\mathcal{G})\right)^2 I_{n_2}\right)$, is (ϵ,δ)-differentially private, where I_{n_2} denotes the $n_2 \times n_2$ identity matrix, and $\Delta_2(\mathcal{G})$ denotes the ℓ_2 sensitivity of $\mathcal{G}$.*

Differential privacy is “resilient to post-processing”, i.e., manipulating a result that is differentially private does not weaken the differential privacy guarantee, as long as the sensitive data is not re-accessed during the manipulation [10], [11, Theorem 1]. Next, we design a differentially private algorithm for detection of outliers in our signal setting with respect to the decision rule (3).



```
if   Eqn. (4) $\geq h$
     return fault
else:
     return no-fault
```

```
if   Eqn. (7) $\geq \widehat{h}$
     return fault
else:
     return no-fault
```

Fig. 2. Overview of our differentially private outlier detection algorithm applied to the privacy-sensitive setting of smart meters monitoring household electricity consumption.

III. Differentially Private Detection of Outliers

By using the resilience to post-processing property, we deduce from the form of the statistical test (3) that basing the decision rule (3) on a differentially private version of $q\left(\mathbf{x}^{(k)}\right)$ will provide us with a differentially private test.

Theorem 2. *A mechanism that publicly releases $\widehat{\mathbf{x}}^{(k)} = \mathbf{x}^{(k)} + \boldsymbol{\zeta}^{(k)}$ with $\boldsymbol{\zeta}^{(k)} \sim \mathcal{N}\left(\mathbf{0}, \kappa_{\delta,\epsilon}^2 \rho^2 I_n\right)$ is (ϵ,δ)-differentially private for the adjacency relation (5).*

Proof. For two observation sequences $\mathcal{O}_m$ and $\widetilde{\mathcal{O}}_m$ that are adjacent with respect to (5), we can bound the sensitivity as follows:

$$\Delta_2 = \sup_{\substack{1 \leq k \leq m \\ \mathcal{O}_m, \widetilde{\mathcal{O}}_m : \operatorname{Adj}\left(\mathcal{O}_m, \widetilde{\mathcal{O}}_m\right)}} \|\mathbf{x} - \tilde{\mathbf{x}}\|_2 ,$$

where $\mathbf{x}$ is a vector-valued signal composed of $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(m)}$, and the ℓ_2-norm $\|s\|_2 = \left(\sum_{k=1}^m |s_k|_2^2\right)^{1/2}$ for a vector-valued signal s. Therefore, we get

$$\begin{aligned} \Delta_2 &= \sup_{\substack{1 \leq k \leq m \\ \mathcal{O}_m, \widetilde{\mathcal{O}}_m : \operatorname{Adj}\left(\mathcal{O}_m, \widetilde{\mathcal{O}}_m\right)}} \left(\sum_{k=1}^m \left|\mathbf{x}^{(k)} - \widetilde{\mathbf{x}}^{(k)}\right|_2^2\right)^{1/2} \\ &= \sup_{\substack{1 \leq k \leq m \\ \mathcal{O}_m, \widetilde{\mathcal{O}}_m : \operatorname{Adj}\left(\mathcal{O}_m, \widetilde{\mathcal{O}}_m\right)}} \left(\sum_{k=1}^m \sum_{i=1}^n \left|x_i^{(k)} - \widetilde{x}_i^{(k)}\right|^2\right)^{1/2} , \\ &= \rho. \end{aligned}$$

The result follows by applying Theorem 1. □

We now consider the statistical test described in Section II with a differential privacy constraint. We first compute

$$\widehat{q}\left(\mathbf{x}^{(k)}\right) = \left(\widehat{\mathbf{x}}^{(k)} - \boldsymbol{\mu}\right)^\intercal \left(\Sigma + \kappa_{\delta,\epsilon}^2 \rho^2 I_n\right)^{-1} \left(\widehat{\mathbf{x}}^{(k)} - \boldsymbol{\mu}\right), \tag{7}$$

recalling that $\widehat{\mathbf{x}}^{(k)} = \mathbf{x}^{(k)} + \boldsymbol{\zeta}^{(k)}$ with $\boldsymbol{\zeta}^{(k)} \sim \mathcal{N}\left(\mathbf{0}, \kappa_{\delta,\epsilon}^2 \rho^2 I_n\right)$ from Theorem 2. Then, we release publicly

$$\widehat{\operatorname{dec}}\left(\mathbf{x}^{(k)}\right) = \begin{cases} 0, & \text{if } \widehat{q}\left(\mathbf{x}^{(k)}\right) < \widehat{h} : \mathrm{H}_0 \text{ is chosen}, \\ 1, & \text{if } \widehat{q}\left(\mathbf{x}^{(k)}\right) \geq \widehat{h} : \mathrm{H}_1 \text{ is chosen}, \end{cases} \tag{8}$$

where $\widehat{h}$ is threshold that is set suitably, independent of the sequence of observations $\left\{\mathbf{x}^{(k)}\right\}_{k=1}^{k=m}$. With the new statistical test in (8), we have the following corollary:

Corollary 1. The statistical test (8) is (ϵ,δ)-differentially private.

Proof. This can be deduced by applying Theorem 2 and the resilience to post-processing property of differential privacy. □

A. Performance Analysis

In this section, we characterize the privacy-utility tradeoff of our privacy-preserving statistical test in (8). For a given test, we denote the probability of incorrectly accepting H_0 (type II error) by P_{II}, and the probability of incorrectly rejecting H_0 (type I error) by P_I. The following definitions are needed for the next theorem: Denote the complementary cumulative distribution function (ccdf) of the chi-squared distribution with k degrees of freedom by $\mathcal{F}(\cdot;k)$ and denote its inverse, defined on $[0,1)$, by $\mathcal{F}^{-1}(\cdot;k)$.

Theorem 3. *By selecting the threshold $\widehat{h}$ as*

$$\widehat{h} \;\; = \;\; \mathcal{F}^{-1}\left(P_I;n\right), \tag{9}$$

the statistical test in (8) *achieves a type I error probability of P_I for each observation $\mathbf{x}^{(k)} \in \mathbb{R}^{n\times 1}$ indexed by $k = 1,\ldots,m$.*

Proof. By applying [33, Section 2.3], we have that the random variable $\widehat{q}\left(\mathbf{x}^{(k)}\right) \sim \chi_n^2$. The probability of type I error of the statistical test (8) is given by $P_I = \mathbb{P}\left(\widehat{q}\left(\mathbf{x}^{(k)}\right) \geq \widehat{h}\right)$ when $\widehat{\mathbf{x}}^{(k)}$ is a realization of a random variable drawn independently and identically from $\mathcal{N}\left(\boldsymbol{\mu}, \Sigma + \kappa_{\delta,\epsilon}^2\rho^2 I_n\right)$ in (7). Consequently, we have that

$$P_I = \mathcal{F}\left(\widehat{h};n\right), \tag{10}$$

and the result (9) follows directly. □

Next, we derive an analytical formula for the probability of true detection for the statistical test (8). We denote by $\mathcal{Q}\left(\cdot;n;a_1,\ldots,a_n;\nu_1,\ldots,\nu_n\right)$ the ccdf for a weighted sum of non-central chi-squared random variables with 1 degree of freedom, scalar weights a_1 through a_n, and non-centrality parameters ν_1 through ν_n. Explicitly, these random variables are of the form $\sum_{i=1}^{n} a_i\chi_{1,\nu_i}^2$, where χ_{1,ν_i}^2 is a non-central chi-squared random variable with 1 degree of freedom and non-centrality parameter ν_i. For tractability, we assume that an outlier is caused by an additive signal $\boldsymbol{f}^{(k)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\boldsymbol{\mu}_f, \Sigma_f\right)$ acting on the corresponding input data. For notational brevity, we will redefine $\mathbf{x}^{(k)} \triangleq \mathbf{x}_{\text{nominal}}^{(k)} + \boldsymbol{f}^{(k)}$, where now $\mathbf{x}_{\text{nominal}}^{(k)} \sim \mathcal{N}\left(\boldsymbol{\mu}, \Sigma\right)$, and $\boldsymbol{f}^{(k)} = \mathbf{0}$ indicates no additive signal. We note that when an observation $\mathbf{x}^{(k)}$ is an outlier, $\widehat{q}\left(\mathbf{x}^{(k)}\right)$ is no longer a χ_n^2 random variable. In particular, when an observation $\mathbf{x}^{(k)}$ is an outlier, we can construct the symmetric matrix $S^{(k)} = \Sigma_{\widehat{\mathbf{x}}^{(k)}}^{\frac{1}{2}}\left(\Sigma + \kappa_{\delta,\epsilon}^2\rho^2 I_n\right)^{-1}\Sigma_{\widehat{\mathbf{x}}^{(k)}}^{\frac{1}{2}}$, and write down its eigenvalue decomposition $S^{(k)} = P^{(k)}\Lambda^{(k)}\left(P^{(k)}\right)^{\mathsf{T}}$, where $\Lambda^{(k)} = \operatorname{diag}\left(\lambda_1^{(k)},\ldots,\lambda_n^{(k)}\right)$ with $\{\lambda_i^{(k)}\}_{i=1}^{i=n}$ denoting the eigenvalues of $S^{(k)}$ and $P^{(k)} \in \mathrm{O}\left(n\right)$ an orthogonal matrix. Such a decomposition is always possible as $S^{(k)}$ is positive definite. We now provide the following result for the true positive rate $P_D = 1 - P_{II}$ of the statistical test in (8) under (ϵ,δ)-differential privacy constraints:

Theorem 4. *Assume that an outlier is caused by an additive signal $\boldsymbol{f}^{(k)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\boldsymbol{\mu}_f, \Sigma_f\right)$ on each corresponding nominal input data $\mathbf{x}_{\text{nominal}}^{(k)}$. For each data observation indexed by $k = 1,\ldots,m$, a differentially private statistical test in* (8) *designed to achieve a type I error probability of P_I, also achieves a true positive rate of*

$$\begin{aligned} P_D^{(k)} &= 1 - P_{II}^{(k)} \\ &= \mathcal{Q}\left(\widehat{h};n;\lambda_1^{(k)},\ldots,\lambda_n^{(k)};\gamma_1^{(k)},\ldots,\gamma_n^{(k)}\right), \end{aligned} \tag{11}$$

where $\gamma_i^{(k)}$ denotes the i^{th} component of the vector $\boldsymbol{\gamma}^{(k)} = \left(P^{(k)}\right)^{\mathsf{T}}\Sigma_{\widehat{\mathbf{x}}^{(k)}}^{-\frac{1}{2}}\boldsymbol{\mu}_f$, and $\widehat{h}$ is given in (9).

Proof. When an observation $\mathbf{x}^{(k)}$ is an outlier, we have that $\widehat{\mathbf{x}}^{(k)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\boldsymbol{\mu} + \boldsymbol{\mu}_f, \Sigma_{\widehat{\mathbf{x}}}\right)$ with $\Sigma_{\widehat{\mathbf{x}}} = \Sigma + \kappa_{\delta,\epsilon}^2\rho^2 I_n + \Sigma_f$. For notational convenience, we define $\widehat{\Sigma} = \Sigma + \kappa_{\delta,\epsilon}^2\rho^2 I_n$. Then, the privacy-preserving query function $\widehat{q}\left(\mathbf{x}^{(k)}\right)$ can be rewritten as follows

$$\begin{aligned} \widehat{q}\left(\mathbf{x}^{(k)}\right) &= \left(\widehat{\mathbf{x}}^{(k)} - \boldsymbol{\mu}\right)^{\mathsf{T}}\widehat{\Sigma}^{-1}\left(\widehat{\mathbf{x}}^{(k)} - \boldsymbol{\mu}\right) \\ &= \left(\widehat{\mathbf{x}}^{(k)} - \boldsymbol{\mu}\right)^{\mathsf{T}}\Sigma_{\widehat{\mathbf{x}}^{(k)}}^{-\frac{1}{2}}\Sigma_{\widehat{\mathbf{x}}^{(k)}}^{\frac{1}{2}}\widehat{\Sigma}^{-1}\Sigma_{\widehat{\mathbf{x}}^{(k)}}^{\frac{1}{2}}\Sigma_{\widehat{\mathbf{x}}^{(k)}}^{-\frac{1}{2}}\left(\widehat{\mathbf{x}}^{(k)} - \boldsymbol{\mu}\right) \\ &= \left(\boldsymbol{\xi}^{(k)}\right)^{\mathsf{T}}\Lambda^{(k)}\boldsymbol{\xi}^{(k)}, \end{aligned}$$

with $\boldsymbol{\xi}^{(k)} = \left(P^{(k)}\right)^{\mathsf{T}}\Sigma_{\widehat{\mathbf{x}}^{(k)}}^{-\frac{1}{2}}\left(\widehat{\mathbf{x}}^{(k)} - \boldsymbol{\mu}\right)$. Consequently, we have that $\boldsymbol{\xi}^{(k)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\boldsymbol{\gamma}^{(k)}, I_n\right)$. Furthermore, by using the fact that $\Lambda^{(k)} = \operatorname{diag}\left(\lambda_1^{(k)},\ldots,\lambda_n^{(k)}\right)$, we get

$$\widehat{q}\left(\mathbf{x}^{(k)}\right) = \sum_{i=1}^{n}\left(\xi_i^{(k)}\right)^2\mathbf{e}_i^{\mathsf{T}}\Lambda^{(k)}\mathbf{1}_n,$$

with $\xi_i^{(k)} \sim \mathcal{N}\left(\gamma_i^{(k)}, 1\right)$ denoting the i^{th} component of $\boldsymbol{\xi}^{(k)}$, $\mathbf{e}_i^{\mathsf{T}}$ the i^{th} standard basis (row) vector of $\mathbb{R}^n$, and $\mathbf{1}_n$ an $n \times 1$ vector containing all 1. Note that a single $\left(\xi_i^{(k)}\right)^2$ is a non-central chi-squared random variable with 1 degree of freedom and non-centrality parameter $\gamma_i^{(k)}$, so $\widehat{q}\left(\mathbf{x}^{(k)}\right)$ is indeed a sum of non-central chi-squared random variables, weighted by $\mathbf{e}_i^{\mathsf{T}}\Lambda^{(k)}\mathbf{1}_n$. The desired result follows by recalling that $P_D^{(k)} = \mathbb{P}\left(\widehat{q}\left(\mathbf{x}^{(k)}\right) \geq \widehat{h}\right)$ when $\mathbf{x}^{(k)}$ is an outlier. □

Unfortunately, the ccdf $\mathcal{Q}$ in (11) cannot be expressed in a closed form, but various series expansions and approximations can be found in [35]. Next, we derive an analytical formula for the probability of detection of the statistical test in (8) when an outlier is caused by a *deterministic* additive signal $\boldsymbol{f}^{(k)}$ at each corresponding nominal input data $\mathbf{x}^{(k)}$. Such a situation represents a change in the mean of the observation, which is ubiquitous when considering the monitoring and control of large-scale systems [36]. First, denote the ccdf for a non-central chi-squared distribution with k degrees of freedom and non-centrality parameter λ by $\mathcal{T}(\cdot; k, \lambda)$.

Corollary 2. Assume that an outlier is caused by a deterministic additive signal $\boldsymbol{f}^{(k)}$ on each corresponding nominal input data $\mathbf{x}^{(k)}_{\text{nominal}}$. For each data observation indexed by $k = 1, \ldots, m$, the differentially private statistical test in (8) designed to achieve a type I error probability of P_I, also achieves a true positive rate of

$$\begin{aligned} P_D^{(k)} &= 1 - P_{II}^{(k)} \\ &= \mathcal{T}\left(\widehat{h}; n; \left(\boldsymbol{f}^{(k)}\right)^{\mathsf{T}} \left(\Sigma + \kappa^2_{\delta,\epsilon}\rho^2 I_n\right)^{-1} \boldsymbol{f}^{(k)}\right), \end{aligned} \tag{12}$$

where $\widehat{h}$ is given in (9).

Proof. When an observation $\mathbf{x}^{(k)}$ is an outlier, we have that $\widehat{\mathbf{x}}^{(k)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\boldsymbol{\mu} + \boldsymbol{f}^{(k)}, \Sigma + \kappa^2_{\delta,\epsilon}\rho^2 I_n\right)$. It can be inferred from [33, Section 2.3] that $\widehat{q}\left(\mathbf{x}^{(k)}\right) \sim \chi^2_{n,\varphi}$, where $\varphi = \left(\boldsymbol{f}^{(k)}\right)^{\mathsf{T}} \left(\Sigma + \kappa^2_{\delta,\epsilon}\rho^2 I_n\right)^{-1} \boldsymbol{f}^{(k)}$ and $\chi^2_{n,\varphi}$ denotes a non-central chi-squared distribution with n degrees of freedom and non-centrality parameter φ. The desired result then follows by using the fact that that $P_D^{(k)} = \mathbb{P}\left(\widehat{q}\left(\mathbf{x}^{(k)}\right) \geq \widehat{h}\right)$ when $\mathbf{x}^{(k)}$ is an outlier. □

IV. Numerical Simulations

We apply our privacy-preserving outlier detection method to the REFIT data set, which contains aggregate household electricity usage data sampled at 8-second intervals from 20 houses in the United Kingdom between 2013-14 [26]. For each of the 20 houses, we average the total energy consumption in watts over a 24-hour period. By averaging over an entire day, we assume independence between observations of consumption. For each day indexed by k, we construct $\mathbf{x}^{(k)} \in \mathbb{R}^{20\times 1}_{\geq 0}$, and collect 255 signals of daily household consumption into $\mathcal{O}^{\text{RFT}}_m := \left\{\mathbf{x}^{(k)}\right\}_{k=1}^{k=m}$. We define a nominal energy consumption range per household as an interval of width 6 standard deviations around the mean, and fit a Gaussian distribution via maximum likelihood estimation to the consumption subset within this nominal range. The consumption histograms and best-fit Gaussian distribution can be seen in Figure 3. From $\mathcal{O}^{\text{RFT}}_m$, we then compute $\boldsymbol{\mu}_{\text{RFT}}$ and Σ_{RFT}.

We examine the scenario where a deterministic additive signal $\boldsymbol{f}^{(k)}$ is added to a subset of $\mathcal{O}^{\text{RFT}}_m$, and analyze the performance of the privacy-preserving statistical test in (8) using Corollary 2. Such an additive signal could represent persistent anomalous increases in household energy consumption. For our experiments, the faulty additive signal (in watts) is $\boldsymbol{f}^{(k)} = 150 \times \mathbf{1}_{20}$. Further, we set $\rho = 0.1$ and $\delta = 0.01$ in our experiments. First, we show that the classification performance deteriorates as more privacy is required, i.e., $\epsilon \to 0$. We fix the acceptable level of type I error via Theorem 3, vary ϵ, and plot the corresponding true positive rate $P_D = 1 - P_{II}$ in Figure 4. Note that the true positive rate approaches 1 for larger ϵ, i.e., less stringent privacy requirements, and deteriorates for $\epsilon \to 0$. This is expected, since $\kappa_{\epsilon,\delta} \propto 1/\sqrt{\epsilon}$.

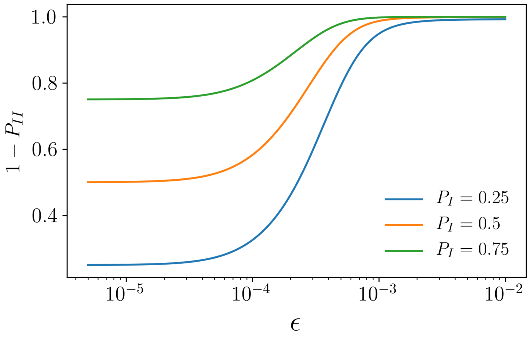


Fig. 4. True positive rate P_D as a function of ϵ in the setting of Corollary 2 with REFIT electricity consumption data, for fixed outlier thresholds $\widehat{h} = \mathcal{F}^{-1}(P_I; n)$ and non-degenerate $\boldsymbol{f}^{(k)} \neq \mathbf{0}$.

A. Receiver operating characteristic curve

We also characterize the performance of (8) across the entire range of thresholds $\widehat{h}$ by fixing the desired privacy level ϵ, varying $P_I \in [0, 1]$, and computing the true positive rate via Corollary 2. For each fixed ϵ, this provides us with the receiver operating characteristic (ROC) curve for our private outlier detection algorithm with any given fixed privacy level. We compute and plot several such ROC curves in Figure 5, each corresponding to a different privacy level ϵ. Again, we see that the performance of the differentially private outlier detection worsens for higher privacy requirements, as the area under the ROC curve for $\epsilon \approx 0$ is nearly half of the $\epsilon \gg 0$ case.

We demonstrated the privacy-accuracy trade-offs of using (8) on the REFIT data set of electricity consumption, i.e., enabling the release of information regarding whether or not certain daily consumption measurements $\mathbf{x}^{(k)} \in \mathcal{O}^{\text{RFT}}_m$ were outliers with respect to a historical baseline distribution $\mathcal{N}(\mu_{\text{RFT}}, \Sigma_{\text{RFT}})$, while guaranteeing privacy for each individual household. Analogous trade-offs and performance analyses can be carried out with a stochastic additive signal $\boldsymbol{f}^{(k)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\boldsymbol{\mu}_f, \Sigma_f)$, using the results in Theorem 4. The privacy-accuracy trade-off is qualitatively

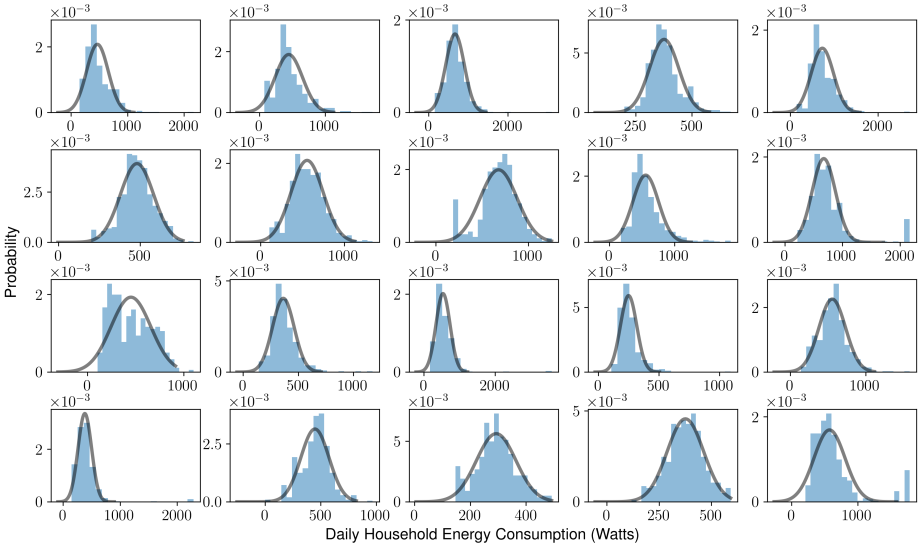


Fig. 3. Histograms of daily household energy consumption in watts for the 20 households in the REFIT data set, with best-fit Gaussian distributions overlaid. Note that houses 1 through 20 are ordered in a left-to-right, top-to-bottom arrangement.

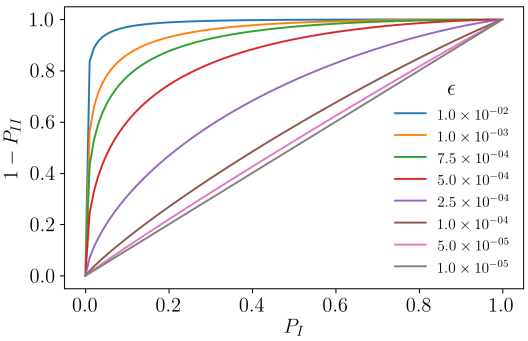


Fig. 5. ROC curves for different privacy levels (i.e., different ϵ) corresponding to the setting of Corollary 2 with REFIT electricity consumption data. Note that these ROC curves correspond to the non-degenerate case of $\boldsymbol{f}^{(k)} \neq \mathbf{0}$.

similar to that of Fig. 4 and 5, i.e., the deterministic additive signal case. Thus, we do not include these results in the paper.

V. Conclusion

We consider the problem of designing a squared Mahalanobis distance-based algorithm for outlier detection in correlated data under a differentially private constraint in this paper. We design an input perturbation architecture to preserve the privacy of individual data-contributing agents when detecting outliers, and we derive analytical formulas for detection thresholds, detection rates, and error rates of the differentially private algorithm. We then analyze the trade-off between detection accuracy and privacy level in a numerical example using a data set of household electricity consumption. Future research will consider the application of the proposed approach to differentially private fault detection in large-scale control systems with uncertainties.

References

[1] J. C. Herrera, D. B. Work, R. Herring, X. Ban, Q. Jacobson, and A. M. Bayen, “Evaluation of traffic data obtained via GPS-enabled mobile phones: The Mobile Century field experiment,” *Transportation Research Part C: Emerging Technologies*, vol. 18, no. 4, pp. 568 – 583, 2010.

[2] A. Molina-Markham, P. Shenoy, K. Fu, E. Cecchet, and D. Irwin, “Private memoirs of a smart meter,” ser. BuildSys ’10. New York, NY, USA: Association for Computing Machinery, 2010, p. 61–66.

[3] F. Xu, Z. Tu, Y. Li, P. Zhang, X. Fu, and D. Jin, “Trajectory recovery from ash: User privacy is not preserved in aggregated mobility data,” in *Proceedings of the 26th International Conference on World Wide Web*, 2017, pp. 1241–1250.

[4] A. Pyrgelis, C. Troncoso, and E. D. Cristofaro, “What does the crowd say about you? evaluating aggregation-based location privacy,” *Proceedings on Privacy Enhancing Technologies*, vol. 4, pp. 156–176, 2017.

[5] L. Sweeney, “k-anonymity: a model for protecting privacy,” *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, vol. 10, pp. 557–570, 2002.

[6] ——, “Only you, your doctor, and many others may know,” *Technology Science*, September 2015.

[7] L. Sweeney, M. Von Loewenfeldt, and M. Perry, “Saying it’s anonymous doesn’t make it so: Re-identifications of “anonymized” law school data,” *Technology Science*, November 2018.

[8] C. Dwork, F. McSherry, K. Nissim, and A. Smith, “Calibrating noise to sensitivity in private data analysis,” in *Proceedings of the Third Theory of Cryptography Conference*, 2006, pp. 265–284.

[9] C. Dwork, “Differential privacy,” in *Proceedings of the 33rd International Colloquium on Automata, Languages and Programming (ICALP)*, ser. Lecture Notes in Computer Science, vol. 4052. Springer-Verlag, 2006.

[10] C. Dwork and A. Roth, *The Algorithmic Foundations of Differential Privacy*. Foundations and Trends in Theoretical Computer Science, 2014, vol. 9, no. 3-4.

[11] J. Le Ny and G. Pappas, "Differential private filtering," *IEEE Transactions on Automatic Control*, vol. 59, no. 2, pp. 341–354, February 2014.

[12] Y. Wang, Z. Huang, S. Mitra, and G. E. Dullerud, "Differential privacy in linear distributed control systems: Entropy minimizing mechanisms and performance tradeoffs," *IEEE Transactions on Control of Network Systems*, vol. 4, no. 1, pp. 118–130, Mar. 2017.

[13] J. Cortés, G. E. Dullerud, S. Han, J. Le Ny, S. Mitra, and G. J. Pappas, "Differential privacy in control and network systems," in *2016 IEEE 55th Conference on Decision and Control (CDC)*. IEEE, 2016, pp. 4252–4272.

[14] J. He and L. Cai, "Differential private noise adding mechanism: Basic conditions and its application," in *2017 American Control Conference (ACC)*. IEEE, 2017, pp. 1673–1678.

[15] E. Nozari, P. Tallapragada, and J. Cortés, "Differentially private average consensus: Obstructions, trade-offs, and optimal algorithm design," *Automatica*, vol. 81, pp. 221–231, 2017.

[16] K. H. Degue and J. Le Ny, "Differentially private interval observer design with bounded input perturbation," in *2020 American Control Conference (ACC)*, 2020, pp. 1465–1470.

[17] J. Le Ny, "Differentially private nonlinear observer design using contraction analysis," *International Journal of Robust and Nonlinear Control*, 2018.

[18] M. Lyu, D. Su, and N. Li, "Understanding the sparse vector technique for differential privacy," *Proceedings of the VLDB Endowment*, vol. 10, no. 6, pp. 637–648, Feb. 2017.

[19] K. H. Degue, K. Gopalakrishnan, M. Z. Li, H. Balakrishnan, and J. Le Ny, "Differentially private outlier detection in multivariate gaussian signals," in *2021 American Control Conference (ACC)*, New Orleans, LA, USA, May 2021 (Accepted).

[20] M. Gaboardi, H. W. Lim, R. Rogers, and S. P. Vadhan, "Differentially Private Chi-Squared Hypothesis Testing: Goodness of Fit and Independence Testing," in *Proceedings of the 33rd International Conference on Machine Learning*, New York City, NY, USA, Jun. 2016, pp. 2111–2120.

[21] Y. Wang, J. Lee, and D. Kifer, "Revisiting differentially private hypothesis tests for categorical data," *ArXiv e-prints*, Mar. 2017.

[22] M. Ghassemi, A. D. Sarwate, and R. Wright, "Differentially private online active learning with applications to anomaly detection," in *Proceedings of the 9th ACM Workshop on Artificial Intelligence and Security*, October 2016.

[23] X. Tong, B. Xi, M. Kantarcioglu, and A. Inan, "Gaussian mixture models for classification and hypothesis tests under differential privacy," in *31st Annual IFIP WG 11.3 Conference on Data and Applications Security and Privacy (DBSec'17)*, Philadelphia, PA, USA, Jul. 2017.

[30] I. T. Jolliffe, *Principal Component Analysis*. Springer-Verlag, 1986.

[24] K. H. Degue and J. Le Ny, "On differentially private Gaussian hypothesis testing," in *Proceedings of the 56th Annual Allerton Conference on Communication, Control, and Computing*, Allerton Park and Retreat Center, Monticello, Illinois, USA, Oct. 2018.

[25] G. Liao, X. Chen, and J. Huang, "Social-aware privacy-preserving mechanism for correlated data," *IEEE/ACM Transactions on Networking*, vol. 28, no. 4, pp. 1671–1683, 2020.

[26] D. Murray, L. Stankovic, and V. Stankovic, "An electrical load measurements dataset of united kingdom households from a two-year longitudinal study," *Scientific Data*, vol. 4, no. 1, p. 160122, Jan 2017.

[27] J. Zheng, D. W. Gao, and L. Lin, "Smart meters in smart grid: An overview," in *2013 IEEE Green Technologies Conference (GreenTech)*, 2013, pp. 57–64.

[28] R. Okada, K. Fukuchi, and J. Sakuma, "Differentially private analysis of outliers," in *Proceedings of the 2015th European Conference on Machine Learning and Knowledge Discovery in Databases (ECML PKDD)*, Porto, Portugal, Sep. 2015, pp. 458–473.

[29] G. M. Mimmack, M. Mason, and J. Galpin, "Choice of distance matrices in cluster analysis: defining regions," *Journal of Climate*, vol. 14, p. 2790–2797, 2001.

[31] P. C. Mahalanobis, "On the generalized distance in statistics," in *Proceedings of the National Institute of Sciences of India*, vol. 2, no. 1, 1936, pp. 49–55.

[32] R. Cummings, S. Krehbiel, Y. Mei, R. Tuo, and W. Zhang, "Differentially private change-point detection," in *Proceedings of the 32nd International Conference on Neural Information Processing Systems (NeurIPS 2018)*, Montreal, Canada, Dec. 2018.

[33] S. M. Kay, *Fundamentals Of Statistical Processing, Volume 2: Detection Theory*, ser. Prentice Hall Signal Processing Series. Upper Saddle River: Prentice-Hall PTR, 2009.

[34] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor, "Our data, ourselves: Privacy via distributed noise generation," *Advances in Cryptology-EUROCRYPT*, vol. 4004, pp. 486–503, 2006.

[35] A. M. Mathai and S. B. Provost, *Quadratic Forms in Random Variables: Theory and Applications*. New York: Marcel Dekker, Inc., 1992.

[36] S. X. Ding, *Model-based Fault Diagnosis Techniques. Design Schemes, Algorithms and Tools*. Berlin: Marcel Dekker, Inc.