IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 15 August 2024, accepted 4 October 2024, date of publication 8 October 2024, date of current version 22 October 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3476209

RESEARCH ARTICLE

A Privacy-Preserving User-Centric Data-Sharing Scheme

VENIAMIN BOIARKIN[1], BRUNO BOGAZ ZARPELAO[2], JAFAR AL-ZAILI[1], AND MUTTUKRISHNAN RAJARAJAN[1], (Senior Member, IEEE)
[1]School of Science and Technology, City, University of London, EC1V 0HB London, U.K.
[2]Department of Computer Science, State University of Londrina, Londrina, Paraná 86057-970, Brazil

Corresponding author: Veniamin Boiarkin (veniamin.boiarkin@city.ac.uk)

ABSTRACT Using raw sensitive data of end-users helps service providers manage their operations efficiently and provide high-quality services to end-users. Although access to sensitive information benefits both parties, it poses several challenges concerning end-user privacy. Most data-sharing schemes based on differential privacy allow control of the level of privacy, which is not straightforward for end-users and leads to unpredictable utility. To address this issue, a novel local differentially private data-sharing scheme is proposed featuring a bimodal probability distribution that allows determining the range of random variables from which the noise is drawn with high probability. Additionally, a local differentially private mechanism is introduced to regulate the amount of noise injected into the data to control data utility. These components are combined to make up a user-centric data-sharing scheme which provides the end-user with control over the utility of their data, with the level of privacy being calculated from individual utility preferences. The simulation results show that the proposed scheme allows keeping the utility within the boundaries defined by the end-user, while providing the maximum possible level of privacy. Furthermore, it allows injecting more noise into the data for the same error in utility compared to the Laplace mechanism.

INDEX TERMS Data utility, local differential privacy, personal data, privacy-preserving, probability distribution.

I. INTRODUCTION

The variety of sensitive data generated by end-users grows over time while the number of data-driven services also increases gradually. To use a service provided by a third party, customers need to share their sensitive information. For example, households need to share their electricity consumption data with the energy supplier that calculates the electricity bill. Another example is when patients share their health information with a hospital to predict genetic diseases [1]. The number of use cases where sensitive end-user data may be used is enormous and constantly expanding.

Although having access to sensitive end-user data benefits both parties - the end-user gets more accurate results, and a service provider can better manage its business - it also poses some privacy-related issues. For example, access to electricity consumption profiles of households can help local authorities make better decisions about where to install charge points for electric vehicles [2], whereas by analyzing these data, an adversary may gain some knowledge about the customers behavior [3]. According to regulations such as General Data Protection Regulation (GDPR) [4] and Health Insurance Portability and Accountability Act (HIPPA) [5], end-users must have control over the use of their sensitive data. Thus, it is essential to preserve end-user privacy when sharing data with third parties.

Differential Privacy (DP) has become one of the most widely used techniques to preserve privacy by injecting noise into sensitive data. The main idea of DP is that end-users report sensitive data to an aggregator through a secure (encrypted) channel. Then, the aggregator adds controllable noise to the aggregated data and shares it with a service provider for processing. The most notable problem of DP is that end-users still share raw data with a third party, which may cause a leakage of sensitive information in case of a key leakage or insider attack on the aggregator's side. Secondly,

The associate editor coordinating the review of this manuscript and approving it for publication was Yassine Maleh.

it cannot be used when a service provider needs access to each end-user's individual data to provide personalized services.

Local Differential Privacy (LDP) is a variation of DP that addresses the issue of sharing raw data with an aggregator. LDP enables end-users to inject noise into the data on their side before sharing it with a third party. As a result, in the event of a key leakage or insider attack on the third party's side, an adversary would only gain access to the noisy data.

Many DP and LDP-based data-sharing schemes utilize the Laplace randomized mechanism to perturb sensitive data with noise drawn from the Laplace probability distribution. These schemes allow end-users to specify the desired level of privacy, based on which a controllable amount of noise is injected into the data. However, they do not allow end-users to specify utility (the result accuracy), which could then determine the level of privacy. Instead of adjusting privacy levels directly, end-users might prefer to set limits (boundaries) on the extra cost incurred from added noise, i.e., the cost they are willing to pay for enhanced privacy. The privacy level would then be determined based on this extra cost or utility preference. This approach would make it easier for end-users to control utility while the level of privacy is automatically adjusted based on their utility preferences.

Another challenge in DP and LDP-based schemes is obtaining more privacy without reducing utility proportionally. In the Laplace mechanism, an intuitive approach to increase noise levels, and consequently enhance privacy for a given privacy budget, is to shift the distribution so that its mean is not centered around zero. However, this can lead to a significant reduction in utility. An alternative approach involves using a multimodal distribution, where random variables with opposite signs are drawn with equal probability. This means that similar amounts of noise with opposite signs, drawn from the probability distribution, can cancel each other out over time. In this paper, we refer to this concept as "noise compensation." Noise compensation can ensure that, over time, the accuracy of queries performed on noisy data does not significantly deviate from the results obtained by querying the original data.

To address the above mentioned challenges, this article proposes a novel user-centric data-sharing scheme that preserves privacy using LDP. The main contributions of this article are as follows:

- To enhance the end-users privacy and inject more noise into the data, a novel bimodal probability distribution is proposed that enables the control over the intervals from which a random variable can be drawn with high probability. This distribution allows to control the spread around the modes, as well as the ratio of probabilities around the modes and the mean.
- To ensure that end-user privacy is preserved, we propose a novel randomized mechanism. This mechanism is based on the proposed bimodal probability distribution and satisfies LDP.
- To enable the end-users to control the utility, a novel user-centric data-sharing scheme is designed that allows the end-user to determine the boundaries for the utility, while the maximum possible level of privacy is provided based on the end-user's utility preference.

The rest of this article is organized as follows. Related works are presented in Section II. Section III describes the theoretical background of this article. A novel LDP mechanism and a novel user-centric data-sharing scheme are presented in Section IV. Simulation results are presented in Section V. Finally, the conclusion is given in Section VI.

II. RELATED WORK

To preserve end-user privacy, DP-based models are widely used across different domains that include but are not limited to machine learning [6], Industrial Internet of Things [7], and healthcare [8]. Most of the privacy preserving data-sharing schemes based on DP focus on injecting noise into end-users' aggregated data on the service provider's side. For example, Wei et al. [1] proposed a DP-based genetic matching scheme to achieve effective genetic matching and protect genetic data before outsourcing it to an untrusted cloud server for diagnosing patients' diseases. End-users share their sensitive data with a gene provider that injects noise into the aggregated data, which means that the end-users' private data may be disclosed in case of an attack on the gene provider's side. In [1], the level of privacy is chosen by following other works.

To protect the location privacy of both workers and tasks in a location-based crowdsourcing service, Wei et al. [9] proposed a novel DP-based location protection scheme. End-users share their sensitive information, including precise locations, with a cellular service provider. The provider then adds noise to the data and sends the noisy data to a server for processing. However, this process carries the risk of personal data leakage. The level of privacy is determined by the service provider, which means that end-users do not have control over their privacy. In [10], a novel traffic estimation scheme using DP is proposed to protect the vehicles' data in vehicular cyber-physical systems. A vehicle shares its location data with the roadside unit after the authentication using Public Key Infrastructure (PKI). The roadside unit perturbs aggregated data and submits it to the central server for future analysis. A key leakage attack on the roadside unit's side may cause the leakage of vehicle's data. In this case, the level of privacy is determined by the roadside unit (service provider). To achieve privacy and high estimation accuracy, the model's parameters must be set properly, but there is still no clear understanding of the relationship between privacy level and accuracy.

Instead of injecting noise on the aggregator's side, the noise can be injected on the end-user's side using a LDP-based mechanism. For example, Zheng et al. [11] proposed a novel recommendation system scheme by combining a matrix factorization algorithm with LDP to prevent the leakage of end-users sensitive information. End-users' sensitive data are perturbed using the Laplace mechanism on the end-user's side based on personalized privacy requirements and then sent to an aggregator. Similar to other works, the accuracy of the model increases with decreasing level of privacy (amount of noise). End-users may adjust their level of privacy, whereas it is still unclear how the change in the end-user's privacy level affects the accuracy of the result. In [12], a novel game-theoretic federated learning framework using DP is proposed to prevent malicious clients from compromising private information of other parties through inference attacks. After performing local training, end-users perturb the trained model parameters and submit these data to a central server. To enhance the global model accuracy, the amount of noise injected on the end-user's side should be reduced, resulting in a low level of privacy. To reduce the risk of industrial data leakage in the process of deep model training, Jiang et al. [13] proposed a new federated edge learning scheme using hybrid DP for industrial data processing. After training a local model, the edge terminal (end-user) generates and injects noise into the parameters of the local model, after which the noisy data are sent to the central server that generates the parameters of the new global model. To achieve better accuracy of the training model, the level of privacy chosen is relatively low compared to other works. As expected, the training loss decreases with the decreasing level of privacy (amount of noise). An optimal level of privacy is chosen (adjusted) based on the results of several simulations. To protect node feature and graph structure information against a malicious data curator, Lin et al. [14] designed a novel privacy-preserving framework for decentralized network graphs based on graph neural networks using edge LDP. The central server sends a query to the end-users, whereas each end-user sends an obfuscated answer (noisy data) back to the server. The proposed model achieves high accuracy for the predefined level of privacy, which means that the level of privacy (model's parameters) is chosen based on the number of simulations. A novel privacy-preserving data aggregation scheme satisfying LDP is presented in [15] to prevent the disclosure of end-user's electricity usage habits and daily activities in the smart grid. A smart meter, which is deployed on the end-user's side, measures the electricity consumption, perturbs electricity usage data using randomized response, and submits noisy data to the aggregator.

The use of DP-based schemes implies that end-users share their sensitive data with a service provider that determines the level of privacy according to which the noise is injected into the aggregated data. Thus, end-users do not have control either over their privacy or utility. Because of sharing raw data with a third party, the end-users privacy may be disclosed due to a key leakage or insider attack on the aggregator's side. On the contrary, the use of LDP schemes allows each end-user to determine an individual level of privacy, based on which the noise is injected into the data on the end-user's side before sharing it with a third party. Both DP and LDP-based data-sharing schemes allow control over the level of privacy. However, since there is no theoretically defined relationship between privacy level and model accuracy, users have no control over the resulting data utility, even when privacy levels are predefined. Most existing schemes rely on the results of simulations and select the most appropriate level of privacy (model's parameters) that should be used to achieve high accuracy.

To address these limitations, we propose a novel data-sharing scheme based on LDP that allows end-users to choose the tolerated error in utility (result accuracy), with the corresponding privacy level being automatically determined based on their preferences. Additionally, this work introduces a new bimodal probability distribution that enables the injection of more noise to perturb raw data while minimizing the impact on data utility through its noise compensation feature.

III. PRELIMINARIES

In this section, the theoretical background and mathematical formulation of LDP, sensitivity of a function, and the Laplace mechanism are presented.

A. LOCAL DIFFERENTIAL PRIVACY

DP is one of the most popular privacy-preserving techniques, where a trusted centralized aggregator (data curator) accesses sensitive data of end-users, aggregates those data, and adds controllable noise to the aggregated data. In a real-world setting, it is challenging to determine whether a centralized aggregator operates honestly or not, as well as to guarantee that end-users' sensitive data are not shared (accessed) with malicious actors during the aggregation process. LDP has become a solution to overcome these limitations of DP, so that end-users locally perturb their sensitive data using a LDP mechanism, after which end-users share noisy data with a centralized aggregator. Thus, an adversary cannot obtain the sensitive data of end-users.

Definition 1 (ϵ-local differential privacy): Let x and y denote two neighboring datasets, where y can be produced by adding, removing, or modifying exactly one entry from x. A randomized mechanism $\mathcal{M} : D \rightarrow S$ satisfies ϵ-local differential privacy iff for any output $s \in S$, and two neighboring datasets $x, y \in D$:

$$\frac{Pr[\mathcal{M}(x) = s]}{Pr[\mathcal{M}(y) = s]} \leq e^{\epsilon} \tag{1}$$

where S is the set of all possible outputs that a mechanism $\mathcal{M}$ can produce, $Pr[\mathcal{M}(x) = s]$ is the probability of a randomized mechanism $\mathcal{M}$ outputting the result s given the input x, and ϵ is the privacy budget (level of privacy) that bounds the probability of $\mathcal{M}$ outputting the same result for any pair of neighboring datasets x, y [16]. A smaller value of ϵ provides stronger privacy guarantee, whereas large ϵ provides weak privacy guarantee [17].

B. LAPLACE MECHANISM

The Laplace mechanism has become one of the most popular techniques to preserve end-users privacy. Let $f : D \rightarrow \mathbb{R}^k$

denote the function that maps datasets (D) to real numbers. For example, $f(\cdot)$ may be a function that takes a dataset $x \in D$ as an input and calculates the mean $f(x) \in \mathbb{R}$. To introduce controllable noise to the result of $f(\cdot)$, the Laplace mechanism relies on the sensitivity ($\mathcal{L}_1$-sensitivity) of $f(\cdot)$.

Definition 2 ($\mathcal{L}_1$-sensitivity): Given a query function $f(\cdot)$, its $\mathcal{L}_1$-sensitivity Δf is the maximum $\mathcal{L}_1$ distance between the results of $f(\cdot)$ over any pair of neighboring datasets x and y, which is defined as follows [16]:

$$\Delta f = \max_{x,y} \|f(x) - f(y)\|_1 \tag{2}$$

The Laplace mechanism uses the Laplace probability distribution to generate noise. The Probability Density Function (PDF) of the Laplace distribution centered around 0 with the scale factor $b = \Delta f/\epsilon$ is defined as follows [16]:

$$Lap(x|b) = \frac{1}{2b} exp\left(-\frac{|x|}{b}\right) \tag{3}$$

where the scale factor b is calibrated according to the Δf. The Laplace mechanism is ϵ-differentially private [18], and has the following definition:

$$\mathcal{M}_L\left(x, f(\cdot), \epsilon\right) = f(x) + Lap\left(\frac{\Delta f}{\epsilon}\right) \tag{4}$$

IV. PROPOSED SCHEME

Preserving end-users' privacy is important, but in real-world applications, it is equally important to maintain control over the utility of the protected data. The main question in the DP domain is recognized to be the trade-off between privacy and utility [19]. Adding more noise to the data increases the level of privacy and may decrease the utility. End-users may find it challenging to understand how added noise will affect data utility. As a result, selecting an optimal privacy budget (ϵ) becomes difficult. Instead of adjusting ϵ, end-users could simply set the maximum change (error) in utility they can bear, which is usually more palpable for them.

This article proposes a novel user-centric data-sharing scheme utilizing LDP. In the proposed scheme, end-users determine the boundaries for utility change, that is the relative error in utility they are willing to tolerate. The noise within the randomized mechanism is generated using a novel probability distribution that allows to determine positive and negative ranges of random variables from which the noise will be drawn with high probability. Keeping the amount of noise within the specified ranges helps to keep the relative error in utility within the boundaries. Thus, based on the end-user's utility preference, the proposed scheme adjusts its parameters to inject the right amount of noise and to provide the maximum possible level of privacy.

A. BOIARKIN PROBABILITY DISTRIBUTION

The proposed privacy-preserving scheme relies on a new probability distribution, named the Boiarkin distribution. When the Laplace distribution is used in differential privacy mechanisms, its mean (μ) is set to 0 [18], [20]. As a result, the noise values drawn from the distribution are centered around zero. In the Boiarkin distribution, we have two non-zero means with opposite signs. For this reason, the noise values drawn from the mechanism based on the Boiarkin distribution are higher than those from the Laplace-based mechanism, but the utility is maintained since the opposite signs compensate each other in the long term. Additionally, the Boiarkin distribution includes some hyperparameters that allow fine-tuning, providing more precise control over the generated values. The rationale behind the proposed probability distribution is to provide control over the probability of a random variable $r = 0$ to be generated ($Pr[r = 0]$), as well as to control the amount of noise drawn from the distribution by increasing the probability $Pr[r \neq 0]$. The PDF of the Boiarkin probability distribution is defined as follows:

$$Boi(x|b, \psi, \mu) = q \cdot exp\left(-\frac{\left|\psi - |x - \mu|\right|}{b}\right) \tag{5}$$

where b is the scale of the distribution, ψ is the spread of the modes, μ is the shift, and q is the normalizing factor, which is defined as follows:

$$q = \frac{1}{2b\left(2 - exp\left(-\frac{\psi}{b}\right)\right)} \tag{6}$$

The Boiarkin distribution is a symmetric bimodal probability distribution centered at μ (mean of the distribution) and has two modes, namely $\lambda_1 = -\psi$ and $\lambda_2 = \psi$. The purpose of having two modes is to determine the ranges from which random variables (with opposite signs) can be drawn with high probability. Since continuously generating non-zero random variables may seriously affect the utility, it is important to have the noise compensation feature. When using the Boiarkin probability distribution with increased $Pr[r \neq 0]$, more non-zero random variables will be generated around the modes λ_1 and λ_2. Since the negative and positive modes are on the same distance (spread ψ) from the mean $\mu = 0$, many non-zero random variables will be generated around λ_1 and λ_2 but with opposite signs, which means that they will compensate each other. Thus, the Boiarkin distribution enables the noise compensation feature, while providing control over the intervals from which random variables can be drawn with high probability.

Let c (Fig. 1) denote the probability of a random variable $r = 0$ to be generated ($c = Pr[r = 0]$), and q denote the probability of a non-zero random variable $r = \lambda_1 = \lambda_2$ to be generated ($q = Pr[r = \lambda_1] = Pr[r = \lambda_2]$). To be able to control the ratio of probabilities of generating zero (c) and non-zero (q) random variables, the ratio p ($p \in (0; 1]$) of these probabilities is introduced. The ratio p allows to fine-tune the amount of noise drawn from the probability distribution and is defined as follows:

$$p = \frac{c}{q} = exp\left(\frac{\lambda_1}{b}\right) \tag{7}$$

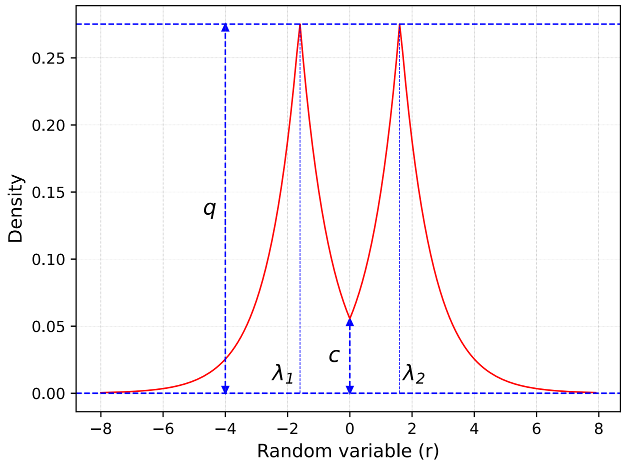

FIGURE 1. **Probability density function of the Boiarkin probability distribution with scale factor $b = 1$ and $p = 0.2$.**

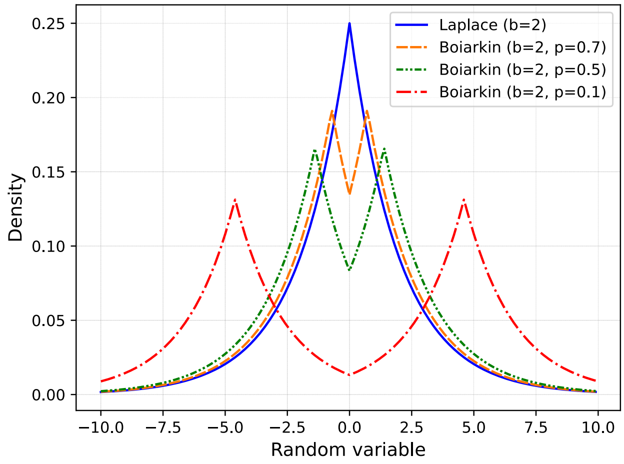

FIGURE 2. **Probability density functions of the Laplace distribution with the scale factor $b = 2$ and Boiarkin distribution with the scale factor $b = 2$ and $p = 0.1, 0.5, 0.7$.**

Fig. 1 shows the PDF of the Boiarkin probability distribution with the scale factor $b = 1$, and $p = 0.2$. It can be seen that the probabilities around the modes λ_1 and λ_2 are higher compared to the mean ($\mu = 0$). More precisely, the probabilities $Pr[r = \lambda_1]$ and $Pr[r = \lambda_2]$ are 5 times higher than $Pr[r = 0]$, which is controlled by p.

Fig. 2 shows the PDF of the Laplace distribution with the scale factor $b = 2$ and Boiarkin distribution with the scale factor $b = 2$ and $p = 0.1, 0.5, 0.7$. By decreasing p, $Pr[r = 0]$ decreases, whereas the probabilities $Pr[r = \lambda_1]$ and $Pr[r = \lambda_2]$ increase. Thus, by decreasing p, more random variables are generated around λ_1 and λ_2. When $p = 1$, i.e., the $Pr[r = \mu] = Pr[r = \lambda_1] = Pr[r = \lambda_2]$, the Boiarkin distribution becomes the Laplace distribution with the scale factor b centered at μ.

B. BOIARKIN MECHANISM

In this section, to preserve end-user privacy, a novel ϵ-LDP mechanism is introduced, which utilizes the proposed Boiarkin probability distribution to generate noise.

Definition 3 (The Boiarkin Mechanism): Given any function $f : D \to \mathbb{R}^k$, the Boiarkin mechanism is defined as follows:

$$\mathcal{M}_{\mathcal{B}}(x, f(\cdot), \epsilon, \psi) = f(x) + (Y_1, \ldots, Y_k) \tag{8}$$

where Y_i are i.i.d. random variables drawn from $Boi(\Delta f/\epsilon, \psi, 0)$ (5).

Theorem 1: The Boiarkin mechanism $\mathcal{M}_{\mathcal{B}}(x, f(\cdot), \epsilon, \psi)$ preserves ϵ-LDP for any end-user with personal privacy budget ϵ.

Proof: Let $x \in \mathcal{D}$ and $y \in \mathcal{D}$ be any neighboring datasets, differing in one entry, and $f(\cdot)$ be some function $f : \mathcal{D} \to \mathbb{R}^k$. Let $P_x(z)$ denote the probability density function of $\mathcal{M}_{\mathcal{B}}(x, f(\cdot), \epsilon, \psi)$, and let $P_y(z)$ denote the probability density function of $\mathcal{M}_{\mathcal{B}}(y, f(\cdot), \epsilon, \psi)$. To prove ϵ-local differential privacy [16], it is shown that the ratio $P_x(z)/P_y(z)$ is bounded by $exp(\epsilon)$ at any arbitrary point $z \in \mathbb{R}^k$.

$$\begin{aligned}
&\frac{P_x(z)}{P_y(z)} \\
&= \prod_{i=1}^{k} \left(\frac{exp\left(-\frac{\epsilon\left|\psi - |z_i - f(x)|\right|}{\Delta f}\right)}{exp\left(-\frac{\epsilon\left|\psi - |z_i - f(y)|\right|}{\Delta f}\right)} \right) \\
&= \prod_{i=1}^{k} exp\Bigg(-\frac{\epsilon}{\Delta f}\left|\psi - |z_i - f(x)|\right| \\
&\quad + \frac{\epsilon}{\Delta f}\left|\psi - |z_i - f(y)|\right|\Bigg) \\
&= \prod_{i=1}^{k} exp\left(\frac{\epsilon}{\Delta f}\left(\left|\psi - |z_i - f(y)|\right| - \left|\psi - |z_i - f(x)|\right|\right)\right) \\
&\leq \prod_{i=1}^{k} exp\left(\frac{\epsilon}{\Delta f}\left|\psi - |z_i - f(y)| - \psi + |z_i - f(x)|\right|\right) \\
&= \prod_{i=1}^{k} exp\left(\frac{\epsilon}{\Delta f}\left||z_i - f(x)| - |z_i - f(y)|\right|\right) \\
&\leq \prod_{i=1}^{k} exp\left(|z_i - f(x) - z_i - f(y)|\right) \\
&= \prod_{i=1}^{k} exp\left(\frac{\epsilon}{\Delta f}|f(y) - f(x)|\right) \\
&= exp\left(\frac{\epsilon}{\Delta f}\sum_{i=1}^{k}|f(y) - f(x)|\right) \\
&= exp\left(\frac{\epsilon}{\Delta f}\|f(y) - f(x)\|_1\right) \\
&\leq exp(\epsilon)
\end{aligned} \tag{9}$$

where the first and second inequalities are triangle inequalities, and the last inequality follows from the definition of sensitivity (2). Therefore, the Boiarkin mechanism ensures that the probability of a given outcome is nearly the same (bounded by $exp(\epsilon)$) for any neighboring datasets $x, y \in D$ differing in one entry.

C. USER-CENTRIC DATA-SHARING SCHEME

Most of DP privacy-preserving mechanisms imply the adjustment or use of a predefined privacy budget ϵ to control the trade-off between the level of privacy and utility. The question regarding this trade-off has been studied by many researchers [21], [22], [23], whereas there is still no clear answer on how to chose the right ϵ.

In this section, a novel user-centric privacy-preserving mechanism is proposed, which enables the end-users to choose an acceptable error in utility, whereas the level of privacy (an optimal privacy budget ϵ) is automatically calculated by the scheme. The proposed scheme utilizes the Boiarkin ϵ-LDP mechanism, which relies on the Boiarkin probability distribution to fine-tune the amount of generated noise. Let $U : D \to \mathbb{R}$ denote the utility function that maps input datasets (D) to real numbers, and it is defined as follows:

$$U(d) = g(d), \quad d \in D \tag{10}$$

where $g(\cdot)$ is the function that takes d as an input and outputs the result for a particular use case by performing mathematical operations on d. For example, $g(\cdot)$ may be a function that just returns d, or it may be a function that outputs the energy usage cost for a household by multiplying the average energy consumption per day (d) by the number of time slots and purchasing price for energy per kWh.

Let δ denote the maximum acceptable relative error in utility, which is expressed as follows:

$$\delta = \frac{U(d') - U(d)}{U(d)} \cdot 100\% \tag{11}$$

where d is the result of a query function $f(\cdot)$ (original data), and d' is the output of a randomized mechanism $\mathcal{M}(\cdot)$. Thus, (11) may be rewritten as follows:

$$\delta = \frac{U(\mathcal{M}(x)) - U(f(x))}{U(f(x))} \cdot 100\% \tag{12}$$

where x is the input dataset. Taking into account that a randomized mechanism adds noise to the result of a query function (8), (12) may be rewritten as follows:

$$\delta = \frac{U(f(x) + r) - U(f(x))}{U(f(x))} \cdot 100\% \tag{13}$$

where r is the noise drawn from a probability distribution used by a randomized mechanism. Expressing the variable r (13), the maximum acceptable noise can be found for the chosen error δ. Note that an equation for r would be different depending on $U(\cdot)$.

To make sure that the noise drawn from a probability distribution does not exceed the maximum acceptable noise, the boundaries for this probability distribution are calculated using its Cumulative Distribution Function (CDF). The CDF of the Boiarkin probability distribution centered at 0 is defined as follows:

$$F_X(x) = \begin{cases} q \cdot b \cdot exp\left(\frac{x - \lambda_1}{b}\right), \; x \leq \lambda_1 \\ \frac{1}{2} - q \cdot b \cdot \left(exp\left(\frac{\lambda_1 - x}{b}\right) - exp\left(\frac{\lambda_1}{b}\right)\right), \; \lambda_1 \leq x \leq 0 \\ \frac{1}{2} + q \cdot b \cdot \left(exp\left(\frac{x - \lambda_2}{b}\right) - exp\left(\frac{-\lambda_2}{b}\right)\right), \; 0 \leq x \leq \lambda_2 \\ 1 - q \cdot b \cdot exp\left(\frac{\lambda_2 - x}{b}\right), \; \lambda_2 \leq x \end{cases} \tag{14}$$

To find a point x at which the CDF of the Boiarkin distribution gives the required level of the CDF's accuracy α ($\alpha = 0.9999$), x has to be expressed from the last equation in (14). Taking into account that $p = exp(\lambda_1/b)$ and $p = exp(-\lambda_2/b)$, x is expressed as follows:

$$x = -b \cdot log(p) - b \cdot log(2(1 - \alpha)(2 - p)) \tag{15}$$

By replacing x with the maximum acceptable noise, scale b or spread ψ of the Boiarkin probability distribution may be adjusted, so that a random variable r drawn from the probability distribution lies within the acceptable interval $[-x, x]$. For the Laplace probability distribution, x would be expressed as follows:

$$x = -b \cdot log(2(1 - \alpha)) \tag{16}$$

Based on the end-user's preference regarding the maximum acceptable error in utility (δ), parameters of the probability distribution used by the randomized mechanism can be adjusted, so that δ (12) is within the acceptable range (e.g., $\delta \leq 50\%$). Note that end-users set the maximum acceptable error δ, which affects the parameters of the randomized mechanism, more specifically the probability distribution used by it. This means that the boundaries for the probability distribution are calculated based on the end-user's preference regarding δ, and this configuration of the probability distribution gives the maximum possible level of privacy (range of random variables).

V. RESULTS

This section presents the simulation results to evaluate the proposed data-sharing scheme. In this work, a smart grid environment is studied as a specific use case, whereas the model can be applied to other scenarios.

In the simulated use case, a smart meter deployed on the end-user's side measures the electricity consumption of a household and submits the average electricity consumption per hour to the energy supplier once per day. In the smart grid, end-users energy usage data may be shared with different third parties, including the electricity system operator that controls the balance between supply and demand. In this case study, the data are only shared with the energy supplier and used exclusively for billing. All other scenarios, including the influence on the network operation, are out of the scope of this article. End-users' individual electricity consumption data is classified as personal data under GDPR. For this reason, end-users must have control over who can access their electricity consumption data, how often, and for what purposes, except when the data access is mandated [24]. In addition, in case of an insider attack or a key leakage attack on the energy supplier's side, an adversary would be able to access the raw electricity consumption data of the end-users.

To make sure that the privacy of end-users is not disclosed, as well as to prevent the consequences of cyber-attacks on the energy supplier's side, households may choose to use a LDP scheme to protect their privacy and data, whereas the final energy usage cost may be increased due to the noise injected into the electricity usage data. Thus, end-users decide the extra cost they are willing to pay to increase their privacy level, based on which the noise will be added to the data. The real electricity consumption data for 100 households in London are taken from [25], which contains smart meter readings from 2011 to 2014. The buying price of energy from the utility grid in the UK is taken from [26] and is equal to 14.37 pence/kWh. The objective of this use case is to evaluate if the error in utility stays within the boundaries determined by the end-user, as well as how the proposed mechanism performs compared to the Laplace mechanism. The simulations are conducted on a machine with Apple M1 CPU @3.2GHz and 8.0GB RAM using Python programming language.

Let $f(x)$ denote the function that calculates the average electricity consumption per hour for a household, which is defined as follows:

$$f(x) = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{17}$$

where x is the vector that contains electricity consumption for each time slot (30 minutes) in a day. Let $g(f(x))$ denote the function that takes the average electricity consumption of a household and calculates the energy usage cost for one day, which is defined as follows:

$$g(f(x)) = f(x) \cdot T \cdot \gamma \tag{18}$$

where T is the number of time slots in a day ($T = 48$), and γ is the buying price of energy per kWh ($\gamma = 14.37$ pence/kWh).

By combining (17) and (18), the maximum acceptable relative error in utility δ (12) is calculated as follows:

$$\begin{aligned}\delta &= \frac{(f(x) + r) \cdot T \cdot \gamma - f(x) \cdot T \cdot \gamma}{f(x) \cdot T \cdot \gamma} \cdot 100\% \\ &= \frac{r}{f(x)} \cdot 100\% \end{aligned} \tag{19}$$

where r is the random variable (noise) drawn from a probability distribution used by a mechanism $\mathcal{M}(\cdot)$. When the end-user decides on the acceptable relative error in the energy usage cost, the boundaries for the probability distribution can be calculated as follows:

$$r = \frac{\delta \cdot f(x)}{100} \tag{20}$$

By combining (15) and (20), an adjusted scale for the Boiarkin probability distribution is calculated as follows:

$$b = -\frac{\delta \cdot f(x)}{100(log(p) + log(2(1-\alpha)(2-p)))} \tag{21}$$

By combining (16) and (20), the scale for the Laplace distribution is adjusted as follows:

$$b = -\frac{\delta \cdot f(x)}{100(log(2(1-\alpha)))} \tag{22}$$

Therefore, based on the end-user's preference regarding δ, the scale of a probability distribution used by a mechanism $\mathcal{M}(\cdot)$ is adjusted to make sure that the noise is within the boundaries, which helps to control the change (error) in utility.

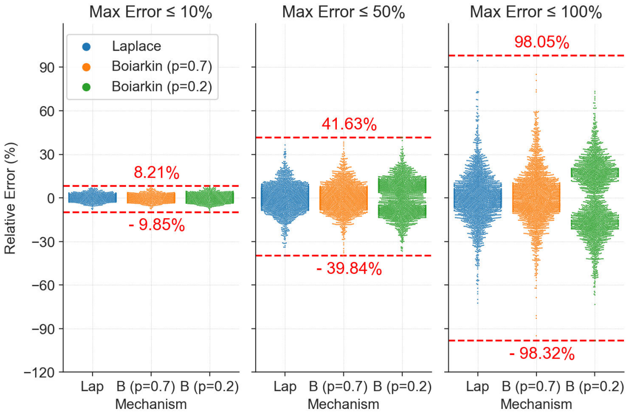

FIGURE 3. **Dependency between the relative error in utility σ and the maximum acceptable error δ for randomly chosen end-users for 1 day using the Laplace mechanism and the Boiarkin mechanism with $p = 0.7$ and $p = 0.2$.**

Let σ denote the actual relative error in utility. The dependency between the relative error in utility σ and the maximum acceptable error δ set by the end-user using the Laplace and Boiarkin mechanisms for randomly chosen end-users is shown in Fig. 3. To demonstrate how p affects the relative error σ, first we set $p = 0.7$. This p value close to 1 makes the modes of the Boiarkin distribution close to 0, which means that most of the noise will be concentrated around 0. The second value picked for p ($p = 0.2$) is closer to zero, which increases the spread of the modes of the Boiarkin distribution. As a result, most of the noise will be concentrated around λ_1 and λ_2. Since end-users send data to the energy supplier once per day, Fig. 3 shows the relative error in the energy usage cost (σ) for one day depending on the maximum acceptable error (δ), 10%, 50%, or 100%. It can be seen that σ does not exceed δ. The Laplace mechanism results in a concentration of relative errors around 0% because the noise (a random variable r) drawn from the Laplace distribution has the highest probability around 0. The Boiarkin mechanism results in similar relative error as the Laplace mechanism when p is set close to 1 because the modes λ_1 and λ_2 of the Boiarkin distribution are close to 0. When decreasing p

for the Boiarkin mechanism, the concentration of the relative errors around 0% decreases because the noise is concentrated around the modes λ_1 and λ_2, which can be clearly observed when the maximum acceptable error $\delta \leq 100\%$ and $p = 0.2$. Compared to the Laplace mechanism, the Boiarkin mechanism allows to increase the spread of relative errors while keeping it within the boundaries by fine-tuning the amount of generated noise using the parameter p. Note that Fig. 3 shows the results for only one day (iteration), so there is no room for noise compensation. When $p = 0.2$, although the noise is frequently not zero, positive and negative values have the same magnitude, which can result in compensation in the long term.

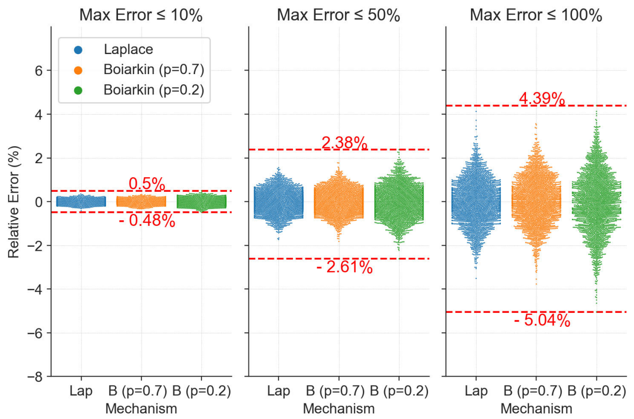

FIGURE 4. **Dependency between the relative error in utility σ and the maximum acceptable error δ for randomly chosen end-users over a period of 300 days using the Laplace mechanism and Boiarkin mechanism with $p = 0.7$ and $p = 0.2$.**

To show the effect of noise compensation, the simulation was conducted for randomly chosen end-users for a period of 300 days using the Laplace mechanism and the Boiarkin mechanism with $p = 0.7$ and $p = 0.2$. The dependency between the relative error in utility σ and the maximum acceptable error δ is shown in Fig. 4. It can be clearly observed that the relative error in utility is smaller than the maximum acceptable error because of the noise compensation for both the Laplace and Boiarkin mechanisms. The concentration of the relative errors for the Laplace and Boiarkin mechanisms is around 0%, whereas the spread of relative errors using the Boiarkin mechanism is slightly higher but does not exceed the maximum acceptable error δ. When a randomized mechanism is executed only once (Fig. 3), $U(\mathcal{M}(x))$ may result in both higher and lower values compared to the result of $U(x)$, which may result in positive or negative relative error. Over time, with the mechanism being executed multiple times, the relative error becomes smaller because of the noise compensation. Unlike the Laplace mechanism, the Boiarkin mechanism allows to fine-tune the amount of noise injected into the data, which results in a slightly higher relative error when the mechanism is executed only once (see Fig. 3), while over time, due to the noise compensation, both the Laplace and Boiarkin mechanisms yield comparable relative errors, all of which fall within the user-defined boundaries (see Fig. 4). Thus, fine-tuning the parameter p of the Boiarkin mechanism allows to inject more noise, while resulting in the same relative error over time compared to the Laplace mechanism.

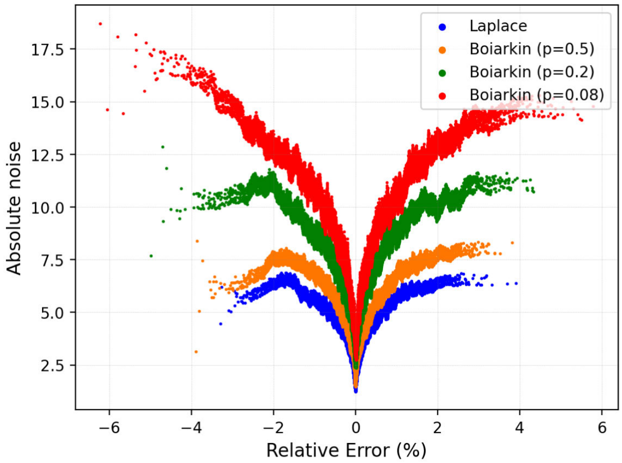

FIGURE 5. **Dependency between the relative error σ and the absolute amount of noise added to the data for randomly chosen end-users over a period of 300 days using the Laplace mechanism and the Boiarkin mechanism with different p, where the maximum acceptable error in utility δ is set to 100%.**

To show the dependency between the amount of noise injected by a randomized mechanism and relative error σ, the simulation was conducted for randomly chosen end-users over a period of 300 days using the Laplace mechanism and the Boiarkin mechanism with $p = 0.5$, $p = 0.2$, and $p = 0.08$ (Fig. 5). The maximum acceptable error in utility δ is set to 100%, which means that the maximum energy usage cost the end-user is willing to pay should not exceed the double of the original cost. It can be observed that the Boiarkin mechanism allows to inject more noise compared to the Laplace mechanism for the same relative error, which may provide better privacy for the end-user with the same utility. When p is set close to 1, the amount of noise is slightly higher compared to the noise injected by the Laplace mechanism. With the decreasing p, the amount of noise increases because the probability of a random variable $Pr[r = 0]$ decreases, and the probabilities around the modes (λ_1 and λ_2) of the Boiarkin distribution increase. Since the Boiarkin probability distribution has two modes, around which the probabilities are concentrated, and taking into account the effect of noise compensation, it is possible to inject more noise and produce the same error in utility as the Laplace mechanism.

The results above showed that the Boiarkin mechanism allows to inject more noise compared to the Laplace, whereas the final error in utility is the same. To check whether more noise means more privacy, the mutual information score between the noisy electricity consumption profiles and the original profiles for randomly chosen end-users over a period of 300 days is evaluated for the Laplace and

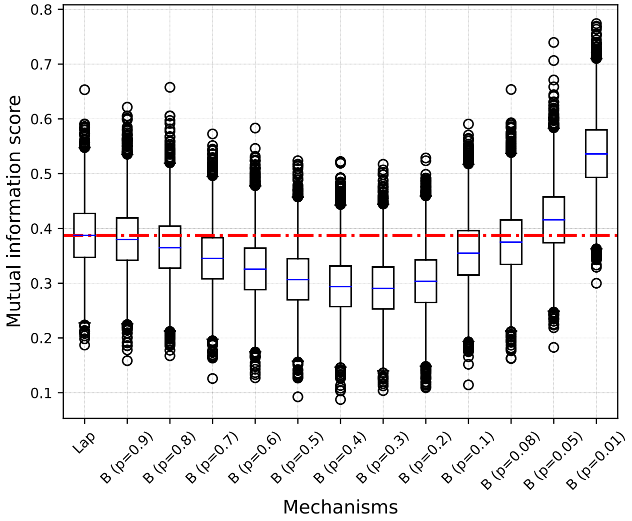

FIGURE 6. Mutual information score between the noisy and original electricity consumption profiles for different end-users over a period of 300 days using the Laplace mechanism and the Boiarkin mechanism with different p, where the maximum acceptable error in utility δ is set to 100%.

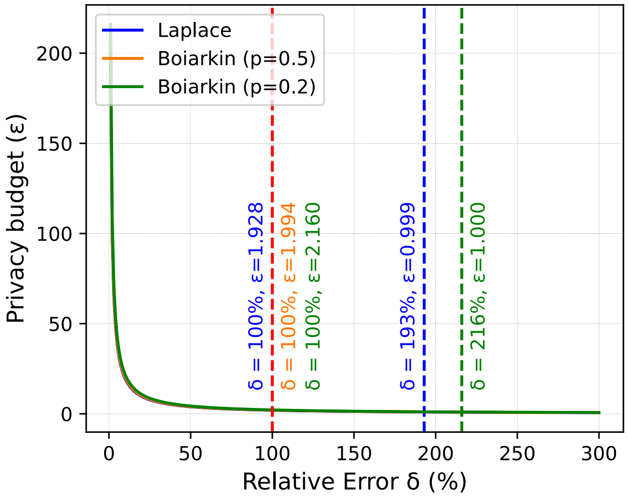

FIGURE 7. Dependency between the privacy budget ϵ and the maximum acceptable relative error in utility δ for a randomly chosen end-user for one day using the Laplace mechanism and the Boiarkin mechanism with $p = 0.5$ and $p = 0.2$.

Boiarkin mechanisms (Fig. 6). The maximum acceptable error in utility δ set by end-users is equal to 100%. The mutual information score reflects the extent to which the noisy electricity consumption profile is similar to the original profile of the end-user. The lower the mutual information score, the more independent electricity consumption profiles are. It can be seen that for the Boiarkin mechanism, the mutual information score decreases when p decreases ($p \geq 0.2$), whereas when p becomes too small ($p < 0.2$), the mutual information score increases compared to the Laplace mechanism. Thus, when increasing the probability around the modes (λ_1 and λ_2) of the Boiarkin distribution, more noise can be injected into the data, which makes noisy and original electricity consumption profiles of the end-user more independent. To keep noisy and original electricity consumption profiles of the end-user more independent and preserve the privacy of end-users, it is suggested to use $p \geq 0.2$. These results also show that the parameter p of the Boiarkin mechanism allows to fine-tune the amount of noise injected into the data, while still keeping the relative error within the boundaries defined by the end-user. For $p = 0.2$, the Boiarkin mechanism provides more privacy compared to the Laplace mechanism since the mutual information score between the noisy and original profiles is lower than for the Laplace mechanism. When $p = 0.9$, the Boiarkin mechanism provides slightly better privacy compared to the Laplace mechanism because the modes λ_1 and λ_2 (Fig. 2) of the Boiarkin distribution are close to 0.

To understand how the end-user's preference regarding δ affects the privacy budget ϵ used by the Laplace and Boiarkin mechanisms, the simulation was conducted for a randomly chosen end-user for one day using the global sensitivity of a query function $f(\cdot)$ (17) (Fig. 7). The global sensitivity reflects the maximum distance between two neighboring datasets. The level of electricity consumption for a typical household is around 4kWh, and by changing one entry in an empty dataset (no electricity consumption) to the maximum level of consumption (4kWh), the global sensitivity is calculated as $\Delta f = 4/48 = 0.08333$, where 48 is the number of elements (time slots) in the dataset. After calculating the sensitivity of the function Δf, our next step is to determine the appropriate scale, denoted as b. This scale will be determined based on the maximum acceptable error in utility δ. Subsequently, we calculate ϵ, taking into account the global sensitivity and the adjusted scale ($\epsilon = \Delta f / b$). It can be seen that to keep the relative error around 0%, the privacy budget has to be very high ($\epsilon \approx 200$) because the scale of the probability distribution used by the randomized algorithm should produce random variables within a very limited range, which depends on the sensitivity and may be different for other applications. If the end-user sets the maximum acceptable error $\delta = 100\%$, the privacy budget for the Laplace mechanism $\epsilon = 1.928$, whereas $\epsilon = 1.994$ and $\epsilon = 2.160$ for the Boiarkin mechanism with $p = 0.5$ and $p = 0.2$ respectively. When $\epsilon \leq 1$, the relative error in utility increases, namely the Laplace mechanism with $\epsilon = 0.999$ results in 193% error, whereas the Boiarkin mechanism with $p = 0.2$ and $\epsilon = 1$ results in 216% error.

Although the privacy budget ϵ is slightly higher for the Boiarkin mechanism, the amount of noise injected by the Boiarkin mechanism is higher, as well as the mutual information score between noisy and original data is lower compared to the Laplace mechanism, which means that the Boiarkin mechanism may provide better privacy for the same ϵ. In the scheme proposed in this work, when the end-user defines the boundaries for utility (δ), the privacy budget ϵ is calculated automatically based on the sensitivity and the scale of the probability distribution. Since the scale of the probability distribution is adjusted based on δ, which allows to control the amount of noise, there is only one ϵ that can be calculated for the given δ and Δf. This provides the maximum possible level of privacy for the defined δ. On the other hand, if the end user had to choose the privacy budget ϵ instead of the maximum tolerated error, the error in the utility they would get in the future would be unclear. By utilizing the proposed data-sharing scheme, end-users have control over the utility, as well as better understanding how the level of privacy affects utility.

VI. CONCLUSION

In this article, a novel user-centric privacy-preserving data sharing scheme is proposed. First, a novel bimodal probability distribution has been proposed that provides control over the ranges of random variables from which the noise is drawn with high probability, and enables the noise compensation. Second, a novel privacy-preserving mechanism that satisfies ϵ-LDP has been introduced. The proposed Boiarkin mechanism allows adding more noise, compared to the Laplace mechanism, while the relative error in utility over time is the same. Finally, a novel user-centric data-sharing scheme has been designed that allows the end-users to specify the boundaries for utility, whereas the maximum possible level of privacy (privacy budget ϵ) is provided by the scheme. The amount of noise injected into the data can be fine-tuned by adjusting the parameter p of the Boiarkin mechanism, which affects the relative error for one iteration of the mechanism. According to the mutual information score analysis, injecting more noise using the Boiarkin mechanism ($p \geq 0.2$) makes the noisy and original data more independent. Thus, the Boiarkin mechanism can provide better privacy compared to the Laplace mechanism, whereas the relative error in utility is always within the boundaries defined by the end-user. A smart grid environment is studied as a particular use case, where the model can be further extended for different application scenarios.

The results obtained in this work can be used to enhance existing and design novel privacy-preserving user-centric data-sharing schemes in different application domains. For example, by utilizing the proposed Boiarkin mechanism, car owners may share the data about their vehicles such as speed and location with insurance companies in a privacy-preserving manner, whereas the level of privacy would affect insurance premiums.

As future work, we plan to study how to design a privacy-preserving user-centric federated learning mechanism, where end-users do not share their privacy preferences with a central server. Specifically, by utilizing the proposed Boiarkin mechanism, we aim to enable personalized privacy settings for each client while maintaining a high level of accuracy of a global model.

REFERENCES

[1] J. Wei, Y. Lin, X. Yao, J. Zhang, and X. Liu, "Differential privacy-based genetic matching in personalized medicine," *IEEE Trans. Emerg. Topics Comput.*, vol. 9, no. 3, pp. 1109–1125, Jul. 2021.

[2] Department for Energy Security & Net Zero (DESNZ). (2020). *Energy White Paper: Powering Our Net Zero Future*. Accessed: Sep. 3, 2023. [Online]. Available: https://www.gov.uk/government/publications/energy-white-paper-powering-our-net-zero-future

[3] European Data Protection Supervisor (EDPS). (2019). *TechDispatch #2: Smart Meters Smart Homes*. Accessed: Aug. 9, 2023. [Online]. Available: https://edps.europa.eu/data-protection/our-work/publications/techdispatch/techdispatch-2-smart-meters-smart-homes_en

[4] Official Journal of the European Union. (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council on the Protection of Natural Persons With Regard To the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation)*. Accessed: Aug. 9, 2023. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN

[5] U.S. Department of Health & Human Services. (1996). *Health Insurance Portability Accountability Act 1996 (HIPAA)*. Accessed: Aug. 20, 2023. [Online]. Available: https://www.cdc.gov/phlp/publications/topic/hipaa.html

[6] T. Zhu, D. Ye, W. Wang, W. Zhou, and P. S. Yu, "More than privacy: Applying differential privacy in key areas of artificial intelligence," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 6, pp. 2824–2843, Jun. 2022.

[7] B. Jiang, J. Li, G. Yue, and H. Song, "Differential privacy for industrial: Opportunities, applications, and challenges," *IEEE Internet Things J.*, vol. 8, no. 13, pp. 10430–10451, Jul. 2021.

[8] A. Krall, D. Finke, and H. Yang, "Mosaic privacy-preserving mechanisms for healthcare analytics," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 6, pp. 2184–2192, Jun. 2021.

[9] J. Wei, Y. Lin, X. Yao, and J. Zhang, "Differential privacy-based location protection in spatial crowdsourcing," *IEEE Trans. Services Comput.*, vol. 15, no. 1, pp. 45–58, Jan. 2022.

[10] Y.-E. Sun, H. Huang, W. Yang, S. Chen, and Y. Du, "Toward differential privacy for traffic measurement in vehicular cyber-physical systems," *IEEE Trans. Ind. Informat.*, vol. 18, no. 6, pp. 4078–4087, Jun. 2022.

[11] X. Zheng, M. Guan, X. Jia, L. Guo, and Y. Luo, "A matrix factorization recommendation system-based local differential privacy for protecting users' sensitive data," *IEEE Trans. Computat. Social Syst.*, vol. 10, no. 3, pp. 1189–1198, Jun. 2023.

[12] L. Zhang, T. Zhu, P. Xiong, W. Zhou, and P. S. Yu, "A robust game-theoretical federated learning framework with joint differential privacy," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 4, pp. 3333–3346, Apr. 2023.

[13] B. Jiang, J. Li, H. Wang, and H. Song, "Privacy-preserving federated learning for industrial edge computing via hybrid differential privacy and adaptive compression," *IEEE Trans. Ind. Informat.*, vol. 19, no. 2, pp. 1136–1144, Feb. 2023.

[14] W. Lin, B. Li, and C. Wang, "Towards private learning on decentralized graphs with local differential privacy," *IEEE Trans. Inf. Forensics Security*, vol. 17, pp. 2936–2946, 2022.

[15] N. Gai, K. Xue, B. Zhu, J. Yang, J. Liu, and D. He, "An efficient data aggregation scheme with local differential privacy in smart grid," *Digit. Commun. Netw.*, vol. 8, no. 3, pp. 333–342, Jun. 2022.

[16] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Found. Trends Theor. Comput. Sci.*, vol. 9, nos. 3–4, pp. 211–407, Aug. 1561.

[17] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," *J. Privacy Confidentiality*, vol. 7, no. 3, pp. 17–51, May 2017.

[18] M. B. Gough, S. F. Santos, T. AlSkaif, M. S. Javadi, R. Castro, and J. P. S. Catalão, "Preserving privacy of smart meter data in a smart grid environment," *IEEE Trans. Ind. Informat.*, vol. 18, no. 1, pp. 707–718, Jan. 2022.

[19] L. Lyu, K. Nandakumar, B. Rubinstein, J. Jin, J. Bedo, and M. Palaniswami, "PPFA: Privacy preserving fog-enabled aggregation in smart grid," *IEEE Trans. Ind. Informat.*, vol. 14, no. 8, pp. 3733–3744, Aug. 2018.

[20] Z. Zheng, T. Wang, A. K. Bashir, M. Alazab, S. Mumtaz, and X. Wang, "A decentralized mechanism based on differential privacy for privacy-preserving computation in smart grid," *IEEE Trans. Comput.*, vol. 71, no. 11, pp. 2915–2926, Nov. 2022.

[21] M. Li, Y. Tian, J. Zhang, D. Fan, and D. Zhao, "The trade-off between privacy and utility in local differential privacy," in *Proc. Int. Conf. Netw. Netw. Appl. (NaNA)*, Oct. 2021, pp. 373–378.

[22] M. A. Mustafa, S. Cleemput, A. Aly, and A. Abidin, ''A secure and privacy-preserving protocol for smart metering operational data collection,'' *IEEE Trans. Smart Grid*, vol. 10, no. 6, pp. 6481–6490, Nov. 2019.

[23] B. Rassouli and D. Gündüz, ''Optimal utility-privacy trade-off with total variation distance as a privacy measure,'' *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 594–603, 2020.

[24] Department for Energy Security & Net Zero (DESNZ). (2018). *Smart Metering Implementation Programme: Review of the Data Access and Privacy Framework*. Accessed: Sep. 10, 2023. [Online]. Available: https://www.gov.uk/government/publications/smart-metering-implementation-programme-review-of-the-data-access-and-privacy-framework

[25] U.K. Power Networks. (2022). *SmartMeter Energy Consumption Data in London Households*. Accessed: Sep. 8, 2023. [Online]. Available: https://data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london-households

[26] UKPower. (2021). *Compare Energy Prices Per KWh*. Accessed: Sep. 8, 2023. [Online]. Available: https://www.ukpower.co.uk/home_energy/tariffs-per-unit-kwh

VENIAMIN BOIARKIN is currently pursuing the Ph.D. degree with the Department of Engineering, City, University of London, London, U.K. His current research interests include cyber-security, blockchain, the Internet of Things, and data privacy.

BRUNO BOGAZ ZARPELAO received the B.Sc. degree in computer science from the State University of Londrina (UEL), Brazil, and the Ph.D. degree in electrical engineering from the University of Campinas, Brazil. In 2012, he joined UEL, where he is currently an Associate Professor with the Computer Science Department. From March 2018 to February 2019, he was a Visiting Postdoctoral Researcher with the City, University of London. His research interests include security analytics, intrusion detection, and the Internet of Things.

JAFAR AL-ZAILI is a member of the Turbomachinery and Energy Systems Research Group and a Senior Lecturer in power and prolusion at City. He held several engineering positions, such as a Design Engineer, a Consultant, a Research and Development Engineer, and the Project Manager prior to joining City, in 2015. Since then, he has been involved in several research projects, mainly in the field of small-scale power generation for renewable energy applications. His current research focus is on low-carbon and hydrogen micro gas turbine power systems for distributed power generation and propulsion system for small air vehicles, impact of the distributed generation in large cities on the level of pollution, pricing and market design for distributed generation and microgrids, high-temperature modular compact thermal energy storage, pollution reduction for combustion of hydrogen as a carbon-free fuel, and the dynamics between low-carbon power technologies and supporting policies.

MUTTUKRISHNAN RAJARAJAN (Senior Member, IEEE) is currently a Professor of security engineering with the City, University of London, U.K., where he leads the Information Security Group. He is the Director of the Institute for Cyber Security, City, University of London. He is a Visiting Researcher with the British Telecommunication's Security Research and Innovation Laboratory. He has published well over 300 articles and continues to be involved in the editorial boards and technical program committees of several international security and privacy conferences and journals. His research interests include privacy-preserving data analytics, cloud computing, the Internet of Things security, and wireless networks. He is an Advisory Board Member of the Institute of Information Security Professionals, U.K.; and acts as an Advisor to the U.K. Government's Identity Assurance Program.