

Noiseless Privacy

Farhad Farokhi

CSIRO's Data61

The University of Melbourne

farhad.farokhi@data61.csiro.au

farhad.farokhi@unimelb.edu.au

Abstract—In this paper, we define noiseless privacy, as a non-stochastic rival to differential privacy, requiring that the outputs of a mechanism (i.e., function composition of a privacy-preserving mapping and a query) can attain only a few values while varying the data of an individual (the logarithm of the number of the distinct values is bounded by the privacy budget). Therefore, the output of the mechanism is not fully informative of the data of the individuals in the dataset. We prove several guarantees for noiselessly-private mechanisms. The information content of the output about the data of an individual, even if an adversary knows all the other entries of the private dataset, is bounded by the privacy budget. The zero-error capacity of memory-less channels using noiselessly private mechanisms for transmission is upper bounded by the privacy budget. The performance of a non-stochastic hypothesis-testing adversary is bounded again by the privacy budget. Finally, assuming that an adversary has access to a stochastic prior on the dataset, we prove that the estimation error of the adversary for individual entries of the dataset is lower bounded by a decreasing function of the privacy budget. In this case, we also show that the maximal information leakage is bounded by the privacy budget. In addition to privacy guarantees, we prove that noiselessly-private mechanisms admit composition theorem and post-processing does not weaken their privacy guarantees. We prove that quantization operators can ensure noiseless privacy if the number of quantization levels is appropriately selected based on the sensitivity of the query and the privacy budget. Finally, we illustrate the privacy merits of noiseless privacy using multiple datasets in energy and transport.

Index Terms—data privacy; noiseless privacy; non-stochastic information theory; hypothesis testing.

I. INTRODUCTION

Big data revolution, equipped with novel tools for data collection, analysis, and reporting, has significant promises for answering societal challenges. These promises however come at the cost of erosion of privacy. Therefore, there is a need for rigorous protection of the privacy of individuals.

Natural candidates for privacy protection, such as differential privacy [1], [2] and information-theoretic privacy [3], [4], require randomized policies for privacy protection. The definition of differential privacy assumes the use of randomized functions as well as the probability of outputs, and conventional information-theoretic tools, such as mutual information and entropy, rely on random variables.

Heuristic-based privacy-preserving methods, such as k -anonymity [5], [6] and ℓ -diversity [7], are however deterministic in nature. They employ deterministic mechanisms, such as suppression and generalization, and do not assume stochastic

properties about the datasets. Popularity of these methods is evident from the availability of toolboxes for implementation¹.

Although providing powerful guarantees, randomized or stochastic privacy-preserving policies sometimes cause problems, such as un-truthfulness [8], that are undesirable in practice [9]. This is touted as a reason behind slow adoption of differential privacy within financial and health sectors [8]. For instance, randomized policies, stemming from differential privacy in financial auditing, complicate fraud detection [10], [11]. Randomized policies can also generate unreasonable and unrealistic outputs that might mislead investors or market operators, e.g., by reporting noisy outputs that point to lack of liquidity in a financial sector while that was not the case. For instance, the slow-decaying nature of the Laplace noise means impossible reports (e.g., negative median income) can occur with a non-negative probability [12]. Randomized privacy-preserving policies have also encountered difficulties in medical, health, or social sciences [13], [14]. Furthermore, the Laplace mechanism, a common approach to ensuring differential privacy, is shown to cause undesirable properties, e.g., the optimal estimation in the presence of Laplace noise is computationally expensive [15]. These motivate the development of non-stochastic privacy metrics and privacy-preserving policies in a rigorous manner.

Although it has been proved that noiseless policies cannot provide the strong guarantees of randomized policies, e.g., it has been proved that differential privacy cannot be delivered without noise [2], the popularity of noiseless privacy-preserving policies justifies investigating metrics for analysis and comparison. This must be done irrespective of their inherent philosophical weaknesses in comparison to stochastic policies because they belong to a different category.

In this paper, we define the new notion of noiseless privacy. Noiseless privacy implies that the outputs of a mechanism can only attain a few distinct values while varying the data of an individual. Therefore, the output of the mechanism is not very informative about the data of the individuals in a dataset. We prove the following guarantees for the noiselessly-private mechanisms:

- The information content of the output about the data of each individual, even if an adversary knows all the other entries of the private dataset, is bounded from above by the privacy budget (a constant similar to the privacy

¹<https://arx.deidentifier.org/overview/related-software/>

budget in differential privacy capturing the amount of the leaked information). As non-stochastic notions of information, we use non-stochastic information leakage in [16] and the maximin information [17]. These are established measures of information in the non-stochastic information theory literature [16]–[19].

- Zero-error capacity of memory-less channels using noiselessly-private mechanisms for data transmission is upper bounded by the privacy budget. Zero-error capacity is the non-stochastic equivalent of normal capacity, also coined by Shannon while investigating non-stochastic communication channels and worst-case behaviours [20].
- The performance of an adversary performing non-stochastic hypothesis tests [21] on the data of an individual, while knowing all the other entries of the private dataset, is bounded again by the privacy budget.
- Assuming that an adversary has access to a stochastic prior about the dataset, we prove that the error of an adversary for estimating the data each individual is lower bounded by a decreasing function of the privacy budget. Therefore, by reducing the privacy budget, the estimation error of the adversary worsens. In this case, we also show that the maximal information leakage (in the sense of [65]) is upper bounded by the privacy budget. Hence, by reducing the privacy budget, we can also reduce the maximal information leakage.

In addition to these privacy guarantees, we prove the following important properties:

- Noiselessly-private mechanisms admit composition theorem, i.e., the privacy budgets of the mechanisms add up when reporting on multiple queries on the same private dataset.
- Post-processing of noiselessly-private mechanisms does not weaken their privacy guarantees, i.e., the privacy budget can only be increased by post-processing.

We also prove that quantization operators can ensure noiseless privacy. We provide a recipe for determining the number of quantization levels based on the sensitivity of the query and the privacy budget. Finally, we illustrate the privacy merits of noiseless privacy using multiple datasets in energy and transport.

A. Related Studies

Anonymization: Anonymization is widely used within public and private sectors for releasing sensitive datasets² for public competitions and analysis. Although popularly adopted, anonymization is often insufficient for privacy preservation [22]–[24] and hence, systematic methods with provable guarantees are required.

Multi-Party Computation and Encryption: We may use secure multi-party computation, for instance based on homomorphic encryption, to compute aggregate statistics or machine learning models [25]–[31]. Secure multi-party computation and homomorphic encryption introduce massive com-

putation and communication overheads. They also do not fully eliminate the risk of privacy breaches, e.g., risks associated with dis-aggregation attacks still remain if these algorithms are not paired with other privacy-preserving techniques.

Differential Privacy: Differential privacy offers provable privacy guarantees [2], [32]–[37]. This method uses randomization to provide plausible deniability for the data of an individual by ensuring that the statistics of privacy-preserving outputs do not change significantly by varying the data of individual. Additive Laplace and Gaussian noise with scales proportional to the sensitivity of the submitted query with respect to the individual entries of the dataset are proved to guarantee differential privacy [2]. By definition, differential privacy requires randomization.

Information-Theoretic Privacy: Information-theoretic privacy, a rival to differential privacy, dates back to studying secrecy [38] and its generalizations [3], [4], [39], [40]. Information-theoretic guarantees have been also used to measure the quantity of leaked private information when using differential privacy [41], [42]. In information-theoretic privacy, entropy, mutual information, Kulback-Leiber divergence, and Fisher information have been repeatedly used as measures of privacy [43]–[49]. Information theory, starting with Shannon [50], assumes that data source and communication channels are random, and is powerful in modelling and analysing communication systems. However, traditional notions in information theory, such as mutual information, are not useful for analysing non-stochastic/noiseless settings and deterministic privacy-preserving policies.

Deterministic Privacy-Preserving Policies: Noiseless privacy-preserving policies are often heuristic-based making them vulnerable to attacks, e.g., k -anonymity is vulnerable to homogeneity attack [7]. This is because we do not possess sensible measures/definitions for privacy that extend to noiseless privacy-preserving policies on deterministic datasets. Therefore, we cannot prove, in any sense, privacy guarantees of noiseless privacy-preserving (even if weak or limited in scope or practice). The popularity of noiseless privacy-preserving policies justifies investigating metrics for analysis and comparison. In this paper, we propose a rival to differential privacy that is noiseless. We use non-stochastic information theory, non-stochastic hypothesis testing, stochastic estimation theory to investigate the merits of this definition. Non-stochastic information theory dates back to early studies of information transmission [17], [51]–[54]. It has been recently used in engineering [55]–[57]. Most recently, non-stochastic information theory was used in [16] for investigating deterministic privacy-preserving policies. Interestingly, in [16], it was easily proved that k -anonymity is not privacy-preserving using non-stochastic information theory, a fact that was only observed using adversarial attacks in [7].

B. Paper Organization

The rest of the paper is organized as follows. Background material on non-stochastic information theory and hypothesis testing are presented in Section II. Noiseless privacy is defined

²See <https://data.gov.au> and <https://www.kaggle.com> for examples.

in Section III. In this section, guarantees and properties of noiseless privacy are also presented. A method for ensuring noiseless privacy is presented in Section IV. Experimental results are presented in Section V. Finally, the paper is concluded in Section VI.

II. UNCERTAIN VARIABLES, HYPOTHESIS TESTING, AND NON-STOCHASTIC INFORMATION THEORY

We start by reviewing necessary concepts from non-stochastic information theory, particularly, uncertain variables, non-stochastic information leakage, and hypothesis testing.

A. Uncertain Variables

Let Ω be an uncertainty set/space whose elements, i.e., $\omega \in \Omega$, model/capture the source of uncertainty. An uncertain variable X is defined as a mapping on Ω . For any uncertain variable $X : \Omega \rightarrow \mathbb{X}$, $X(\omega)$ is the realization of uncertain variable X (corresponding to the realization of uncertainty $\omega \in \Omega$). *Marginal range* of uncertain variable X is $\llbracket X \rrbracket := \{X(\omega) : \omega \in \Omega\} \subseteq \mathbb{X}$. *Joint range* of uncertain variables $X : \Omega \rightarrow \mathbb{X}$ and $Y : \Omega \rightarrow \mathbb{Y}$ is defined as $\llbracket X, Y \rrbracket := \{(X(\omega), Y(\omega)) : \omega \in \Omega\} \subseteq \mathbb{X} \times \mathbb{Y}$. *Conditional range* of uncertain variable X , conditioned on the realizations of uncertain variable Y belonging to the set $\mathcal{Y}$, i.e., $Y(\omega) \in \mathcal{Y} \subseteq \llbracket Y \rrbracket$, is given by $\llbracket X|\mathcal{Y} \rrbracket := \{X(\omega) : \exists \omega \in \Omega \text{ such that } Y(\omega) \in \mathcal{Y}\} \subseteq \llbracket X \rrbracket$. If $\mathcal{Y}$ is a singleton, i.e., $\mathcal{Y} = \{y\}$, we use $\llbracket X|y \rrbracket$ instead of $\llbracket X|\{y\} \rrbracket = \llbracket X|\mathcal{Y} \rrbracket$. The definition of uncertain variables and their properties are similar to those of random variables with the exception of not requiring a measure on Ω . Finally, if the marginal range $\llbracket X \rrbracket$ is uncountably infinite for an uncertain variable X , we refer to X as a *continuous* uncertain variable, similar to a continuous random variable. If the marginal range $\llbracket X \rrbracket$ is countable for an uncertain variable X , we call X a *discrete* uncertain variable.

B. Non-Stochastic Information Theory

Non-stochastic entropy of discrete uncertain variable X is

$$H_0(X) := \log_2(|\llbracket X \rrbracket|) \in \mathbb{R} \cup \{\pm\infty\}. \quad (1)$$

This is commonly referred to as the Hartley entropy [17], [51]. For continuous uncertain variable X , the non-stochastic (differential) entropy is given by

$$h_0(X) := \log_e(\mu(|\llbracket X \rrbracket|)) \in \mathbb{R} \cup \{\pm\infty\}, \quad (2)$$

where $\mu(\cdot)$ is the Lebesgue measure. This is sometimes referred to as Rényi differential 0-entropy [17]. The authors of [17], [58] define the non-stochastic conditional entropy of uncertain variable X , conditioned on uncertain variable Y , as

$$H_0(X|Y) := \max_{y \in \llbracket Y \rrbracket} \log_2(|\llbracket X|y \rrbracket|), \quad (3)$$

for discrete uncertain variables X and Y . Similarly, for continuous uncertain variables X and Y , we get

$$h_0(X|Y) := \text{ess sup}_{y \in \llbracket Y \rrbracket} \log_e(\mu(\llbracket X|y \rrbracket)). \quad (4)$$

Now, we can define non-stochastic information between uncertain variables X and Y as the difference of the entropy of X with and without access to realizations of Y . Hence, for discrete uncertain variables, non-stochastic information can be defined as

$$\begin{aligned} I_0(X; Y) &:= H_0(X) - H_0(X|Y) \\ &= \min_{y \in \llbracket Y \rrbracket} \log_2 \left(\frac{|\llbracket X \rrbracket|}{|\llbracket X|y \rrbracket|} \right). \end{aligned} \quad (5)$$

For continuous uncertain variables, non-stochastic information can be similarly defined as $I_0(X; Y) := h_0(X) - h_0(X|Y)$. It is clear that the non-stochastic information is in fact not symmetric, i.e., $I_0(X; Y) \neq I_0(Y; X)$ in general.

With slight adaptation, Kolmogorov had previously defined ‘combinatorial’ conditional entropy using $\log(|\llbracket X|y \rrbracket|)$ and the information gain as $|\llbracket X \rrbracket|/|\llbracket X|y \rrbracket|$ in [52]. The combinatorial conditional entropy and the information gain are only defined for a fixed realization $Y(\omega) = y$ while (5) is based on the worst-case scenario.

In [16], it was observed that, in the context of information-theoretic privacy, the non-stochastic information (5) is not a good measure of information leakage and therefore, the non-stochastic information leakage was proposed as

$$L_0(X; Y) := \max_{y \in \llbracket Y \rrbracket} \log_2 \left(\frac{|\llbracket X \rrbracket|}{|\llbracket X|y \rrbracket|} \right), \quad (6)$$

for discrete uncertain variables. Similarly, for continuous uncertain variables, the non-stochastic information leakage was defined as

$$L_0(X; Y) := \text{ess sup}_{y \in \llbracket Y \rrbracket} \log_e \left(\frac{\mu(\llbracket X \rrbracket)}{\mu(\llbracket X|y \rrbracket)} \right). \quad (7)$$

In general, the non-stochastic information I_0 and non-stochastic information leakage L_0 are not equal, i.e., $I_0(X; Y) \neq L_0(Y; X)$. In fact, from the definition, it is easy to see that $I_0(X; Y) \leq L_0(X; Y)$. Further, $L_0(X; Y)$ is not symmetric. We propose the symmetrized non-stochastic information leakage as

$$L_0^s(X; Y) := \min(L_0(X; Y), L_0(Y; X)). \quad (8)$$

Note that, by construction, $L_0^s(X; Y) = L_0^s(Y; X)$.

a) *Maximin Information:* In [17], the maximin information was introduced as a symmetric measure of information and its relationship with zero-error capacity was explored. To present the definition of the maximin information, we need to introduce the notion of overlap partitions:

- $x, x' \in \llbracket X \rrbracket$ are $\llbracket X|Y \rrbracket$ -overlap connected, or in short $x \rightsquigarrow x'$, if there exists a finite sequence of conditional ranges $\{\llbracket X|y_i \rrbracket\}_{i=1}^n$ such that $x \in \llbracket X|y_1 \rrbracket$, $x' \in \llbracket X|y_n \rrbracket$, and $\llbracket X|y_i \rrbracket \cap \llbracket X|y_{i+1} \rrbracket \neq \emptyset$ for all $i = 1, \dots, n-1$;
- $\mathcal{A} \subseteq \llbracket X \rrbracket$ is $\llbracket X|Y \rrbracket$ -overlap connected if all $x, x' \in \mathcal{A}$ are $\llbracket X|Y \rrbracket$ -overlap connected;
- $\mathcal{A}, \mathcal{B} \subseteq \llbracket X \rrbracket$ are $\llbracket X|Y \rrbracket$ -overlap isolated if there does not exist $x \in \mathcal{A}, x' \in \mathcal{B}$ such that x, x' are $\llbracket X|Y \rrbracket$ -overlap connected;

- An $\llbracket X|Y \rrbracket$ -overlap partition is a partition of $\llbracket X \rrbracket$ such that each member set is $\llbracket X|Y \rrbracket$ -overlap connected and all two member sets are $\llbracket X|Y \rrbracket$ -overlap isolated.

Symmetry, i.e., $x \rightsquigarrow x'$ implies that $x' \rightsquigarrow x$, and transitivity, i.e., $x \rightsquigarrow x'$ and $x' \rightsquigarrow x''$ implies that $x \rightsquigarrow x''$, guarantee that a unique $\llbracket X|Y \rrbracket$ -overlap partition always exists [17]. The unique $\llbracket X|Y \rrbracket$ -overlap partition is shown by $\llbracket X|Y \rrbracket_\star$ in what follows. The maximin information is

$$I_\star(X; Y) := \log_2(|\llbracket X|Y \rrbracket_\star|). \quad (9)$$

In [17], it was proved that $|\llbracket X|Y \rrbracket_\star| = |\llbracket Y|X \rrbracket_\star|$ and thus $I_\star(X; Y) = I_\star(Y; X)$. We now prove an important result regarding the relationship between non-stochastic information leakage and maximin information.

Proposition 1. *For discrete uncertain variable Y , $I_\star(X; Y) \leq L_0^s(X; Y)$.*

Proof. See Appendix A. $\square$

An uncertain time series X is a sequence of uncertain variables $X[k] : \Omega \rightarrow \mathbb{X}$ for all $k \in \mathbb{N}$. Alternatively, we can think of uncertain time series X as a mapping from the sample space Ω to the set of discrete-time functions $\mathbb{X}^\infty := \{\forall x : \mathbb{N} \rightarrow \mathbb{X}\}$.

Now, we can define a memory-less uncertain communication channel. A memory-less uncertain channel maps any uncertain time series X to uncertain time series Y such that

$$\begin{aligned} \llbracket Y[k], \dots, Y[1] | X[k](\omega) = x[k], \dots, X[1](\omega) = x[1] \rrbracket \\ = \llbracket Y[k] | X[k](\omega) = x[k] \rrbracket \times \dots \times \llbracket Y[1] | X[1](\omega) = x[1] \rrbracket, \end{aligned}$$

for all $(x[k], \dots, x[1]) \in \llbracket X[k], \dots, X[1] \rrbracket$ and all $k \in \mathbb{N}$. A code of length k is a finite set $\mathcal{F} \subseteq \mathbb{X}^k$ with each codeword $f \in \mathcal{F}$ denoting a distinct message. Define

$$\begin{aligned} \mathcal{X}(y[k], \dots, y[1]) \\ := \llbracket X[k], \dots, X[1] | Y[k](\omega) = y[k], \dots, Y[1](\omega) = y[1] \rrbracket. \end{aligned}$$

The zero-error capacity is

$$C_0 := \lim_{k \rightarrow \infty} \sup_{\substack{\mathcal{F} \subseteq \mathbb{X}^k : \\ |\mathcal{F} \cap \mathcal{X}(y[k], \dots, y[1])| \leq 1, \\ \forall (y[k], \dots, y[1]) \in \mathbb{Y}^k}} \frac{\log_2(|\mathcal{F}|)}{k}. \quad (10)$$

In what follows, we only consider sequence of discrete uncertain variables $Y[k]$. Now, we are ready to relate the symmetrized non-stochastic information leakage to zero-error capacity of memory-less uncertain channels.

Proposition 2. *Any memory-less uncertain channel satisfies $C_0 \leq \sup_{\llbracket X[k] \rrbracket \subseteq \mathbb{X}} L_0^s(X[k]; Y[k])$.*

Proof. See Appendix B. $\square$

C. Non-Stochastic Hypothesis Testing

Consider uncertain variable X denoting the original uncertain variable. An adversary is interested in testing the validity of a hypothesis for the realizations of X . The adversary does not have access to realizations of this uncertain variable as

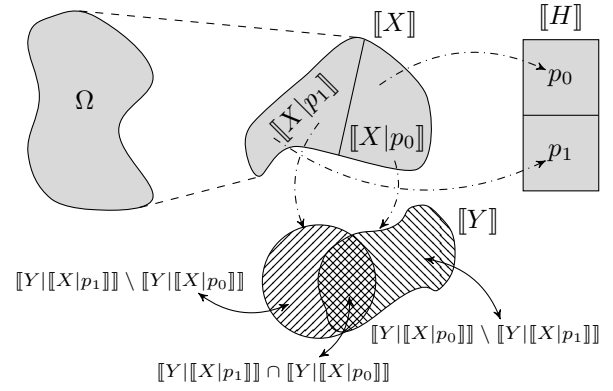


Fig. 1: Relationship between uncertain variables in non-stochastic hypothesis testing based on uncertain measurements. If the realization of uncertain measurement Y belongs to $\llbracket Y|p_0 \rrbracket \cap \llbracket Y|p_1 \rrbracket$, there is not enough evidence to accept or reject the null hypothesis p_0 or the alternative hypothesis p_1 . However, if the realization of uncertain measurement Y belongs to $(\llbracket Y|p_0 \rrbracket \setminus \llbracket Y|p_1 \rrbracket) \cup (\llbracket Y|p_1 \rrbracket \setminus \llbracket Y|p_0 \rrbracket)$, we can confidently accept (reject) the null hypothesis p_0 and reject (accept) the alternative hypothesis p_1 .

otherwise hypothesis testing is trivial. Instead, it has access to an *uncertain measurement* of this variable denoted by Y . This is captured by that $Y(\omega) = g_Y(X(\omega))$ for a mapping $g_Y : \llbracket X \rrbracket \rightarrow \llbracket Y \rrbracket$. Recalling that uncertain variables are mappings from the uncertainty set, it must be that $Y = g_Y \circ X$, where $\circ$ denotes composition of mappings. Similarly, we may define the *hypothesis* as an uncertain variable H with binary range $\llbracket H \rrbracket = \{p_0, p_1\}$, where p_0 denotes the *null hypothesis* and p_1 denotes the *alternative hypothesis*. We assume that there exists a mapping $g_H : \llbracket X \rrbracket \rightarrow \llbracket H \rrbracket$ such that $H = g_H \circ X$; the hypothesis is constructed based on the uncertain variable X as $H(\omega) = g_H(X(\omega))$. This setup and the relationship between all uncertain variables is summarized in Figure 1.

A *test* is a function $T : \llbracket Y \rrbracket \rightarrow \llbracket H \rrbracket = \{p_0, p_1\}$. If $T(Y) = p_1$, the test rejects the null hypothesis in favour of the alternative hypothesis; however, if $T(Y) = p_0$, the test accepts the null hypothesis (and rejects the alternative hypothesis). The set of all tests is given by $\mathcal{C}(\llbracket H \rrbracket, \llbracket Y \rrbracket)$ which captures the set of all functions from $\llbracket Y \rrbracket$ to $\llbracket H \rrbracket$. Following [21], we say that a test $T \in \mathcal{C}(\llbracket H \rrbracket, \llbracket Y \rrbracket)$ is correct at a particular realization of uncertain variable Y , $Y(\omega) = y \in \llbracket Y \rrbracket$, if $\llbracket H|\llbracket X|y \rrbracket \rrbracket = \{T(y)\}$. The set of all outputs at which test T is correct is equal to $\aleph(T) := \{y \in \llbracket Y \rrbracket : \llbracket H|\llbracket X|y \rrbracket \rrbracket = \{T(y)\}\}$. Based on this definition of correctness, we can define a performance measure for tests [21]. If Y is a continuous uncertain variable, the performance is

$$\mathcal{P}(T) := \log_e(\mu(\aleph(T))). \quad (11)$$

Similarly, if Y is a discrete uncertain variable, the performance

is equal to

$$\mathcal{P}(T) := \log_2(|\mathcal{R}(T)|). \quad (12)$$

In the following result, Δ denotes the symmetric difference operator on the sets, i.e., $\mathcal{A}\Delta\mathcal{B} = (\mathcal{A} \setminus \mathcal{B}) \cup (\mathcal{B} \setminus \mathcal{A})$.

Proposition 3 ([21]). *The performance of any test $T \in \mathcal{C}(\llbracket H \rrbracket, \llbracket Y \rrbracket)$ is bounded by $\mathcal{P}(T) \leq \log_e(\mu(\llbracket Y|p_0 \rrbracket \Delta \llbracket Y|p_1 \rrbracket))$ if Y is a continuous uncertain variable, and by $\mathcal{P}(T) \leq \log_2(|\llbracket Y|p_0 \rrbracket \Delta \llbracket Y|p_1 \rrbracket|)$ if Y is a discrete uncertain variable.*

Note that, for any realization of uncertain variable Y in the set $\llbracket Y|p_0 \rrbracket \cap \llbracket Y|p_1 \rrbracket$, there is not enough evidence to accept or reject either the null hypothesis or the alternative hypothesis. This is because these realizations can be caused by realizations of X that are consistent with the null hypothesis p_0 or realizations of X that are consistent with the alternative hypothesis p_1 . On the other hand, if the realization of the measurement Y is in the set $(\llbracket Y|p_0 \rrbracket \setminus \llbracket Y|p_1 \rrbracket) \cup (\llbracket Y|p_1 \rrbracket \setminus \llbracket Y|p_0 \rrbracket) = \llbracket Y|p_0 \rrbracket \Delta \llbracket Y|p_1 \rrbracket$, we can confidently reject or accept the null hypothesis or the alternative hypothesis. Proposition 3 can be thought of as a non-stochastic equivalent of the Chernoff-Stein Lemma; see, e.g., [59, Ch. 11] for randomized hypothesis testing. The size of the set $\llbracket Y|p_0 \rrbracket \Delta \llbracket Y|p_1 \rrbracket$ essentially captures the difference between the ranges $\llbracket Y|p_0 \rrbracket$ and $\llbracket Y|p_1 \rrbracket$ resembling the Kullback–Leibler divergence in a non-stochastic framework.

III. NOISELESS PRIVACY:

DEFINITION, GUARANTEES, AND PROPERTIES

We model a *private dataset* by a realization of a vector-valued uncertain variable $X : \Omega \rightarrow \mathbb{R}^n$ with n denoting the number of individuals whose data is in the dataset. The dataset is therefore in the form of

$$X(\omega) = \begin{bmatrix} X_1(\omega) \\ X_2(\omega) \\ \vdots \\ X_n(\omega) \end{bmatrix},$$

where $X_i(\omega) \in \mathbb{R}$ is the data of the i -th individual. Evidently, each $X_i : \Omega \rightarrow \mathbb{R}$, $1 \leq i \leq n$, is itself an uncertain variable.

A data curator, in possession of the realization of uncertain variable X , i.e., the private dataset $X(\omega)$, must return a response to a query $f : \llbracket X \rrbracket \rightarrow \mathbb{R}^q$ for some $q \in \mathbb{N}$. The curator employs a mechanism $\mathfrak{M} : \mathbb{R}^q \rightarrow \mathbb{R}^q$ to generate a privacy-preserving response. Throughout this paper, $\mathfrak{M} \circ f$ is referred to as the mechanism of the curator. This is the same language used in the privacy literature albeit without the presence of the randomness [1], [60]. Therefore, the curator provides the response $Y(\omega) = \mathfrak{M} \circ f(X(\omega))$. By definition, $Y = \mathfrak{M} \circ f \circ X$ is an uncertain variable. In the remainder of this paper, we use the notation x_{-i} to denote $(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$ for vectors and X_{-i} to denote $(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)$ for uncertain variables alike. In this notation, $-i$ refers essentially refers to the set of all individuals except the i -th one.

Definition 1 (Noiseless Privacy). *A mechanism $\mathfrak{M} \circ f$ is ϵ -noiselessly private, for $\epsilon > 0$, if*

$$|\llbracket Y|X_{-i}(\omega) = x_{-i} \rrbracket| \leq 2^\epsilon, \quad \forall x_{-i} \in \llbracket X_{-i} \rrbracket, \forall i. \quad (13)$$

Note that this definition is akin to a noiseless differential privacy. This is because, instead of bounding the information leakage as in information-theoretic privacy [16], the output realizations are restricted if one individual entry of the dataset changes. Note that, when the data of i -th individual changes, the output can take all the values within the set $\llbracket Y|X_{-i}(\omega) = x_{-i} \rrbracket$. If this set is not informative, i.e., it does not contain many elements, reverse engineering the data of i -th individual changes with knowledge of $X(\omega)$ even in the presence of side-channel information is a difficult task. In what follows, we use non-stochastic information theory to establish the extend of the privacy guarantees from noiseless privacy. Similar to differential privacy, we can also define a local version of noiseless privacy.

Definition 2 (Local Noiseless Privacy). *Assume that $f_i : (x_i)_{i=1}^n \mapsto x_i$ for each i . A mechanism $\mathfrak{M}$ is ϵ -locally noiselessly private if $\mathfrak{M} \circ f_i$ is ϵ -noiselessly private for all i .*

A. Guarantee: Non-Stochastic Information Leakage

We define a function $\psi_{i,v_{-i}} : \llbracket X \rrbracket \rightarrow \llbracket X_i \rrbracket \times \{v_{-i}\}$ replacing the value of $X_j(\omega)$ or the realization of X_j , for all $1 \leq j \leq n$ except for $j = i$, with given constants v_j , i.e., $\psi_{i,v_{-i}}(X(\omega)) = (v_1, \dots, v_{i-1}, X_i(\omega), v_{i+1}, \dots, v_n)$. Let $\Psi_i = \{\psi_{i,v_{-i}} \mid v_{-i} \in \llbracket X_{-i} \rrbracket\}$ be the set of all such functions for $i \in \{1, \dots, n\}$. The uncertain variable $\psi_{i,v_{-i}} \circ X$ becomes unrelated (in the sense of [17]) to X_{-i} for all $\psi_{i,v_{-i}} \in \Psi_i$.

This definition allows us to measure the amount of the information that the curator's mechanism leaks about the data of the i -th individual $X_i(\omega)$. For a given $\psi_{i,v_{-i}} \in \Psi_i$, let us define $Y = \mathfrak{M} \circ f \circ \psi_{i,v_{-i}} \circ X$. Now, the information between Y and X_i captures how much more information can an adversary extract from Y knowing the data of all the individuals except the i -th individual. This is because, here, we let the adversary to select any possible $\psi_{i,v_{-i}}$.

Theorem 1 (Non-Stochastic Information vs Noiseless Privacy). *Assume X is a discrete uncertain variable, $Y = \mathfrak{M} \circ f \circ \psi_{i,v_{-i}} \circ X$, and $\mathfrak{M} \circ f$ is ϵ -noiselessly private. For any $\psi_{i,v_{-i}} \in \Psi_i$,*

$$0 \leq I_\star(X_i; Y) \leq L_0^s(X_i; Y) \leq L_0(Y; X_i) \leq \epsilon. \quad (14)$$

Proof. See Appendix C. □

Theorem 1 shows that, by reducing ϵ , we can reduce the amount of the leaked information about each individual. This makes sense. Consider the case where $\epsilon = +\infty$. In this case, the curator can report the output of the query $f(\psi_{i,v_{-i}} \circ X(\omega))$ completely (i.e., $\mathfrak{M}$ can be chosen to be equal identity) and

the adversary, knowing v_{-i} , can compute the data of the data of the i -th individual $X_i(\omega)$ (at least if the adversary select the query to be linear with non-zero weight for the i -th individual). On the other hand, if $\epsilon = 0$, the output becomes a constant that is independent of $X_i(\omega)$ and thus the adversary learns nothing new about the data of the i -th individual $X_i(\omega)$.

B. Guarantee: Zero-Error Capacity

Let us consider a memory-less noiselessly-private communication channel. This can be seen as a non-stochastic equivalent of differentially-private communication channels in [61].

Let $\mathfrak{M} \circ f$ be a ϵ -noiselessly private for some $\epsilon > 0$. For any given sequence of mappings $\{\psi_{i,v_{-i}}^t\}_{t \in \mathbb{N}}$ with $\psi_{i,v_{-i}}^t \in \Psi_i$, a memory-less ϵ -noiselessly-private channel maps any uncertain time series $X = (X[k], \dots, X[1])$ to uncertain time series $Y = (Y[k], \dots, Y[1])$ such that $Y[\ell](\omega) = \mathfrak{M} \circ f \circ \psi_{i,v_{-i}}^\ell(X(\omega))$ for all $1 \leq \ell \leq k$ and $k \in \mathbb{N}$.

This setup can be seen as a case in which the curator is reporting on a stream of data from the individuals. We can assume that an extremely strong adversary can set the realizations of the data of all individuals except the i -th individual. The capacity of the channel captures the amount of information that passes through a ϵ -noiselessly private mechanism over time.

Theorem 2 (Zero-Error Capacity vs Noiseless Privacy). *Assume, for all k , $X[k]$ is a discrete uncertain variable, $Y[k] = \mathfrak{M} \circ f \circ \psi_{i,v_{-i}}^\ell \circ X[k]$, and $\mathfrak{M} \circ f$ is ϵ -noiselessly private. For any $\psi_{i,v_{-i}} \in \Psi_i$, the zero-error capacity of memory-less ϵ -noiselessly-private channel is bounded by*

$$C_0 \leq \epsilon. \quad (15)$$

Proof. The rest of the proof follows from Proposition 2 and Theorem 1. $\square$

C. Guarantee: Non-Stochastic Hypothesis Testing

In this part, our analysis is motivated by the definition of semantic security or indistinguishability under chosen plaintext attack [62]. Assume that an adversary selects $i \in \{1, \dots, n\}$, $x_i, x'_i \in \llbracket X_i \rrbracket$, and provides this information to the curator. The curator uses uncertain variable $\overline{X}_i : \Omega \rightarrow \llbracket \overline{X}_i \rrbracket := \{x_i, x'_i\}$ to constructs uncertain variable $\overline{X} = (X_{-i}, \overline{X}_i)$. Fix $\psi_{i,v_{-i}} \in \Psi_i$. The curator then generates a realization $\overline{X}(\omega)$, computes $Y(\omega) = \mathfrak{M} \circ f \circ \psi_{i,v_{-i}}(\overline{X}(\omega))$, and provides $Y(\omega)$ to the adversary. The adversary tests whether the realization of the data of individual i is equal to x_i or x'_i knowing that it is bound to be one of those values and knowing that the value of the data of all the other individuals is fixed to v_{-i} . We define the hypothesis uncertain variable H using $g_H : \overline{X}(\omega) \mapsto H(\omega)$ as

$$H(\omega) = g_H(\overline{X}(\omega)) = \begin{cases} p_0, & \overline{X}_i(\omega) = x_i, \\ p_1, & \overline{X}_i(\omega) = x'_i. \end{cases}$$

The following theorem bounds the performance of the adversary for performing its hypothesis test.

Theorem 3 (Hypothesis Testing vs Noiseless Privacy). *Assume $Y = \mathfrak{M} \circ f \circ \psi_{i,v_{-i}} \circ \overline{X}$ and $\mathfrak{M} \circ f$ is ϵ -noiselessly private. Then, for any test T and any $\psi_{i,v_{-i}} \in \Psi_i$, the performance of the adversary is bounded by*

$$\mathcal{P}(T) \leq \epsilon. \quad (16)$$

Proof. See Appendix D. $\square$

Bounding the performance of a hypothesis-testing adversary is in essence close to identifiability [63], [64] for which privacy preservation relates to the potential of an adversary identifying the private data of individuals based on the received outputs.

D. Guarantee: Performance of Adversaries with Stochastic Priors

In this subsection, we briefly assume that the dataset is randomly distributed according to the probability density function p , i.e., for any Lebesgue-measurable set $\mathcal{X} \subseteq \llbracket X \rrbracket$, $\mathbb{P}\{X \in \mathcal{X}\} = \int_{x \in \llbracket X \rrbracket} \xi(x) \mu(x)$. We also consider an adversary that knows the realizations of all the entries of the dataset except the entry of the i -th individual. It constructs an estimate of the missing entry X_i using an estimator $\hat{X}_i(X_{-i}, \mathfrak{M} \circ f(X))$ using its prior information X_{-i} and the response $\mathfrak{M} \circ f(X)$.

Theorem 4 (Stochastic Prior vs Noiseless Privacy). *Assume that $\rho = \inf_{x \in \llbracket X \rrbracket} \xi(x) > 0$, $\mathfrak{M} \circ f$ is ϵ -noiselessly private, and $f^{-1} \circ \mathfrak{M}^{-1}(y)$ is a connected set for any $y \in \llbracket Y \rrbracket$. For any $p \in \mathbb{N}$,*

$$\mathbb{E}\{(X_i - \hat{X}_i(X_{-i}, \mathfrak{M} \circ f(X)))^p | X_{-i}\} \geq \left(\frac{\rho \mu(\llbracket X_i \rrbracket)^{p+1}}{2^{2p+2}} \right) 2^{-\epsilon(p+1)}. \quad (17)$$

Proof. See Appendix E. $\square$

The lower bound on the adversary's estimation performance in Theorem 4 is an decreasing function of ϵ . Therefore, as expected and in-line with the earlier results, by decreasing ϵ , we can reduce the adversary's ability to infringe on the privacy of any individual in the dataset even if the adversary knows the data of all the other individuals.

E. Guarantee: Stochastic Maximal Leakage

In this section, we can recreate the stochastic framework for information leakage in [65] by again endowing all the uncertain variables with a measure. This way, we can define the maximal stochastic leakage from X to Y as

$$\mathcal{L}_c(X \rightarrow Y) = \sup_{U=X-Y-\hat{U}} \log \left(\frac{\mathbb{P}\{U = \hat{U}\}}{\max_{u \in \llbracket U \rrbracket} P_U(u)} \right)$$

where the supremum is taken over all random variables $U, \hat{U}$ taking values in the same finite arbitrary alphabets. Here, $U -$

$X - Y - \hat{U}$ states that these variables from a Markov chain in the introduced order. It was shown in [65] that

$$\begin{aligned}\mathcal{L}_c(X \rightarrow Y) &= \log \left(\sum_{y \in \llbracket Y \rrbracket} \max_{x \in \llbracket X \rrbracket : P_X(x) > 0} P_{Y|X}(y|x) \right) \\ &= I_\infty(X; Y).\end{aligned}$$

Theorem 5 (Maximal Leakage vs Noiseless Privacy). *Assume $Y = \mathfrak{M} \circ f \circ \psi_{i,v-i} \circ X$ and $\mathfrak{M} \circ f$ is ϵ -noiselessly private. Then, $\mathcal{L}_c(X_i \rightarrow Y) \leq \epsilon$.*

Proof: Note that $\sup_{P_{Y|X}} \mathcal{L}_c(X \rightarrow Y) \leq H_0(Y)$ because of [65, Lemma 1 & Example 6]. Furthermore, $\|\llbracket \mathfrak{M} \circ f \circ \psi_{i,v-i} \circ X \rrbracket\| = \|\llbracket Y|X_{-i}(\omega) = v_{-i} \rrbracket\| \leq 2^\epsilon$. ■

Evidently, the amount of the leaked information is upper bounded by the privacy budget. Hence, by reducing the privacy budget, we can minimize the amount of the leaked information.

F. Property: Composition of Noiselessly-Private Mechanisms

Composition of differentially-private mechanisms [2], [66], [67] is an important result showing that the privacy budgets add up when reporting on multiple queries on the same private dataset. In what follows, we show that the same also applies to noiseless privacy.

Theorem 6 (Composition of Noiselessly-Private Mechanisms). *Let $\mathfrak{M}_1$ and $\mathfrak{M}_2$ be such that $\mathfrak{M}_1 \circ f$ and $\mathfrak{M}_2 \circ f$ are ϵ_1 -noiseless private and ϵ_2 -noiseless private, respectively. Then, $(\mathfrak{M}_1, \mathfrak{M}_2) \circ f$ is $(\epsilon_1 + \epsilon_2)$ -noiseless private.*

Proof. See Appendix F. □

G. Property: Post-Processing of Noiselessly-Private Mechanisms

Finally, an important property of differentially-private mechanisms and information-theoretic privacy is that the privacy guarantees do not weaken by post-processing privacy-preserving outputs [2]. In what follows, this also holds for noiselessly-private mechanisms as well.

Theorem 7 (Post-Processing of Noiselessly-Private Mechanisms). *Let $\mathfrak{M}$ be such that $\mathfrak{M} \circ f$ is ϵ -noiseless private. Then, $g \circ \mathfrak{M} \circ f$ is also ϵ -noiseless private for any mapping g .*

Proof. See Appendix G. □

IV. NOISELESS PRIVACY: SATISFACTION

We can ensure noiseless privacy using non-stochastic approaches, such as binning or quantization. To do so, first, we define linear quantizers.

Adversary	Curator
select i_0 and i_1	i_0 and i_1
$\xrightarrow{\quad}$	$j \leftarrow \{0, 1\}$
	select $i_2, \dots, i_n$
$\xleftarrow{y, t}$	$y = \mathfrak{M} \circ f(x_{i_j, t}, x_{i_2, t}, \dots, x_{i_n, t})$
estimate j	$\hat{j}$
$\xrightarrow{\quad}$	return $j = \hat{j}$

Fig. 2: The timing of a game used for evaluating the ability of an adversary in guessing if the data of a particular individual belongs to a publicly-released noiselessly-private aggregate statistics.

Definition 3 (Linear Quantizer). *A q -level quantizer $\mathcal{Q} : [x_{\min}, x_{\max}] \rightarrow \{b_1, \dots, b_q\}$ is a piecewise constant function defined as*

$$\mathcal{Q}(x) = \begin{cases} b_1, & x \in [x_1, x_2), \\ b_2, & x \in [x_2, x_3), \\ \vdots & \vdots \\ b_{q-1}, & x \in [x_{q-1}, x_q), \\ b_q, & x \in [x_q, x_{q+1}], \end{cases}$$

where $(b_i)_{i=1}^q$ are distinct symbols and $x_1 \leq x_2 \leq \dots \leq x_q$ are real numbers such that $x_1 = x_{\min}$, $x_{q+1} = x_{\max}$, $x_{i+1} - x_i = (x_{\max} - x_{\min})/q$ for all $1 \leq i \leq q$.

We can show that linear quantizers can achieve noiseless privacy for any query on private datasets. This is proved in the next theorem.

Theorem 8. *Let $\llbracket f(X)|X \rrbracket \subseteq [y_{\min}, y_{\max}]$. Define sensitivity of query f as*

$$\begin{aligned}\mathcal{S}(f) &:= \sup_{x_{-i} \in \llbracket X_{-i} \rrbracket} \mu(f(\llbracket X_i \rrbracket \times \{x_{-i}\})) \\ &= \sup_{x_{-i} \in \llbracket X_{-i} \rrbracket} \sup_{x_i, x'_i \in \llbracket X_i \rrbracket} |f(x'_i, x_{-i}) - f(x_i, x_{-i})|.\end{aligned}$$

The mechanism $\mathcal{M} \circ f$ is ϵ -noiseless private if $\mathcal{M}$ is a q -level quantizer with

$$q \leq \frac{2^\epsilon (y_{\max} - y_{\min})}{\mathcal{S}_f}.$$

Proof. See Appendix H. □

V. EXPERIMENTS

A. Energy Data: Reporting Aggregate

For this part, we use a publicly available dataset from the Ausgrid³ containing half-hour smart meter measurements for 300 randomly-selected homes with rooftop solar systems over the period from 1 July 2010 to 30 June 2013. In this paper, we use the data over July 2012 to June 2013.

³<https://www.ausgrid.com.au/Industry/Innovation-and-research/Data-to-share/Solar-home->

Let $x_{i,t}$ denote the consumption of house i at day t . We consider reporting aggregate statistics

$$y_t = f((x_{i,t})_{i=1}^n) = \frac{1}{n}(x_{1,t} + \dots + x_{n,t}), \forall t.$$

Here, f denotes the query. We particularly use the mechanism in Theorem 8 to report noiselessly-private outputs. In this experiment, we test the ability of an adversary for inferring if a particular household has contributed to the aggregate or not as in [68]. We use a game, as in [68], [69], to evaluate the ability of the adversary. The setup of the game is summarized in Figure 2. At first, the adversary can select two households i_0, i_1 . The curator select one of those households uniformly at random i_j . It also selects an additional $n-1$ households. Then it reports the privacy-preserving aggregate outputs. Based on the reported output, the adversary guesses the participating household $i_{\hat{j}}$. The adversary's success or advantage is then defined as

$$\text{Adv} := 2|\mathbb{P}\{j = \hat{j}\} - 1/2|.$$

Small Adv means that the adversary is as successful as randomly guessing and large Adv implies that the adversary is successful in recognizing the household participating in the aggregate.

Similar to [68], we use three adversary policies. The first one is based on correlation. In this case, the adversary selects $j \in \{0, 1\}$ based on the correlation between $(x_{i_j,t})_t$ and $(y_t)_t$. The second policy is based on mean square error. In this case, the adversary selects $j \in \{0, 1\}$ by minimizing the square error $\|(x_{i_j,t})_t - (y_t)_t\|_2$. Finally, the last policy uses the relative peaks of each load profile $(x_{i_0,t})_t$, $(x_{i_1,t})_t$, and $(y_t)_t$. In this case, the adversary selects $j \in \{0, 1\}$ based on the most common peaks between $(x_{i_0,t})_t$ and $(y_t)_t$, or $(x_{i_1,t})_t$ and $(y_t)_t$.

Figure 3 illustrates the advantage of the adversary Adv when using the correlation-based policy (left), the mean square error policy (center), and the peak-based policy (right). As ϵ grows larger, the adversary's advantage tends toward the non-private case in [68]. Clearly, even for moderate ϵ when considering small groups, the adversary's advantage is very small. This is not the case for non-private outputs as observed in [68]. For instance, for small groups and moderate privacy budgets, such as $n = 4$ and $\epsilon = 2$ or $n = 8$ and $\epsilon = 3$, the adversary's advantage is negligible (almost zero). This shows that combining noiseless privacy with aggregation is an excellent tool for providing privacy to individuals.

B. Energy Data: Reporting Single Consumption

Non-intrusive load monitoring provides tools for extracting appliance-specific energy consumption statistics from the smart meter readings of a household and is one of privacy concerns behind releasing energy data [70], [71]. In this section, we use Theorem 8 to report high-frequency energy consumption of a household in a privacy-preserving manner using local noiseless privacy. We then proceed to see the effect of privacy budget on an adversary performing non-intrusive load monitoring.

We use the low frequency data from the first house in the REDD database⁴ database [72], which contains the consumption of various appliances in the house every 3-4 seconds. This data in conjunction with the consumption of the entire house is used for training and verification of a non-intrusive load monitoring algorithm. The consumption of the entire house is measured every second. The data is for the period of April 23–May 21, 2011. The part of the data prior to April 30th is used for training and the rest for validation purposes. We select the top 5 appliances in energy consumption for disaggregation purposes, namely, fridge, microwave, socket (in the kitchen), light, and dish washer. For non-intrusive load monitoring, we have used the NILMTK⁵ toolbox in Python [73]. We have used a frequently utilized combinatorial optimization method for non-intrusive load monitoring. We report the success of the non-intrusive load monitoring using the f -score.

Figure 4 shows the f -score of the non-intrusive load monitoring algorithm based on combinatorial optimization versus the privacy budget. As we can see the f -score gets rapidly bad as ϵ decrease. This means an adversary would not be able to identify the appliances that are used within the household reliably. This illustrates the power of local noiseless privacy in reporting energy consumption of households for analysis while protecting the privacy of the households.

C. Transport Data: Reporting Individual Source-Destinations

Finally, we use New York City Taxi Cab trips⁶ for 2014. We use the first million trips and focus on trips that begin and end within the New York City in Figure 5. Here, we consider reporting the start-end point of the taxi rides in a locally noiselessly private manner. In this subsection, we again use Theorem 8 to report start-end points of taxi rides in the New York City in a privacy-preserving manner using local noiseless privacy. In this case, for each ϵ , we split the latitude and the longitude into $2^{\epsilon/2}$ boxes. Therefore, the privacy of the total privacy budget for the reported outputs is 2ϵ , following Theorem 6.

Figure 6 illustrates the portion of unique start-end points of taxi rides versus the privacy budget. As we can see, the portion of unique start-end points is negligible for small ϵ . This means an adversary would not be able to attribute a specific taxi ride to an individual, thus protecting the privacy of contributing individuals.

VI. DISCUSSIONS AND FUTURE WORK

In this paper, we defined noiseless privacy, as a non-stochastic rival to differential privacy, requiring that the outputs of the mechanism to attain very few values while varying the data of an individual remains. We proved that noiseless-private mechanisms admit composition theorem and post-processing does not weaken their privacy guarantees. We proved that quantization operators can ensure noiseless privacy. We finally illustrated the privacy merits of noiseless

⁴<http://redd.csail.mit.edu/>

⁵<https://github.com/nilmtnk/nilmtnk>

⁶<https://www.kaggle.com/kentonnlp/2014-new-york-city-taxi-trips>

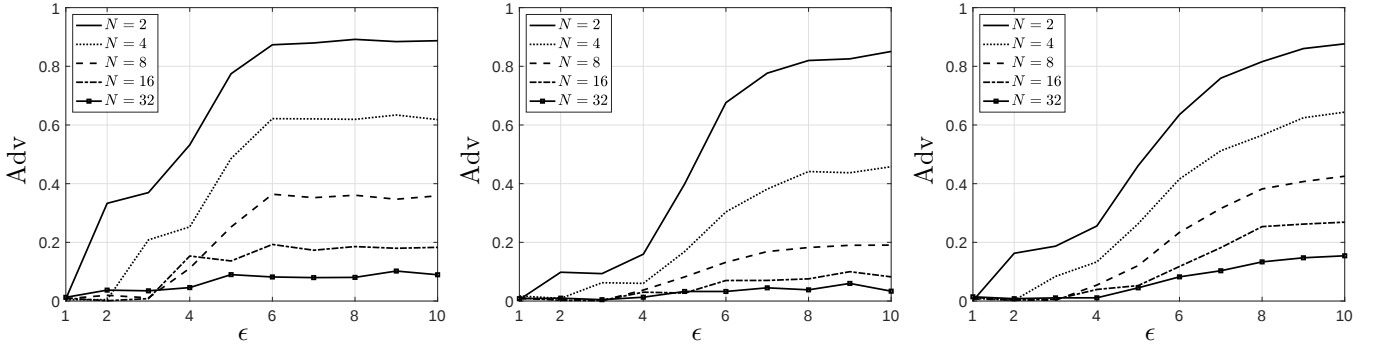


Fig. 3: The advantage of the adversary Adv when using the correlation-based policy (left), the mean square error policy (center), and the peak-based policy (right).

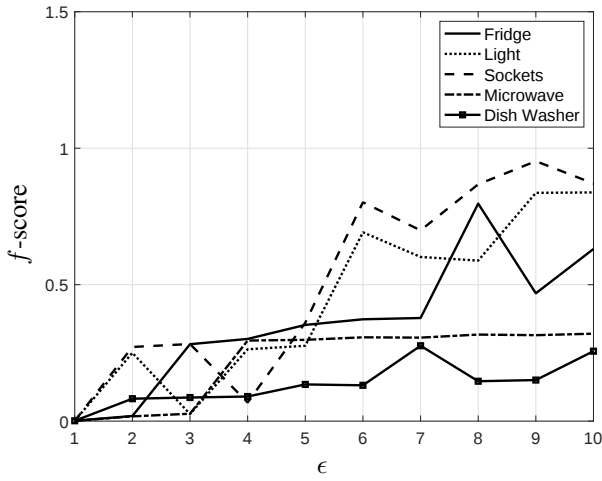


Fig. 4: f -score of the non-intrusive load monitoring algorithm based on combinatorial optimization versus the privacy budget.

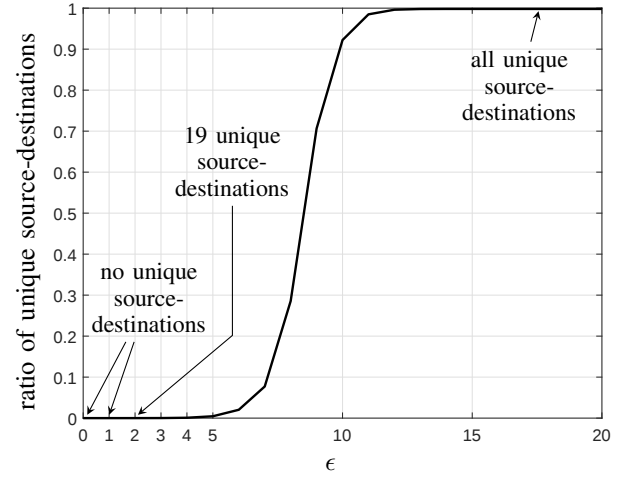


Fig. 6: Portion of taxi rides with unique start-end points versus the privacy budget.



Fig. 5: Map of New York City: latitude in $[40.92^\circ, 40.49^\circ]$ and longitude in $[-74.27^\circ, -73.68^\circ]$

privacy and local noiseless privacy using multiple datasets in energy and transport.

REFERENCES

- [1] C. Dwork, F. McSherry, K. Nissim, and A. Smith, “Calibrating noise to sensitivity in private data analysis,” in *Theory of Cryptography Conference*, pp. 265–284, Springer, 2006.
- [2] C. Dwork and A. Roth, “The algorithmic foundations of differential privacy,” *Foundations and Trends in Theoretical Computer Science*, vol. 9, no. 3–4, pp. 211–407, 2014.
- [3] L. Sankar, S. R. Rajagopalan, and H. V. Poor, “Utility-privacy tradeoffs in databases: An information-theoretic approach,” *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 6, pp. 838–852, 2013.
- [4] H. Yamamoto, “A source coding problem for sources with additional outputs to keep secret from the receiver or wiretappers,” *IEEE Transactions on Information Theory*, vol. 29, no. 6, pp. 918–923, 1983.
- [5] P. Samarati, “Protecting respondents identities in microdata release,” *IEEE transactions on Knowledge and Data Engineering*, vol. 13, no. 6, pp. 1010–1027, 2001.
- [6] L. Sweeney, “ k -anonymity: A model for protecting privacy,” *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, no. 05, pp. 557–570, 2002.

- [7] A. Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkatasubramanian, " ℓ -diversity: privacy beyond k -anonymity," in *22nd International Conference on Data Engineering (ICDE'06)*, pp. 24–24, 2006.
- [8] R. Bild, K. A. Kuhn, and F. Prasser, "SafePub: A truthful data anonymization algorithm with strong privacy guarantees," *Proceedings on Privacy Enhancing Technologies*, vol. 2018, no. 1, pp. 67–87, 2018.
- [9] G. Poulis, A. Gkoulalas-Divanis, G. Loukides, S. Skiadopoulos, and C. Tryfonopoulos, *SECRETA: A Tool for Anonymizing Relational, Transaction and RT-Datasets*, pp. 83–109. Springer International Publishing, 2015.
- [10] R. Bhaskar, A. Bhowmick, V. Goyal, S. Laxman, and A. Thakurta, "Noiseless database privacy," in *International Conference on the Theory and Application of Cryptology and Information Security*, pp. 215–232, Springer, 2011.
- [11] S. U. Nabar, B. Marthi, K. Kenthapadi, N. Mishra, and R. Motwani, "Towards robustness in query auditing," in *Proceedings of the 32nd international conference on Very large data bases*, pp. 151–162, VLDB Endowment, 2006.
- [12] J. Bambauer, K. Muralidhar, and R. Sarathy, "Fool's gold: an illustrated critique of differential privacy," *Vanderbilt Journal of Entertainment & Technology Law*, vol. 16, p. 701, 2013.
- [13] F. K. Dankar and K. El Emam, "Practicing differential privacy in health care: A review," *Transactions on Data Privacy*, vol. 6, no. 1, pp. 35–67, 2013.
- [14] J. Mervis, "Researchers object to census privacy measure," *Science*, vol. 363, no. 6423, pp. 114–114, 2019.
- [15] F. Farokhi, J. Milosevic, and H. Sandberg, "Optimal state estimation with measurements corrupted by Laplace noise," in *Decision and Control (CDC), 2016 IEEE 55th Conference on*, pp. 302–307, IEEE, 2016.
- [16] F. Farokhi, "Development and analysis of deterministic privacy-preserving policies using non-stochastic information theory," *IEEE Transactions on Information Forensics and Security*, vol. 14, pp. 2567–2576, Oct 2019.
- [17] G. N. Nair, "A nonstochastic information theory for communication and state estimation," *IEEE Transactions on Automatic Control*, vol. 58, no. 6, pp. 1497–1510, 2013.
- [18] P. Duan, F. Yang, S. L. Shah, and T. Chen, "Transfer zero-entropy and its application for capturing cause and effect relationship between variables," *IEEE Transactions on Control Systems Technology*, vol. 23, no. 3, pp. 855–867, 2014.
- [19] T. J. Lim and M. Franceschetti, "A deterministic view on the capacity of bandlimited functions," in *2014 52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 691–698, IEEE, 2014.
- [20] C. E. Shannon, "The zero error capacity of a noisy channel," *IRE Transactions on Information Theory*, vol. 2, no. 3, pp. 8–19, 1956.
- [21] F. Farokhi, "Non-stochastic hypothesis testing with application to privacy against hypothesis-testing adversary," in *Decision and Control (CDC), 2019 IEEE 58th Conference on*, 2019. <https://arxiv.org/abs/1904.07377>.
- [22] A. Narayanan and V. Shmatikov, "Robust de-anonymization of large sparse datasets," in *Security and Privacy, 2008. SP 2008. IEEE Symposium on*, pp. 111–125, IEEE, 2008.
- [23] J. Su, A. Shukla, S. Goel, and A. Narayanan, "De-anonymizing web browsing data with social networks," in *Proceedings of the 26th International Conference on World Wide Web*, pp. 1261–1269, 2017.
- [24] Y.-A. De Montjoye, C. A. Hidalgo, M. Verleysen, and V. D. Blondel, "Unique in the crowd: The privacy bounds of human mobility," *Scientific reports*, vol. 3, p. 1376, 2013.
- [25] R. A. Popa, A. J. Blumberg, H. Balakrishnan, and F. H. Li, "Privacy and accountability for location-based aggregate statistics," in *Proceedings of the 18th ACM conference on Computer and Communications Security*, pp. 653–666, ACM, 2011.
- [26] F. Li, B. Luo, and P. Liu, "Secure information aggregation for smart grids using homomorphic encryption," in *2010 First IEEE International Conference on Smart Grid Communications*, pp. 327–332, IEEE, 2010.
- [27] Y. Lindell and B. Pinkas, "Privacy preserving data mining," in *Advances in Cryptology — CRYPTO 2000* (M. Bellare, ed.), pp. 36–54, Springer Berlin Heidelberg, 2000.
- [28] W. Du, Y. S. Han, and S. Chen, "Privacy-preserving multivariate statistical analysis: Linear regression and classification," in *Proceedings of the 2004 SIAM International Conference on Data Mining*, pp. 222–233, SIAM, 2004.
- [29] J. Vaidya and C. Clifton, "Privacy preserving association rule mining in vertically partitioned data," in *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 639–644, ACM, 2002.
- [30] J. Vaidya, M. Kantarcioğlu, and C. Clifton, "Privacy-preserving naive bayes classification," *The VLDB Journal*, vol. 17, no. 4, pp. 879–898, 2008.
- [31] G. Jagannathan and R. N. Wright, "Privacy-preserving distributed k-means clustering over arbitrarily partitioned data," in *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, pp. 593–599, ACM, 2005.
- [32] C. Dwork, "Differential privacy: A survey of results," in *Theory and Applications of Models of Computation: 5th International Conference, TAMC 2008, Xi'an, China, April 25-29, 2008. Proceedings* (M. Agrawal, D. Du, Z. Duan, and A. Li, eds.), pp. 1–19, Berlin, Heidelberg: Springer Berlin Heidelberg, 2008.
- [33] J. C. Duchi, M. I. Jordan, and M. J. Wainwright, "Local privacy and statistical minimax rates," in *Foundations of Computer Science (FOCS), 2013 IEEE 54th Annual Symposium on*, pp. 429–438, IEEE, 2013.
- [34] P. Kairouz, S. Oh, and P. Viswanath, "Extremal mechanisms for local differential privacy," in *Advances in Neural Information Processing Systems*, pp. 2879–2887, 2014.
- [35] A. Machanavajjhala, D. Kifer, J. Abowd, J. Gehrke, and L. Vilhuber, "Privacy: Theory meets practice on the map," in *Proceedings of the 2008 IEEE 24th International Conference on Data Engineering*, pp. 277–286, IEEE Computer Society, 2008.
- [36] R. Hall, A. Rinaldo, and L. Wasserman, "Random differential privacy," *Journal of Privacy and Confidentiality*, vol. 4, no. 2, pp. 43–59, 2012.
- [37] A. Padakandla, P. Kumar, and W. Szpankowski, "Preserving privacy and fidelity via Ehrhart theory," in *2018 IEEE International Symposium on Information Theory (ISIT)*, pp. 696–700, IEEE, 2018.
- [38] A. D. Wyner, "The wire-tap channel," *Bell System Technical Journal*, vol. 54, no. 8, pp. 1355–1387, 1975.
- [39] T. Courtade, "Information masking and amplification: The source coding setting," in *Proceedings of the IEEE International Symposium on Information Theory Proceedings (ISIT)*, pp. 189–193, 2012.
- [40] H. Yamamoto, "A rate-distortion problem for a communication system with a secondary decoder to be hindered," *IEEE Transactions on Information Theory*, vol. 34, no. 4, pp. 835–842, 1988.
- [41] M. S. Alvim, M. E. Andrés, K. Chatzikokolakis, and C. Palamidessi, "On the relation between differential privacy and quantitative information flow," in *Automata, Languages and Programming* (L. Aceto, M. Henzinger, and J. Sgall, eds.), vol. 6756 of *Lecture Notes in Computer Science*, pp. 60–76, Springer Berlin Heidelberg, 2011.
- [42] F. du Pin Calmon and N. Fawaz, "Privacy against statistical inference," in *Proceedings of the 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 1401–1408, 2012.
- [43] F. Farokhi, H. Sandberg, I. Shames, and M. Cantoni, "Quadratic Gaussian privacy games," in *Proceedings of the 54th IEEE Conference on Decision and Control*, pp. 4505–4510, 2015.
- [44] M. J. Wainwright, M. I. Jordan, and J. C. Duchi, "Privacy aware learning," in *Proceedings of Advances in Neural Information Processing Systems (NIPS)*, pp. 1430–1438, 2012.
- [45] Y. Liang, H. V. Poor, and S. Shamai, "Information theoretic security," *Foundations and Trends in Communications and Information Theory*, vol. 5, no. 4–5, pp. 355–580, 2009.
- [46] L. Lai, S.-W. Ho, and H. V. Poor, "Privacy–security trade-offs in biometric security systems—Part I: Single use case," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 1, pp. 122–139, 2011.
- [47] Z. Li and T. Oechtering, "Privacy on hypothesis testing in smart grids," in *IEEE Information Theory Workshop (ITW) 2015, Jeju, Korea, Oct. 11-15, 2015*, pp. 337–341, IEEE, 2015.
- [48] G. Bassi, M. Skoglund, and P. Piantanida, "Lossy communication subject to statistical parameter privacy," in *2018 IEEE International Symposium on Information Theory (ISIT)*, pp. 1031–1035, IEEE, 2018.
- [49] F. Farokhi and H. Sandberg, "Fisher information as a measure of privacy: Preserving privacy of households with smart meters using batteries," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 4726–4734, 2018.
- [50] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.
- [51] R. V. L. Hartley, "Transmission of information," *Bell System Technical Journal*, vol. 7, no. 3, pp. 535–563, 1928.
- [52] A. N. Kolmogorov and V. M. Tikhomirov, " ϵ -entropy and ϵ -capacity of sets in function spaces," *Uspekhi Matematicheskikh Nauk*, vol. 14, no. 2,

pp. 3–86, 1959. English translation American Mathematical Society Translations, series 2, vol. 17, pp. 277364.

- [53] A. Renyi, “On measures of entropy and information,” in *Proc. of the Fourth Berkeley Symp. on Math. Statist. and Prob.*, vol. 1, pp. 547–561, 1961.
- [54] D. Jagerman, “ ε -entropy and approximation of bandlimited functions,” *SIAM Journal on Applied Mathematics*, vol. 17, no. 2, pp. 362–377, 1969.
- [55] G. N. Nair, “A nonstochastic information theory for feedback,” in *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*, pp. 1343–1348, IEEE, 2012.
- [56] P. Duan, F. Yang, S. L. Shah, and T. Chen, “Transfer zero-entropy and its application for capturing cause and effect relationship between variables,” *IEEE Transactions on Control Systems Technology*, vol. 23, no. 3, pp. 855–867, 2015.
- [57] M. Wiese, K. H. Johansson, T. J. Oechtering, P. Papadimitratos, H. Sandberg, and M. Skoglund, “Uncertain wiretap channels and secure estimation,” in *Information Theory (ISIT), 2016 IEEE International Symposium on*, pp. 2004–2008, IEEE, 2016.
- [58] H. Shingir and Y. Ohta, “Disturbance rejection with information constraints: Performance limitations of a scalar system for bounded and gaussian disturbances,” *Automatica*, vol. 48, no. 6, pp. 1111–1116, 2012.
- [59] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, 2012.
- [60] K. Chatzikokolakis, M. E. Andrés, N. E. Bordenabe, and C. Palamidessi, “Broadening the scope of differential privacy using metrics,” in *International Symposium on Privacy Enhancing Technologies Symposium*, pp. 82–102, Springer, 2013.
- [61] G. Barthe and B. Kopf, “Information-theoretic bounds for differentially private mechanisms,” in *2011 IEEE 24th Computer Security Foundations Symposium*, pp. 191–204, IEEE, 2011.
- [62] J. Katz and Y. Lindell, *Introduction to Modern Cryptography, Second Edition*. Chapman & Hall/CRC Cryptography and Network Security Series, Taylor & Francis, 2 ed., 2014.
- [63] W. Wang, L. Ying, and J. Zhang, “On the relation between identifiability, differential privacy, and mutual-information privacy,” *IEEE Transactions on Information Theory*, vol. 62, no. 9, pp. 5018–5029, 2016.
- [64] A. Bkakra, N. Cuppens-Boulahia, and F. Cuppens, “Linking differential identifiability with differential privacy,” in *International Conference on Information and Communications Security*, pp. 232–247, Springer, 2018.
- [65] I. Issa, A. B. Wagner, and S. Kamath, “An operational approach to information leakage,” *arXiv preprint arXiv:1807.07878*, 2018.
- [66] C. Dwork, G. N. Rothblum, and S. Vadhan, “Boosting and differential privacy,” in *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, pp. 51–60, IEEE, 2010.
- [67] P. Kairouz, S. Oh, and P. Viswanath, “The composition theorem for differential privacy,” *IEEE Transactions on Information Theory*, vol. 63, no. 6, pp. 4037–4049, 2017.
- [68] N. Buescher, S. Boukoro, S. Bauregger, and S. Katzenbeisser, “Two is not enough: Privacy assessment of aggregation schemes in smart metering,” *Proceedings on Privacy Enhancing Technologies*, vol. 2017, no. 4, pp. 198–214, 2017.
- [69] J.-M. Bohli, C. Sorge, and O. Ugu, “A privacy model for smart metering,” in *2010 IEEE International Conference on Communications Workshops*, pp. 1–5, IEEE, 2010.
- [70] A. Zoha, A. Gluhak, M. A. Imran, and S. Rajasegarar, “Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey,” *Sensors*, vol. 12, no. 12, pp. 16838–16866, 2012.
- [71] O. Parson, S. Ghosh, M. Weal, and A. Rogers, “Non-intrusive load monitoring using prior models of general appliance types,” in *Twenty-Sixth AAAI Conference on Artificial Intelligence*, 2012.
- [72] J. Z. Kolter and M. J. Johnson, “REDD: A public data set for energy disaggregation research,” in *Workshop on Data Mining Applications in Sustainability (SIGKDD)*, vol. 25, pp. 59–62, 2011.
- [73] N. Batra, J. Kelly, O. Parson, H. Dutta, W. Knottenbelt, A. Rogers, A. Singh, and M. Srivastava, “NILMTK: An open source toolkit for non-intrusive load monitoring,” in *Fifth International Conference on Future Energy Systems (ACM e-Energy)*, 2014.

APPENDIX A

PROOF OF PROPOSITION 1

First, we prove that $I_*(X; Y) \leq L_0(X; Y)$. Let $I_*(X; Y) = m$. Then, $\llbracket X|Y \rrbracket_* = \{P_1, \dots, P_{2^m}\}$. Each P_i is non-empty.

Therefore, there exists at least one x such that $x \in P_i$. Note that x must also belong to $\llbracket X|y \rrbracket$ for any $y \in \llbracket Y|x \rrbracket$. We prove that $\llbracket X|y \rrbracket \subseteq P_i$. Assume that this not the case. Therefore, there exist an element of $x' \in \llbracket X|y \rrbracket$, distinct from x , that belongs to another P_j , $j \neq i$, because $\{P_1, \dots, P_{2^m}\}$ covers $\llbracket X \rrbracket$. We know that P_i and P_j are $\llbracket X|Y \rrbracket$ -overlap isolated by the definition of partition $\llbracket X|Y \rrbracket_*$. On the other hand, we evidently have $x \rightsquigarrow x'$ (by the definition of $\llbracket X|Y \rrbracket$ -overlap connectedness). This is a contradiction and thus $\llbracket X|y \rrbracket$ must be a subset of P_i . This results in $|\llbracket X|y \rrbracket| \leq |P_i|$ and hence

$$\min_{y \in \llbracket Y \rrbracket} |\llbracket X|y \rrbracket| \leq |P_i|, \quad \forall i \in \{1, \dots, 2^m\}. \quad (18)$$

On the other hand, $\bigcup_{i=1}^{2^m} P_i = \llbracket X \rrbracket$ because $\{P_1, \dots, P_{2^m}\}$ is a partition for $\llbracket X \rrbracket$. Because of the non-overlapping nature of the sets $\{P_1, \dots, P_{2^m}\}$, we get

$$\sum_{i=1}^{2^m} |P_i| = |\llbracket X \rrbracket|. \quad (19)$$

Combining (18) and (19) results in $2^m \min_{y \in \llbracket Y \rrbracket} |\llbracket X|y \rrbracket| \leq |\llbracket X \rrbracket|$. This implies that $I_*(X; Y) \leq L_0(X; Y)$. Similarly, we can show that $I_*(Y; X) \leq L_0(Y; X)$. By symmetry of the maximin information [17], i.e., $I_*(X; Y) = I_*(Y; X)$, we get that $I_*(X; Y) \leq L_0(X; Y)$ and $I_*(X; Y) \leq L_0(Y; X)$. This concludes the proof.

APPENDIX B

PROOF OF PROPOSITION 2

Let $\aleph(y[k], \dots, y[1])$ denote the statement $Y[k](\omega) = y[k], \dots, Y[1](\omega) = y[1]$. Note that

$$\begin{aligned} & 2^{L_0(X[k], \dots, X[1]; Y[k], \dots, Y[1])} \\ &= \frac{\max_{(y[i])_{i=1}^k \in \llbracket Y \rrbracket^k} |\llbracket X[k], \dots, X[1] \rrbracket \aleph(y[k], \dots, y[1])|}{|\llbracket X[k], \dots, X[1] \rrbracket|} \\ &= \frac{|\llbracket X[k], \dots, X[1] \rrbracket|}{\min_{(y[i])_{i=1}^k \in \llbracket Y \rrbracket^k} |\llbracket X[k], \dots, X[1] \rrbracket \aleph(y[k], \dots, y[1])|}} \\ &= \frac{\prod_{\ell=1}^k |\llbracket X[\ell] \rrbracket|}{\prod_{\ell=1}^k \min_{y[\ell] \in \llbracket Y \rrbracket} |\llbracket X[\ell] \rrbracket Y[\ell](\omega) = y[\ell]|}} \\ &= \prod_{\ell=1}^k \frac{|\llbracket X[\ell] \rrbracket|}{\min_{y[\ell] \in \llbracket Y \rrbracket} |\llbracket X[\ell] \rrbracket Y[\ell](\omega) = y[\ell]|}} \\ &= \prod_{\ell=1}^k L_0(X[\ell]; Y[\ell]) \\ &= (L_0(X[k]; Y[k]))^k. \end{aligned}$$

Therefore,

$$L_0(X[k], \dots, X[1]; Y[k], \dots, Y[1]) = k L_0(X[k]; Y[k]),$$

and, as a result,

$$L_0(X[k], \dots, X[1]; Y[k], \dots, Y[1])/k = L_0(X[k]; Y[k]).$$

Similarly, we can show that

$$L_0(Y[k], \dots, Y[1]; X[k], \dots, X[1])/k = L_0(Y[k]; X[k]).$$

Combining these inequalities with Proposition 1 in this paper and Theorem 4.1 in [17] proves the result.

APPENDIX C
PROOF OF PROPOSITION 1

Note that

$$\begin{aligned}
2^{L_0^s(X_i; Y)} &\leq 2^{L_0(Y; X_i)} \\
&= \sup_{x_i \in \llbracket X_i \rrbracket} \frac{|\llbracket Y \rrbracket|}{|\llbracket Y|X_i(\omega) = x_i \rrbracket|} \\
&= |\llbracket Y \rrbracket| \\
&= \left| \bigcup_{x'_i \in \llbracket X_i \rrbracket} \llbracket Y|X_i(\omega) = x'_i, X_{-i}(\omega) = v_{-i} \rrbracket \right| \\
&= |\llbracket Y|X_{-i}(\omega) = v_{-i} \rrbracket| \\
&\leq 2^\epsilon,
\end{aligned}$$

where the second equality follows from that the realization of Y can be uniquely determined based on the realization of X_i , i.e., $\llbracket Y|X_i(\omega) = x_i \rrbracket$ is a singleton. Therefore, $L_0^s(X_i; Y) \leq L_0(Y; X_i) \leq \epsilon$. The rest follows from Proposition 1.

APPENDIX D
PROOF OF THEOREM 3

Since Y is a discrete uncertain variable, for any test T , Theorem 3 states that the performance of the adversary is bounded from the above by $\mathcal{P}(T) \leq \log_2(|\llbracket Y|X_i(\omega) = x_i \rrbracket \Delta \llbracket Y|X_i(\omega) = x'_i \rrbracket|)$. Now, note that $\llbracket Y|X_i(\omega) = x_i \rrbracket \Delta \llbracket Y|X_i(\omega) = x'_i \rrbracket \subseteq \llbracket Y|X_{-i}(\omega) = v_{-i} \rrbracket$. Therefore, $|\llbracket Y|X_i(\omega) = x_i \rrbracket \Delta \llbracket Y|X_i(\omega) = x'_i \rrbracket| \leq |\llbracket Y|X_{-i}(\omega) = v_{-i} \rrbracket| \leq 2^\epsilon$. This concludes the proof.

APPENDIX E
PROOF OF THEOREM 4

Define function $g(\cdot)$ such that $g(X_i) = \mathfrak{M} \circ f(X)$. There must exists $y \in \llbracket Y \rrbracket$ such that $\mu(g^{-1}(y)) \geq \mu(\llbracket X_i \rrbracket)2^{-\epsilon}$. As otherwise, $\mu(g^{-1}(y)) < \mu(\llbracket X_i \rrbracket)2^{-\epsilon}$ for all $y \in \llbracket Y \rrbracket$ and thus

$$\begin{aligned}
\mu(\llbracket X_i \rrbracket) &= \mu\left(\bigcup_{y \in \llbracket Y \rrbracket} g^{-1}(y)\right) \\
&= \sum_{y \in \llbracket Y \rrbracket} \mu(g^{-1}(y)) \\
&< \sum_{y \in \llbracket Y \rrbracket} \mu(\llbracket X_i \rrbracket)2^{-\epsilon} \\
&= \mu(\llbracket X_i \rrbracket).
\end{aligned}$$

This is a contradiction. Hence, we get

$$\begin{aligned}
&\mathbb{E}\{(X_i - \hat{X}_i(X_{-i}, \mathfrak{M} \circ f(X))^p | X_{-i}\} \\
&\geq \rho \int_{g^{-1}(y)} (X_i - \hat{X}_i(X_{-i}, \mathfrak{M} \circ f(X))^p d\mu(X_i),
\end{aligned}$$

where $\rho = \inf_{X \in \llbracket X \rrbracket} \xi(X)$. Since $g^{-1}(y)$ is a connected set, there must exists $\underline{x}_i, \bar{x}_i$ such that closure of the $g^{-1}(y)$ is equal to $[\underline{x}_i, \bar{x}_i]$. Hence, we get

$$\begin{aligned}
&\int_{g^{-1}(y)} (X_i - \hat{X}_i(X_{-i}, \mathfrak{M} \circ f(X))^p d\mu(X_i) \\
&= \int_{\underline{x}_i}^{\bar{x}_i} (X_i - \hat{X}_i(X_{-i}, \mathfrak{M} \circ f(X))^p d\mu(X_i) \\
&\geq \int_{\underline{x}_i}^{\bar{x}_i} (z - (\underline{x}_i + \bar{x}_i)/2)^p dz \\
&\geq 2 \left(\frac{\bar{x}_i - \underline{x}_i}{4}\right)^p \frac{\bar{x}_i - \underline{x}_i}{4} \\
&= \frac{\mu(g^{-1}(y))^{p+1}}{4^{p+1/2}} \\
&\geq \mu(\llbracket X_i \rrbracket)^{p+1} 2^{-\epsilon(p+1)} / 2^{2p+1}.
\end{aligned}$$

This concludes the proof.

APPENDIX F
PROOF OF THEOREM 6

Note that

$$\begin{aligned}
&\llbracket (\mathfrak{M}_1, \mathfrak{M}_2) \circ f(X) | X_{-i}(\omega) = x_{-i} \rrbracket \\
&= \llbracket (\mathfrak{M}_1 \circ f(X), \mathfrak{M}_2 \circ f(X)) | X_{-i}(\omega) = x_{-i} \rrbracket \\
&\subseteq \llbracket \mathfrak{M}_1 \circ f(X) | X_{-i}(\omega) = x_{-i} \rrbracket \\
&\quad \times \llbracket \mathfrak{M}_2 \circ f(X) | X_{-i}(\omega) = x_{-i} \rrbracket,
\end{aligned}$$

and as a result

$$\begin{aligned}
&\mu(\llbracket (\mathfrak{M}_1, \mathfrak{M}_2) \circ f(X) | X_{-i}(\omega) = x_{-i} \rrbracket) \\
&\leq \mu(\llbracket \mathfrak{M}_1 \circ f(X) | X_{-i}(\omega) = x_{-i} \rrbracket) \\
&\quad \times \mu(\llbracket \mathfrak{M}_2 \circ f(X) | X_{-i}(\omega) = x_{-i} \rrbracket).
\end{aligned}$$

Hence,

$$\begin{aligned}
&\log_e(\mu(\llbracket (\mathfrak{M}_1, \mathfrak{M}_2) \circ f(X) | X_{-i}(\omega) = x_{-i} \rrbracket)) \\
&\leq \log_e(\mu(\llbracket \mathfrak{M}_1 \circ f(X) | X_{-i}(\omega) = x_{-i} \rrbracket)) \\
&\quad \times \mu(\llbracket \mathfrak{M}_2 \circ f(X) | X_{-i}(\omega) = x_{-i} \rrbracket) \\
&= \log_e(\mu(\llbracket \mathfrak{M}_1 \circ f(X) | X_{-i}(\omega) = x_{-i} \rrbracket)) \\
&\quad + \log_e(\mu(\llbracket \mathfrak{M}_2 \circ f(X) | X_{-i}(\omega) = x_{-i} \rrbracket)).
\end{aligned}$$

The proof for discrete uncertain variables follow the same approach.

APPENDIX G
PROOF OF THEOREM 7

The proof follows from that $\llbracket g \circ \mathfrak{M} \circ f(X) | X_{-i}(\omega) = x_{-i} \rrbracket \leq \llbracket \mathfrak{M} \circ f(X) | X_{-i}(\omega) = x_{-i} \rrbracket$ for any x_{-i} .

APPENDIX H
PROOF OF THEOREM 8

For any given $x_{-i} \in \llbracket X_{-i} \rrbracket$, due to continuity of f , we know that $f(\llbracket X_i \rrbracket \times \{x_{-i}\}) = f \circ \psi_{i, x_{-i}}(\llbracket X_i \rrbracket) \subseteq [y_{\min}, y_{\max}]$ is a connected set (because $\llbracket X_i \rrbracket$ is connected). Therefore, if $\mathfrak{M}$ is a q -level quantizer, $\llbracket Y|X_{-i}(\omega) = x_{-i} \rrbracket = \mathfrak{M} \circ f(\llbracket X_i \rrbracket \times \{x_{-i}\})$ can at most contain $q\mu(f(\llbracket X_i \rrbracket \times \{x_{-i}\}))/ (y_{\max} - y_{\min})$ points. Therefore, $|\llbracket Y|X_{-i}(\omega) = x_{-i} \rrbracket| \leq q\mathcal{S}_f / (y_{\max} - y_{\min})$. This concludes the proof.