Author/s:
Farokhi, F;Nair, G

Title:
Non-Stochastic Private Function Evaluation

Date:
2021-04-11

Citation:
Farokhi, F. & Nair, G. (2021). Non-Stochastic Private Function Evaluation. 2020 IEEE Information Theory Workshop (ITW), 00, IEEE. https://doi.org/10.1109/itw46852.2021.9457645.

Persistent Link:
https://hdl.handle.net/11343/279425

Non-Stochastic Private Function Evaluation

Farhad Farokhi and Girish Nair
Department of Electrical and Electronic Engineering The University of Melbourne, Parkville, Australia
{ffarokhi,gnair}@unimelb.edu.au

***Abstract*—We consider private function evaluation to provide query responses based on private data of multiple untrusted entities in such a way that each cannot learn something substantially new about the data of others. First, we introduce perfect non-stochastic privacy in a two-party scenario. Perfect privacy amounts to conditional unrelatedness of the query response and the private uncertain variable of other individuals conditioned on the uncertain variable of a given entity. We show that perfect privacy can be achieved for queries that are functions of the common uncertain variable, a generalization of the common random variable. We compute the closest approximation of the queries that do not take this form. To provide a trade-off between privacy and utility, we relax the notion of perfect privacy. We define almost perfect privacy and show that this new definition equates to using conditional disassociation instead of conditional unrelatedness in the definition of perfect privacy. Then, we generalize the definitions to multi-party function evaluation (more than two data entities). We prove that uniform quantization of query responses, where the quantization resolution is a function of privacy budget and sensitivity of the query (cf., differential privacy), achieves function evaluation privacy.**

I. Introduction

Privacy research in information theory [1] and computer science [2] often deals with the problem of reporting privacy-preserving query responses based on private data available on a secure server. That is, when computing the privacy-preserving responses, the server has access to the entire private dataset and generates a noisy output with desired utility and privacy against a third-party adversary. Those studies fail to investigate private information leakage from the query responses to individuals whose data is used for responding to the query, which amounts to privacy analysis in the presence of side-channel information. This is an important problem when multiple untrusted parties must come together to compute the response to an aggregate query.

In this work, we consider providing query responses based on the data of multiple untrusted entities in such a way that they cannot learn something substantially new about the data of others. We refer to this as *private function evaluation*. At first, we introduce and investigate perfect privacy in a two-party scenario. This is the non-stochastic counterpart of perfect privacy in stochastic literature [3]–[5]. Perfect privacy equates to conditional unrelatedness of the query response and the private uncertain variable of other individuals conditioned on the uncertain variable of a given entity. We show that perfect privacy can be achieved for queries that are functions of the common uncertain variable, a generalization of the common random variable in the sense of [6]. For queries that do not take this form, we can approximate the query to compute the response with the most utility, i.e., closest worst-case response. To provide a trade-off between privacy and utility, we relax the notion of privacy. We define almost perfect non-stochastic privacy and show that this new definition equates to using conditional disassociation, borrowed from [7], instead of conditional unrelatedness in the definition of perfect privacy. We investigate the family of functions that can achieve almost perfect privacy. Then, we generalize our definition to multi-party function evaluation (more than two data entities). We prove that private function evaluation can be achieved by uniform quantization of the query responses, where the quantization resolution is a function of privacy budget and sensitivity of the query (cf., scale of the Laplace mechanism differential privacy [8]). We also investigate the utility of these reporting policies to establish utility-privacy trade-off.

Private function evaluation, at first glance, might seem related to private computation [9] and private information retrieval [10]. However, in private function evaluation, the objective is not to hide the function to be computed (cf., private computation) or the datasets on which the function is evaluated (cf., private information retrieval). We are rather interested in ensuring that the parties contributing to the private dataset cannot infer the private information of the other parties. This links the problem more intimately to privacy-preserving distributed decision making [11]. However, all those papers consider stochastic mechanisms for ensuring privacy while, in this paper, we are interested in a non-stochastic notion of information leakage and privacy.

Despite their shortcomings, due to heuristic-based development, non-stochastic privacy-preserving policies have remained popular, e.g., [12]. Those studies are motivated by concerns about the use of randomization in popular stochastic approaches. For instance, randomized policies, stemming from differential privacy in financial auditing, can potentially complicate fraud detection [13] and cause difficulties in medical, health, or social research [14]. This has motivated the use of information-theoretic tools to investigate non-stochastic privacy recently [15], [16]. Nonetheless, those studies do not consider the private function computation setup in this paper.

II. Uncertain Variables

In this section, we present necessary preliminaries from non-stochastic information theory from [15], [17].

A sample space Ω models the set of uncertainties. An uncertain variable is a mapping from the sample space to a set of interest, such as $X : \Omega \to \mathbb{X}$. Here, $X(\omega)$ denotes a

The work of F. Farokhi is funded by the Melbourne School of Engineering. The work of G. Nair was supported by the Australian Research Council grant FT140100527.

realization of uncertain variable X corresponding to sample $\omega \in \Omega$. If $\mathbb{X}$ is finite, the uncertain variable X is discrete. In this paper, we focus on discrete uncertain variables. The range of an uncertain variable X is the set of all its realizations, e.g., $[\![X]\!] := \{X(\omega) : \omega \in \Omega\} \subseteq \mathbb{X}$. The joint range of any two uncertain variables $X : \Omega \to \mathbb{X}$ and $Y : \Omega \to \mathbb{Y}$ is $[\![X, Y]\!] := \{(X(\omega), Y(\omega)) : \omega \in \Omega\} \subseteq \mathbb{X} \times \mathbb{Y}$. The conditional range of uncertain variable X, conditioned on a realization of another uncertain variable $Y(\omega) = y$, is $[\![X|Y(\omega) = y]\!] := \{X(\omega) : \exists \omega \in \Omega$ such that $Y(\omega) = y\} = X(Y^{-1}(y)) \subseteq [\![X]\!]$, where Y^{-1} denotes the pre-image or inverse image of Y. Whenever it is evident from the context, $[\![X|y]\!]$ is used instead of $[\![X|Y(\omega) = y]\!]$ for the sake of brevity.

Uncertain variables $X_1, \ldots, X_n$ are unrelated if $[\![X_1, \ldots, X_n]\!] = [\![X_1]\!] \times \cdots \times [\![X_n]\!]$. They are conditionally unrelated (conditioned on observations of Y) if $[\![X_1, \ldots, X_n|Y(\omega) = y]\!] = [\![X_1|Y(\omega) = y]\!] \times \cdots \times [\![X_n|Y(\omega) = y]\!], \forall y \in [\![Y]\!]$. For two uncertain variables, this is equivalent to stating that X_1 and X_2 are unrelated if $[\![X_1|X_2(\omega) = x_2]\!] = [\![X_1]\!], \forall x_2 \in [\![X_2]\!]$, and *vice versa* [17]. For two uncertain variables, X_1 and X_2 are conditionally unrelated (conditioned on realizations of Y) if $[\![X_1|X_2(\omega) = x_2, Y(\omega) = y]\!] = [\![X_1|Y(\omega) = y]\!], \forall (x_2, y) \in [\![X_2, Y]\!]$ [17].

The non-stochastic entropy of uncertain variable X is $H_0(X) := \log_2(|[\![X]\!]|)$. This is commonly referred to as the Hartley entropy [17], [18] and coincides with the Rényi entropy of order 0 for discrete variables [19]. Conditional (or relative) entropy of uncertain variable X given uncertain variable Y is $H_0(X|Y) := \max_{y \in [\![Y]\!]} \log_2(|[\![X|Y(\omega) = y]\!]|)$. This coincides with the Arimoto-Rényi conditional entropy of order 0 [20]. So, the non-stochastic information between two uncertain variables X and Y can be defined as the difference of the entropy of X with and without access to realizations of Y: $I_0(X;Y) := H_0(X) - H_0(X|Y) = \min_{y \in [\![Y]\!]} \log_2 \left(|[\![X]\!]|/|[\![X|y]\!]|\right)$. The zeroth order information, defined above, is not symmetric in general. It is related to Kolmogorov's information gain $|[\![X]\!]|/|[\![X|y]\!]|$ and the 'combinatorial' conditional entropy $\log_2(|[\![X|y]\!]|)$ [21]. However, the combinatorial conditional entropy and the information gain are defined for a given realization while this is for the least informative realization.

In [15], it was observed that, in the context of information-theoretic privacy, I_0 is not an appropriate measure of information leakage. This is because I_0 focuses on least informative observations while a privacy-intrusive adversary is interested in most informative realizations. Therefore, in [15], an alternative non-stochastic information leakage is proposed as

$$L_0(X;Y) := \max_{y \in [\![Y]\!]} \log_2 \left(\frac{|[\![X]\!]|}{|[\![X|y]\!]|} \right). \tag{1}$$

Non-stochastic information leakage $L_0(X;Y)$ captures the worst-case reduction in the complexity of brute-force guessing X after observing Y [22]. In general, I_0 and L_0 are not equal, i.e., $I_0(X;Y) \neq L_0(Y;X)$. In fact, $I_0(X;Y) \leq L_0(X;Y)$. Again, $L_0(X;Y)$ is not symmetric. Note that $L_0(X;Y) \geq 0$ with equality achieved if and only if X and Y are unrelated. We can define conditional non-stochastic information leakage:

$$L_0(X;Y|Z) := \max_{(y,z) \in [\![Y,Z]\!]} \log_2 \left(\frac{|[\![X|Z(\omega) = z]\!]|}{|[\![X|Y(\omega) = y, Z(\omega) = z]\!]|} \right). \tag{2}$$

Note that $L_0(Y;X|Z) \geq 0$ with equality achieved if and only if X and Y are unrelated conditioned on Z .

In [17], maximin or non-stochastic information is introduced as a symmetric measure of information and its relationship with zero-error capacity is explored. To present its definition, we need to introduce overlap partitions.

Definition 1 (Overlap Partition)**.**

- *$x, x' \in [\![X]\!]$ are overlap connected (via $[\![X|Y]\!]$), $x \leftrightsquigarrow x'$, if there exists a finite sequence of conditional ranges $\{[\![X|y_i]\!]\}_{i=1}^n$ such that $x \in [\![X|y_1]\!]$, $x' \in [\![X|y_n]\!]$, and $[\![X|y_i]\!] \cap [\![X|y_{i+1}]\!] \neq \emptyset$ for all $i = 1, \ldots, n-1$;*
- *$\mathcal{A} \subseteq [\![X]\!]$ is overlap connected if all $x, x' \in \mathcal{A}$ are overlap connected;*
- *$\mathcal{A}, \mathcal{B} \subseteq [\![X]\!]$ are overlap isolated if there do not exist points $x \in \mathcal{A}$ and $x' \in \mathcal{B}$ such that $x \leftrightsquigarrow x'$;*
- *An overlap partition of $[\![X]\!]$ is a set of sets $[\![X|Y]\!]_\star := \{\mathcal{A}_i\}_{i=1}^n$ such that $[\![X]\!] \subseteq \bigcup_{i=1}^n \mathcal{A}_i$, $\mathcal{A}_i, \mathcal{A}_j$ are overlap isolated if $j \neq i$, and $\mathcal{A}_i$ is overlap connected;*

There always exists a unique overlap partition [17]. The maximin information is

$$I_\star(X;Y) := \log_2(|[\![X|Y]\!]_\star|). \tag{3}$$

Note that $I_\star(X;Y) \geq 0$ and $I_\star(X;Y) = 0$ if and only if uncertain variables X and Y are unrelated. The maximin information is symmetric [17], i.e., $I_\star(X;Y) = I_\star(Y;X)$. The maximin information is related to the non-stochastic information leakage $I_\star(X;Y) \leq L_0(X;Y)$ [23]. Due to symmetry of maximin information, $I_\star(X;Y) = I_\star(Y;X) \leq L_0(Y;X)$.

III. Common Uncertain Variable, Information, and Perfect Privacy

We first discuss extension of common random variables in [6] to uncertain variables in line with the approach of [17]. This extends the use of common information [24], also known as the Gács-Körner common information [25], in perfect privacy [26] to the non-stochastic framework.

A. Common Uncertain Variable and Information

We start by introducing common uncertain variables and relating it to overlap partitions and maximin information.

Definition 2 (Common Uncertain Variable)**.** *Let X_1 and X_2 be any two uncertain variables with disjoint[1] ranges.*

- *$\mathcal{G}$ is a bipartite graph with the vertex set $\mathcal{V} = [\![X_1]\!] \cup [\![X_2]\!]$ and the edge set $\mathcal{E} = [\![X_1, X_2]\!]$;*

[1]The disjoint assumption is just to simplify definition of the bipartite graph by making vertexes associated with alphabet of X_1 and X_2 distinguishable. This is clearly without loss of generality as changing the event sets/alphabets of uncertain variables does not change their properties.

- $f_1 : [\![X_1]\!] \to 2^{[\![X_1]\!] \cup [\![X_2]\!]}$ *is a function that maps* $x_1 \in [\![X_1]\!] \subseteq \mathcal{V}$ *to the set of vertices in* $[\![X_1]\!] \cup [\![X_2]\!]$ *that are in the connected component of* $\mathcal{G}$ *containing* x_1*;*
- $f_2 : [\![X_2]\!] \to 2^{[\![X_1]\!] \cup [\![X_2]\!]}$ *is a function that maps* $x_2 \in [\![X_2]\!] \subseteq \mathcal{V}$ *to the set of vertices in* $[\![X_1]\!] \cup [\![X_2]\!]$ *that are in the connected component of* $\mathcal{G}$ *containing* x_2*.*

The common uncertain variable is $X_1 \wedge X_2 = f_1 \circ X_1 = f_2 \circ X_2$.

Similar to [6], we should note that the common uncertain variable $X_1 \wedge X_2$ is the "largest" uncertain variable that can be extracted from uncertain variables X_1 and X_2.

Proposition 1. *Assume that uncertain variables* X_1*,* X_2*, and* C *exist such that* $C = \bar{f}_1 \circ X_1 = \bar{f}_2 \circ X_2$ *for functions* $\bar{f}_1 : [\![X_1]\!] \to [\![C]\!]$ *and* $\bar{f}_1 : [\![X_2]\!] \to [\![C]\!]$*. There exists* $g : 2^{[\![X_1]\!] \cup [\![X_2]\!]} \to [\![C]\!]$ *such that* $C = g(X_1 \wedge X_2)$.

Proof: The proof follows the same line of reasoning as in the proof of Lemma 1 in [6]. ∎

For the unique overlap partition of $[\![X_1]\!]$, $[\![X_1|X_2]\!]_\star = \{\mathcal{A}_i\}_{i=1}^{n_1}$ with $n_1 \in \mathbb{N}$, define $\mathfrak{i}_1 : [\![X_1]\!] \to \{1, \ldots, n_1\}$ such that $\mathfrak{i}_1(x_1) = i$ for which $x_1 \in \mathcal{A}_i$. Similarly, for the unique overlap partition of $[\![X_2]\!]$, $[\![X_2|X_1]\!]_\star = \{\mathcal{B}_i\}_{i=1}^{n_2}$ with $n_2 \in \mathbb{N}$, define $\mathfrak{i}_2 : [\![X_2]\!] \to \{1, \ldots, n_2\}$ such that $\mathfrak{i}_2(x_2) = i$ for which $x_2 \in \mathcal{B}_i$. The mappings $\mathfrak{i}_1$ and $\mathfrak{i}_1$ are well-defined because $[\![X_1|X_2]\!]_\star$ and $[\![X_2|X_1]\!]_\star$ are partitions.

Definition 3 (Equivalence)**.** *Two uncertain variables* X *and* Y *are equivalent,* $X \equiv Y$*, if there exists a one-to-one correspondence*[2] $f : [\![X]\!] \to [\![Y]\!]$ *such that* $Y = f \circ X$.

The notion of "equivalence" between two uncertain variables is weaker than "equality"; see [27] for random variables. Entropy is invariant under the equivalence relationship.

Proposition 2. $X_1 \wedge X_2 \equiv \mathfrak{i}_1 \circ X_1 \equiv \mathfrak{i}_2 \circ X_2$ *and* $I_\star(X_1; X_2) = H_0(X_1 \wedge X_2)$.

Proof: Due to page limits, the proofs are presented in a technical note online [28]. ∎

B. Perfect Privacy

Perfect privacy [3], defined by adapting Shannon's perfect secrecy [29] to the privacy framework, states that an observations is perfectly private if it is statistically independent of the private random variable. This concept has been recently investigated [4], [5] to provide a fundamental understanding of utility-privacy trade-off. In the non-stochastic case, independence can be replaced with unrelatedness. We can tailor this definition to the case of private function computation by assuming that, conditioned on the realization of the uncertain variable of each party, the outcome should not leak any information about the uncertain variable of the other party.

Definition 4 (Perfect Privacy in Two-Party Function Evaluation)**.** *Let* X_1 *and* X_2 *be any two uncertain variables. The mapping* $f : [\![X_1, X_2]\!] \to \mathbb{R}^m$ *provides perfect privacy if* $f(X_1, X_2)$ *is unrelated to* X_1 *conditioned on* X_2 *and* $f(X_1, X_2)$ *is unrelated to* X_2 *conditioned on* X_1.

Note that Definition 4 implies that the mapping f provides perfect privacy if $[\![X_2|X_1(\omega) = x_1, Z(\omega) = z]\!] = [\![X_2|X_1(\omega) = x_1]\!]$ and $[\![X_1|X_2(\omega) = x_2, Z(\omega) = z]\!] = [\![X_1|X_2(\omega) = x_2]\!]$ with $Z = f(X_1, X_2)$. Perfect privacy for two-party function evaluation can be equivalently defined using conditional non-stochastic information leakage. This is proved in the next proposition.

Proposition 3. *Mapping* f *provides perfect privacy if and only if* $L_0(f(X_1, X_2); X_1|X_2) = L_0(f(X_1, X_2); X_2|X_1) = 0$.

Proof: See online technical note [28]. ∎

Proposition 4. *The following statements hold:*

- $X_1 \wedge X_2$ *is unrelated to* X_1 *conditioned on* X_2*;*
- $X_1 \wedge X_2$ *is unrelated to* X_2 *conditioned on* X_1.

Proof: See online technical note [28]. ∎

Hence, all functions of the common uncertain variable also provide perfect privacy. The inverse is however also true. In fact, any function that provides perfect privacy must only be computable based on the common uncertain variable. This is explored in the following proposition.

Proposition 5. *For any* $f : [\![X_1, X_2]\!] \to \mathbb{R}^m$ *providing perfect privacy, there exists* $g : 2^{[\![X_1]\!] \cup [\![X_2]\!]} \to \mathbb{R}^m$ *such that* $f(X_1, X_2) = g(X_1 \wedge X_2)$.

Proof: See online technical note [28]. ∎

Not all functions provides perfect privacy as not all functions can be rewritten in terms of the common uncertain variable. For instance, $f(X_1, X_2) = X_1$ cannot be written in terms of the common uncertain variable. This function also does not provide perfect privacy. Thus, we might need to approximate such a function with one that provides perfect privacy; see online technical note for more detail [28]. In general, the condition for perfect privacy does not offer a systematic way for trading-off utility and privacy. In the remainder of this paper, we relax this notion of privacy.

IV. Almost Perfect Privacy

Recall that Proposition 3 implies f provides perfect privacy if and only if $\max\{L_0(f(X_1, X_2); X_1|X_2), L_0(f(X_1, X_2); X_2|X_1)\} = 0$. Therefore, we can relax perfect privacy by requiring that $\max\{L_0(f(X_1, X_2); X_1|X_2), L_0(f(X_1, X_2); X_2|X_1)\}$ is small rather than zero.

Definition 5 (γ-Privacy in Two-Party Function Evaluation)**.** *Let* X_1 *and* X_2 *be any two uncertain variables. For* $\gamma \geq 0$*, the mapping* $f : [\![X_1, X_2]\!] \to \mathbb{R}^m$ *provides* γ*-privacy if* $\max\{L_0(f(X_1, X_2); X_1|X_2), L_0(f(X_1, X_2); X_2|X_1)\} \leq \gamma$.

In what follows, we borrow disassociation from [7] as a relaxation of unrelatedness. This way, we can investigate γ-privacy in more depth by casting it in terms of disassociation rather than conditional information leakage.

[2]The partitions of sample space Ω induced by X and Y are the same though their labeling may be different.

Definition 6 (Disassociated Uncertain Variables)**.** *For $\delta \in [0,1]$, two uncertain variables X and Y are δ-disassociated if*

$$\frac{|[\![X|Y(\omega)=y_1]\!] \cap [\![X|Y(\omega)=y_2]\!]|}{|[\![X]\!]|} \geq \delta, \quad \forall y_1, y_2 \in [\![Y]\!] : y_1 \neq y_2, \tag{4a}$$

$$\frac{|[\![Y|X(\omega)=x_1]\!] \cap [\![Y|X(\omega)=x_2]\!]|}{|[\![Y]\!]|} \geq \delta, \quad \forall x_1, x_2 \in [\![X]\!] : x_1 \neq x_2. \tag{4b}$$

If only (4a) *holds, X is partially δ-disassociated with Y.*

As δ increases, any two δ-disassociated uncertain variables "appear more unrelated" and 1-disassociated implies unrelatedness between two uncertain variables [28]. Therefore, we can think of δ-disassociation as a relaxation of unrelatedness.

Proposition 6. *For δ-disassociated uncertain variables X and Y, $L_0(X;Y) \leq -\log_2(\delta)$, $L_0(Y;X) \leq -\log_2(\delta)$, and $I_\star(X;Y) \leq -\log_2(\delta)$.*

Proof: See online technical note [28]. ∎

Proposition 6 shows that the information between any two δ-disassociated uncertain variables X and Y reduces as δ gets larger. For $\delta = 1$, $L_0(X;Y) = L_0(Y;X) = I_\star(X;Y) = 0$. This shows X and Y are unrelated if they are 1-disassociated.

Definition 7 (Conditionally Disassociated Uncertain Variables)**.** *For $\delta \in [0,1]$, two uncertain variables X and Y are δ-disassociated conditioned on uncertain variable Z if*

$$\frac{|[\![X|Y(\omega)=y_1, Z(\omega)=z]\!] \cap [\![X|Y(\omega)=y_2, Z(\omega)=z]\!]|}{|[\![X|Z(\omega)=z]\!]|} \geq \delta, \quad \forall y_1, y_2 \in [\![Y|Z(\omega)=z]\!] : y_1 \neq y_2, \forall z \in [\![Z]\!], \tag{5a}$$

$$\frac{|[\![Y|X(\omega)=x_1, Z(\omega)=z]\!] \cap [\![Y|X(\omega)=x_2, Z(\omega)=z]\!]|}{|[\![Y|Z(\omega)=z]\!]|} \geq \delta, \quad \forall x_1, x_2 \in [\![X|Z(\omega)=z]\!] : x_1 \neq x_2, \forall z \in [\![Z]\!]. \tag{5b}$$

If only (5a) *holds, X is partially δ-disassociated with Y conditioned on Z.*

Similarly, two uncertain variables X and Y are unrelated conditioned on uncertain variable Z if they are 1-disassociated conditioned on Z. Hence, we can think of conditional disassociation as a relaxation of conditional unrelatedness.

Proposition 7. *Assume that two uncertain variables X and Y are δ-disassociated conditioned on uncertain variable Z for some $\delta \in [0,1]$. Then, $L_0(X;Y|Z) \leq -\log_2(\delta)$.*

Proof: See online technical note [28]. ∎

Proposition 8. *Let X_1 and X_2 be any two uncertain variables. For $\gamma \geq 0$, the mapping $f : [\![X_1, X_2]\!] \to \mathbb{R}^m$ provides γ-privacy if $f(X_1, X_2)$ and X_1 are $e^{-\gamma}$-disassociated conditioned on X_2, and $f(X_1, X_2)$ and X_2 are $e^{-\gamma}$-disassociated conditioned on X_1.*

Proof: It follows form Proposition 7. ∎

Proposition 8 shows that we can relax the definition of perfect privacy by requiring conditional disassociation instead of conditional unrelatedness.

Definition 8 (δ-Overlap Connectivity)**.** *For $\delta \in [0,1]$,*

- *$x, x' \in [\![X]\!]$ are δ-overlap connected (via $[\![X|Y]\!]$), $x \leftrightsquigarrow_\delta x'$, if there exists a sequence of points $\{y_i\}_{i=1}^n \subseteq [\![Y]\!]$ such that $x \in [\![X|Y(\omega)=y_1]\!]$, $x' \in [\![X|Y(\omega)=y_n]\!]$, and $|[\![X|Y(\omega)=y_i]\!] \cap [\![X|Y(\omega)=y_{i-1}]\!]|/|[\![X]\!]| \geq \delta$, for all $i \in \{2,\dots,n\}$;*
- *if $x, x' \in [\![X]\!]$ are δ-overlap connected with $n = 1$, they are singly δ-overlap connected;*
- *$\mathcal{A} \subseteq [\![X]\!]$ is (singly) δ-overlap connected if all points in $[\![X]\!]$ are (singly) δ-overlap connected;*
- *$\mathcal{A}, \mathcal{B} \subseteq [\![X]\!]$ are δ-overlap isolated if there do not exist points $x \in \mathcal{A}$ and $x' \in \mathcal{B}$ such that $x \leftrightsquigarrow_\delta x'$;*
- *A δ-overlap partition of $[\![X]\!]$ is a set of sets $[\![X|Y]\!]_\delta := \{\mathcal{A}_i\}_{i=1}^n$ such that $[\![X]\!] \subseteq \bigcup_{i=1}^n \mathcal{A}_i$, $\mathcal{A}_i, \mathcal{A}_j$ are δ-overlap isolated if $j \neq i$, and $\mathcal{A}_i$ is δ-overlap connected;*
- *A δ-overlap family of $[\![X]\!]$, denoted by $[\![X|Y]\!]_\star^\delta$, is the largest δ-overlap partition of $[\![X]\!]$ such that each set in the family contains a singly δ-overlap connected set of the form $[\![X|Y(\omega)=y]\!]$, there exists a set containing any two singly δ-overlap connected points, and the measure of overlap between any two sets in the family is at most $\delta|[\![X]\!]|$.*

For uncertain variables X and Y, there always exists a δ-overlap family of $[\![X]\!]$ while the uniqueness is guaranteed if X and Y are δ-disassociated [7, Theorems 3-4]. For a given δ-overlap family of $[\![X_1]\!]$, $[\![X_1|X_2]\!]_\star^\delta = \{\mathcal{A}_i\}_{i=1}^{n_1}$ with $n_1 \in \mathbb{N}$, define $\mathfrak{i}_1^\delta : [\![X_1]\!] \to \{1,\dots,n_1\}$ such that $\mathfrak{i}_1^\delta(x_1) = i$ for which $x_1 \in \mathcal{A}_i$. Similarly, for a given δ-overlap family of $[\![X_2]\!]$, $[\![X_2|X_1]\!]_\star^\delta = \{\mathcal{B}_i\}_{i=1}^{n_2}$ with $n_2 \in \mathbb{N}$, define $\mathfrak{i}_2^\delta : [\![X_2]\!] \to \{1,\dots,n_2\}$ such that $\mathfrak{i}_2^\delta(x_2) = i$ for which $x_2 \in \mathcal{B}_i$. The mappings $\mathfrak{i}_1^\delta$ and $\mathfrak{i}_1^\delta$ are well-defined because $[\![X_1|X_2]\!]_\star^\delta$ and $[\![X_2|X_1]\!]_\star^\delta$ partition $[\![X_1]\!]$ and $[\![X_2]\!]$, respectively.

Proposition 9. *If X_1 and X_2 are δ-disassociated for $\delta \in [0,1]$, $\mathfrak{i}_1^\delta \circ X \equiv \mathfrak{i}_2^\delta \circ X_2$.*

Proof: See online technical note [28]. ∎

Proposition 10. *For $\delta \in [0,1]$, the following statements hold:*

- *$\mathfrak{i}_1^\delta \circ X$ and Y are δ-disassociated conditioned on X;*
- *$\mathfrak{i}_2^\delta \circ Y$ and X are δ-disassociated conditioned on Y.*

Proof: See online technical note [28]. ∎

Propositions 9 and 10 show that equivalent uncertain variables $\mathfrak{i}_1^\delta \circ X$ and $\mathfrak{i}_2^\delta \circ X_2$ are akin to relaxations of the common uncertain variable (c.f., Proposition 2).

Corollary 11. *Let $f : [\![X_1, X_2]\!] \to \mathbb{R}^m$ be any mapping for which there exists $g_1 : [\![\mathfrak{i}_1^\delta \circ X]\!] \to \mathbb{R}^m$ or $g_2 : [\![\mathfrak{i}_2^\delta \circ Y]\!] \to \mathbb{R}^m$ such that $f(X_1, X_2) = g_1 \circ \mathfrak{i}_1^\delta \circ X$ or $f(X_1, X_2) = g_2 \circ \mathfrak{i}_2^\delta \circ Y$. Then, f provides $-\log_2(\delta)$-privacy.*

V. Private Multi-Party Function Evaluation

Now, we generalize the earlier sections to more than two entities. Consider $n \geq 2$ entities, each possessing uncertain variable X_i. Let $X = (X_i)_{i=1}^n$. We want to publish the evaluation of $f : [\![X]\!] \to [\![Y]\!]$ in a privacy-preserving manner.

Definition 9 (γ-Privacy in Multi-Party Function Evaluation)**.** *Let $X_1, X_2, \ldots, X_n$ be any $n \geq 1$ uncertain variables. For $\gamma \geq 0$, the mapping $f : [\![X_1, \ldots, X_n]\!] \to \mathbb{R}^m$ provides γ-privacy if $\max_{1 \leq i \leq n} L_0(f(X_1, \ldots, X_n); X_{-i}|X_i) \leq \gamma$.*

We might not be able to evaluate any function f in a privacy-preserving manner. Therefore, we might need to approximate function f with one that can be privately evaluated $f' : [\![X]\!] \to [\![Z]\!]$, where $Z = f' \circ X$. Doing so, we publish the result of evaluating another function f' instead f. The error in the function evaluation is measured by $\mathcal{E}(f', f) = \sup_{x \in [\![X]\!]} \rho(f(x), f'(x))$, where $\rho : [\![Y]\!] \times [\![Z]\!] \to \mathbb{R}$ is a distance function. We use $\rho(y, z) = \|y - z\|$ for $[\![Y]\!], [\![Z]\!] \subseteq \mathbb{R}^m$.

Definition 10 (Accuracy)**.** *Any f' is said to be β-accurate for $\beta > 0$ if $\mathcal{E}(f', f) \leq \beta$.*

We show linear quantizers, defined below, can provide privacy. They have been previously used to provide privacy in the sense of non-stochastic information leakage [15].

Definition 11 (Linear Quantizer)**.** *A q-level quantizer $\mathcal{Q} : [x_{\min}, x_{\max}] \to \{b_1, \ldots, b_q\}$ is a piecewise constant function defined as $Q(x) = b_i$ if $x_i \in [x_i, x_{i+1})$, where $(b_i)_{i=1}^q$ are distinct symbols and $x_1 \leq x_2 \leq \cdots \leq x_q$ are real numbers such that $x_1 = x_{\min}$, $x_{q+1} = x_{\max}$, $x_{i+1} - x_i = (x_{\max} - x_{\min})/q$. It is a mid-point linear quantizer if $b_i = (x_i + x_{i+1})/2, \forall i$.*

Theorem 12. *Assume that f is Lipschitz continuous, i.e., there exists $L > 0$ such that $|f(x) - f(x')| \leq L\|x - x'\|_\infty$ for all $x, x' \in [\![X]\!]$, and $[\![X_i]\!] \subseteq [x_{\min}, x_{\max}]$ for all i. Then, $f' = \mathcal{M} \circ f$, where $\mathcal{M}$ is a $\lfloor \exp(\gamma) - 1 \rfloor$-level mid-point linear quantizer over $[\![f(X)]\!]$, is γ-private and β-accurate with $\beta \geq L(x_{\max} - x_{\min})/\lfloor \exp(\gamma) - 1 \rfloor$.*

Proof: See online technical note [28]. ∎

Theorem 12 shows that private function evaluation can be achieved by uniform quantization of the query responses, where the quantization resolution is a function of privacy budget γ and sensitivity of the query $L(x_{\max} - x_{\min})$ (cf., scale of the Laplace mechanism differential privacy [8]). Note that $L(x_{\max} - x_{\min})$ captures the sensitivity of f, i.e., how much the output of the function f varies if one of its entries change. For the mechanism in Theorem 12, $\beta \exp(\gamma) \geq L(x_{\max} - x_{\min})$. This inequality provides a utility-privacy trade-off for non-stochastic private function evaluation.

VI. Conclusions

We consider private function evaluation to provide query responses based on private data of multiple untrusted entities in such a way that no entity can learn something substantially new about the data of others. We prove that uniform quantization of the query responses achieves privacy.

References

[1] F. du Pin Calmon and N. Fawaz, "Privacy against statistical inference," in *2012 50th annual Allerton conference on communication, control, and computing (Allerton)*, pp. 1401–1408, 2012.

[2] 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, 2006.

[3] G. Miklau and D. Suciu, "A formal analysis of information disclosure in data exchange," *Journal of Computer and System Sciences*, vol. 73, no. 3, pp. 507–534, 2007.

[4] F. P. Calmon, A. Makhdoumi, and M. Médard, "Fundamental limits of perfect privacy," in *2015 IEEE International Symposium on Information Theory (ISIT)*, pp. 1796–1800, 2015.

[5] B. Rassouli and D. Gunduz, "On perfect privacy," in *IEEE International Symposium on Information Theory (ISIT)*, pp. 2551–2555, 2018.

[6] S. Wolf and J. Wultschleger, "Zero-error information and applications in cryptography," in *Information Theory Workshop*, pp. 1–6, 2004.

[7] A. Rangi and M. Franceschetti, "Towards a non-stochastic information theory," in *IEEE International Symposium on Information Theory (ISIT)*, pp. 997–1001, 2019.

[8] C. Dwork, "Differential privacy: A survey of results," in *Theory and Applications of Models of Computation* (M. Agrawal, D. Du, Z. Duan, and A. Li, eds.), (Berlin, Heidelberg), pp. 1–19, Springer, 2008.

[9] B. Tahmasebi and M. A. Maddah-Ali, "Private sequential function computation," in *IEEE International Symposium on Information Theory (ISIT)*, pp. 1667–1671, 2019.

[10] K. Banawan and S. Ulukus, "Private information retrieval from multiple access channels," in *IEEE Information Theory Workshop (ITW)*, 2018.

[11] M. J. Wainwright, M. I. Jordan, and J. C. Duchi, "Privacy aware learning," in *Advances in Neural Information Processing Systems*, pp. 1430–1438, 2012.

[12] 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.

[13] 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*, 2011.

[14] 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.

[15] F. Farokhi, "Development and analysis of deterministic privacy-preserving policies using non-stochastic information theory," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 10, pp. 2567–2576, 2019.

[16] N. Ding and F. Farokhi, "Developing non-stochastic privacy-preserving policies using agglomerative clustering," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 3911–3923, 2020.

[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] R. V. L. Hartley, "Transmission of information," *Bell System Technical Journal*, vol. 7, no. 3, pp. 535–563, 1928.

[19] A. Rényi, "On measures of entropy and information," in *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, 1961.

[20] S. Arimoto, "Information measures and capacity of order α for discrete memoryless channels," in *Proceedings of the 2nd Colloquium on Topics on Information Theory, Keszthely, Hungary*, vol. 16, 1975.

[21] 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.

[22] F. Farokhi and N. Ding, "Measuring information leakage in non-stochastic brute-force guessing," *arXiv preprint arXiv:2004.10911*, 2020.

[23] N. Ding and F. Farokhi, "Developing non-stochastic privacy-preserving policies using agglomerative clustering," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 3911–3923, 2020.

[24] C. Shannon, "The lattice theory of information," *Transactions of the IRE professional Group on Information Theory*, vol. 1, no. 1, pp. 105–107, 1953.

[25] P. Gács and J. Körner, "Common information is far less than mutual information," *Problems of Control and Information Theory*, vol. 2, no. 2, pp. 149–162, 1973.

[26] S. Asoodeh, F. Alajaji, and T. Linder, "Notes on information-theoretic privacy," in *2014 52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 1272–1278, 2014.

[27] H. Li and E. K. Chong, "On a connection between information and group lattices," *Entropy*, vol. 13, no. 3, pp. 683–708, 2011.

[28] F. Farokhi and G. Nair, "Non-stochastic private function evaluation," arXiv:2010.09968 [cs.IT], https://arxiv.org/abs/2010.09968.

[29] C. E. Shannon, "Communication theory of secrecy systems," *Bell System Technical Journal*, vol. 28, no. 4, pp. 656–715, 1949.