A Differentially Private Encryption Scheme

Carlo Brunetta, Christos Dimitrakakis, Bei Liang, and Aikaterini Mitrokotsa

Chalmers University of Technology, Gothenburg, Sweden
{brunetta,chrdimi,lbei,aikmitr}@chalmers.se

Abstract Encrypting data with a semantically secure cryptosystem guarantees that nothing is learned about the plaintext from the ciphertext. However, querying a database about individuals or requesting for summary statistics can leak information. *Differential privacy* (DP) offers a formal framework to bound the amount of information that an adversary can discover from a database with private data, when statistical findings of the stored data are communicated to an untrusted party. Although both encryption schemes and differential private mechanisms can provide important privacy guarantees, when employed in isolation they do not guarantee full privacy-preservation.
This paper investigates how to efficiently combine DP and an encryption scheme to prevent leakage of information. More precisely, we introduce and instantiate *differentially private encryption schemes* that provide both DP and confidentiality. Our contributions are five-fold, we: *(i)* define an encryption scheme that is **not** correct with some probability α_{m_1,m_2} *i.e.,* an α_{m_1,m_2}-correct encryption scheme and we prove that it satisfies the DP definition; *(ii)* prove that combining DP and encryption, is equivalent to using an α_{m_1,m_2}-correct encryption scheme and provide a construction to build one from the other; *(iii)* prove that an encryption scheme that belongs in the DP-then-Encrypt class is at least as computationally secure as the original base encryption scheme; *(iv)* provide an α_{m_1,m_2}-correct encryption scheme that achieves both requirements (*i.e.,* DP and confidentiality) and relies on Dijk *et al.*'s homomorphic encryption scheme (EUROCRYPT 2010); and *(v)* perform some statistical experiments on our encryption scheme in order to empirically check the correctness of the theoretical results.

Keywords: Differential privacy, Encryption, Homomorphic encryption

1 Introduction

The Internet has evolved into a powerful platform interconnecting billions of users and has changed the way we do business, communicate with our friends, and perform our financial transactions. In this new communication paradigm, we leave our digital fingerprints everywhere: medical records, financial records, web search histories, and social network data. There is no doubt that the privacy implications of this increased connectivity can lead to oppressive electronic data surveillance.

Let us consider a real-world scenario: a company sells electricity to different customers in large geographical areas. The company owns and distributes a smart-metering grid [6] in order to offer the lowest price possible for its customers. Alice, that wants to pay as less as possible for her electrical consumption, signs a contract with the company by providing her personal information and accepts to install in her home different *sensors* that will measure the electrical consumption during the day and transmit this data to the electricity company. The company collects data from all its customers in an entire geographical region and, by performing statistical analysis on the collected data, is able to optimize the electrical supply distribution. Alice worries that her data may be used in a malicious way and wants to get guarantees that her privacy will be respected. She is aware that by analysing the data of her power consumption, someone may deduce private information such as when she is at home and what habits she may have. She wants her personal information to be confidential (encrypted) when they are used by a third party but she accepts that the company may use her data for statistical analysis in order to optimise the supply distribution.

This particular problem might raise different privacy concerns that we categorize into two classes, as represented in Figure (1):

- An *individual privacy breach* can be described as the act of deducing private information for an individual from some public information.
  In this case, the electricity company can deduce Alice's habits just by observing her power consumption measurements.
- A *group privacy breach* can be defined as the act of deducing a single individual private information from public statistical information of groups of people.
  Let us suppose that the electricity company offers an open-source interface where everyone can query and obtain statistical information about the company's customers. The only limitation is that the statistics are not computed if the sample of customers is lower than five people.
  Eve wants to find out Alice's habits for malicious reasons. To achieve that she checks on every social network and finds out that Alice is a *student* and she lives in a *one-room apartment*. Eve starts querying the company's database by asking for the *"average daily power consumption of a student that lives in an one-room apartment"* and does not obtain any information because the sample is too small. Then, Eve asks for the *"average daily consumption of people that live in an one-room apartment"* and the *"average daily consumption of people that live in an one-room apartment that are* ***not*** *students"*. Thus, Eve can deduce some approximation of Alice's habits by computing the difference between the two values and obtain the *"average daily consumption of a student that lives in an one-room apartment"* in which Alice is contained.

In this paper, we do not deal with the problem of inferring some private information about an individual (such as habits) from other private data, such as consumption, from a trusted third party (*e.g.,* a company). However, we care

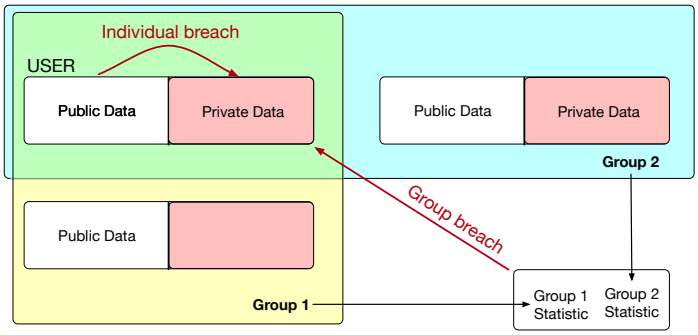


Figure 1: Individual and group privacy breaches.

about inferring private information from publicly available data published by a third party (*e.g.,* the billing information). To protect against either of the two types of privacy breaches, different notions of privacy and methodologies that preserve privacy have been defined in the literature such as *t-closeness* [11], *k-anonymity* [5], *ℓ-diversity* [8]. However, these notions of privacy have been proven to be weak, since even when they are employed information leakage and de-anomyization attacks can still be performed.

Differential privacy (DP) introduced by Dwork *et al.* [3], addresses the problem of learning as little as possible about an individual, while learning useful information about a population. It offers a formal framework that can be used to bound the amount of info that an adversary can discover from a database that contains private data, when statistical findings of the stored data are communicated to an untrusted party. More precisely, DP assumes the existence of a data aggregator, who is publishing statistics about a population. In other words, DP is a formalism that allows statistical analysis of private datasets while minimizing a group privacy breach. Informally, by employing a *DP-mechanism* to respond to a query, we are publishing noisy statistics about a dataset. The amount of noise should depend on the sensitivity of the queried statistic to the input, *i.e., "how much the query result would change if one single entry is changed or removed?"*. This means that if the query result will change a lot, we have to introduce more noise in order to *"hide"* the influence of the changed/removed entry in the query result. Otherwise, a drop in the query result will reveal partial information on the modified entry.

Complementary, a *semantically secure encryption* scheme guarantees the *confidentiality* of the encrypted information *i.e.,* no-one can decrypt and obtain the original message of a ciphertext. As a plus, an *homomorphic encryption* scheme [9,12] allows the computation of particular functions on the encrypted data. Informally, we can encrypt our messages and then compute a particular function on the ciphertext and obtain a new ciphertext that will be decrypted to the function computed on the original plaintext messages.

The solution required to avoid any possible information leakage should guarantee privacy breach *resistance* (provided by the DP framework) **and** confidentiality of the encrypted data (provided by a semantically secure encryption scheme). Each of these frameworks, if employed alone, does not provide full privacy guarantees. In this paper, we investigate for the first time, how we may achieve both differential privacy and confidentiality and introduce the concept of a differentially private encryption scheme.

Related Work: Privacy-preservation has received a lot of attention in the literature and multiple semantically secure crypto systems as well as differential private mechanisms have been proposed. However, existing work on encrypted computation and differential privacy has proceeded mainly in isolation. In order to avoid all possible information leakage, while guaranteeing both *confidentiality* and *differential privacy*, the most common solution is to process the plaintext data in a DP-mechanism and then encrypt the result using a secure homomorphic encryption scheme. The ciphertext will guarantee confidentiality until the decryption phase, while the plaintext message will satisfy the DP definition. In the literature, it is possible to find different solutions [10,1,7] that use this paradigm: a DP-mechanism and an encryption scheme; used sequentially. We will define these solutions that combine a *DP-framework* and an *Encryption-framework* as an element in the DP-then-Encrypt class (formally defined in Def. (5)). Our solution has as a starting point Dwork *et al.*'s definition of an α-correct encryption scheme [4] *i.e.,* an encryption scheme that can *wrongly decrypt* (or encrypt) a message with some probability bounded by α. Dwork *et al.* [4] defined an algorithm that takes an α-correct encryption scheme and returns a new encryption scheme, built using the α-correct one, that is correct (or almost-correct). We provide a more detailed definition of α-correctness, where we are interested in the precise probability of encrypting a message m_1 and obtaining a message m_2. Our definition is the first result that provides the sufficient conditions for an α-correct encryption scheme in order to achieve ϵ-DP. In order to build a concrete instantiation of a differentially private encryption scheme, we rely on Dijk *et al.*'s [2] homomorphic public-key encryption scheme over the integers.

Our Contributions: Our main idea is defining the class Encrypt+DP that contains all the encryption schemes that are differential private and achieve privacy and confidentiality *atomically*, as represented in Figure (2). As a starting point, we define an α_{m_1,m_2}-correct encryption scheme (Def. (4)) that will permit an encryption scheme to be **not** correct, *i.e.,* the decryption of the encryption of a specific message m_1 can be a different message m_2 with probability α_{m_1,m_2}. From this definition, we prove that an α_{m_1,m_2}-correct 1-bit encryption scheme satisfies the Dwork's DP definition [3] with $\epsilon(\alpha_{m_1,m_2})$-DP, *i.e.,* the DP parameter ϵ will be strongly related to the probabilities α_{m_1,m_2} of the encryption scheme. Then, we prove in Proposition (2) that the more general N-element encryption scheme achieves $\epsilon(\alpha_{m_1,m_2})$-DP.

Furthermore, we formally define the DP-then-Encrypt and Encrypt+DP classes. As our main result, we prove in Proposition (4) that the two classes are equiv-

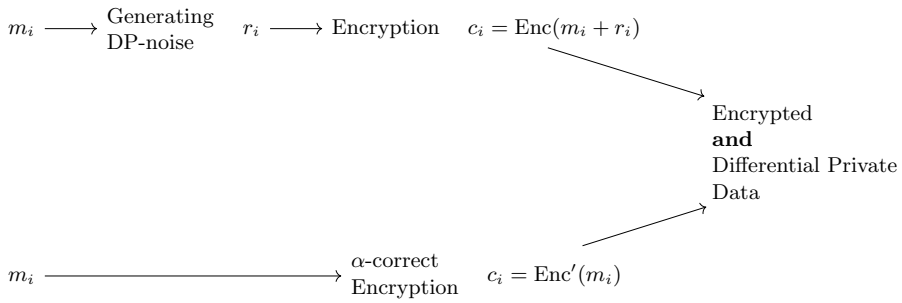


Figure 2: The difference between the DP-then-Encrypt (on the top) and our solution (at the bottom).

alent and provide a construction to switch between them. This means that our solution of an α_{m_1,m_2}-correct encryption scheme can be re-written with a DP-then-Encrypt encryption scheme.

As the second main contribution, in Lemma 1, we reduce the security of a DP-then-Encrypt encryption scheme to the security of the correct encryption scheme framework. The considered security-computational model is built around a non-interactive adversary that has access only to the public key and a particular ciphertext and it guesses the original plaintext. This security model is a necessary condition in order to satisfy more complex security models like $\mathrm{IND}-\mathrm{CPA}$, $\mathrm{IND}-\mathrm{CCA}$, etc.

The last contribution is a concrete $\alpha_{m,m}$-correct encryption scheme inside Encrypt+DP . We modify the Dijk *et al.* [2] integer homomorphic encryption scheme and we show how to compute the probability $\alpha_{m,m}$. As a final point, we exploit the structure of the scheme and obtain the correspondent DP-then-Encrypt encryption scheme that relies on Dijk *et al.*'s homomorphic encryption scheme.

Paper Organisation: The paper is organised as follows. In Section (2), we describe the notation used throughout the paper and the definitions we are based on. In Section (3), we give our definition of α_{m_1,m_2}-correct encryption schemes and prove that it has $\epsilon(\alpha_{m_1,m_2})$-DP. In Section (4), we show the equality between our framework, Encrypt+DP , and the DP-then-Encrypt . The proof will sketch an algorithm that transforms a correct encryption scheme into an α_{m_1,m_2}-correct encryption scheme. We define the security-hardness model and prove the security-hardness of a DP-then-Encrypt encryption scheme with respect to the corresponding base (correct) encryption scheme. In Section (5), we provide an instantiation of an $\alpha_{m,m}$-correct encryption scheme starting from Dijk *et al.*'s [2] encryption scheme and we prove its security.

2 Preliminaries

In this section, we will define the notation used in the paper and the basic definitions of the notions we employ in the rest of the paper.

2.1 Notation

We always denote with $\mathcal{M}$ the message-space. We denote with $\mathcal{K} = \mathcal{K}_{\text{sk}} \times \mathcal{K}_{\text{pk}}$ the key-space where $\mathcal{K}_{\text{sk}}$ is the secret-key-space and $\mathcal{K}_{\text{pk}}$ is the public key-space and with $\mathcal{C}$ the ciphertext-space. $\mathbb{N}$ is the set of natural numbers (*i.e.,* integers $z \geq 0$). Then we define intervals with $[a, b] = \{a, a+1, \cdots, b\}$ and $(a, b) = [a, b] \setminus \{a, b\}$. We denote with $\mathbb{1}_A$ the identity function on the set A. We define with the symbol $\simeq$, a *probabilistic* equality between functions, *i.e.,* $f(x) \simeq g(x)$ means $\mathbb{P}(f(x) = g(x)) = p$ for some $p \in [0, 1]$. We denote with $\text{negl}(n)$ a negligible function. We denote with $a \pmod n$ the modulo n of a in the interval $\left(-\frac{n}{2}, \frac{n}{2}\right]$. We denote with U_A the uniform distribution over the set A. We denote M times the cartesian product of a set A as A^M and the range of a function f with domain X as $\text{Rg}(f) \coloneqq \{f(x) : x \in X\}$. For a set X, we define with $\mathcal{P}(X)$ the power-set of X, *i.e.,* the set of all the subset of X.

2.2 Basic Definitions

In order to define differential privacy, we will define a data-set:

Definition 1 (Dataset). *A dataset D is defined on an alphabet A so that either $D \in A^n$ for a fixed dataset size n, or $D \in A^*$ with $A^* = \bigcup_{i=0}^{\infty} A^i$ being the union of all product sets of A.*

Definition 2 (ϵ-differential privacy [3]). *A randomized function $\mathcal{Q}$ is ϵ-differentially private if for all data-sets D_1 and D_2 differing on at most one element, i.e., the ℓ_0-distance between D_1 and D_2 is at most 1, and all $S \subseteq Rg(\mathcal{Q})$, it holds*

$$\mathbb{P}(\mathcal{Q}(D_1) \in S) \leq \exp(\epsilon) \cdot \mathbb{P}(\mathcal{Q}(D_2) \in S)$$

Remark 1. For finite ϵ, we must have that the distribution of a DP-mechanism has always the same range, *i.e.*, for every $D_0, D_1 \subset \mathcal{M}$ it holds $\text{Rg}(\mathcal{Q}(D_0)) = \text{Rg}(\mathcal{Q}(D_1))$.

In our construction, we will use messages as databases and we will always use the ℓ_0-distance; for two different messages m, m', the distance is always 1.

Below we provide Dwork *et al.*'s [4] definition of an α-correct (public-key) encryption scheme:

Definition 3 (Dwork *et al.*'s α-correct public-key encryption scheme [4]). *Let (G, E, D) be any public-key encryption scheme and $\alpha : \mathbb{N} \to [0, 1]$ an arbitrary function.*

(a) (G, E, D) *is* all-keys α-correct *if for every pair* $(\mathrm{sk}, \mathrm{pk})$ *generated by* G *on input* 1^λ*, it holds that* $\mathbb{P}(D_{\mathrm{sk}}(E_{\mathrm{pk}}(m)) \neq m) \leq 1-\alpha(\lambda)$*, where the probability is taken over the choice of* $m \in U_n$*, and over the random coins of* E *and* D*.*

(b) (G, E, D) *is* almost-all-keys α-correct *if with probability* $1 - \mathrm{negl}(\lambda)$ *over the random coins of* G *used to generate* $(\mathrm{sk}, \mathrm{pk})$ *on input* 1^λ*, it holds that* $\mathbb{P}(D_{\mathrm{sk}}(E_{\mathrm{pk}}(m)) \neq m) \leq 1 - \alpha(\lambda)$ *where the probability is taken over the choice of* $m \in U_n$ *and over the random coins of* E *and* D*.*

(c) (G, E, D) *is* almost-all-keys perfectly correct *if with probability* $1 - \mathrm{negl}(\lambda)$ *over the random coins of* G *used to generate* $(\mathrm{sk}, \mathrm{pk})$ *on input* 1^λ*, it holds that* $\mathbb{P}(D_{\mathrm{sk}}(E_{\mathrm{pk}}(m) \neq m) = 0$*, where the probability is taken over the choice of* $m \in U_n$ *and over the random coins of* E *and* D*.*

3 Our Definition of α_{m_1,m_2}-correct Encryption Scheme

In this section, we define an α_{m_1,m_2}-correct encryption scheme and compare it to the Dwork *et al.*'s Definition (3). Then, we prove that an α_{m_1,m_2}-correct encryption scheme satisfies the definition of differential privacy with respect to the function $\mathcal{Q} \coloneqq D \circ E \simeq \mathbb{1}_{\mathcal{M}}$. We start by presenting and describing the main constructions and properties for the case of a 1-bit encryption scheme, as the simplest example possible, and after that we generalize the result to an N-element encryption scheme.

3.1 Definition

Our goal is to formally define the possibility that an encryption scheme can wrongly decrypt a message with some well defined probability.

Definition 4 (α_{m_1,m_2}-correctness encryption scheme). *Let* (G, E, D) *be an encryption scheme defined over* $(\mathcal{M}, \mathcal{K}, \mathcal{C})$ *as*

- ***Generation algorithm****: let* $\lambda \in \mathbb{N}$ *be a security parameter.* G *is defined as a probabilistic algorithm that given a security parameter* 1^λ*, returns a key-pair* $(\mathrm{sk}, \mathrm{pk}) \in \mathcal{K}$*.*
- ***Encryption algorithm****: let* $m \in \mathcal{M}$*,* $\mathrm{pk} \in \mathcal{K}_{\mathrm{pk}}$ *and* $c \in \mathcal{C}$*.* E *is defined as an algorithm that takes as input a public key* pk *and a message* m*, and returns a ciphertext* c*.*
- ***Decryption algorithm****: let* $m \in \mathcal{M}$*,* $\mathrm{sk} \in \mathcal{K}_{\mathrm{sk}}$ *and* $c \in \mathcal{C}$*.* D *is defined as an algorithm that given a secret key* sk *and a ciphertext* c*, returns a plaintext* m*.*

(G, E, D) *is said to be an* α_{m_1,m_2}-correct encryption scheme *if, for all* $m_1, m_2 \in \mathcal{M}$*, a fixed* $\lambda \in \mathbb{N}$ *and a fixed key-pair* $(\mathrm{sk}, \mathrm{pk}) \leftarrow\!\shortmid G(1^\lambda)$*, it holds*

$$\alpha_{m_1,m_2}((\mathrm{sk}, \mathrm{pk})) \coloneqq \mathbb{P}(D(\mathrm{sk}, E(\mathrm{pk}, m_1)) = m_2)$$

If for all $m \in \mathcal{M}$ *it holds* $\alpha_{m,m} = 1$*, then* (G, E, D) *is said to be a* correct *encryption scheme.*

In simple words, in an α_{m_1,m_2}-*correct encryption scheme*, the probability of encrypting m_1 and decrypting into m_2 using the key-pair $(\mathrm{sk}, \mathrm{pk})$ is equal to α_{m_1,m_2}.

Remark 2. From the definition above, it is easy to see that every encryption scheme is an α_{m_1,m_2} encryption scheme.

Remark 3. The $\alpha_{m_1,m_2}((\mathrm{sk}, \mathrm{pk}))$ values are strongly connected with the choice of $(\mathrm{sk}, \mathrm{pk})$. We will abuse notation and drop the key-pair since in our arguments, we will always fix some key-pair $(\mathrm{sk}, \mathrm{pk})$.

Remark 4. Our α_{m_1,m_2}-correctness (Def. 4) and Dwork *et al.*'s definition (Def. 3) describe the same encryption schemes.

Proof. – *Our definition* $\Rightarrow$ *Dwork* et al.*'s definition*:
Let (G, E, D) be any α_{m_1,m_2}-correct public-key encryption scheme. Let us consider

$$\alpha = \max_{m \in \mathcal{M}, (\mathrm{sk},\mathrm{pk}) \in \mathcal{K}} \alpha_{m,m}((\mathrm{sk}, \mathrm{pk}))$$

Let $(\mathrm{sk}, \mathrm{pk}) \in \mathcal{K}$ be any possible random key and $m \in \mathcal{M}$ any possible random message.

$$1 - \mathbb{P}(D_{\mathrm{sk}}(E_{\mathrm{pk}}(m)) \neq m) = \mathbb{P}(D_{\mathrm{sk}}(E_{\mathrm{pk}}(m)) = m) = \alpha_{m,m}((\mathrm{sk}, \mathrm{pk})) \leq \alpha$$

And so, we have that (G, E, D) is an α-correct encryption scheme in Dwork *et al.*'s Definition (3).

– *Our definition* $\Leftarrow$ *Dwork* et al.*'s definition*: Follows directly from Remark (2)

□

Dwork *et al.*'s definition describes a global upper bound on the correctness probability of an encryption scheme, while our definition defines the precise values of α_{m_1,m_2} of the encryption scheme.

3.2 Construction of an α_{m_1,m_2}-correct 1-bit Encryption Scheme

Fix $\mathcal{M} = \{0, 1\}$. Let (G, E, D) be an α_{m_1,m_2}-correct encryption scheme defined over $(\mathcal{M}, \mathcal{K}, \mathcal{C})$. Let us fix a key pair $(\mathrm{sk}, \mathrm{pk}) \leftarrow\!\!\shortmid G(1^\lambda)$ and let $\mathcal{Q}(m) = D(\mathrm{sk}, E(\mathrm{pk}, m))$. It holds:

$$\mathrm{Rg}(\mathcal{Q}) = \{0, 1\} \qquad D_0 = \{0\}, D_1 = \{1\}$$
$$S \in \mathcal{P}(\mathrm{Rg}(\mathcal{Q})) = \{\emptyset,\ \{0\} = S_0,\ \{1\} = S_1,\ \{0, 1\} = \mathcal{M}\}$$
$$\mathcal{Q}(m) = D(\mathrm{sk}, E(\mathrm{pk}, m)) \simeq m \qquad \forall_{m_1,m_2 \in \mathcal{M}}\, \mathbb{P}(\mathcal{Q}(m_1) = m_2) = \alpha_{m_1,m_2}$$

Proposition 1. *An* α_{m_1,m_2}*-correct* 1*-bit encryption scheme such that for all* $m_1, m_2 \in \mathcal{M}$ *it holds that* $\mathbb{P}(D(\mathrm{sk}, E(\mathrm{pk}, m_1)) = m_2) = \alpha_{m_1,m_2}$*, achieves* $\epsilon(\alpha_{m_1,m_2})$*-differential privacy where*

$$\epsilon(\alpha_{m_1,m_2}) := \inf \left\{ \epsilon : \begin{array}{c} e^\epsilon \geq \frac{\alpha_{0,0}}{\alpha_{1,0}}\ ,\ e^\epsilon \geq \frac{\alpha_{0,1}}{\alpha_{1,1}} \\ \\ e^\epsilon \geq \frac{\alpha_{1,0}}{\alpha_{0,0}}\ ,\ e^\epsilon \geq \frac{\alpha_{1,1}}{\alpha_{0,1}} \end{array} \right\}$$

Proof. Let us prove that any α_{m_1,m_2}-correct encryption scheme satisfies the ϵ-DP definition.

From the Definition (2), we can state that $\mathbb{P}(\mathcal{Q}(D_i) \in S_j)$ means that we encrypt the bit i and we decrypt it into the bit j. We can impose the DP definition in all possible cases in order to study the differential privacy coefficient ϵ:

- *If $S = \emptyset$, all the probabilities are 0, and so the ϵ-DP definition holds for every $\epsilon \in \mathbb{R}$ since $0 \leq 0$*
- *If $S = \{0,1\} = \mathcal{M}$, all the probabilities are 1, and so the ϵ-DP definition holds since $1 \leq e^\epsilon$ and $\epsilon \geq 0$*
- *If $S = \{0\} = S_0$:*
  - *$\mathbb{P}(\mathcal{Q}(D_0) \in S_0) \leq e^\epsilon\, \mathbb{P}(\mathcal{Q}(D_1) \in S_0)$ becomes $\alpha_{0,0} \leq e^\epsilon \alpha_{1,0} \Longrightarrow e^\epsilon \geq \frac{\alpha_{0,0}}{\alpha_{1,0}}$*
  - *$\mathbb{P}(\mathcal{Q}(D_1) \in S_0) \leq e^\epsilon\, \mathbb{P}(\mathcal{Q}(D_0) \in S_0)$ becomes $\alpha_{1,0} \leq e^\epsilon \alpha_{0,0} \Longrightarrow e^\epsilon \geq \frac{\alpha_{1,0}}{\alpha_{0,0}}$*
- *If $S = \{1\} = S_1$:*
  - *$\mathbb{P}(\mathcal{Q}(D_1) \in S_1) \leq e^\epsilon\, \mathbb{P}(\mathcal{Q}(D_0) \in S_1)$ becomes $\alpha_{1,1} \leq e^\epsilon \alpha_{0,1} \Longrightarrow e^\epsilon \geq \frac{\alpha_{1,1}}{\alpha_{0,1}}$*
  - *$\mathbb{P}(\mathcal{Q}(D_0) \in S_1) \leq e^\epsilon\, \mathbb{P}(\mathcal{Q}(D_1) \in S_1)$ becomes $\alpha_{0,1} \leq e^\epsilon \alpha_{1,1} \Longrightarrow e^\epsilon \geq \frac{\alpha_{0,1}}{\alpha_{1,1}}$*

We can conclude that for every $\alpha_{m_1,m_2} \in [0,1]$, we achieve ϵ-DP where ϵ has to be in the convex solution set $\mathcal{E}(\alpha_{m_1,m_2})$ defined as:

$$\text{for } \alpha_{m_1,m_2} \in [0,1] \qquad \mathcal{E}(\alpha_{m_1,m_2}) \coloneqq \left\{ \epsilon : \begin{array}{c} e^\epsilon \geq \frac{\alpha_{0,0}}{\alpha_{1,0}}\ e^\epsilon \geq \frac{\alpha_{0,1}}{\alpha_{1,1}} \\ \\ e^\epsilon \geq \frac{\alpha_{1,0}}{\alpha_{0,0}}\ e^\epsilon \geq \frac{\alpha_{1,1}}{\alpha_{0,1}} \end{array} \right\}$$

from which we can define the curve

$$\epsilon(\alpha_{m_1,m_2}) = \inf \mathcal{E}(\alpha_{m_1,m_2})$$

that defines the minimum ϵ such that the ϵ-DP definition holds for the encryption scheme. □

Proposition (1) is a special case of Proposition (2).

3.3 Construction of an α_{m_1,m_2}-correct N-Elements Encryption Scheme

Let $\#\mathcal{M} = N$ be the message space with uniform distribution of being transmitted, *i.e.,* for all $m \in \mathcal{M}$, $\mathbb{P}(M \in \{m\}) = \frac{1}{\#\mathcal{M}}$. Fix a key-pair $(\mathrm{sk}, \mathrm{pk})$ and then for all $m_1, m_2 \in \mathcal{M}$ it holds

$$\alpha_{m_1,m_2} = \mathbb{P}(D(\mathrm{sk}, E(\mathrm{pk}, m_1)) = m_2 \mid m_1)$$

Proposition 2. *An N-element α_{m_1,m_2}-correct encryption scheme such that for all $m_1, m_2 \in \mathcal{M}$ it holds that $\mathbb{P}(D(\mathrm{sk}, E(\mathrm{pk}, m_1)) = m_2) = \alpha_{m_1,m_2}$. Then, the encryption scheme achieves $\epsilon(\alpha_{m_1,m_2})$-differential privacy where*

$$\epsilon(\alpha_{m_1,m_2}) \coloneqq \inf \left\{ \epsilon \;\middle|\; \forall D_0, D_1 \in \mathcal{M},\ S \subseteq \mathcal{M}.\ \frac{\sum_{m_2 \in S} \alpha_{D_0,m_2}}{\sum_{m_2 \in S} \alpha_{D_1,m_2}} \leq e^\epsilon \right\}$$

Proof. *Let $\mathcal{Q} = D \circ E$ and $S \subseteq \mathcal{M}$ as before. Then, $\mathbb{P}(\mathcal{Q}(D_0) \in S) = \sum_{m_2 \in S} \alpha_{D_0,m_2}$.*

Imposing the DP definition, we have that for all $D_0, D_1 \in \mathcal{M}$ such that the two elements are different and for every $S \subseteq \mathcal{M}$ it holds:

$$\mathbb{P}(\mathcal{Q}(D_0) \in S) \leq e^{\epsilon}\, \mathbb{P}(\mathcal{Q}(D_1) \in S) \implies \sum_{m_2 \in S} \alpha_{D_0,m_2} \leq e^{\epsilon} \left(\sum_{m_2 \in S} \alpha_{D_1,m_2} \right)$$

We can manipulate the equation and obtain $\frac{\sum_{m_2 \in S} \alpha_{D_0,m_2}}{\sum_{m_2 \in S} \alpha_{D_1,m_2}} \leq e^{\epsilon}$

We define the convex set

$$\mathcal{E}(\alpha_{m_1,m_2}) \coloneqq \left\{ \epsilon \;\middle|\; \forall D_0, D_1 \in \mathcal{M}, S \subseteq \mathcal{M}.\ \frac{\sum_{m_2 \in S} \alpha_{D_0,m_2}}{\sum_{m_2 \in S} \alpha_{D_1,m_2}} \leq e^{\epsilon} \right\}$$

The value $\epsilon(\alpha_{m_1,m_2}) = \inf \mathcal{E}(\alpha_{m_1,m_2})$ will satisfy the DP-definition. □

3.4 Fix ϵ, find α_{m_1,m_2}

The parameters ϵ and α_{m_1,m_2} are dependent one from the other since for all $D_0, D_1 \in \mathcal{M}$ and for all $S \subseteq \mathcal{M}$, it holds

$$\frac{\sum_{m_2 \in S} \alpha_{D_0,m_2}}{\sum_{m_2 \in S} \alpha_{D_1,m_2}} \leq e^{\epsilon} \tag{1}$$

The goal of finding the best α_{m_1,m_2} that achieves a fixed ϵ-DP depends on practical requirements and conditions that we want to impose on the probabilities α_{m_1,m_2}, *i.e., "maximizing the difference between two different messages"* or *"having a specific probability distribution"*.

For completeness, we will provide a simple solution in a particular case.

Proposition 3. *Let α_{m_1,m_2} be the probabilities of an N-element encryption scheme, where for all $m \in \mathcal{M}$, it holds $\alpha_{m,m} = \alpha$ and for all $m' \in \mathcal{M}$ with $m' \neq m$, it holds $\alpha_{m,m'} = \beta < \alpha$. If $\alpha \geq (N-1)\beta$, then the scheme achieves $\log\left(\frac{\alpha}{\beta}\right)$-DP.*

Proof. *In order to prove the thesis, we have to find the D_0, D_1, S that maximize the left side of Equation (1). We can consider the polynomials $f_\alpha(x) = \alpha + x\beta$ and $f_\beta(x) = \beta + x\beta$. From the hypothesis, we have that $f_\alpha(x) \geq f_\beta(x)$ for all $x \in \mathbb{R}$ and $x \geq 0$. In particular, this is true for the integer values between 0 and $N-1$. Since $\frac{f_\alpha(x)}{f_\beta(x)}$ is a decreasing function for all $x \in \mathbb{R}$ and $x \geq 0$, we can conclude that for $i \in [0, N-1]$ integers, it holds:*

$$\begin{aligned} \frac{\alpha}{\beta} = \frac{f_\alpha(0)}{f_\beta(0)} \geq \frac{f_\alpha(i)}{f_\beta(i)} \geq \frac{f_\alpha(i+1)}{f_\beta(i+1)} \geq \cdots \geq \frac{f_\alpha(N-1)}{f_\beta(N-1)} \\ \frac{\beta}{\alpha} = \frac{f_\beta(0)}{f_\alpha(0)} \leq \frac{f_\beta(i)}{f_\alpha(i)} \leq \frac{f_\beta(i+1)}{f_\alpha(i+1)} \leq \cdots \leq \frac{f_\beta(N-1)}{f_\alpha(N-1)} = \frac{(N-1)\beta}{(N-2)\beta + \alpha} \end{aligned} \tag{2}$$

From Equation (2) and since $\frac{\alpha}{\beta} \geq \frac{\beta}{\alpha}$ *from the hypothesis, we have*

$$\frac{(N-1)\beta}{(N-2)\beta+\alpha} \leq \frac{\alpha}{(N-2)\beta+\alpha} \leq \frac{\alpha}{\beta}$$

and, in Equation (1)

$$\frac{\sum_{m_2 \in S} \alpha_{D_0,m_2}}{\sum_{m_2 \in S} \alpha_{D_1,m_2}} \leq \frac{\alpha}{\beta} \leq e^{\epsilon} \tag{3}$$

We can so conclude that the minimal ϵ *for which the equation holds is* $\log\left(\frac{\alpha}{\beta}\right)$ *and so the* N*-element encryption scheme will achieve* $\log\left(\frac{\alpha}{\beta}\right)$*-DP.*

□

4 Equality Between DP-then-Encrypt and Encrypt+DP

In this section, we define the two main methods of combining an encryption scheme with a differential private mechanism: *(i)* the DP-then-Encrypt and *(ii)* the Encrypt+DP . We then prove a proposition on the equivalence between the DP-then-Encrypt and the Encrypt+DP classes. After this, we prove that combining a differential privacy framework with a correct encryption scheme is at least as computationally secure as the relying encryption scheme.

Definition 5. *Define the* DP-then-Encrypt *class as the set of all the encryption schemes* (G', E', D') *such that*

$$G'(1^\lambda) \coloneqq G(1^\lambda) \qquad E'(\mathrm{pk}, m) \coloneqq E(\mathrm{pk}, \mathcal{Q}(m)) \qquad D'(\mathrm{sk}, c) \coloneqq D(\mathrm{sk}, c)$$

for some (G, E, D) *correct encryption scheme on* $(\mathcal{M}, \mathcal{K}, \mathcal{C})$ *and* $\mathcal{Q} \simeq \mathbb{1}_{\mathcal{M}}$ *a DP-mechanism.*

It is trivial that $D'(\mathrm{sk}, E'(\mathrm{pk}, m)) = \mathcal{Q}(m)$.

Definition 6. *Define the* Encrypt+DP *class as the set of all the* α_{m_1,m_2}*-correct encryption schemes* $(\hat{G}, \hat{E}, \hat{D})$ *on* $(\mathcal{M}, \mathcal{K}, \mathcal{C})$*. From the Proposition (2), we have that* $(\hat{G}, \hat{E}, \hat{D})$ *is* $\epsilon(\alpha_{m_1,m_2})$*-DP and it holds* $\hat{D}(\mathrm{sk}, \hat{E}(\mathrm{pk}, m)) \simeq \mathbb{1}_{\mathcal{M}}(m)$.

In a nutshell, the DP-then-Encrypt class contains all the different combinations of the identity map as a DP-mechanism and a correct encryption scheme. On the other hand, the Encrypt-then-DP achieves the identity map as a DP-mechanism directly in the α_{m_1,m_2}-correct encryption scheme used.

In order to prove the equality between the two classes, we define a *probability "permutation"* as:

Definition 7. *Let* $m_1, m_2 \in \mathcal{M}$*. Let us denote a* probability "permutation" π *as the random variable on* $\mathcal{M}$ *with measure probability of the event* "permute the message m_1 into the message m_2" *defined as* $\mathbb{P}(\pi(m_1) = m_2) = \alpha_{m_1,m_2}$.

Remark 5. Let π be a probability permutation. Then, π is a DP-mechanism. This means it is a $\epsilon(\alpha_{m_1,m_2})$-DP mechanism (or it achieves ∞-DP).

Proposition 4. *The* DP-then-Encrypt *class is equivalent to the* Encrypt+DP *class.*

Proof. – DP-then-Encrypt $\subseteq$ Encrypt+DP

Let (G', E', D') be a DP-then-Encrypt *encryption scheme. Let us fix a key pair $(\mathrm{sk}, \mathrm{pk}) \leftarrow\!\!\shortmid G'(1^\lambda)$. Trivially using Remark (2), there exists an $\alpha_{m_1,m_2} \in [0,1]$ such that for all $m_1, m_2 \in \mathcal{M}$ it holds:*

$$\mathbb{P}((D'(\mathrm{sk}, E'(\mathrm{pk}, m_1)) = m_2) = \mathbb{P}(\mathcal{Q}(m_1) = m_2) = \alpha_{m_1,m_2}$$

From the Definition (4), (G', E', D') is an α_{m_1,m_2}-correct encryption scheme and so from Proposition (2), we have that (G', E', D') is contained in the class Encrypt+DP *of Definition (6).*

– DP-then-Encrypt $\supseteq$ Encrypt+DP

Let $(\hat{G}, \hat{E}, \hat{D})$ be an α_{m_1,m_2}-correct encryption scheme such that $\hat{D}(\mathrm{sk}, \hat{E}(\mathrm{pk}, m)) \simeq \mathbb{1}_{\mathcal{M}}(m)$. For every $m_1, m_2 \in \mathcal{M}$, we define the random variable $\pi : \mathcal{M} \to \mathcal{M}$ as

$$\mathbb{P}(\pi(m_1) = m_2) \coloneqq \mathbb{P}(\hat{D}(\mathrm{sk}, \hat{E}(\mathrm{pk}, m_1)) = m_2) = \alpha_{m_1,m_2}$$

π is a probability permutation as in Definition (7) and for Remark (5), we have that π is a DP-mechanism.

Let us define $(\hat{G}, E, D)$ a correct encryption scheme such that:

- *$\hat{G}$ is the same key generator as the α_{m_1,m_2}-correct encryption scheme*
- *$E : \mathcal{K} \times \mathcal{M} \to \mathcal{C}$ is an encryption algorithm*
- *$D : \mathcal{K} \times \mathcal{C} \to \mathcal{M}$ is a decryption algorithm*

and for all $(\mathrm{sk}, \mathrm{pk}) \leftarrow\!\!\shortmid \hat{G}(1^\lambda)$, it holds that for all $m \in \mathcal{M}$

$$\mathbb{P}(D(\mathrm{sk}, E(\mathrm{pk}, m)) = m) = 1$$

We can claim that E, D always exist and we can consider any injective function $\phi : \mathcal{M} \to \mathcal{C}$ with left inverse ϕ^{-1}. Let us define:

$$E(\mathrm{pk}, m) \coloneqq \phi(m) \qquad D(\mathrm{sk}, c) \coloneqq \phi^{-1}(c)$$

For $(\hat{G}, E, D)$, we have

$$\mathbb{P}(D(\mathrm{sk}, E(\mathrm{pk}, m)) = m) = \mathbb{P}(\phi^{-1}(\phi(m)) = m) = \mathbb{P}(m = m) = 1$$

In order to conclude, we need to prove that $(\hat{G}, E, D)$ with π as in Definition (5), acts like an encryption scheme (G', E', D') that is contained in the Encrypt+DP *class of Definition (6). Fix a key pair $(\mathrm{sk}, \mathrm{pk}) \leftarrow\!\!\shortmid \hat{G}(1^\lambda)$:*

$$\begin{aligned}
\mathbb{P}(\hat{D}(\mathrm{sk}, \hat{E}(\mathrm{pk}, m_1)) = m_2) &= \alpha_{m_1,m_2} \\
&= \mathbb{P}(\pi(m_1) = m_2) \\
&= \mathbb{P}(\phi^{-1}(\phi(\pi(m_1))) = m_2) \\
&= \mathbb{P}(D(\mathrm{sk}, E(\mathrm{pk}, \pi(m_1))) = m_2) \\
&= \mathbb{P}(D'(\mathrm{sk}, E'(\mathrm{pk}, m_1)) = m_2)
\end{aligned}$$

□

We will now define a concept of *security-hardness* with respect to an adversary without specifying the computational model used.

Definition 8. *The adversary $\mathcal{A}$ for an encryption scheme (G, E, D) is an algorithm that takes the public key*[1] *and a ciphertext and it outputs a guess m' for the message m.*

$$\mathcal{A} : \mathcal{K}_{\mathrm{pk}} \times \mathcal{C} \to \mathcal{M} \qquad \mathcal{A}(\mathrm{pk}, E(\mathrm{pk}, m)) \mapsto m'$$

An encryption scheme (G, E, D) is said to be security-hard *with respect to the adversary $\mathcal{A}$ (in some computational model) if*

$$\mathbb{P}(\mathcal{A}(\mathrm{pk}, E(\mathrm{pk}, m)) = m) \leq \frac{1}{\#\mathcal{M}} + \mathrm{negl}$$

Informally, we defined the simplest adversary possible whose goal is to guess the correct decryption of a ciphertext given all the public information possible. In order to obtain a general result, we do not impose any complexity-hardness assumption. The security-hardness adversary is a weaker adversary with respect to the ones from $\mathrm{IND-CPA}, \mathrm{IND-CCA}$ (and so on). On the other hand, for an encryption scheme, being security-hard is a necessary condition in order to achieve any security requirement: the security-hardness adversary can be used as a distinguisher in a more structured security model.

Lemma 1. *Let (G, E, D) be a correct encryption scheme which is security-hard. Let $\mathcal{Q} \simeq \mathbb{1}_{\mathcal{M}}$ DP-mechanism. Then the combination of $\mathcal{Q}$ with (G, E, D), which is in the* DP-then-Encrypt *class, is security-hard. In other word, the security-hardness of the combination $\mathcal{Q}$ with (G, E, D) is at least computationally hard as the security-hardness of (G, E, D).*

Proof. We have to show and prove:

1. Reduce every instance of a (G, E, D) correct encryption scheme to an instance in the DP-then-Encrypt class.
2. We prove the lemma by contradiction and *Reductio ad absurdum*: If there exists an adversary $\mathcal{A}$ with non-negligible advantage for the DP-then-Encrypt instance, there will exist an adversary $\mathcal{B}$ with non-negligible advantage for the (G, E, D) correct encryption scheme. Let us suppose that there exists $\mathcal{A}$ with non-negligible advantage, and let us suppose that all $\mathcal{B}$ have negligible advantage. Then we prove that it is a contradiction, and so we conclude.

The reduction is trivial: we can just consider as the instance in the DP-then-Encrypt class, (G, E, D) encryption scheme with the deterministic identity map as the DP-mechanism.

For a fixed key $(\mathrm{sk}, \mathrm{pk}) \leftarrow\!\shortmid G(1^\lambda)$, suppose there exists an adversary $\mathcal{A}$ for the DP-then-Encrypt scheme, it means $\mathcal{A}(m) \coloneqq \mathcal{A}(\mathrm{pk}, E(\mathrm{pk}, \mathcal{Q}(m)))$ will output

[1] It is possible to give a pure symmetric key encryption scheme definition but we do not need it.

the guess m' and the guess will be correct with probability $\frac{1}{\#\mathcal{M}} + \delta$ with $\delta > 0$ non-negligible. Formally $\mathbb{P}(\mathcal{A}(\mathrm{pk}, E(\mathrm{pk}, \mathcal{Q}(m))) = m) = \frac{1}{\#\mathcal{M}} + \delta$

Let us suppose that for all the adversaries $\mathcal{B}$ of the original scheme such that $\mathcal{B}(m) \coloneqq \mathcal{B}(\mathrm{pk}, E(\mathrm{pk}, m))$, we have $\mathbb{P}(\mathcal{B}(\mathrm{pk}, E(\mathrm{pk}, m)) = m) = \frac{1}{\#\mathcal{M}} + \epsilon$ where $\epsilon > 0$ is negligible.

From the probability independence between the DP-mechanism $\mathcal{Q}$ and the encryption scheme (G, E, D) we have

$$\begin{aligned}\frac{1}{\#\mathcal{M}} + \delta = \mathbb{P}(\mathcal{A}(m) = m) &= \mathbb{P}(\mathcal{B}(m) = m \mid \mathcal{Q}(m) = m) \\ &= \mathbb{P}(\mathcal{B}(m) = m)\,\mathbb{P}(\mathcal{Q}(m) = m) \\ &\leq \mathbb{P}(\mathcal{B}(m) = m) = \frac{1}{\#\mathcal{M}} + \epsilon\end{aligned}$$

Absurd. So there exists an adversary $\mathcal{B}$ with non-negligible advantage.[2]

□

5 Example of an α_{m_1,m_2}-Correct Homomorphic Encryption Scheme

In this section, we introduce a variation of the Dijk's *et al.* public key integer homomorphic encryption scheme [2] by only introducing a new parameter ξ that will be used to increase the noisy randomness of the encryption scheme. Then, we show how to compute the probabilities α_{m_1,m_2} that will prove that the scheme is α_{m_1,m_2}-correct. At the end, we show the connection between the original and the modified scheme and prove the security-hardness of the modified one.

Definition 9 (Variation of the Dijk *et al.* public key homomorphic encryption scheme). *Let $\mathcal{M} = \{0, 1\}$ and let γ, η, ρ, τ be the four parameters defined in the original scheme such that all the security constraints hold. Let ξ be an additional parameter required for the variation.*
Let (G, E, D) be defined as:

- $G(1^\lambda)$ *: randomly pick $p \in [2^{\eta-1}, 2^\eta)$ and p odd.*
  For the public key, for all $i \in 0..\tau$ sample

$$x_i \in \mathcal{D}_{\gamma,\rho}(p) = \left\{ pq + r : q \in U\left(\mathbb{Z} \cap \left[0, \frac{2^\gamma}{p}\right)\right), r \in U(\mathbb{Z} \cap (-2^\rho, 2^\rho)) \right\}$$

  and relabel so that x_0 is the greatest. Restart until x_0 is odd and $(x_0 \pmod p) \in \left(-\frac{p}{2}, \frac{p}{2}\right]$ is even.
  Define $\mathrm{pk} \coloneqq \{x_0, \ldots, x_\tau\}$ *as the public key and* $\mathrm{sk} \coloneqq p$ *as the secret key.*
- $E(\mathrm{pk}, m)$*: choose at random $S \subseteq [1, \tau]$ and a random integer $r \in (-2^{\rho'+\xi}, 2^{\rho'+\xi})$.*
  The difference with respect to the original scheme is that ξ is present in the interval-bounds exponents. Output the ciphertext $c = \left(m + 2r + 2\sum_{i \in S} x_i\right) \pmod{x_0}$

[2] Take for example adversary $\mathcal{A}$.

– $D(p, c)$: *output* $(c \pmod p) \pmod 2$

In order to prove that the scheme achieves some α-correctness with $\alpha \neq 1$, fix a random S and observe that

$$\begin{aligned} m + 2r + 2\sum_{i\in S} x_i &= m + 2r + 2\sum_{i\in S} pq_i + r_i \\ &= m + 2\left(r + \sum_{i\in S} r_i\right) + p \cdot 2\sum_{i\in S} q_i = m + 2R + pQ \end{aligned}$$

where $Q \in \mathbb{Z}$ and R will be contained in a subset of the integers

$$A_S \coloneqq \left(-(\#S \cdot 2^{\rho} + 2^{\rho'+\xi}), (\#S \cdot 2^{\rho} + 2^{\rho'+\xi})\right) \subseteq \mathbb{Z}$$

For this reason, for a fixed S, we can reduce the computation of $\alpha_{m,m}$ as a combinatorial problem:

$$\alpha \coloneqq \frac{\#\left\{r : r \in \left(-2^{\rho'+\xi}, 2^{\rho'+\xi}\right) \mid \left|2\left(r + \sum_{i\in S} r_i\right)\right| < \frac{p}{2}\right\}}{\#S \cdot 2^{\rho+1} + 2^{\rho'+\xi+1}}$$

For the right parameter ξ, we can obtain that the encryption scheme is an $\alpha_{m,m}$-correct encryption scheme.

Remark 6. It is important to notice that using a different S will change the probability $\alpha_{m,m}$. You can think of it as *using a different public key* for the encryption algorithm.

Consider a fixed S and the function $\lfloor x \rceil$ = closest integer to x. We can compute $\Delta = 2 \cdot \sum_{i\in S} r_i$ and if we consider $\tilde{\xi}$ as the bound for the noise r, we can define the function

$$F(\tilde{\xi}, \Delta) = \frac{\displaystyle\int_{-\tilde{\xi}+\Delta}^{\tilde{\xi}+\Delta} \left\lfloor \frac{x}{p} \right\rceil \pmod 2 \, dx}{2 \cdot \tilde{\xi}} \in [0, 1]$$

that represents the correctness probability. We have the trivial properties

$$F(\tilde{\xi}, 0) = \frac{1}{2} \qquad \lim_{\tilde{\xi}\to\infty} F(\tilde{\xi}, \Delta) = \frac{1}{2} \tag{4}$$

In order to prove that our modified scheme is secure, we reduce the security-hardness of our scheme to the security of the original Dijk *et al.*'s encryption scheme. From the Proposition (4) on the class equality between Encrypt+DP and DP-then-Encrypt we will transform our modified scheme into the Dijk *et al.*'s encryption scheme in the DP-then-Encrypt class.

Remark 7. We can observe that r is randomly picked from $\left(-2^{\rho'+\xi}, 2^{\rho'+\xi}\right)$. We will now consider a random $r' \in (-2^{\rho'}, 2^{\rho'})$ and rewrite $r = r' + \hat{r}$ for some $\hat{r} \in \mathbb{Z}$. At this point, we can rewrite the general encrypted message as

$$m + 2r + 2\sum_{i \in S} x_i = m + 2(r' + \hat{r}) + 2\sum_{i \in S} x_i = (m + 2\hat{r}) + 2r' + 2\sum_{i \in S} x_i \quad (5)$$

where r' and x_i are regular values from the original encryption scheme. During the decryption phase, we will obtain:

$$\left(m + 2r + 2\sum_{i \in S} x_i\right) \pmod p \pmod 2 =$$

$$\text{Equation } (5) = \left((m + 2\hat{r}) + \left(2r' + 2\sum_{i \in S} x_i\right)\right) \pmod p \pmod 2$$

$$\begin{aligned} \text{Original scheme's values} &= (m + 2\hat{r}) \pmod p \pmod 2 \\ &= m \oplus (2\hat{r} \pmod p \pmod 2) \end{aligned}$$

From this equality, the message m can be decrypted in a different message $\hat{m}$ just by looking at the value $\hat{r}$.

This is exactly a DP-then-Encrypt scheme, where we can define a probability permutation π as in Definition (7) with $\mathbb{P}(\pi(m_1) = m_2) = \alpha_{m_1,m_2}$ and the original Dijk's encryption scheme.

Remark 8. As in the Remark (6), changing S will change the probability permutation π since the probability α will change. For this reason, the random subset S, the probability permutation π, the probability α and the new parameter ξ are dependent one from the others.

Proposition 5. *Given an $\alpha_{m,m}$-correct public key modified Dijk* et al. *'s encryption scheme with fixed parameters $(\rho, \rho', \eta, \gamma, \tau, \xi)$.*
Any adversary $\mathcal{A}$ with non-negligible advantage ϵ on the $\alpha_{m,m}$-correct encryption scheme can be converted into an adversary $\mathcal{B}$ with non-negligible advantage ϵ on the original Dijk et al.*'s encryption scheme with parameter $(\rho, \rho', \eta, \gamma, \tau)$.*

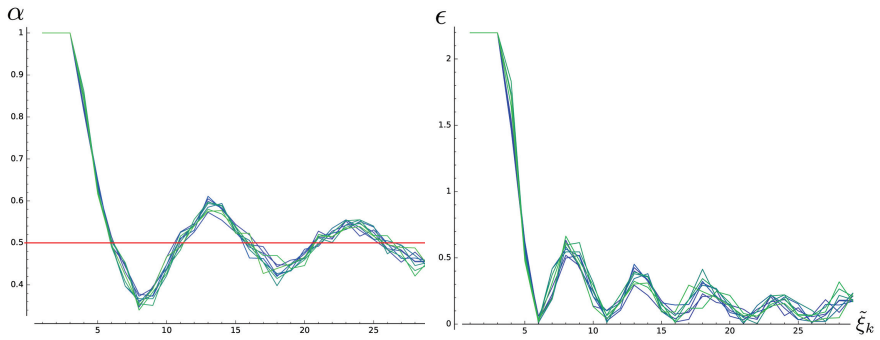
Proof. Follows from Lemma (1). □

5.1 Implementation and Statistics

In order to empirically study the dependency between the parameters ξ, α and ϵ, we implemented the modified Dijk *et al.* 's encryption scheme of Section 5 in Sage. Considering $\lambda = 10$ as a general security parameter, we started from the scheme with parameters:

$$\rho = \lambda \qquad \rho' = 2 \cdot \lambda \qquad \eta = \lambda^2 \qquad \gamma = \lambda^5 \qquad \tau = \lambda \qquad \xi = 0$$

and then we consider the k-th variation where we add a factor of $\tilde{\xi}_k = \frac{k \cdot p}{10}$ to the noise interval $2^{\rho'} + \tilde{\xi}_k$. In Figure 3a and Figure 3b, we have the measured value for α and ϵ with respect to k. For every $k \in [1, 30]$, we tested λ different choice of S, we executed $N = 100$ experiments and retrieved an empirical value for α. In order to obtain the ϵ, we just took the $\epsilon = \sup\left\{\frac{\alpha}{1-\alpha}, \frac{1-\alpha}{\alpha}\right\}$. We tested different random keys S and the empirical difference between the plots is barely visible, but it can easily be described as a *"really small translation of the plot to the left or right"*. In the chosen key used for the test, if we want to have a $\alpha = 0.8$ correctness probability, we have to use $\tilde{\xi}_4 = \frac{2 \cdot p}{5}$ and the scheme will have $\epsilon = 1.38$ -DP.



(a) Different α with respect to $\tilde{\xi}_k$ (b) Different ϵ with respect to $\tilde{\xi}_k$

Figure 3: Empirical measurements of α and ϵ from the implementation.

6 Conclusions & Future Work

This paper bridges concepts in cryptography and differential privacy and we propose the first differentially private encryption scheme. More precisely, we show how to constructively combine differential privacy with an encryption framework in a single scheme, contained in the Encrypt+DP class, and vice versa. This construction is not limited to homomorphic encryption schemes and can be used in order to define an encryption scheme that can guarantee both privacy and confidentiality.

So far we have only examined this link in an abstract way. An open question is the trade-off between α_{m_1,m_2}-correctness and $\epsilon(\alpha_{m_1,m_2})$-DP for specific homomorphic operations, with a particular attention to the *bootstrap* procedure. This might lead to interesting practical applications, such as faster, α-correct homomorphic encryption schemes with differential privacy guarantees.

Acknowledgment. This paper was partially funded by the VR project "PRECIS: Privacy and Security in Wearable Computing devices" and the STINT project "Secure, Private and Efficient Healthcare with Wearable Computing".

References

1. Beimel, A., Nissim, K., Omri, E.: Distributed Private Data Analysis: On Simultaneously Solving How and What. arXiv:1103.2626 [cs] (Mar 2011), http://arxiv.org/abs/1103.2626, arXiv: 1103.2626
2. Dijk, M.v., Gentry, C., Halevi, S., Vaikuntanathan, V.: Fully Homomorphic Encryption over the Integers. In: Advances in Cryptology – EUROCRYPT 2010. pp. 24–43. Springer, Berlin, Heidelberg (May 2010), http://link.springer.com/chapter/10.1007/978-3-642-13190-5_2, dOI: 10.1007/978-3-642-13190-5_2
3. Dwork, C.: Differential Privacy, vol. 4052 (Jul 2006), https://www.microsoft.com/en-us/research/publication/differential-privacy/
4. Dwork, C., Naor, M., Reingold, O.: Immunizing Encryption Schemes from Decryption Errors. In: Advances in Cryptology - EUROCRYPT 2004. pp. 342–360. Springer, Berlin, Heidelberg (May 2004), http://link.springer.com/chapter/10.1007/978-3-540-24676-3_21, dOI: 10.1007/978-3-540-24676-3_21
5. El Emam, K., Dankar, F.K.: Protecting Privacy Using k-Anonymity. J Am Med Inform Assoc 15(5), 627–637 (2008), http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2528029/
6. Erkin, Z., Troncoso-Pastoriza, J.R., Lagendijk, R., Pérez-González, F.: Privacy-Preserving Data Aggregation in Smart Metering Systems: An Overview. IEEE Signal Processing Magazine 30(2), 75–86 (2013)
7. Garcia, F.D., Jacobs, B.: Privacy-Friendly Energy-Metering via Homomorphic Encryption. In: Security and Trust Management. pp. 226–238. Springer, Berlin, Heidelberg (Sep 2010), http://link.springer.com/chapter/10.1007/978-3-642-22444-7_15, dOI: 10.1007/978-3-642-22444-7_15
8. Gehrke, J., Kifer, D., Machanavajjhala, A.: l-Diversity. In: Tilborg, H.C.A.v., Jajodia, S. (eds.) Encyclopedia of Cryptography and Security, pp. 707–709. Springer US (2011), http://link.springer.com/referenceworkentry/10.1007/978-1-4419-5906-5_899, dOI: 10.1007/978-1-4419-5906-5_899
9. Gentry, C.: A fully homomorphic encryption scheme. Ph.D. thesis, Stanford University (2009)
10. Ji, Z., Lipton, Z.C., Elkan, C.: Differential Privacy and Machine Learning: a Survey and Review. arXiv:1412.7584 [cs] (Dec 2014), http://arxiv.org/abs/1412.7584, arXiv: 1412.7584
11. Li, N., Li, T., Venkatasubramanian, S.: t-Closeness: Privacy Beyond k-Anonymity and l-Diversity. In: 2007 IEEE 23rd International Conference on Data Engineering. pp. 106–115 (Apr 2007)
12. Meissen, R.: A Mathematical Approach to Fully Homomorphic Encryption. Ph.D. thesis, Worcester Polytechnic Institute (2012)