Ciphertext Expansion in Limited-Leakage Order-Preserving Encryption: A Tight Computational Lower Bound

Gil Segev[*] Ido Shahaf[*]

Abstract

Order-preserving encryption emerged as a key ingredient underlying the security of practical database management systems. Boldyreva et al. (EUROCRYPT '09) initiated the study of its security by introducing two natural notions of security. They proved that their first notion, a "best-possible" relaxation of semantic security allowing ciphertexts to reveal the ordering of their corresponding plaintexts, is not realizable. Later on Boldyreva et al. (CRYPTO '11) proved that any scheme satisfying their second notion, indistinguishability from a random order-preserving function, leaks about half of the bits of a random plaintext.

This unsettling state of affairs was recently changed by Chenette et al. (FSE '16), who rigorously relaxed the above "best-possible" notion and constructed a scheme satisfying it based on any pseudorandom function. In addition to revealing the ordering of any two encrypted plaintexts, ciphertexts in their scheme reveal only the position of the most significant bit on which the plaintexts differ. A significant drawback of their scheme, however, is its substantial ciphertext expansion: Encrypting plaintexts of length m bits results in ciphertexts of length $m \cdot \ell$ bits, where ℓ determines the level of security (e.g., $\ell = 80$ in practice).

In this work we prove a lower bound on the ciphertext expansion of any order-preserving encryption scheme satisfying the "limited-leakage" notion of Chenette et al. with respect to non-uniform polynomial-time adversaries, matching the ciphertext expansion of their scheme up to lower-order terms. This improves a recent result of Cash and Zhang (TCC '18), who proved such a lower bound for schemes satisfying this notion with respect to *computationally-unbounded* adversaries (capturing, for example, schemes whose security can be proved in the random-oracle model without relying on cryptographic assumptions). Our lower bound applies, in particular, to schemes whose security is proved in the standard model.

[*] School of Computer Science and Engineering, Hebrew University of Jerusalem, Jerusalem 91904, Israel. Email: {segev,ido.shahaf}@cs.huji.ac.il. Supported by the European Union's Horizon 2020 Framework Program (H2020) via an ERC Grant (Grant No. 714253), by the Israel Science Foundation (Grant No. 483/13), by the Israeli Centers of Research Excellence (I-CORE) Program (Center No. 4/11), and by the US-Israel Binational Science Foundation (Grant No. 2014632).

1 Introduction

An order-preserving encryption (OPE) scheme is a private-key encryption scheme whose ciphertexts preserve the numerical ordering of their corresponding plaintexts. Such schemes were introduced in the database community by Agrawal et al. [AKS$^+$04] for enabling efficient indexing of encrypted data and efficient range queries over encrypted databases. By now, order-preserving encryption has become a key cryptographic ingredient underlying the security of database management systems (see [FVY$^+$17] for a long list of OPE-based commercial systems).

The security of OPE. Given that the ciphertexts of any order-preserving encryption scheme reveal the numerical ordering of their corresponding plaintexts, such schemes clearly cannot satisfy the standard notion of semantic security. This motivated Boldyreva, Chenette, Lee and O'Neill [BCL$^+$09, BCL$^+$12] to initiate a foundational study of the security of order-preserving encryption. They introduced two notions of security for such schemes. Their first notion is a "best-possible" relaxation of the standard semantic security notion, allowing ciphertexts to reveal only the numerical ordering of their corresponding plaintexts. Informally, their notion asks that the encryptions of any two sequences of plaintexts should be indistinguishable as long as the two sequences share the same order pattern. Unfortunately, Boldyreva et al. then proved that such a notion cannot be satisfied.

Their second notion asks that an order-preserving encryption scheme should be indistinguishable from a random order-preserving function (similarly to the standard notion of pseudorandomness for pseudorandom functions). Boldyreva et al. provided an efficient scheme that satisfies this notion, but it was later on demonstrated by Boldyreva, Chenette and O'Neill [BCO11, BCO12] that a random order-preserving function may in fact reveal substantial information on its input (specifically, about half of the bits of a random message) – and thus this notion may not be sufficiently strong for most applications.

Limited-leakage OPE. The absence of a strong (and realizable) notion of security has somewhat questioned our confidence in the potential security guarantees of order-preserving encryption. This state of affairs, however, has recently changed due to the work of Chenette, Lewi, Weis and Wu [CLW$^+$16]. They rigorously relaxed the "best-possible" notion introduced by Boldyreva et al. [BCL$^+$09, BCL$^+$12] to allow a *limited amount of well-defined "leakage"*, and constructed a practical scheme that satisfies it, based on pseudorandom functions. Concretely, in addition to revealing the relative ordering of any two encrypted plaintexts, ciphertexts in their scheme reveal the position of the most significant bit on which they differ – but no additional information is revealed. We refer to this specific leakage as "CLWW-leakage", and to schemes that satisfy their notion as $\mathcal{L}_{\mathsf{CLWW}}$-secure schemes.

Drawback: Ciphertext expansion. Incorporating the limited-leakage scheme of Chenette et al. in practical OPE-based systems finally enables to rigorously reason about their security. However, a significant drawback of their scheme is its ciphertext expansion. Roughly speaking, encrypting plaintexts of length m bits using their scheme results in ciphertexts of length $m \cdot \ell$ bits, where ℓ determines the level of security (i.e., "ℓ bits of security" – we discuss the relation between the ciphertext expansion and the security of their scheme in more detail in Section 1.1).

In fact, Chenette et al. first constructed an *order-revealing* encryption scheme [BLR$^+$15] with ciphertexts of length only $\lceil \log_2 3 \cdot m \rceil$ bits, and then showed that the main ideas underlying their scheme can be used to construct an order-preserving encryption scheme – but with significantly longer ciphertexts (see Section 1.2 for more details on the less-strict notion of order-revealing encryption). Given the practical importance of order-preserving encryption, this poses the question of whether or not such a significant expansion is inherent.

Initial evidence indicating that such an expansion is inherent was recently provided by Cash and Zhang [CZ18]. They introduced an *information-theoretic* variant of the limited-leakage notion of security considered by Chenette et al. (that is, a notion of security with respect to computationally-unbounded adversaries and CLWW-leakage), and showed that any scheme satisfying it must suffer from a significant ciphertext expansion, matching the ciphertext expansion in the scheme of Chenette et al. up to lower-order terms.

As discussed by Cash and Zhang, although no scheme can satisfy their information-theoretic notion in the standard model, they nevertheless capture schemes whose security can be proved in the random-oracle model without relying on any cryptographic assumption. They do not capture, however, schemes whose security is proved in the standard model based on cryptographic assumptions (such as the existence of pseudorandom functions, and specific number-theoretic or combinatorial assumptions).

1.1 Our Contributions

In this paper we prove a tight lower bound on the ciphertext expansion of any order-preserving encryption scheme that satisfies the "limited-leakage" notion of security considered by Chenette et al. [CLW$^+$16]. In its weakest form, this notion asks that the encryptions of any two sequences of plaintexts should be indistinguishable as long as the two sequences share the same CLWW-leakage, as discussed above (see Section 2 for the formal definition). We prove the following theorem:

Theorem (informal). *Let Π be an order-preserving encryption scheme with m-bit plaintexts and n-bit ciphertexts. Then, there exists a non-uniform polynomial-time adversary $\mathcal{A}$ that breaks the $\mathcal{L}_{\mathsf{CLWW}}$-security of the scheme with probability at least $2^{-n/m} \cdot m^{-1}$.*

Under the minimal requirement that the success probability of any efficient adversary in breaking the $\mathcal{L}_{\mathsf{CLWW}}$-security of the scheme should be negligible, our theorem implies that ciphertexts must be of length at least $n = m \cdot \omega(\log \lambda)$ bits, where $\lambda \in \mathbb{N}$ is the security parameter. Practically, when aiming at (say) 80 bits of security (and focusing, for simplicity, on the significant $2^{-n/m}$ term), this implies that ciphertexts must be of length at least roughly $n = 80m$ bits.

Comparison to the Cash-Zhang lower bound. When compared to the lower bound proved by Cash and Zhang [CZ18], our lower bound and their lower bound are identical in terms of the attacker's success probability (and, thus, in terms of the implications on the ciphertext expansion). As discussed above, however, their lower bound applies to an information-theoretic variant of the notion of security to which our lower bound applies. Concretely, Cash and Zhang prove their lower bound by analyzing the statistical distance between ciphertext distributions (which translates into a *computationally-unbounded* adversary), whereas we prove our lower bound by presenting a non-uniform *polynomial-time* adversary[1]. Thus, our lower bound applies to any $\mathcal{L}_{\mathsf{CLWW}}$-secure order-preserving encryption scheme, and most notably to such schemes whose security is proved in the standard model.

[1] Although utilizing non-uniformity in cryptographic constructions, reductions, and impossibility proofs is rather standard [Gol04, GW11, GK16] (e.g., given the practical benefits of preprocessing-based attacks [Hel80, FN99, BL13, CDG$^+$18, CK18]), an interesting open question is whether or not the above theorem can even be proved via a *uniform* polynomial-time adversary.

The tightness of our lower bound. Looking into the security of the scheme provided by Chenette et al. [CLW^{+}16] (when adapted to offer perfect correctness as suggested by Cash and Zhang), we observe that our lower bound is in fact tight up to low-order terms. Specifically, their scheme is based on the existence of any pseudorandom function F mapping inputs of length at most $m = m(\lambda)$ bits to outputs of length $\ell = \ell(\lambda)$ bits, and encrypting plaintexts of length m bits using their scheme results in ciphertexts of length $n = m \cdot \ell$ bits. An analysis of the security of their construction shows that the advantage $\mathsf{Adv}^{\mathsf{OPE}}_{\Pi,\mathcal{L}_{\mathsf{CLWW}},\mathcal{A}}$ of any adversary $\mathcal{A}$ in breaking the $\mathcal{L}_{\mathsf{CLWW}}$-security of their scheme can be upper bounded as

$$\mathsf{Adv}^{\mathsf{OPE}}_{\Pi,\mathcal{L}_{\mathsf{CLWW}},\mathcal{A}} \leq \mathsf{Adv}^{\mathsf{PRF}}_{\mathsf{F},\mathcal{B}} + \frac{m \cdot q}{2^{\ell}},$$

where $q = q(\lambda)$ denotes the number of encryption queries made by $\mathcal{A}$, $\mathsf{Adv}^{\mathsf{PRF}}_{\mathsf{F},\mathcal{B}}$ denotes the advantage of an algorithm $\mathcal{B}$ (efficiently derived from $\mathcal{A}$) in breaking the pseudorandomness of F, and recall that $\ell = n/m$. The above theorem provides a lower bound on the advantage of our specific adversary (which issues only $q = 2$ encryption queries), and this yields

$$\frac{1}{2^{n/m} \cdot m} \leq \mathsf{Adv}^{\mathsf{OPE}}_{\Pi,\mathcal{L}_{\mathsf{CLWW}},\mathcal{A}}(\lambda) \leq \mathsf{Adv}^{\mathsf{PRF}}_{\mathsf{F},\mathcal{B}}(\lambda) + \frac{m \cdot 2}{2^{n/m}}.$$

Up to the lower-order terms in the above expression[2], our lower bound and the security of the scheme constructed by Chenette et al. match.

1.2 Related Work

Boneh et al. [BLR^{+}15] introduced the notion of an order-revealing encryption (ORE) scheme, which is a less-strict variant of order-preserving encryption scheme. Such schemes allow to compare plaintexts by invoking a publicly-computable comparison algorithm on their ciphertexts (no secret key is required), and can be viewed as a specific form of multi-input functional encryption [GGG^{+}14, AJ15, BKS16, KS17]. The notion of order-preserving encryption is then obtained by requiring, in addition, that the comparison algorithm is simply a numerical comparison. Based on assumptions involving multi-linear maps, Boneh et al. presented a (rather theoretical) construction of an ORE scheme that satisfies the aforementioned "best-possible" security notion of Boldyreva et al. [BCL^{+}09]. This stands in contrast to the impossibility of Boldyreva et al. for constructing an order-preserving encryption scheme satisfying the same "best-possible" security notion.

As for ORE schemes that satisfy weaker notions of security, as mentioned above Chenette et al. [CLW^{+}16] constructed an efficient $\mathcal{L}_{\mathsf{CLWW}}$-secure ORE scheme that has ciphertexts of length only $\lceil \log_2 3 \cdot m \rceil$, where m is the length of their corresponding plaintexts, and their construction is based on pseudorandom functions.

Finally, when dealing with encryption schemes that inherently leak non-trivial information, one should always pay attention to potential attacks that may be enabled by such leakage. Indeed, such attacks on order-revealing encryption are known in some specific settings (e.g., [DDC16, GSB^{+}17]), but this does not rule out their deployment in other settings.

1.3 Overview of Our Approach

In this section we provide a brief overview of the main ideas underlying the proof of our lower bound. In what follows, let $\Pi = (\mathsf{KeyGen}, \mathsf{Enc})$ be an order-preserving encryption scheme with plaintexts

[2] Assuming, in addition, that the security of the pseudorandom function is not the main bottleneck (for example, by choosing a sufficiently large security parameter, or by using AES in practice).

of length m bits and ciphertexts of length n bits (both m and n may be functions of the security parameter $\lambda \in \mathbb{N}$ – see Section 2 for the formal definition of such a scheme). For any plaintext $i \in \{0,1\}^m$, viewed an integer $0 \leq i \leq 2^m - 1$, we denote by $X_i = \mathsf{Enc}_K(i)$ the random variable corresponding to an encryption of i with respect to a randomly-generated key $K \leftarrow \mathsf{KeyGen}(1^\lambda)$. Each such random variable X_i is distributed over $\{0,1\}^n$, and is viewed as an integer $0 \leq X_i \leq 2^n - 1$. In addition, we let $\epsilon = 2^{-n/m} \cdot m^{-1}$ (note that this is the success probability stated by our theorem), and let $\Delta(X, Y)$ denote the statistical distance between the distributions X and Y.

The proof of Cash and Zhang. Cash and Zhang [CZ18] observed that for every $1 \leq j \leq m-1$ it holds that $\mathcal{L}_{\mathsf{CLWW}}(0, 2^{j+1} - 1) = \mathcal{L}_{\mathsf{CLWW}}(2^j - 1, 2^j)$, where $\mathcal{L}_{\mathsf{CLWW}}$ is the CLWW-leakage as discussed above[3]. Assuming towards a contradiction that a scheme Π is $\mathcal{L}_{\mathsf{CLWW}}$-secure in the *statistical* sense that no computationally-unbounded adversary has advantage larger than ϵ, then the distributions $(\mathsf{Enc}_K(0), \mathsf{Enc}_K(2^{j+1}-1))$ and $(\mathsf{Enc}_K(2^j-1), \mathsf{Enc}_K(2^j))$ must be statistically close, as both $(0, 2^{j+1} - 1)$ and $(2^j - 1, 2^j)$ have the same CLWW-leakage. That is, it must hold that

$$\Delta((X_0, X_{2^{j+1}-1}), (X_{2^j-1}, X_{2^j})) \leq \epsilon.$$

Therefore, denoting $G_1^j = X_{2^{j+1}-1} - X_0$ and $G_2^j = X_{2^j} - X_{2^j-1}$, and noting that applying the same function to two distributions cannot increase their statistical distance, it also holds that $\Delta(G_1^j, G_2^j) \leq \epsilon$. By the order-preserving property of the scheme, it holds that $G_1^j \geq 0$, $G_2^j \geq 0$, and that

$$G_1^j = X_{2^{j+1}-1} - X_0 \geq (X_{2^j} - X_{2^j-1}) + (X_{2^j-1} - X_0) = G_2^j + G_1^{j-1}.$$

This shows that G_1^j is ϵ-statistically-close to G_2^j, and that G_1^j is larger than G_2^j by at least G_1^{j-1}. Equipped with this observation, Cash and Zhang inductively proved that the support of G_1^{j-1} must contain "large" values, and that the support of G_1^j must contain even larger values. As a final step, note that $X_{2^m-1} = X_0 + G_1^{m-1}$ and also $\Delta(X_0, X_{2^m-1}) \leq \epsilon$ as it trivially holds that $\mathcal{L}_{\mathsf{CLWW}}(0) = \mathcal{L}_{\mathsf{CLWW}}(2^m - 1)$. Using their reasoning once again, they deduced that the support of X_{2^m-1} must contain values larger than $2^n - 1$, which contradicts the definition of X_{2^m-1} as an integer in the range $\{0, \ldots, 2^n - 1\}$.

Our approach: A non-uniform polynomial-time adversary. When considering schemes that are $\mathcal{L}_{\mathsf{CLWW}}$-secure in the standard *computational* sense, we cannot take advantage of the fact that $\Delta(G_1^j, G_2^j) \leq \epsilon$ and apply the reasoning of Cash and Zhang. Instead, we show that if the *consequence* of the reasoning of Cash and Zhang does not hold (specifically, if the support of G_1^j does not contain large values), then there exists a polynomial-time test that distinguishes between G_1^j and G_2^j: Given a sample y from either G_1^j or G_2^j, our distinguisher checks whether $y \leq t$ for some fixed threshold value $0 \leq t \leq 2^n - 1$.

Then, assuming that the consequence of the reasoning of Cash and Zhang does hold for every step $1 \leq j \leq m-1$, we can then prove via an additional step that either there is a threshold test for distinguishing between X_0 and X_{2^m-1}, or it holds that support of X_{2^m-1} contains values larger than $2^n - 1$. Since the second case contradicts the definition of X_{2^m-1} as an integer in the range $\{0, \ldots, 2^n - 1\}$, it must be that the first case holds.

[3] Specifically, for any distinct two plaintexts m_i and m_j it holds that $\mathcal{L}_{\mathsf{CLWW}}(m_i, m_j) = (\mathsf{ind}_{\mathsf{diff}}(m_i, m_j), \mathbf{1}(m_i < m_j))$, where $\mathsf{ind}_{\mathsf{diff}}(m_i, m_j) \in \{1, \ldots, m, \perp\}$ is the index of the most significant bit on which m_i and m_j differ, and $\mathbf{1}(m_i < m_j) \in \{0,1\}$ indicates whether or not $m_i < m_j$. For the full definition of $\mathcal{L}_{\mathsf{CLWW}}$, which takes as input an arbitrary number of messages, see Section 2.

As a result, either there exist $1 \leq j \leq m-1$ and $0 \leq t \leq 2^n - 1$ such that given ciphertexts $(c_1, c_2) \in \{(X_0, X_{2^{j+1}-1}), (X_{2^j-1}, X_{2^j})\}$, the test $c_1 - c_2 \leq t$ distinguishes between the two cases, or there exists $0 \leq t \leq 2^n - 1$ such that given a ciphertext $c \in \{X_0, X_{2^m-1}\}$, the test $c \leq t$ distinguishes between the two cases. This translates into a non-uniform polynomial-time adversary that breaks the $\mathcal{L}_{\mathsf{CLWW}}$-security of any given scheme with probability at least ϵ, where the non-uniform advice specifies which test out of the m possible tests to perform, as well as which threshold value $0 \leq t \leq 2^n - 1$ to use. We refer the reader to Section 3 for our proof.

1.4 Paper Organization

The remainder of this paper is organized as follows. In Section 2 we present the notation and definitions that are used in this work. Then, in Section 3 we prove our lower bound for order-preserving encryption schemes. Finally, in Section 4 we prove two general statements (regarding certain joint distributions of random variables) on which our proof relies.

2 Preliminaries

In this section we present the notation and definitions that are used in this work. We denote by $\lambda \in \mathbb{N}$ the security parameter. For a distribution X we denote by $x \leftarrow X$ the process of sampling a value x from the distribution X. Similarly, for a set $\mathcal{X}$ we denote by $x \leftarrow \mathcal{X}$ the process of sampling a value x from the uniform distribution over $\mathcal{X}$. A function $\mathsf{negl} : \mathbb{N} \to \mathbb{R}_{\geq 0}$ is *negligible* if for every constant $c > 0$ there exists an integer N_c such that $\mathsf{negl}(n) < n^{-c}$ for all $n > N_c$. All logarithms in this paper are to the base of 2. The *statistical distance* between two random variables X and Y over a finite domain Ω is $\Delta(X, Y) = \frac{1}{2} \sum_{\omega \in \Omega} |\Pr[X = \omega] - \Pr[Y = \omega]|$.

Order-preserving encryption [AKS+04, BCL+09]. An order-preserving encryption scheme Π is a pair $(\mathsf{KeyGen}, \mathsf{Enc})$ of probabilistic polynomial-time algorithms satisfying the following requirements for parameters $m = m(\lambda)$ and $n = n(\lambda)$:

- The key-generation algorithm KeyGen takes as input the security parameter $\lambda \in \mathbb{N}$ in unary representation and outputs a secret key K.
- The encryption algorithm Enc takes as input a secret key K and a plaintext $x \in \{0,1\}^m$ interpreted as a numerical value $0 \leq x \leq 2^m - 1$, and outputs ciphertext $c \in \{0,1\}^n$ interpreted as a numerical value $0 \leq c \leq 2^n - 1$.

Note that a decryption algorithm is not required by this definition. We say that Π is *correct* if for all $\lambda \in \mathbb{N}$ and $0 \leq i < j \leq 2^{m(\lambda)} - 1$ it holds that $\Pr[\mathsf{Enc}_K(i) < \mathsf{Enc}_K(j)] = 1$, where $K \leftarrow \mathsf{KeyGen}(1^\lambda)$.

Remark. It is also possible to consider a relaxed game-based correctness notion, where a probabilistic polynomial-time adversary (without explicit access to the secret key) should not be able to come up with plaintexts $0 \leq i < j \leq 2^{m(\lambda)} - 1$ such that $\mathsf{Enc}_K(i) \geq \mathsf{Enc}_K(j)$, expect with a negligible probability. In Section 3, we discuss the effect of such a relaxation on our lower bound.

Security. We prove our lower bound for any scheme that satisfies the following non-adaptive indistinguishability-based security notion. This notion is (tightly) implied by its (stronger) adaptive and/or simulation-based variants, and thus our lower bound applies to those as well.

More concretely, given a scheme $\Pi = (\mathsf{KeyGen}, \mathsf{Enc})$, a leakage function $\mathcal{L}$, an algorithm $\mathcal{A}$, a bit $b \in \{0,1\}$, and a security parameter λ, we consider the following experiment.

The experiment $\mathsf{Ind}^{\mathsf{OPE}}_{\Pi,\mathcal{L},\mathcal{A},b}(\lambda)$

1. $K \leftarrow \mathsf{KeyGen}(1^\lambda)$.
2. $(\mathbf{m}_0, \mathbf{m}_1, \mathsf{state}) \leftarrow \mathcal{A}(1^\lambda)$, where $\mathbf{m}_0$ and $\mathbf{m}_1$ are plaintext vectors of the same length (which we denote by q).
3. $\mathbf{c} = (c_1, \ldots, c_q)$, where $c_i \leftarrow \mathsf{Enc}_K(\mathbf{m}_b[i])$ for every $1 \leq i \leq q$.
4. $b' \leftarrow \mathcal{A}(\mathsf{state}, \mathbf{c})$.
5. If $\mathcal{L}(\mathbf{m}_0) = \mathcal{L}(\mathbf{m}_1)$ then the experiment outputs b', and otherwise it outputs 0.

The advantage of $\mathcal{A}$ is defined as

$$\mathsf{Adv}^{\mathsf{OPE}}_{\Pi,\mathcal{L},\mathcal{A}}(\lambda) = \left| \Pr[\mathsf{Ind}^{\mathsf{OPE}}_{\Pi,\mathcal{L},\mathcal{A},1}(\lambda) = 1] - \Pr[\mathsf{Ind}^{\mathsf{OPE}}_{\Pi,\mathcal{L},\mathcal{A},0}(\lambda) = 1] \right| .$$

As discussed above, in this paper we consider security with respect to non-uniform polynomial-time adversaries, captured by the following definition:

Definition 2.1. An order-preserving encryption scheme Π is $\mathcal{L}$*-secure* if for every non-uniform polynomial-time algorithm $\mathcal{A}$ it holds that $\mathsf{Adv}^{\mathsf{OPE}}_{\Pi,\mathcal{L},\mathcal{A}}(\lambda)$ is negligible.

In this work we consider the leakage function introduced by Chenette et al. [CLW+16]:

$$\mathcal{L}_{\mathsf{CLWW}}(x_1, \ldots, x_q) = \{(i, j, \mathsf{ind}_{\mathsf{diff}}(x_i, x_j), \mathbf{1}(x_i < x_j)) : 1 \leq i < j \leq q\},$$

where $\mathsf{ind}_{\mathsf{diff}}(x_i, x_j) \in \{1, \ldots, m, \perp\}$ is the index of the most significant bit on which x_i and x_j differ (and is set to $\perp$ if $x_i = x_j$), and $\mathbf{1}(x_i < x_j) \in \{0, 1\}$ indicates whether or not $x_i < x_j$.

3 Our Lower Bound

In this section we prove the following theorem, and then show that it can be extended to schemes without perfect correctness.

Theorem 3.1. *Let Π be an order-preserving encryption scheme with plaintext length $m = m(\lambda)$ bits and ciphertext length $n = n(\lambda)$ bits, where $\lambda \in \mathbb{N}$ is the security parameter. Then, there exists a non-uniform polynomial-time adversary $\mathcal{A}$ such that $\mathsf{Adv}^{\mathsf{OPE}}_{\Pi,\mathcal{L}_{\mathsf{CLWW}},\mathcal{A}}(\lambda) \geq 2^{-n/m} \cdot m^{-1}$ for all $\lambda \in \mathbb{N}$.*

Proof. For any $1 \leq j(\lambda) \leq m(\lambda) - 1$ and $0 \leq t(\lambda) \leq 2^{n(\lambda)} - 1$, we define an adversary $\mathcal{A}_{j,t}$ that participates in the experiment $\mathsf{Ind}^{\mathsf{OPE}}_{\Pi,\mathcal{L},\mathcal{A},b}(\lambda)$ (see Section 2) as follows:

The adversary $\mathcal{A}_{j,t}$:

- Given a security parameter 1^λ as input, $\mathcal{A}_{j,t}$ outputs $(\mathbf{m}_0, \mathbf{m}_1, \mathsf{state}) = ((0, 2^{j+1} - 1), (2^j - 1, 2^j), \perp)$.
- Given a state $\mathsf{state} = \perp$ and ciphertexts $\mathbf{c} = (c_1, c_2)$ as input, $\mathcal{A}_{j,t}$ outputs 1 if $c_2 - c_1 \leq t$ and 0 otherwise.

Additionally, we define an adversary $\mathcal{B}_t$ as follows:

The adversary $\mathcal{B}_t$:

- Given a security parameter 1^λ as input, $\mathcal{B}_t$ outputs $(\mathbf{m}_0, \mathbf{m}_1, \mathsf{state}) = ((2^m - 1), (0), \perp)$.
- Given a state $\mathsf{state} = \perp$ and a single ciphertext $\mathbf{c} = (c_1)$ as input, $\mathcal{B}_t$ outputs 1 if $c_1 \leq t$ and 0 otherwise.

It is easy to verify that both $\mathcal{A}_{j,t}$ and $\mathcal{B}_t$ output plaintext vectors with the same CLWW-leakage, and thus these are valid adversaries.

From this point on we fix a security parameter $\lambda \in \mathbb{N}$ and omit it for ease of notation. Denoting $\epsilon = 2^{-n/m} \cdot m^{-1}$, we show that either there exist $1 \leq j \leq m-1$ and $0 \leq t \leq 2^n - 1$ such that $\mathsf{Adv}^{\mathsf{OPE}}_{\Pi, \mathcal{L}_{\mathsf{CLWW}}, \mathcal{A}_{j,t}} \geq \epsilon$ or there exists $0 \leq t \leq 2^n - 1$ such that $\mathsf{Adv}^{\mathsf{OPE}}_{\Pi, \mathcal{L}_{\mathsf{CLWW}}, \mathcal{B}_t} \geq \epsilon$. This guarantees that the following non-uniform polynomial-time adversary $\mathcal{A}$ satisfies $\mathsf{Adv}^{\mathsf{OPE}}_{\Pi, \mathcal{L}_{\mathsf{CLWW}}, \mathcal{A}} \geq \epsilon$ as claimed: Given a non-uniform advise $j \in \{1, \ldots, m-1, \perp\}$ and $0 \leq t \leq 2^n - 1$, if $j \neq \perp$ then $\mathcal{A}$ invokes $\mathcal{A}_{j,m}$, and if $j = \perp$ it invokes $\mathcal{B}_t$.

For any $0 \leq i \leq 2^m - 1$ let $X_i = \mathsf{Enc}_K(i)$ where $K \leftarrow \mathsf{KeyGen}(1^\lambda)$. Then, by the definition of the above adversaries it holds that

$$\mathsf{Adv}^{\mathsf{OPE}}_{\Pi, \mathcal{L}_{\mathsf{CLWW}}, \mathcal{A}_{j,t}} = \left| \Pr[X_{2^j} - X_{2^j - 1} \leq t] - \Pr[X_{2^{j+1}-1} - X_0 \leq t] \right|$$

and

$$\mathsf{Adv}^{\mathsf{OPE}}_{\Pi, \mathcal{L}_{\mathsf{CLWW}}, \mathcal{B}_t} = \left| \Pr[X_0 \leq t] - \Pr[X_{2^m - 1} \leq t] \right| .$$

For a parameter $1 \leq j \leq m-1$, consider the following property:

$\mathsf{Property}(j)$: For each $t \in \mathbb{N}$ it holds that $\Pr[X_{2^{j+1}-1} \leq X_0 + t] \leq (t \cdot j!)^{1/j} \cdot \epsilon$

We proceed to consider three cases, according to what values $1 \leq j \leq m-1$ (if any) satisfy $\mathsf{Property}(j)$.

Case I: $\mathsf{Property}(1)$ does not hold. In this case we rely on the following lemma, which we prove in Section 4.

Lemma 3.2. *Let (X, Y) be jointly distributed random variables, taking values in $\mathbb{Z}_{\geq 0}$, such that $X > Y$, and let $\epsilon \geq 0$. Then, at least one of the following must hold:*

1. *For every $t \in \mathbb{Z}_{\geq 0}$ it holds that $\Pr[X \leq t] \leq t \cdot \epsilon$.*
2. *There exists $t \in \mathbb{Z}_{\geq 0}$ such that $\Pr[Y \leq t] - \Pr[X \leq t] \geq \epsilon$.*

Observing that $X_3 - X_0 > X_2 - X_1$ and applying Lemma 3.2 for $X = X_3 - X_0$ and $Y = X_2 - X_1$, since $\mathsf{Property}(1)$ does not hold the second case of the lemma must hold. That is, there exists $t \in \mathbb{Z}_{\geq 0}$ such that $\Pr[X_2 - X_1 \leq t] - \Pr[X_3 - X_0 \leq t] \geq \epsilon$, and so $\mathcal{A}_{1,t}$ is an adversary with an advantage of at least ϵ.

Case II: $\mathsf{Property}(m-1)$ holds. In this case we rely on the following lemma, which we prove in Section 4.

Lemma 3.3. *Let (X,Y) be jointly distributed random variables, taking values in $\mathbb{Z}_{\geq 0}$, such that $X \geq Y$. Suppose there exist $i \in \mathbb{N}$ and $\epsilon \geq 0$ such that for every $k \in \mathbb{Z}_{\geq 0}$ it holds that $\Pr[X \leq Y+k] \leq (k \cdot i!)^{1/i} \cdot \epsilon$. Then, at least one of the following must hold:*

1. *For every $t \in \mathbb{Z}_{\geq 0}$ it holds that $\Pr[X \leq t] \leq (t \cdot (i+1)!)^{1/(i+1)} \cdot \epsilon$.*
2. *There exists $t \in \mathbb{Z}_{\geq 0}$ such that $\Pr[Y \leq t] - \Pr[X \leq t] \geq \epsilon$.*

Applying Lemma 3.3 for $X = X_{2^m-1}$ and $Y = X_0$, the conditions hold since $X_{2^m-1} \geq X_0$ and $\mathsf{Property}(m-1)$ holds, and we obtain that either there exists $t \in \mathbb{Z}_{\geq 0}$ such that $\Pr[X_0 \leq t] - \Pr[X_{2^m-1} \leq t] \geq \epsilon$, or for every $t \in \mathbb{Z}_{\geq 0}$ it holds that $\Pr[X_{2^m-1} \leq t] \leq (t \cdot m!)^{1/m} \cdot \epsilon$. In the first case we get that $\mathcal{B}_t$ is an adversary with an advantage of at least ϵ. In the second case, for $t = 2^n - 1$ we get that $1 = \Pr[X_{2^m-1} \leq 2^n - 1] < (2^n \cdot m!)^{1/m} \cdot \epsilon$. But then, using the bound $m! \leq m^m$ which holds for every positive m, we obtain that

$$\begin{aligned} \epsilon &> 2^{-n/m} \cdot (m!)^{-1/m} \\ &\geq 2^{-n/m} \cdot m^{-1}, \end{aligned}$$

which contradicts our definition of ϵ.

Case III: $\mathsf{Property}(1)$ holds but $\mathsf{Property}(m-1)$ does not hold. In this case let $2 \leq j \leq m-1$ be the smallest j for which $\mathsf{Property}(j)$ does not hold. Observing that $X_{2^{j+1}-1} - X_0 \geq (X_{2^j} - X_{2^j-1}) + (X_{2^j-1} - X_0)$ and applying Lemma 3.3 for $X = X_{2^{j+1}-1} - X_0$ and $Y = X_{2^j} - X_{2^j-1}$, the conditions hold since $\mathsf{Property}(j-1)$ holds, and we obtain that since $\mathsf{Property}(j)$ does not hold then the second case of the lemma must hold. That is, there exists $t \in \mathbb{Z}_{\geq 0}$ such that $\Pr[Y \leq t] - \Pr[X \leq t] \geq \epsilon$, so $\mathcal{A}_{j,t}$ is an adversary with advantage of at least ϵ. ■

Extending the proof to schemes without prefect correctness. We note that our lower bound is only based on the correctness of the scheme with respect to a polynomial number of pairs of plaintexts, that is, all pairs of plaintexts from the set $\{2^j : 1 \leq j \leq m-1(\lambda)\} \cup \{2^j - 1 : 0 \leq j \leq m(\lambda)\}$. Therefore, even for a scheme that satisfies a relaxed game-based correctness notion, it must hold that the scheme is correct for all those pairs of plaintexts with probability $1 - \mathsf{negl}(\lambda)$, where negl is a fixed negligible function. Hence, similarly to Theorem 3.1, there must exist a non-uniform polynomial-time adversary $\mathcal{A}$ such that $\mathsf{Adv}^{\mathsf{OPE}}_{\Pi,\mathcal{L}_{\mathsf{CLWW}},\mathcal{A}}(\lambda) \geq 2^{-n/m} \cdot m^{-1} - \mathsf{negl}(\lambda)$ for all $\lambda \in \mathbb{N}$.

4 Proofs of Lemma 3.2 and Lemma 3.3

We restate and prove Lemma 3.2.

Lemma 3.2. *Let (X,Y) be jointly distributed random variables, taking values in $\mathbb{Z}_{\geq 0}$, such that $X > Y$, and let $\epsilon \geq 0$. Then, at least one of the following must hold:*

1. *For every $t \in \mathbb{Z}_{\geq 0}$ it holds that $\Pr[X \leq t] \leq t \cdot \epsilon$.*
2. *There exists $t \in \mathbb{Z}_{\geq 0}$ such that $\Pr[Y \leq t] - \Pr[X \leq t] \geq \epsilon$.*

Proof. Assume that there exists $t \in \mathbb{N}$ such that $\Pr[X \leq t] > t \cdot \epsilon$ (the case $t = 0$ is impossible since then $Y < 0$), and let t_0 be the first such t. Then, it holds that $\Pr[X \leq t_0 - 1] \leq (t_0 - 1) \cdot \epsilon$, but $\Pr[Y \leq t_0 - 1] \geq \Pr[X \leq t_0] > t_0 \cdot \epsilon$, so it holds that $\Pr[Y \leq t_0 - 1] - \Pr[X \leq t_0 - 1] \geq \epsilon$. ■

Next, we restate and prove Lemma 3.3.

Lemma 3.3. *Let (X, Y) be jointly distributed random variables, taking values in $\mathbb{Z}_{\geq 0}$, such that $X \geq Y$. Suppose there exist $i \in \mathbb{N}$ and $\epsilon \geq 0$ such that for every $k \in \mathbb{Z}_{\geq 0}$ it holds that $\Pr[X \leq Y + k] \leq (k \cdot i!)^{1/i} \cdot \epsilon$. Then, at least one of the following must hold:*

1. *For every $t \in \mathbb{Z}_{\geq 0}$ it holds that $\Pr[X \leq t] \leq (t \cdot (i+1)!)^{1/(i+1)} \cdot \epsilon$.*
2. *There exists $t \in \mathbb{Z}_{\geq 0}$ such that $\Pr[Y \leq t] - \Pr[X \leq t] \geq \epsilon$.*

Proof. We make use of the following lemma.

Lemma 4.1. *Let (X, Y) be jointly distributed random variables, taking values in $\mathbb{Z}_{\geq 0}$, such that $X \geq Y$, and let $\epsilon \geq 0$. Then, at least one of the following must hold:*

1. *For every $t \in \mathbb{Z}_{\geq 0}$ and (possibly non-integer) $s > 0$ it holds that*
$$\Pr[X \leq t] \leq \frac{t}{s} \cdot \epsilon + \frac{1}{s} \int_0^s \Pr[X \leq Y + k] dk.$$
2. *There exists $t \in \mathbb{Z}_{\geq 0}$ such that $\Pr[Y \leq t] - \Pr[X \leq t] \geq \epsilon$.*

Assume for now the correctness of Lemma 4.1. We obtain that either there exists $t \in \mathbb{Z}_{\geq 0}$ such that $\Pr[Y \leq t] - \Pr[X \leq t] \geq \epsilon$, or that for every $t \in \mathbb{Z}_{\geq 0}$ it holds that

$$\begin{aligned}\Pr[X \leq t] &\leq \frac{t}{s} \cdot \epsilon + \frac{1}{s} \int_0^s \Pr[X \leq Y + k] dk \\ &\leq \frac{t}{s} \cdot \epsilon + \frac{1}{s} \int_0^s (k \cdot i!)^{1/i} \cdot \epsilon dk \\ &= \left(\frac{t}{s} + \frac{i}{i+1}(s \cdot i!)^{1/i}\right) \cdot \epsilon,\end{aligned}$$

and by choosing $s = (i+1)/(i+1)!^{1/(i+1)} \cdot t^{i/(i+1)}$ (which minimizes the above term), we obtain that $\Pr[X \leq t] \leq (t \cdot (i+1)!)^{1/(i+1)} \cdot \epsilon$ as claimed. ■

We now prove Lemma 4.1.

Proof of Lemma 4.1. First, for $t = 0$ it always holds that

$$\begin{aligned}\Pr[X \leq 0] &\leq \Pr[X \leq Y] \\ &= \frac{1}{s} \int_0^s \Pr[X \leq Y] dk \\ &\leq \frac{1}{s} \int_0^s \Pr[X \leq Y + k] dk.\end{aligned}$$

Now, for every $t \in \mathbb{N}$ we show that either it holds that

$$\Pr[X \leq t] \leq \frac{t}{s} \cdot \epsilon + \frac{1}{s} \int_0^s \Pr[X \leq Y + k] dk,$$

or there exists $0 \le k < t$ such that $\Pr[Y \le k] - \Pr[X \le k] \ge \epsilon$. We define the random variables (W, Z) as follows

$$(W, Z) = \begin{cases} (X, Y) & X \le t \\ (0, 0) & X > t. \end{cases}$$

We bound $\mathbb{E}(W - Z)$ both from above and below. For the lower bound, it holds that

$$\begin{aligned} \mathbb{E}(W - Z) &= \sum_{k=0}^{t} k \cdot \Pr[X = Y + k, X \le t] \\ &\ge s \cdot \sum_{k=0}^{t} \Pr[X = Y + k, X \le t] - \sum_{k=0}^{\lfloor s \rfloor} (s - k) \cdot \Pr[X = Y + k, X \le t] \\ &\ge s \cdot \Pr[X \le t] - \sum_{k=0}^{\lfloor s \rfloor} (s - k) \cdot \Pr[X = Y + k] \\ &= s \cdot \Pr[X \le t] - \int_0^s \Pr[X \le Y + \lfloor k \rfloor] dk \\ &= s \cdot \Pr[X \le t] - \int_0^s \Pr[X \le Y + k] dk. \end{aligned}$$

For the upper bound, we make use of the following lemma (a similar lemma appears in [CZ18]).

Lemma 4.2. *Let (W, Z) be jointly distributed random variables, taking values in $\{0, \ldots, t\}$. Then, there exists $0 \le k < t$ such that $\Pr[Z \le k] - \Pr[W \le k] \ge \mathbb{E}(W - Z)/t$.*

Assume for now the correctness of Lemma 4.2. We obtain that there exists $0 \le k < t$ such that $\Pr[Z \le k] - \Pr[W \le k] \ge \mathbb{E}(W - Z)/t$. Note that

$$\Pr[Z \le k] - \Pr[W \le k] = \Pr[Y \le k, X \le t] - \Pr[X \le k] \le \Pr[Y \le k] - \Pr[X \le k].$$

If $\Pr[Y \le k] - \Pr[X \le k] \ge \epsilon$ then we are done. Otherwise, it holds that

$$t \cdot \epsilon > s \cdot \Pr[X \le t] - \int_0^s \Pr[X \le Y + k] dk,$$

and the lemma follows. ■

We finish by proving Lemma 4.2.

Proof of Lemma 4.2. It holds that

$$\begin{aligned} \mathbb{E}(W - Z) &= \mathbb{E}W - \mathbb{E}Z \\ &= \sum_{k=1}^{t} \Pr[W \ge k] - \sum_{k=1}^{t} \Pr[Z \ge k] \\ &= \sum_{k=1}^{t} (1 - \Pr[W < k]) - \sum_{k=1}^{t} (1 - \Pr[Z < k]) \\ &= \sum_{k=0}^{t-1} \left(\Pr[Z \le k] - \Pr[W \le k] \right). \end{aligned}$$

Hence, there exists $0 \le k < t$ such that $\Pr[Z \le k] - \Pr[W \le k] \ge \mathbb{E}(W - Z)/t$ as claimed. ■

Acknowledgments

We thank Gili Schul-Ganz and the anonymous referees for various useful comments.

References

[AJ15] P. Ananth and A. Jain. Indistinguishability obfuscation from compact functional encryption. In *Advances in Cryptology – CRYPTO '15*, pages 308–326, 2015.

[AKS$^+$04] R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu. Order-preserving encryption for numeric data. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pages 563–574, 2004.

[BCL$^+$09] A. Boldyreva, N. Chenette, Y. Lee, and A. O'Neill. Order-preserving symmetric encryption. In *Advances in Cryptology – EUROCRYPT '09*, pages 224–241, 2009.

[BCL$^+$12] A. Boldyreva, N. Chenette, Y. Lee, and A. O'Neill. Order-preserving symmetric encryption. Cryptology ePrint Archive, Report 2012/624, 2012.

[BCO11] A. Boldyreva, N. Chenette, and A. O'Neill. Order-preserving encryption revisited: Improved security analysis and alternative solutions. In *Advances in Cryptology – CRYPTO '11*, pages 578–595, 2011.

[BCO12] A. Boldyreva, N. Chenette, and A. O'Neill. Order-preserving encryption revisited: Improved security analysis and alternative solutions. Cryptology ePrint Archive, Report 2012/625, 2012.

[BKS16] Z. Brakerski, I. Komargodski, and G. Segev. Multi-input functional encryption in the private-key setting: Stronger security from weaker assumptions. In *Advances in Cryptology – EUROCRYPT '16*, pages 852–880, 2016.

[BL13] D. J. Bernstein and T. Lange. Non-uniform cracks in the concrete: The power of free precomputation. In *Advances in Cryptology – ASIACRYPT '13*, pages 321–340, 2013.

[BLR$^+$15] D. Boneh, K. Lewi, M. Raykova, A. Sahai, M. Zhandry, and J. Zimmerman. Semantically secure order-revealing encryption: Multi-input functional encryption without obfuscation. In *Advances in Cryptology – EUROCRYPT '15*, pages 563–594, 2015.

[CDG$^+$18] S. Coretti, Y. Dodis, S. Guo, and J. P. Steinberger. Random oracles and non-uniformity. In *Advances in Cryptology – EUROCRYPT '18*, pages 227–258, 2018.

[CK18] H. Corrigan-Gibbs and D. Kogan. The discrete-logarithm problem with preprocessing. In *Advances in Cryptology – EUROCRYPT '18*, pages 415–447, 2018.

[CLW$^+$16] N. Chenette, K. Lewi, S. A. Weis, and D. J. Wu. Practical order-revealing encryption with limited leakage. In *Proceedings of the 23rd International Conference on Fast Software Encryption*, pages 474–493, 2016.

[CZ18] D. Cash and C. Zhang. A ciphertext-size lower bound for order-preserving encryption with limited leakage. To appear in *Proceedings of the 16th Theory of Cryptography Conference*, 2018.

[DDC16] F. B. Durak, T. M. DuBuisson, and D. Cash. What else is revealed by order-revealing encryption? In *Proceedings of the 2016 ACM Conference on Computer and Communications Security*, pages 1155–1166, 2016.

[FN99] A. Fiat and M. Naor. Rigorous time/space trade-offs for inverting functions. *SIAM Journal on Computing*, 29(3):790–803, 1999.

[FVY$^+$17] B. Fuller, M. Varia, A. Yerukhimovich, E. Shen, A. Hamlin, V. Gadepally, R. Shay, J. D. Mitchell, and R. K. Cunningham. SoK: Cryptographically protected database search. In *Proceedings of the 38th IEEE Symposium on Security and Privacy*, pages 172–191, 2017.

[GGG$^+$14] S. Goldwasser, S. D. Gordon, V. Goyal, A. Jain, J. Katz, F.-H. Liu, A. Sahai, E. Shi, and H.-S. Zhou. Multi-input functional encryption. In *Advances in Cryptology – EUROCRYPT '14*, pages 578–602, 2014.

[GK16] S. Goldwasser and Y. T. Kalai. Cryptographic assumptions: A position paper. In *Proceedings of the 13th Theory of Cryptography Conference*, pages 505–522, 2016.

[Gol04] O. Goldreich. Foundations of Cryptography – Volume 2: Basic Applications. Cambridge University Press, 2004.

[GSB$^+$17] P. Grubbs, K. Sekniqi, V. Bindschaedler, M. Naveed, and T. Ristenpart. Leakage-abuse attacks against order-revealing encryption. In *Proceedings of the 38th IEEE Symposium on Security and Privacy*, pages 655–672, 2017.

[GW11] C. Gentry and D. Wichs. Separating succinct non-interactive arguments from all falsifiable assumptions. In *Proceedings of the 43rd ACM Annual Symposium on Theory of Computing*, pages 99–108, 2011.

[Hel80] M. E. Hellman. A cryptanalytic time-memory trade-off. *IEEE Transanctions on Information Theory*, 26(4):401–406, 1980.

[KS17] I. Komargodski and G. Segev. From Minicrypt to Obfustopia via private-key functional encryption. In *Advances in Cryptology – EUROCRYPT '17*, pages 122–151, 2017.