A Note for the Ideal Order-Preserving Encryption Object and Generalized Order-Preserving Encryption

Liangliang Xiao
University of Texas at Dallas
xll052000@utdallas.edu

I-Ling Yen
University of Texas at Dallas
ilyen@utdallas.edu

Abstract

Order-preserving encryption (OPE) preserves the order of data in their ciphertexts and, hence, allows range search on the encrypted data without needing to decrypt them. Security analysis of OPE schemes is very important because OPE is not a perfect encryption algorithm (the ciphertexts leak the ordering information of the plaintexts). Most of the existing security analysis for the OPE schemes are informal: they are either based on author-defined attacks or experiments. The authors in [4] initiates the cryptographic study of the OPE scheme. They define the security notion POPF-CCA to qualify the security of OPE. In POPF-CCA, the "ideal" OPE object is defined where the encryption function is uniformly randomly selected from all order-preserving functions (generally the "ideal" OPE object is not computationally feasible), and a (constructed) "real" OPE scheme is secure under POPF-CCA if it is computationally indistinguishable from the ideal object. In other words, although the "ideal" OPE object is not computationally feasible, it is used as the security goal, and a (constructed) "real" OPE scheme is secure if it is as secure as the "ideal" OPE object. Such approach conceives the assumption (but not clearly stated and proved) that the "ideal" OPE object is the most secure OPE. But the correctness of the assumption is an easily ignored problem.

In this paper, we investigate the security of the OPE in more depth. We first give example to show that the "ideal" OPE object may not always be the most secure OPE. It indicates that we need to use the "ideal" encryption object more cautiously in the security analysis of OPE. Additionally we extend the concept of OPE to generalized OPE (GOPE). Unlike OPE, the ciphertexts of GOPE may not be numbers, but GOPE still enables the comparisons on the encrypted data without needing to decrypt them. We present two GOPEs in polynomial-sized and superpolynomial-sized domains that satisfy stronger notions of security than that of the ideal OPE object, respectively.

Key Words: Order-preserving encryption; range query processing; ideal OPE object and generalized OPE; big jump attack and small jump attach; IND-OCPA and IND-OLCPA.

1 Introduction

Order preserving encryption (OPE) [1, 3, 4, 7, 10] is a very important technique for database related applications due to its capability of supporting range query processing [2, 6, 8, 9, 11, 12] directly

on encrypted data without needing to decrypt them and expose them to potential attackers who may have compromised the system. The OPEs do not have perfect security since the ciphertexts leak the ordering information of the plaintexts. But on the other hand, when it is desirable to have a reasonable performance for range query processing while achieving a reasonable degree of security protection, the OPE scheme can be used as long as there is a good understanding of its security risks. However, how secure is the OPE scheme has not been sufficiently analyzed and further research is needed to investigate its security properties.

There are various constructions of the OPE scheme. In [3], the proposed OPE algorithm first generates a sequence of random numbers and then encrypts an integer x to the sum of the first x random numbers. In [10], a sequence of strictly increasing polynomial functions are used to construct the OPE algorithm. The encryption of an integer x is the outcome of the iterative operations of those functions on x. In [7], the OPE algorithm is constructed by using a mapping function composed of partition and identification functions. The partition function divides the range into multiple partitions, and the identification function assigns an identifier to each partition. Then, the mapping function maps an integer x to an identifier. Since different integers may be mapped to the same identifier, the OPE algorithm may output false comparison results. In [1], the authors construct the OPE algorithm following three steps: modeling the input and target distributions, flattening the plaintext database into a flat database, and transforming the flat database into the cipher database. However, security analysis for these and other OPE algorithms has not been fully investigated.

Some partial security analysis has been performed on some OPE algorithms. In [1], the authors construct an OPE scheme and analyze its security, but the analysis has some limitations: (1) It assumes that the adversaries can only view ciphertexts. (2) The analysis is not based on cryptographic analysis, but based on experiments, i.e., they use Kolmogorov-Smirnov test to show that the distribution of the ciphertexts and the target distribution cannot be distinguished. The authors in [4] initiate the cryptographic study of the OPE scheme. They define the security notion IND-OCPA where the adversary can only query the left-or-right encryption oracle with ordered plaintext pairs. An encryption scheme is secure under IND-OCPA if the advantage of an efficient adversary (probability to distinguish whether the returned ciphertexts are encrypted from the left or the right plaintexts) is negligible. IND-OCPA is the highest security notion (with respect to indistinguishability and left-or-right encryption oracle) for OPE algorithms. However, it can be shown that the OPE scheme is susceptible to the big jump attack, and cannot be secure under IND-OCPA unless its ciphertext-space is exponential in the size of the plaintext-space. Then the paper takes an alternative approach: It defines the security notion POPF-CCA and constructs an OPE scheme that is secure under POPF-CCA. In POPF-CCA, the "ideal" OPE object is defined where the encryption function is uniformly randomly selected from all order-preserving functions (the "ideal" OPE object is not computationally feasible), and a "real" OPE scheme is secure if it satisfies the security implied by the ideal OPE object. In other words, although the "ideal" OPE object is not computationally feasible, it is used as the security goal, and a "real" OPE scheme is secure if it is computationally indistinguishable from the "ideal" OPE object. Such approach conceives the assumption (but not clearly stated and proved) that the "ideal" OPE object is the most secure OPE. But unfortunately, the assumption of the ideal OPE object has not been proved.

In this paper, we fist show the negative of the assumption, i.e., the ideal OPE object may not be the most secure OPE. We consider a specific plaintext domain $[m] = \{1, 2\}$ and ciphertext range $[n] = \{j \mid 1 \leq j \leq 2^\lambda\}$, construct a real OPE scheme $\mathcal{SE} = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ for $[m]$ and $[n]$ and prove that

$\mathcal{SE}$ is secure under IND-OCPA, and prove that the ideal OPE object $\mathcal{SE}^*$ for $[m]$ and $[n]$ is not secure under IND-OCPA. For $\mathcal{SE}$ the encryption function $\mathcal{E}$ maps 1 to a random element r in $[1, n-1]$ and maps 2 to $r+1$. We compute the statistical distance between the probability distribution of ciphertexts for plaintext 1 and the probability distribution of ciphertexts for plaintext 2, and prove that it is negligibly small. Based on this fact, we prove that the success probability of every attack against $\mathcal{SE}$ in IND-OCPA is also negligibly small. For the ideal OPE object $\mathcal{SE}^*$, we compute the statistical distance between the probability distribution of ciphertexts for plaintext 1 and the probability distribution of ciphertexts for plaintext 2, and prove that it is significant (greater than a positive constant). Based on this fact, we design an attack to distinguish left plaintext 1 and right plaintext 2 according to the returned ciphertext y by comparing the conditional probabilities $\Pr[y|1]$ and $\Pr[y|2]$. We therefore prove that the success probability of the attack against $\mathcal{SE}^*$ in IND-OCPA is non-negligible. Our proofs indicate that we need to use the "ideal" OPE object more cautiously in the security analysis of OPE.

Then we present two generalized OPE (GOPE) algorithms in polynomial-sized and superpolynomial-sized domains that satisfy stronger notions of security than the ideal OPE object, respectively. The difference between OPE and GOPE lies in that the ciphertexts of OPE are numbers while the ciphertexts of GOPE are allowed to be general mathematical objects. Hence, the GOPE scheme requires a special comparison algorithm to compare the ciphertexts. First, we analyze the security of the ideal OPE object and show that it is not secure under IND-OCPA even in polynomial-sized domains. To demonstrate the existence of a GOPE algorithm that is secure under IND-OCPA in polynomial-sized domains, we construct $\mathcal{SE}_2$, in which the ciphertext y_i for plaintext x_i is a "set". An element in y_i is a share of the relation between x_i and x_j, for all $j \neq i$. When comparing x_i and x_j, the matching pair of shares from x_i and x_j, namely, $s_{i,j}$ and $s_{j,i}$, can be retrieved to reconstruct the relation ($x_i < x_j$ or $x_i > x_j$). We show that $\mathcal{SE}_2$ is secure under IND-OCPA. Next, we weaken the security notion from IND-OCPA to IND-OLCPA for OPE and GOPE schemes in superpolynomial-sized domains. To prevent an adversary from launching the big jump attack, IND-OLCPA has one more constraint to the adversary compared to IND-OCPA, the range of plaintexts in the oracle queries is bounded by a polynomial g_1, i.e., the difference between the largest and the smallest plaintexts in the oracle query is less than or equal to g_1. However, it can be shown that an efficient adversary can still have a non-negligible advantage against any OPE algorithm under IND-OLCPA. We design a small jump attack to prove this. Unlike the big jump attack where the range of plaintexts in the oracle queries is $m-1$, in the small jump attack, the range of plaintexts in the oracle queries is 3. Nevertheless, the lower bound on the advantage of an adversary against any OPE algorithm under IND-OCPA is 1; while the lower bound on the advantage of an adversary against any OPE algorithms under IND-OLCPA decreases to $\frac{1}{g}$, where g is a polynomial. With IND-OLCPA and considering superpolynomial-sized domains, we show that the ideal OPE object cannot achieve the lower bound on the advantage of an adversary. Hence, we construct another GOPE algorithm, $\mathcal{SE}_3$, which achieves the lower bound on the advantage of an adversary under IND-OLCPA. $\mathcal{SE}_3$ is constructed based on two building blocks $\mathcal{SE}_4$ and $\mathcal{SE}_5$. $\mathcal{SE}_4$ is adapted from $\mathcal{SE}_2$ such that the ciphertext of a plaintext x_i is a set, including the shares of the relations between x_i and $g_2 - 1$ plaintexts that are closest to x_i, where g_2 is a polynomial. We prove that $\mathcal{SE}_4$ is secure under IND-OLCPA if $g_2 \geq 2g_1 + 1$. Note that $\mathcal{SE}_4$ can only support comparison between two plaintexts whose difference is bounded by $\frac{g_2-1}{2}$. $\mathcal{SE}_5$ is designed to facilitate the comparison between two plaintexts x_i and x_j, where $|x_i - x_j| \geq \frac{g_2-1}{2}$. Thus, ciphertexts y_i and y_j in $\mathcal{SE}_5$ should preserve the order of the corresponding plaintexts x_i and x_j, when $|x_i - x_j| \geq \frac{g_2-1}{2}$. Note that under IND-OLCPA, the adversary can query plaintexts within the range g_1. Thus, $\mathcal{SE}_5$ should

also guarantee that the ciphertexts y_i and y_j have a small statistical distance if $|x_i - x_j| \leq g_1$. With these two requirements, we construct $\mathcal{SE}_5$ as follows: The ciphertexts of the first l (for some l, $l > g_1$) plaintexts x_1 to x_l are randomly selected numbers. The ciphertext of x_j, $j > l$, is $y_j = y_{j-l} + 1$. This way, the two requirements are satisfied. Since $\mathcal{SE}_3$ includes $\mathcal{SE}_4$ and $\mathcal{SE}_5$, for any pair of plaintexts, either $\mathcal{SE}_4$ or $\mathcal{SE}_5$ will fulfill the comparison task. Also, since the attacker can only query plaintexts within the range g_1, the ciphertexts from $\mathcal{SE}_4$ are indistinguishable and the ciphertexts from $\mathcal{SE}_5$ have a small statistical distance. Thus, $\mathcal{SE}_3$ achieves the lower bound on the advantage of an adversary.

The rest of the paper is organized as follows. In Section 2 we introduce the primitives and how [4] proceeds the security analysis of OPE. In Section 3, we construct an example to prove that the ideal OPE object is not the most secure OPE. In Sections 4, 5, 6 we present two generalized OPE (GOPE) algorithms in polynomial-sized and superpolynomial-sized domains that satisfy stronger notions of security than the ideal OPE object, respectively. Specifically, in Section 4, we prove that the ideal OPE object is not secure under IND-OCPA in polynomial-sized domains. Then, the concept of the GOPE scheme and its construction is introduced. Also, we prove that GOPE is secure under IND-OCPA in polynomial-sized domains. In Section 5, we define the security notion IND-OLCPA, design the small jump attack, and derive a lower bound on the advantage of an adversary under IND-OLCPA in superpolynomial-sized domains. In Section 6, we show that the ideal OPE object does not achieve the lower bound on the advantage of an adversary under IND-OLCPA in superpolynomial-size domains. Also, a GOPE algorithm which can achieve the derived lower bound is constructed. Finally, we conclude the paper in Section 7.

2 Preliminaries

Let λ be the security parameter and ν be a negligible function. Let $x \xleftarrow{\$} A$ denote that x is uniformly randomly selected from set A, $x \xleftarrow{\$} \mathcal{X}$ denote that randomized algorithm $\mathcal{X}$ returns value x, and $\mathcal{X}^{\mathcal{Y}}$ denote that algorithm $\mathcal{X}$ is accessible to oracle $\mathcal{Y}$. For positive integers m and n satisfying $m \leq n$, let $[m] = \{i | 1 \leq i \leq m\}$ denote the domain of plaintexts and $[n] = \{i | 1 \leq i \leq n\}$ denote the range of ciphertexts. The definition of the order-preserving encryption (OPE) scheme [4] is presented as follows.

Definition 2.1 (OPE scheme [4])**.** An OPE scheme $\mathcal{SE} = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ is a deterministic symmetric-key encryption scheme, where $\mathcal{K} : \{0,1\}^* \to \{0,1\}^*$ is a key generation algorithm, $\mathcal{E} : [m] \times \{0,1\}^* \to [n]$ is a deterministic encryption algorithm, and $\mathcal{D} : [n] \times \{0,1\}^* \to [m]$ is a decryption algorithm. $\mathcal{SE}$ satisfies that

$$\Pr[\mathcal{D}(\mathcal{E}(x,k),k) = x] > 1 - \nu(\lambda)$$

for any $x \in [m]$ and key k, and

$$\mathcal{E}(x,k) < \mathcal{E}(x',k)$$

for any $x < x'$. □

Various security notions are defined attempting to qualify the security of OPE. We start from the basic security notion IND-CPA (indistinguishability under chosen-plaintext attack) and define it in Definition 2.2.

Definition 2.2 (IND-CPA)**.** Let $\mathcal{SE} = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ be a symmetric-key encryption scheme and $b \in \{0, 1\}$. Let $\mathcal{E}_k(\mathcal{LR}(\cdot, \cdot, b))$ be a left-or-right encryption oracle such that for queries $\{(x_0^u, x_1^u)\}_{u=1}^h$, it returns

$$\mathcal{E}(x_b^u, k) \xleftarrow{\$} \mathcal{E}_k(\mathcal{LR}(x_0^u, x_1^u, b))$$

for $1 \leq u \leq h$. Let $\mathcal{A}$ be an adversary that can access $\mathcal{E}_k(\mathcal{LR}(\cdot, \cdot, b))$ and finally returns a bit b' as a guess of b. Consider the following experiment.

$$\begin{aligned} &\textbf{\textit{Experiment}}\ \mathbf{Exp}_{\mathcal{SE},\mathcal{A}}^{\text{IND-CPA-}b} \\ &\quad k \xleftarrow{\$} \mathcal{K}; b' \xleftarrow{\$} \mathcal{A}^{\mathcal{E}_k(\mathcal{LR}(\cdot,\cdot,b))}; \text{Return } b' \end{aligned}$$

The encryption scheme $\mathcal{SE}$ is said to be secure under IND-CPA if for every probabilistic polynomial time (PPT) adversary $\mathcal{A}$, the advantage of $\mathcal{A}$, defined by

$$\mathbf{Adv}_{\mathcal{SE},\mathcal{A}}^{\text{IND-CPA}} = \Pr[\mathbf{Exp}_{\mathcal{SE},\mathcal{A}}^{\text{IND-CPA-1}} = 1] - \Pr[\mathbf{Exp}_{\mathcal{SE},\mathcal{A}}^{\text{IND-CPA-0}} = 1],$$

is bounded by a negligible function of the security parameter. □

OPE schemes are not secure under IND-CPA because the ciphertexts leaks the ordering information of plaintexts. Consider the adversary queries (x_0^1, x_1^1) and (x_0^2, x_1^2), where $x_0^1 < x_0^2$ and $x_1^1 \geq x_1^2$. If $b = 0$, x_0^1 and x_0^2 will be encrypted, where $x_0^1 < x_0^2$; if $b = 1$, x_1^1 and x_1^2 will be encrypted where $x_1^1 \geq x_1^2$. Since OPE preserves order, the adversary can distinguish whether the plaintexts are x_0^1 and x_0^2 or x_1^1 and x_1^2 by comparing the corresponding ciphertexts. Thus, the advantage of such adversary is 1.

Thus, the security notion is necessarily weakened to IND-OCPA (indistinguishability under ordered chosen-plaintext attack) [4], where the adversary is forbidden to query plaintexts with different orders.

Definition 2.3 (IND-OCPA [4])**.** IND-OCPA has the same definition as that of IND-CPA except that the adversary is only allowed to query $\{(x_0^u, x_1^u)\}_{u=1}^h$, where the condition $x_0^u < x_0^v \Leftrightarrow x_1^u < x_1^v, 1 \leq u, v \leq h$ is satisfied. □

IND-OCPA is the highest security notion (with respect to indistinguishability and left-or-right encryption oracle) for OPE algorithms. However, it has been shown in [4] that OPE schemes are susceptible to the following the big jump attack under IND-OCPA.

Definition 2.4 (Big jump attack [4])**.** Consider the following PPT adversary $\mathcal{A}_{BJ}$ with three oracle queries in the experiment of security notion IND-OCPA.

$$\begin{aligned} &\textbf{\textit{Adversary}}\ {\mathcal{A}_{BJ}}^{\mathcal{E}_k(\mathcal{LR}(\cdot,\cdot,b))} \\ &\quad x \xleftarrow{\$} \{1, ..., m-1\} \\ &\quad y_1 \longleftarrow \mathcal{E}_k(\mathcal{LR}(1, x, b)) \\ &\quad y_2 \longleftarrow \mathcal{E}_k(\mathcal{LR}(x, x+1, b)) \\ &\quad y_3 \longleftarrow \mathcal{E}_k(\mathcal{LR}(x+1, m, b)) \\ &\quad \text{Return 1 if } y_3 - y_2 > y_2 - y_1; \text{else return 0} \end{aligned}$$ □

In the big jump attack, the attacker chooses left plaintexts 1, x, and $x+1$, and the right plaintexts x, $x+1$, and m, where x is randomly selected from $\{1, ..., m-1\}$. From the ciphertexts, if $y_3 - y_2 > y_2 - y_1$, then the attacker can guess that the right plaintexts were encrypted; if $y_3 - y_2 \leq y_2 - y_1$, then the attacker can guess that the left plaintexts were encrypted. Since the distance between two ciphertexts can reflect, to some extent, the distance between the corresponding two plaintexts, such guess could have a high probability of being correct. The lower bound on advantage of the adversary has been derived in [4] and is cited in Lemma 2.5.

Lemma 2.5. $\mathbf{Adv}^{IND\text{-}OCPA}_{\mathcal{SE},\mathcal{A}_{BJ}} \geq 1 - \frac{2\log n}{m-1}$.

Remark 1. Note that for efficient OPE, both $\log m$ and $\log n$ should be bounded by a polynomial of λ. Therefore $\mathbf{Adv}^{\text{IND-OCPA}}_{\mathcal{SE},\mathcal{A}_{BJ}} \geq 1 - \nu(\lambda)$ if m is a superpolynomial of λ, which implies that it is impossible to construct an OPE that is secure under IND-OCPA if m is a superpolynomial of λ. However, the lower bound on advantage of the adversary does not eliminate the possibility for designing an OPE scheme that is secure under IND-OCPA if m is bounded by a polynomial of λ.

Because of the big jump attack, the authors in [4] take an alternative approach: They define the security notion POPF-CCA (pseudorandom order-preserving function under chosen-ciphertext attack) based on the ideal OPE object defined as follows.

Definition 2.6 (Ideal OPE Object [4])**.** Let $[m]$ be the plaintext domain and $[n]$ be the ciphertext range. The ideal OPE object $\mathcal{SE}^* = (\mathcal{K}^*, \mathcal{E}^*, \mathcal{D}^*)$ is defined as follows.

- $\mathcal{K}^*$: It uniformly randomly selects $f \in OPE_{m,n} = \{f : [m] \to [n] \mid x < x' \Leftrightarrow f(x) < f(x')\}$;
- $\mathcal{E}^*$: For plaintext x, it returns $f(x)$;
- $\mathcal{D}^*$: For ciphertext y, it returns $f^{-1}(y)$. □

For a "real" OPE scheme $\mathcal{SE} = (\mathcal{K}, \mathcal{E}, \mathcal{D})$, it is secure under POPF-CCA if it is computationally indistinguishable from the ideal OPE object $\mathcal{SE}^* = (\mathcal{K}^*, \mathcal{E}^*, \mathcal{D}^*)$. Formally, the security notion POPF-CCA is defined as follows.

Definition 2.7 (POPF-CCA [4])**.** Let the advantage of the adversary in POPF-CCA be

$$\mathbf{Adv}^{\text{POPF-CCA}}_{\mathcal{SE},\mathcal{A}} = \Pr[k \xleftarrow{\$} \mathcal{K} : \mathcal{A}^{\mathcal{E}(k,\cdot),\mathcal{D}(k,\cdot)} = 1] - \Pr[f \xleftarrow{\$} \mathcal{K}^* : \mathcal{A}^{\mathcal{E}^*(\cdot),\mathcal{D}^*(\cdot)} = 1].$$

The encryption scheme $\mathcal{SE}$ is said to be secure under POPF-CCA if $\mathbf{Adv}^{\text{POPF-CCA}}_{\mathcal{SE},\mathcal{A}}$ is bounded by a negligible function of the security parameter for every PPT adversary $\mathcal{A}$. □

Based on the security notion POPF-CCA, the authors in [4] construct a real OPE scheme and prove that it is secure under POPF-CCA. In other words, in their approach the ideal OPE object is used as the security goal and construct real OPE scheme to achieve that security goal. However, the problem is: is the ideal OPE object always the most secure OPE. We construct a counterexample to show the negative conclusion in the next section.

3 Counterexample and the Security Analysis

In this section we show that there exists situation such that the ideal OPE object is not the most secure OPE. We consider a specific plaintext domain $[m]$ and ciphertext range $[n]$, construct a real

OPE scheme $\mathcal{SE} = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ for $[m]$ and $[n]$ and prove that $\mathcal{SE}$ is secure under IND-OCPA, and prove that the ideal OPE object $\mathcal{SE}^*$ for $[m]$ and $[n]$ is not secure under IND-OCPA.

Plaintext domain and ciphertext range: In this section, let $m = 2$ and $n = 2^\lambda$ where λ is the security parameter. Then the plaintext domain is $[m] = \{1, 2\}$ and the ciphertext range is $[n] = \{j \mid 1 \leq j \leq 2^\lambda\}$.

The real OPE scheme: First we construct a real OPE scheme $\mathcal{SE} = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ as follows.

- $\mathcal{K}$: It uniformly randomly selects $f \in \{f : [m] \to [n] \mid f(2) = f(1) + 1\}$;

- $\mathcal{E}$: For plaintext x, it returns $f(x)$;

- $\mathcal{D}$: For ciphertext y, it returns $f^{-1}(y)$.

Unlike the ideal OPE object, in the real OPE scheme $\mathcal{SE}$ the encryption function is uniformly randomly selected from a subset of order-preserving functions. The encryption function $\mathcal{E}$ has the property such that 1 is encrypted to a random element r in $[1, n-1]$ while 2 is encrypted to $r+1$. To show that the real OPE scheme $\mathcal{SE}$ is secure under IND-OCPA, we compute the statistical distance between the probability distribution of ciphertexts for plaintext 1 and the probability distribution of ciphertexts for plaintext 2, and prove that it is negligibly small. Based on this fact, we show that the success probability of every attack in IND-OCPA is also negligibly small.

Lemma 3.1. *Let Δ be the statistical distance between $\mathcal{E}(1)$ and $\mathcal{E}(2)$. Then $\Delta = \nu(\lambda)$.*

Proof. According to the definition of $\mathcal{E}$, $\mathcal{E}(i) \in [n]$ subjects to the probability distribution such that

$$\Pr[\mathcal{E}(1) = j] = \begin{cases} \frac{1}{n-1} & \text{for} \quad 1 \leq j < n \\ 0 & \text{for} \quad j = n \end{cases} \quad \text{and} \quad \Pr[\mathcal{E}(2) = j] = \begin{cases} 0 & \text{for} \quad j = 1 \\ \frac{1}{n-1} & \text{for} \quad 1 < j \leq n \end{cases}$$

Thus

$$\begin{aligned} \Delta &= \frac{1}{2} \sum_j |\Pr[\mathcal{E}(1) = j] - \Pr[\mathcal{E}(2) = j]| \\ &= \frac{1}{n-1} = \frac{1}{2^\lambda - 1} = \nu(\lambda). \end{aligned}$$

□

Proposition 3.2. *$\mathcal{SE}$ is secure under IND-OCPA. Specifically, $\mathbf{Adv}^{IND\text{-}OCPA}_{\mathcal{SE},\mathcal{A}} = \nu(\lambda)$ for every PPT adversary $\mathcal{A}$.*

Proof. Note that the adversary has to query ordered plaintext pairs to $\mathcal{LR}$ in IND-OCPA and here are the all possible queries of the adversary: $\{(1,1)\}$, $\{(2,2)\}$, $\{(1,1),(2,2)\}$, $\{(1,2)\}$, and $\{(2,1)\}$. We analyze the security of $\mathcal{SE}$ according to these queries.

(1) The adversary queries $\{(1,1)\}$ to $\mathcal{LR}$. In this case, since the left plaintext equals to the right plaintext, the returned ciphertexts cannot help the adversary to decide whether the left plaintext or right plaintext is encrypted. Hence $\mathbf{Adv}^{\text{IND-OCPA}}_{\mathcal{SE},\mathcal{A}} = 0$.

(2) The adversary queries $\{(2,2)\}$ or $\{(1,1),(2,2)\}$ to $\mathcal{LR}$. The situation is similar to that in (1) and hence $\mathbf{Adv}^{\text{IND-OCPA}}_{\mathcal{SE},\mathcal{A}} = 0$.

(3) The adversary queries $\{(1,2)\}$ to $\mathcal{LR}$. According to Lemma 3.1,

$$\mathbf{Adv}^{\text{IND-OCPA}}_{\mathcal{SE},\mathcal{A}} = \Pr[\mathbf{Exp}^{\text{IND-CPA-1}}_{\mathcal{SE},\mathcal{A}} = 1] - \Pr[\mathbf{Exp}^{\text{IND-CPA-0}}_{\mathcal{SE},\mathcal{A}} = 1] = \Delta = \nu(\lambda).$$

(4) The adversary queries $\{(2,1)\}$ to $\mathcal{LR}$. The situation is similar to that in (3) and hence $\mathbf{Adv}^{\text{IND-OCPA}}_{\mathcal{SE},\mathcal{A}} = \nu(\lambda)$.

According to (1)-(4), $\mathbf{Adv}^{\text{IND-OCPA}}_{\mathcal{SE},\mathcal{A}} = \nu(\lambda)$ for every PPT adversary $\mathcal{A}$. □

The ideal OPE object: According to Definition 2.6, the ideal OPE object $\mathcal{SE}^* = (\mathcal{K}^*, \mathcal{E}^*, \mathcal{D}^*)$ is defined as follows.

- $\mathcal{K}^*$: It uniformly randomly selects $f \in \{f : [m] \to [n] \mid f(1) < f(2)\}$;

- $\mathcal{E}^*$: For plaintext x, it returns $f(x)$;

- $\mathcal{D}^*$: For ciphertext y, it returns $f^{-1}(y)$.

To show that the ideal OPE object $\mathcal{SE}^*$ is not secure under IND-OCPA, we compute the statistical distance between the probability distribution of ciphertexts for plaintext 1 and the probability distribution of ciphertexts for plaintext 2, and prove that it is significant (greater than a positive constant). Based on this fact, we design an attack to distinguish left plaintext 1 and right plaintext 2 according to the returned ciphertext y by comparing the conditional probabilities $\Pr[y|1]$ and $\Pr[y|2]$. It can be shown that the success probability of the attack is non-negligible (greater than a positive constant).

Lemma 3.3. *Let Δ^* be the statistical distance between $\mathcal{E}^*(1)$ and $\mathcal{E}^*(2)$. Then $\Delta^* = \Omega(1)$.*

Proof. Since $|OPE_{m,n}| = \binom{n}{m}$ and $|\{f \in OPE_{m,n} \mid f(i) = j\}| = \binom{j-1}{i-1}\binom{n-j}{m-i}$, for $i \in [m]$, $\mathcal{E}^*(i) \in [n]$ subjects to the negative hypergeometric distribution

$$\frac{\binom{j-1}{i-1}\binom{n-j}{m-i}}{\binom{n}{m}}, 1 \le j \le n.$$

Thus

$$\begin{aligned}
\Delta^* &= \frac{1}{2}\sum_j \left| \frac{\binom{j-1}{0}\binom{n-j}{m-1}}{\binom{n}{m}} - \frac{\binom{j-1}{1}\binom{n-j}{m-2}}{\binom{n}{m}} \right| \\
&= \frac{1}{2}\sum_j \left| \frac{\binom{j-1}{0}\binom{n-j}{1}}{\binom{n}{2}} - \frac{\binom{j-1}{1}\binom{n-j}{0}}{\binom{n}{2}} \right| \\
&= \frac{\sum_j |n-2j+1|}{2\binom{n}{2}} \\
&= \frac{n}{2(n-1)} \ge \frac{1}{2} = \Omega(1).
\end{aligned}$$

□

For the ideal OPE object, if 1 is encrypted to j, then 2 must be encrypted to $[j+1, n]$, and hence there is more choices of the encryption of 2 if j is small; similarly if 2 is encrypted to j, then 1 must be encrypted to $[1, j-1]$, and hence there is more choices of the encryption of 1 if j is large. Since the encryption function of the ideal OPE object is uniformly randomly selected from all order-preserving functions, 1 is **more likely** to be encrypted to $[1, \frac{n+1}{2}]$ and 2 is **more likely** to be encrypted to $[\frac{n+1}{2}, n]$. Lemma 4.1 indicates that the difference of the encryptions of 1 and 2 is significant. Such significant difference can be used to design the attack, and based on the attack we prove that the ideal OPE object is not secure under IND-OCPA in Proposition 4.2 .

Proposition 3.4. *For the ideal OPE object $\mathcal{SE}^*$ with the plaintext domain $[m]$ and the ciphertext range $[n]$, there exists an adversary $\mathcal{A}$ who can distinguish plaintexts 1 and 2 with one oracle query under IND-OCPA such that $\mathbf{Adv}^{IND\text{-}OCPA}_{\mathcal{SE}^*,\mathcal{A}} = \Omega(1)$. In other words, the ideal OPE object $\mathcal{SE}^*$ is not secure under IND-OCPA.*

Proof. Since $|OPE_{m,n}| = \binom{n}{m}$ and $|\{f \in OPE_{m,n} \mid f(i) = j\}| = \binom{j-1}{i-1}\binom{n-j}{m-i}$, for $i \in [m]$, $\mathcal{E}^*(i) \in [n]$ subjects to the negative hypergeometric distribution

$$\frac{\binom{j-1}{i-1}\binom{n-j}{m-i}}{\binom{n}{m}}, 1 \leq j \leq n.$$

Note that

$$\begin{aligned}
\frac{\binom{j-1}{0}\binom{n-j}{m-1}}{\binom{n}{m}} > \frac{\binom{j-1}{1}\binom{n-j}{m-2}}{\binom{n}{m}} &\Longleftrightarrow \binom{n-j}{m-1} > (j-1)\binom{n-j}{m-2} \\
&\Longleftrightarrow n-j-m+2 > (j-1)(m-1) \\
&\overset{m=2}{\Longleftrightarrow} n-j > j-1 \\
&\Longleftrightarrow j < \frac{n+1}{2}.
\end{aligned}$$

Thus we construct the PPT adversary $\mathcal{A}$ with one oracle query in the experiment of security notion IND-OCPA as follows (note that $y \neq \frac{n+1}{2}$ since $n = 2^\lambda$).

$$\begin{aligned}
&\boldsymbol{Adversary}\ \mathcal{A}^{\mathcal{E}^*_k(\mathcal{LR}(\cdot,\cdot,b))} \\
&\quad y \longleftarrow \mathcal{E}^*_k(\mathcal{LR}(1,2,b)) \\
&\quad \text{Return 0 if } y < \frac{n+1}{2} \\
&\quad \text{Return 1 if } y > \frac{n+1}{2}
\end{aligned}$$

Then

$$\begin{aligned}
\mathbf{Adv}^{\text{IND-OCPA}}_{\mathcal{SE}^*,\mathcal{A}} &= \Pr[\mathbf{Exp}^{\text{IND-OCPA-1}}_{\mathcal{SE}^*,\mathcal{A}} = 1] - \Pr[\mathbf{Exp}^{\text{IND-OCPA-0}}_{\mathcal{SE}^*,\mathcal{A}} = 1] \\
&= \Delta^* = \Omega(1).
\end{aligned}$$

□

Remark 2. The proofs in Lemma 4.1 and Proposition 4.2 can be generalized to show that the ideal OPE object is not secure under IND-OCPA for any plaintext domain $[m]$ and ciphertext range $[n]$.

We conclude the results in this section in the following theorem.

Theorem 3.5. *The ideal OPE object $\mathcal{SE}^*$ is not the most secure OPE for $m = 2$ and $n = 2^\lambda$. Specifically, there exists a real OPE scheme $\mathcal{SE}$ secure under IND-OCPA while the ideal OPE object $\mathcal{SE}^*$ is not secure under IND-OCPA.*

4 Ideal OPE and GOPE in Polynomial-sized Domain

In Subsection 4.1, we extend the proof in Section 3 to show that $\mathcal{SE}^*$ is not secure under IND-OCPA in the polynomial-sized domain. Then, in Subsection 4.2, we construct a generalized OPE scheme in the polynomial-sized domain and show that it is secure under IND-OCPA.

4.1 Ideal OPE Object in Polynomial-sized Domain

Let $\mathcal{SE}^* = (\mathcal{K}^*, \mathcal{E}^*, \mathcal{D}^*)$ be the ideal OPE object [4]. Analogous to the proofs in Section 3, we compute the statistical distance between the probability distribution of ciphertexts for plaintext 1 and the probability distribution of ciphertexts for plaintext 2, and prove that it is greater than a positive constant. Based on this fact, we design an attack to distinguish left plaintext 1 and right plaintext 2 according to the returned ciphertext y by comparing the conditional probabilities $\Pr[y|1]$ and $\Pr[y|2]$. It can be shown that the success probability of the attack is a positive constant.

Lemma 4.1. *Let $\mathcal{SE}^* = (\mathcal{K}^*, \mathcal{E}^*, \mathcal{D}^*)$ be the ideal OPE object with the plaintext domain $[m]$ and the ciphertext range $[n]$, where $2 \leq m \leq n$. Let Δ be the statistical distance between $\mathcal{E}^*(1)$ and $\mathcal{E}^*(2)$. Then $\Delta = \Omega(1)$.*

Proof. Since $|SIF_{m,n}| = \binom{n}{m}$ and $|\{f \in SIF_{m,n} \mid f(i) = j\}| = \binom{j-1}{i-1}\binom{n-j}{m-i}$, for $i \in [m]$, $\mathcal{E}^*(i) \in [n]$ subjects to the negative hypergeometric distribution

$$\frac{\binom{j-1}{i-1}\binom{n-j}{m-i}}{\binom{n}{m}}, 1 \leq j \leq n.$$

Thus

$$\begin{aligned}\Delta &= \frac{1}{2}\sum_j \left| \frac{\binom{j-1}{0}\binom{n-j}{m-1}}{\binom{n}{m}} - \frac{\binom{j-1}{1}\binom{n-j}{m-2}}{\binom{n}{m}} \right| = \frac{1}{2}\sum_j \left| 1 - \frac{(j-1)(m-1)}{n-j-m+2} \right| \frac{\binom{n-j}{m-1}}{\binom{n}{m}} \\ &= \frac{1}{2}\sum_j \left| \frac{n-jm+1}{n-j-m+2} \right| \frac{m}{n}\frac{n-j}{n-1}\cdots\frac{n-j-m+2}{n-m+1}.\end{aligned}$$

Note that

$$\begin{aligned}\frac{n-jm+1}{n-j-m+2} \geq \frac{1}{2} &\Leftrightarrow 2n-2jm+2 \geq n-j-m+2 \Leftrightarrow j \leq \frac{n+m}{2m-1}; \\ \frac{n-j}{n-1} \geq \frac{m-1}{m} &\Leftrightarrow mn-mj \geq mn-m-n+1 \Leftrightarrow j \leq \frac{n+m-1}{m}; \\ &\cdots \\ \frac{n-j-m+2}{n-m+1} \geq \frac{m-1}{m} &\Leftrightarrow mn-mj-m^2+2m \geq mn-m^2+m-n+m-1 \Leftrightarrow j \leq \frac{n+1}{m}.\end{aligned}$$

If $\frac{n}{2m} \geq 1$, it can be verified that all the above inequalities hold for $j \leq \frac{n}{2m}$. Hence

$$\Delta \geq \frac{1}{2}\cdot\frac{n}{2m}\cdot\frac{1}{2}\cdot\frac{m}{n}\cdot(\frac{m-1}{m})^{m-1} \geq \frac{1}{8e},$$

where $(\frac{m-1}{m})^{m-1} = (1+\frac{1}{m-1})^{-(m-1)} \geq \frac{1}{e}$ for $m \geq 2$. If $\frac{n}{2m} < 1$, it can be verified that $\frac{n-jm+1}{n-j-m+2} = \frac{n-j}{n-1} = \cdots = \frac{n-j-m+2}{n-m+1} = 1$ for $j = 1$. Hence

$$\Delta \geq \frac{1}{2}\cdot\frac{m}{n} > \frac{1}{4}.$$

Consequently $\Delta = \Omega(1)$. □

Proposition 4.2. *For the ideal OPE object $\mathcal{SE}^*$ with the plaintext domain $[m]$ and the ciphertext range $[n]$, where $2 \leq m \leq n$, there exists an adversary $\mathcal{A}_1$ who can distinguish plaintexts 1 and 2 with one oracle query under IND-OCPA such that $\mathbf{Adv}^{IND\text{-}OCPA}_{\mathcal{SE}^*,\mathcal{A}_1} = \Omega(1)$.*

Proof. Since $|SIF_{m,n}| = \binom{n}{m}$ and $|\{f \in SIF_{m,n} \mid f(i) = j\}| = \binom{j-1}{i-1}\binom{n-j}{m-i}$, for $i \in [m]$, $\mathcal{E}^*(i) \in [n]$ subjects to the negative hypergeometric distribution

$$\frac{\binom{j-1}{i-1}\binom{n-j}{m-i}}{\binom{n}{m}}, 1 \leq j \leq n.$$

Note that

$$\frac{\binom{j-1}{0}\binom{n-j}{m-1}}{\binom{n}{m}} > \frac{\binom{j-1}{1}\binom{n-j}{m-2}}{\binom{n}{m}} \Leftrightarrow \binom{n-j}{m-1} > (j-1)\binom{n-j}{m-2} \Leftrightarrow n-j-m+2 > (j-1)(m-1).$$

Consider the PPT adversary $\mathcal{A}_1$ with one oracle query in the experiment of security notion IND-OCPA.

Adversary $\mathcal{A}_1^{\mathcal{E}_k^*(\mathcal{LR}(\cdot,\cdot,b))}$
$y \longleftarrow \mathcal{E}_k^*(\mathcal{LR}(1,2,b))$
Return 0 if $n-y-m+2 > (y-1)(m-1)$
Return $b' \xleftarrow{\$} \{0,1\}$ if $n-y-m+2 = (y-1)(m-1)$
Return 1 if $n-y-m+2 < (y-1)(m-1)$

Then

$$\begin{aligned}\mathbf{Adv}^{\text{IND-OCPA}}_{\mathcal{SE}^*,\mathcal{A}_1} &= \Pr[\mathbf{Exp}^{\text{IND-OCPA-1}}_{\mathcal{SE}^*,\mathcal{A}_1} = 1] - \Pr[\mathbf{Exp}^{\text{IND-OCPA-0}}_{\mathcal{SE}^*,\mathcal{A}_1} = 1] \\ &= \Delta = \Omega(1).\end{aligned}$$

□

Note that for encryption $\mathcal{E}^* : [m] \to [n]$, if the adversary retrieves plaintext ciphertext pair (x, y), then $\mathcal{E}^*|_{\{i|x+1\leq i\leq m\}} : \{i \mid x+1 \leq i \leq m\} \to \{j \mid y+1 \leq j \leq n\}$ remains to be ideal OPE encryption. Thus $\mathcal{A}_1$ can be extended to PPT adversary $\mathcal{A}_2$ to distinguish plaintexts $x+1$ and $x+2$, $1 \leq x \leq m-2$, with two oracle queries under IND-OCPA.

Corollary 4.3. *For the ideal OPE object $\mathcal{SE}^*$ with the plaintext domain $[m]$ and the ciphertext range $[n]$, where $2 \leq m \leq n$, there exists an adversary $\mathcal{A}_2$ who can distinguish plaintexts $x+1$ and $x+2$, $1 \leq x \leq m-2$, with two oracle queries under IND-OCPA such that $\mathbf{Adv}^{IND\text{-}OCPA}_{\mathcal{SE}^*,\mathcal{A}_2} = \Omega(1)$.*

Proof. Consider the PPT adversary $\mathcal{A}_2$ with two oracle queries in the experiment of security notion

IND-OCPA.

Adversary $\mathcal{A}_2^{\mathcal{E}_k^*(\mathcal{LR}(\cdot,\cdot,b))}$

$x \xleftarrow{\$} \{1, ..., m-2\}$
$y \longleftarrow \mathcal{E}_k^*(\mathcal{LR}(x, x, b))$
$y' \longleftarrow \mathcal{E}_k^*(\mathcal{LR}(x+1, x+2, b))$
Return 0 if $n - y' - (m-x) + 2 > (y' - y - 1)(m - x - 1)$
Return $b' \xleftarrow{\$} \{0, 1\}$ if $n - y' - (m-x) + 2 = (y' - y - 1)(m - x - 1)$
Return 1 if $n - y' - (m-x) + 2 < (y' - y - 1)(m - x - 1)$

Then $\mathbf{Adv}^{\text{IND-OCPA}}_{\mathcal{SE}^*,\mathcal{A}_2} = \Omega(1)$ based on the same proof of Proposition 4.2. □

Proposition 4.2 and Corollary 4.3 imply that the advantages of adversaries $\mathcal{A}_1$ and $\mathcal{A}_2$ against $\mathcal{SE}^*$ are greater than a positive constant, where $2 \leq m \leq n$. Therefore, $\mathcal{SE}^*$ is not secure under IND-OCPA in polynomial-sized domains.

4.2 Generalized OPE in the Polynomial-sized Domain

We define the concept of the generalized OPE (GOPE) scheme. Unlike OPE whose ciphertext-space is $[n]$, GOPE adopts general mathematical objects as ciphertexts. Hence a special comparison algorithm is needed to compare the ciphertexts.

Definition 4.4 (GOPE scheme)**.** A GOPE scheme $\mathcal{SE} = (\mathcal{K}, \mathcal{E}, \mathcal{D}, \mathcal{C})$ is a symmetric-key encryption scheme, where $\mathcal{K} : \{0,1\}^* \to \{0,1\}^*$ is a key generation algorithm, $\mathcal{E} : [m] \times \{0,1\}^* \to R$ is an encryption algorithm, $\mathcal{D} : R \times \{0,1\}^* \to [m]$ is a decryption algorithm, and $\mathcal{C} : R \times R \to \{=, >, <\}$ is a comparison algorithm. $\mathcal{SE}$ satisfies that

$$\Pr[\mathcal{D}(\mathcal{E}(x, k), k) = x] > 1 - \nu(\lambda)$$

for any $x \in [m]$ and key k, and

$$\Pr[\mathcal{C}(\mathcal{E}(x, k), \mathcal{E}(x', k)) = w] > 1 - \nu(\lambda)$$

for any xwx' and $w \in \{=, >, <\}$. □

Next we construct the GOPE scheme $\mathcal{SE}_2 = (\mathcal{K}_2, \mathcal{E}_2, \mathcal{D}_2, \mathcal{C}_2)$ with m being a polynomial of λ, and prove that it is secure under IND-OCPA. In $\mathcal{SE}_2$ the ciphertext y for plaintext x is a "set". An element in y is a share of the relation between x and x', for all other plaintexts x'. When comparing x and x', the matching pair of shares from x and x' can be retrieved to reconstruct the relation ($x < x'$ or $x > x'$). Let the symbol "<" encoded to $1 \in Z_3$ and the symbol ">" encoded to $2 \in Z_3$. $\mathcal{SE}_2$ is constructed as follows.

- $\mathcal{K}_2$: Given the domain size m, it randomly picks a permutation π of the set $\{(x, x') \mid 1 \leq x < x' \leq m\}$, and randomly generates $r_{xx'} \in Z_3$ for $1 \leq x < x' \leq m$. It returns $\{\pi, r_{xx'} \mid 1 \leq x < x' \leq m\}$;

- $\mathcal{E}_2$: For plaintext x, it returns the ciphertext $y = \{(\pi(x', x), r_{x'x}) \mid x' < x\} \cup \{(\pi(x, x'), 1 + r_{xx'}) \mid x' > x\}$;

- $\mathcal{D}_2$: For ciphertext y, it retrieves (any) two elements (i, s) and (i', s') from the set y, and returns plaintext x which appears in both $\pi^{-1}(i)$ and $\pi^{-1}(i')$;

- $\mathcal{C}_2$: For ciphertexts y and y', if $y = y'$, it returns $=$. Otherwise, it retrieves (i, s) from the set y and (i, s') from the set y', if $s - s' = 1$, it returns $<$; if $s - s' = 2$, it returns $>$.

The efficiency, correctness, and security of $\mathcal{SE}_2$ are presented in Lemma 4.5 and Theorem 4.6.

Lemma 4.5. *$\mathcal{SE}_2$ is efficient and correct.*

Proof. The efficiency of $\mathcal{SE}_2$ and correctness of decryption algorithm can be easily verified. It suffices to verify the correctness of comparison algorithm. For $x = x'$, since $\mathcal{E}_2(x, k) = \mathcal{E}_2(x', k)$, it is correct for the comparison algorithm to return $=$. For $x \neq x'$, there exist unique i, s, s' such that $(i, s) \in \mathcal{E}_2(x, k)$ and $(i, s') \in \mathcal{E}_2(x', k)$. If $x < x'$, $\mathcal{E}_2(x, k) = \{..., (\pi(x, x'), 1+r_{xx'}), ...\}$ and $\mathcal{E}_2(x', k) = \{..., (\pi(x, x'), r_{xx'}), ...\}$, thus $(1 + r_{xx'}) - r_{xx'} = 1$, hence it is correct for the comparison algorithm to return $<$; if $x > x'$, $\mathcal{E}_2(x, k) = \{..., (\pi(x', x), r_{x'x}), ...\}$ and $\mathcal{E}_2(x', k) = \{..., (\pi(x', x), 1 + r_{x'x}), ...\}$, thus $r_{xx'} - (1 + r_{xx'}) = -1 = 2$, hence it is correct for the comparison algorithm to return $>$. □

Theorem 4.6. *$\mathcal{SE}_2$ is secure under IND-OCPA. Specifically,* $\mathbf{Adv}^{\textit{IND-OCPA}}_{\mathcal{SE}_2,\mathcal{A}} = 0$.

Proof. Assume that the adversary queries $\{(x_0^u, x_1^u)\}_{u=1}^h$ under IND-OCPA. According to the restriction under IND-OCPA, $x_0^u = x_0^v \Leftrightarrow x_0^u = x_0^v$. Since it will not increase the advantage by querying two identical plaintexts pairs, it suffices to consider $x_0^1 < x_0^2 < ... < x_0^h$ and $x_1^1 < x_1^2 < ... < x_1^h$. Hence, the adversary views $(\mathcal{E}_2(x_0^1, k), ..., \mathcal{E}_2(x_0^h, k))$ for $b = 0$, and the adversary views $(\mathcal{E}_2(x_1^1, k), ..., \mathcal{E}_2(x_1^h, k))$ otherwise. It suffices to prove that the above two probability distributions are identical because it implies that $\mathbf{Adv}^{\text{IND-OCPA}}_{\mathcal{SE}_2,\mathcal{A}} = 0$.

We use mathematical induction on h to prove that the two probability distributions $(\mathcal{E}_2(x_0^1, k), ..., \mathcal{E}_2(x_0^h, k))$ and $(\mathcal{E}_2(x_1^1, k), ..., \mathcal{E}_2(x_1^1, k))$ are identical. For $h = 1$, it is necessary to show that the probability distribution $\mathcal{E}_2(x_0^1, k)$ equals to the probability distribution $\mathcal{E}_2(x_1^1, k)$. We denote $\Pi = \{(x, x') \mid 1 \leq x < x' \leq m\}$. Let I_j, $1 \leq j \leq m - 1$, be the probability distribution such that $\Pr[I_1 = i_1, ..., I_{m-1} = i_{m-1}] = \frac{1}{\prod_{j=0}^{m-2}(|\Pi|-j)}$ for $(i_1, ..., i_{m-1}) \in \Pi^{m-1}$ and $i_j \neq i_{j'}$ if $j \neq j'$. Let S_j, $1 \leq j \leq m - 1$, be the uniform distribution on Z_3. Then according to the construction of $\mathcal{E}_2$,

$$\mathcal{E}_2(x_0^1, k) = \{(I_j, S_j) \mid 1 \leq j \leq m - 1\} = \mathcal{E}_2(x_1^1, k).$$

We assume that the two probability distributions are identical for $h < h'$. For $h = h'$, we consider the following two conditional probability distributions

$$X = \mathcal{E}_2(x_0^{h'}, k) \mid \mathcal{E}_2(x_0^1, k) = y_1, ..., \mathcal{E}_2(x_0^{h'-1}, k) = y_{h'-1}$$

and

$$Y = \mathcal{E}_2(x_1^{h'}, k) \mid \mathcal{E}_2(x_1^1, k) = y_1, ..., \mathcal{E}_2(x_1^{h'-1}, k) = y_{h'-1},$$

where $y_u = \{(i_j^u, s_j^u) \in \Pi \times Z_3 \mid 1 \leq j \leq m - 1\}$, $1 \leq u \leq h' - 1$. $y_1, ..., y_{h'-1}$ will affect $\mathcal{E}_2(x_0^{h'}, k)$ $(\mathcal{E}_2(x_1^{h'}, k))$. First, for $1 \leq u \leq h' - 1$, there exists unique i_j^u (for some j) appears in $\mathcal{E}_2(x_0^{h'}, k)$ $(\mathcal{E}_2(x_1^{h'}, k))$ according to the construction of $\mathcal{E}_2$. On the other hand, there exists unique $i_{j'}^u$ (for some

j') appears in $y_{u'}$, $1 \le u' \ne u \le h'-1$; hence those $i^u_{j'}$ will not appear in $\mathcal{E}_2(x_0^{h'},k)$ ($\mathcal{E}_2(x_1^{h'},k)$). Thus, let

$$\bar{\Pi}_u = \{i^u_j \mid i^u_j \text{ appears in } y_u \text{ but does not appear in } y_{u'} \text{ for any } 1 \le u' \ne u \le h'-1\},$$

$1 \le u \le h'-1$. Then there exists $i^u_j \in \bar{\Pi}_u$ such that $(i^u_j, s^u_j - 1)$ appears in $\mathcal{E}_2(x_0^{h'},k)$ ($\mathcal{E}_2(x_1^{h'},k)$), $1 \le u \le h'-1$. Note that the elements of a set do not have orders, without loss of generality, for $1 \le u \le h'-1$, let (I_u, S_u) be the probability distribution such that $\Pr[(I_u,S_u) = (i^u_j, s^u_j - 1)] = \frac{1}{|\bar{\Pi}_u|}$. The rest probability distributions are similar to those for the situation of $h = 1$. Let

$$\bar{\Pi} = \{(x,x') \in \Pi \mid (x,x') \text{ does not appear in } y_u, 1 \le u \le h'-1\}.$$

For $h' \le u \le m-1$, let I_u be the probability distribution such that $\Pr[I_{h'} = i_{h'}, ..., I_{m-1} = i_{m-1}] = \frac{1}{\prod_{j=0}^{m-h'-1}(|\bar{\Pi}|-j)}$ for $(i_{h'}, ..., i_{m-1}) \in \bar{\Pi}^{m-h'}$ and $i_u \ne i_{u'}$ if $u \ne u'$. Let S_u, $h' \le u \le m-1$, be the uniform distribution on Z_3. Then

$$X = \{(I_u, S_u) \mid 1 \le u \le m-1\} = Y. \tag{1}$$

Consequently,

$$\begin{aligned}
&\Pr[\mathcal{E}_2(x_0^1,k) = y_1, ..., \mathcal{E}_2(x_0^{h'},k) = y_{h'}] \\
&= \Pr[\mathcal{E}_2(x_0^1,k) = y_1, ..., \mathcal{E}_2(x_0^{h'},k) = y_{h'}] \cdot \Pr[X = y_{h'}] \\
&= \Pr[\mathcal{E}_2(x_1^1,k) = y_1, ..., \mathcal{E}_2(x_1^{h'},k) = y_{h'}] \cdot \Pr[Y = y_{h'}] \qquad \text{(induction hypothesis and (1))} \\
&= \Pr[\mathcal{E}_2(x_1^1,k) = y_1, ..., \mathcal{E}_2(x_1^{h'},k) = y_{h'}].
\end{aligned}$$

Hence it implies that the two probability distributions are identical for $h = h'$, which completes induction. □

Remark 3. In order to improve the efficiency of $\mathcal{SE}_2$, π and π^{-1} can be substituted with deterministic symmetric-key encryption and decryption algorithms, and $r_{xx'}$ can be generated by a pseudorandom number generator. It is obvious that the improved scheme remains secure under IND-OCPA. □

5 IND-OLCPA

Now we consider security notion for OPE schemes in superpolynomial-sized domains. According to the big jump attack, if m is a superpolynomial of λ, then the adversary $\mathcal{A}_{BJ}$ can have $\mathbf{Adv}^{\text{IND-OCPA}}_{\mathcal{SE},\mathcal{A}_{BJ}} \ge 1 - \nu(\lambda)$ by using three oracle queries. From this we can conclude that IND-OCPA is too strong a security notion for OPE schemes in the superpolynomial-sized domain. Thus, we further weaken IND-OCPA and define the security notion IND-OLCPA (indistinguishability under ordered and local chosen-plaintext attack), where the range of the oracle queries is bounded by a polynomial of λ (to prevent the adversary from launching the big jump attack). The definition of IND-OLCPA is given as follows.

Definition 5.1 (IND-OLCPA)**.** The security notion IND-OLCPA has the same definition as that of IND-CPA except that the adversary is restricted so that it can only query $\{(x_0^u, x_1^u)\}_{u=1}^h$ where

$$x_0^u < x_0^v \Leftrightarrow x_1^u < x_1^v \tag{2}$$

for $1 \leq u, v \leq h$, and there exists a polynomial g_1 such that

$$|x_i^u - x_j^v| \leq g_1(\lambda) \tag{3}$$

for $1 \leq u, v \leq h$ and $0 \leq i, j \leq 1$. □

We design the following attack (and call it the small jump attack) against OPE schemes under IND-OLCPA. Similar to the big jump attack, the small jump attack also decides whether the ciphertexts are encrypted from the left or the right plaintexts based on the differences in distances between the ciphertexts.

Definition 5.2 (Small jump attack)**.** Consider the following PPT adversary $\mathcal{A}_{SJ}$ with three oracle queries in the experiment of security notion IND-OLCPA.

Adversary $\mathcal{A}_{SJ}{}^{\mathcal{E}_k(\mathcal{LR}(\cdot,\cdot,b))}$

$x \xleftarrow{\$} \{1, ..., m-3\}$
$y_1 \longleftarrow \mathcal{E}_k(\mathcal{LR}(x, x, b))$
$y_2 \longleftarrow \mathcal{E}_k(\mathcal{LR}(x+1, x+2, b))$
$y_3 \longleftarrow \mathcal{E}_k(\mathcal{LR}(x+3, x+3, b))$
Return 1 if $y_3 - y_2 < y_2 - y_1$; else return 0 □

In the small jump attack given above, the left plaintexts are x, x+1, and x+3, and the corresponding right plaintexts are x, $x+2$, and $x+3$, where x is randomly selected from $\{1, ..., m-3\}$. The following lemma show that the small jump attack can distinguish these two cases with non-negligible probability.

Lemma 5.3. *There is no efficient OPE scheme that is secure under IND-OLCPA (because of $\mathcal{A}_{SJ}$) if m is a superpolynomial of λ. Specifically, there exists a polynomial g such that* $\mathbf{Adv}^{IND\text{-}OLCPA}_{\mathcal{SE},\mathcal{A}_{SJ}} \geq \frac{1}{g(\lambda)\cdot g_1(\lambda)}$.

Proof. Let $d_i = \mathcal{E}(i+1, k) - \mathcal{E}(i, k)$ be the distance of the two ciphertexts, $1 \leq i < m$. Suppose that the adversary selects $x = i$ in the small jump attack. Then $y_3 - y_2 = d_{i+1} + d_{i+2}$ and $y_2 - y_1 = d_i$ if $b = 0$; $y_3 - y_2 = d_{i+2}$ and $y_2 - y_1 = d_i + d_{i+1}$ if $b = 1$. Therefore adversary $\mathcal{A}_{SJ}$ returns correct b if the following condition holds.

$$d_i + d_{i+1} > d_{i+2} \qquad \text{and} \qquad d_i < d_{i+1} + d_{i+2} \tag{4}$$

Consequently, adversary $\mathcal{A}_{SJ}$ may return incorrect b if either of the following two conditions (called small jump and small reverse-jump) holds.

$$d_i + d_{i+1} \leq d_{i+2} \tag{5}$$

$$d_i \geq d_{i+1} + d_{i+2} \tag{6}$$

Note that condition (5) implies that the distance series increases faster than Fibonacci numbers, and condition (6) implies that the reversed distance series increases faster than Fibonacci numbers. Since the formula of Fibonacci Numbers is

$$\frac{1}{\sqrt{5}}\Big[\Big(\frac{1+\sqrt{5}}{2}\Big)^i - \Big(\frac{1-\sqrt{5}}{2}\Big)^i\Big],$$

and $\log d_i$ must be bounded by a polynomial, it implies that condition (5) (resp. condition (6)) cannot consecutively happen superpolynomial times. Moreover, condition (6) cannot happen consecutively after condition (5). Otherwise

$$\begin{aligned} d_i + d_{i+1} \leq d_{i+2} \text{ and } d_{i+1} \geq d_{i+2} + d_{i+3} &\Rightarrow d_i + d_{i+1} + d_{i+2} + d_{i+3} \leq d_{i+2} + d_{i+1} \\ &\Rightarrow d_i + d_{i+3} \leq 0, \end{aligned}$$

which causes contradiction.

Consider $\{(d_i, d_{i+1}, d_{i+2})\}_{i=1}^{m-3}$. Suppose that (d_i, d_{i+1}, d_{i+2}) satisfies condition (5) or condition (6), and $m-3-i$ is a superpolynomial. Since condition (5) (resp. condition (6)) cannot consecutively happen superpolynomial times and condition (6) cannot happen consecutively after condition (5), there must exist polynomial g_i such that $(d_{i+g_i}, d_{i+1+g_i}, d_{i+2+g_i})$ satisfies condition (4). Hence the points in the set

$$\{i \mid (d_i, d_{i+1}, d_{i+2}) \text{ satisfy condition (4)}\}$$

partition $[m]$ into polynomial-sized segments. Let $g \cdot g_1$ be the maximum polynomial. Then there are at least $\frac{m-3}{g \cdot g_1}$ many i's such that (d_i, d_{i+1}, d_{i+2}) satisfies condition (4). Since adversary $\mathcal{A}_{SJ}$ returns correct b if it selects $x = i$ and (d_i, d_{i+1}, d_{i+2}) satisfies condition (4),

$$\mathbf{Adv}^{\text{IND-OLCPA}}_{\mathcal{SE}, \mathcal{A}_{SJ}} \geq \frac{1}{m-3} \cdot \frac{m-3}{g \cdot g_1} = \frac{1}{g \cdot g_1}.$$

□

Proposition 5.4. *If the adversary repeats the small jump attack, then the lower bound on the advantage of the adversary will become* $\frac{1}{g}$.

Proof. Since the range of plaintexts in the oracle queries is bounded by g_1, the probability for some i in the set $\{i \mid (d_i, d_{i+1}, d_{i+2})$ satisfy condition (4)$\}$ in the proof of Lemma 5.3 will fall into the range is at least $\frac{g_1}{m-3} \cdot \frac{m-3}{g \cdot g_1} = \frac{1}{g}$. Therefore the lower bound on the advantage of the adversary will increase to $\frac{1}{g}$. □

6 Ideal OPE and GOPE in the Superpolynomial-sized Domain

According to the same adversaries $\mathcal{A}_1$ in Proposition 4.2 and $\mathcal{A}_2$ in Corollary 4.3, the ideal OPE object $\mathcal{SE}^*$ does not achieve the lower bound on the advantage of the adversary $\frac{1}{g}$ (Proposition 5.4) under IND-OLCPA in superpolynomial-sized domains. Next We design a GOPE scheme $\mathcal{SE}_3 = (\mathcal{K}_3, \mathcal{E}_3, \mathcal{D}_3, \mathcal{C}_3)$ in the superpolynomial-sized domain, and prove that it achieves that lower bound. $\mathcal{SE}_3$ is constructed based on two building blocks: $\mathcal{SE}_4$ and $\mathcal{SE}_5$. $\mathcal{SE}_4$ is adapted from $\mathcal{SE}_1$ and it is secure under IND-OLCPA; but it can only support "local" comparisons (i.e. comparisons for ciphertexts whose plaintexts are closeby). The ciphertexts of $\mathcal{SE}_5$ have proper remote order to support "remote" comparisons (i.e. comparisons for ciphertexts whose plaintexts are far apart).

First we design $\mathcal{SE}_4 = (\mathcal{K}_4, \mathcal{E}_4, \mathcal{D}_4, \mathcal{C}_4)$. Let g_2 denote a polynomial. In $\mathcal{SE}_4$ the ciphertext of x can be compared with $g_2(\lambda) - 1$ (instead of $m-1$) other ciphertexts whose plaintexts are close to x.

- $\mathcal{K}_4$: Given the domain size m, it randomly picks a permutation π of the set $\{(x, x') \mid 1 \leq x < x' \leq m\}$, and randomly generates $r_{xx'} \in Z_3$ for $1 \leq x < x' \leq m$. It returns $\{\pi, r_{xx'} \mid 1 \leq x < x' \leq m\}$;

- $\mathcal{E}_4$: For plaintext x, it returns the ciphertext

$y = \{(\pi(x', x), r_{x'x}) \mid x' < x\} \cup \{(\pi(x, x'), 1 + r_{xx'}) \mid x < x' \leq g_2(\lambda)\}$ if $x \leq \frac{g_2(\lambda)}{2}$;

$y = \{(\pi(x', x), r_{x'x}) \mid x - \frac{g_2(\lambda)}{2} < x' < x\} \cup \{(\pi(x, x'), 1 + r_{xx'}) \mid x < x' \leq x + \frac{g_2(\lambda)}{2}\}$ if $\frac{g_2(\lambda)}{2} < x < m - \frac{g_2(\lambda)}{2}$;

$y = \{(\pi(x', x), r_{x'x}) \mid m - g_2(\lambda) \leq x' < x\} \cup \{(\pi(x, x'), 1 + r_{xx'}) \mid x < x'\}$ if $x \geq m - \frac{g_2(\lambda)}{2}$;

- $\mathcal{D}_4$: For ciphertext y, it retrieves (any) two elements (i, s) and (i', s') from the set y, and returns plaintext x which appears in both $\pi^{-1}(i)$ and $\pi^{-1}(i')$;

- $\mathcal{C}_4$: For ciphertexts y and y', if $y = y'$, it returns $=$. Otherwise, it retrieves (i, s) from the set y and (i, s') from the set y'. If $s - s' = 1$, it returns $<$. If $s - s' = 2$, it returns $>$.

The correctness, security, and efficiency of $\mathcal{SE}_4$ are presented in Lemmas 6.1 and 6.2.

Lemma 6.1. *The decryption of $\mathcal{SE}_4$ is correct. Also for plaintexts $x_1, x_2 \in [m]$, the comparison of $\mathcal{E}_4(x_1, k)$ and $\mathcal{E}_4(x_2, k)$ is correct if $|x_1 - x_2| \leq \frac{g_2(\lambda)-1}{2}$.*

Proof. The correctness of the decryption can be easily verified. Note that ciphertext of x_1 (resp. x_2) can compare with other $g_2(\lambda) - 1$ ciphertexts whose plaintexts are close to x_1 (resp. x_2). Hence $\mathcal{E}_4(x_1, k)$ and $\mathcal{E}_4(x_2, k)$ are comparable if $|x_1 - x_2| \leq \frac{g_2(\lambda)-1}{2}$. The comparison is correct referring to the proof of Lemma 4.5. □

Lemma 6.2. *Suppose that the range of oracle queries under IND-OLCPA is bounded by polynomial g_1. Then $\mathcal{SE}_4$ is secure under IND-OLCPA if $g_2 \geq 2g_1 + 1$. Furthermore, $\mathcal{SE}_4$ can be revised to achieve efficiency and remain secure under IND-OLCPA.*

Proof. The security proof is analogous to that of Theorem 4.6. It is worthy to note that the condition $g_2 \geq 2g_1 + 1$ will be used in the inductive step to guarantee two conditional probability distributions are identical. The detailed proof is presented as follows.

Assume that the adversary queries $\{(x_0^u, x_1^u)\}_{u=1}^h$ under IND-OLCPA. According to the restriction condition (2) under IND-OLCPA, $x_0^u = x_0^v \Leftrightarrow x_0^u = x_0^v$. Since it will not increase the advantage by querying two identical plaintexts pairs, it suffices to consider $x_0^1 < x_0^2 < ... < x_0^h$ and $x_1^1 < x_1^2 < ... < x_1^h$. Hence, the adversary views $(\mathcal{E}_4(x_0^1, k), ..., \mathcal{E}_4(x_0^h, k))$ for $b = 0$, and the adversary views $(\mathcal{E}_4(x_1^1, k), ..., \mathcal{E}_4(x_1^h, k))$ otherwise. It suffices to prove that the above two probability distributions are identical because it implies that $\mathbf{Adv}^{\text{IND-OLCPA}}_{\mathcal{SE}_4, \mathcal{A}} = 0$.

We use mathematical induction on h to prove that the two probability distributions $(\mathcal{E}_4(x_0^1, k), ..., \mathcal{E}_4(x_0^h, k))$ and $(\mathcal{E}_4(x_1^1, k), ..., \mathcal{E}_4(x_1^1, k))$ are identical. For $h = 1$, it is necessary to show that the probability distribution $\mathcal{E}_4(x_0^1, k)$ equals to the probability distribution $\mathcal{E}_4(x_1^1, k)$. We denote $\Pi = \{(x, x') \mid 1 \leq x < x' \leq m\}$. Let I_j, $1 \leq j \leq g_2(\lambda) - 1$, be the probability distribution such that $\Pr[I_1 = i_1, ..., I_{g_2(\lambda)-1} = i_{g_2(\lambda)-1}] = \frac{1}{\prod_{j=0}^{g_2(\lambda)-2}(|\Pi|-j)}$ for $(i_1, ..., i_{g_2(\lambda)-1}) \in \Pi^{g_2(\lambda)-1}$ and $i_j \neq i_{j'}$ if $j \neq j'$. Let S_j, $1 \leq j \leq g_2(\lambda) - 1$, be the uniform distribution on Z_3. Then according to the construction of $\mathcal{E}_4$,

$$\mathcal{E}_4(x_0^1, k) = \{(I_j, S_j) \mid 1 \leq j \leq g_2(\lambda) - 1\} = \mathcal{E}_4(x_1^1, k).$$

We assume that the two probability distributions are identical for $h < h'$. For $h = h'$, we consider

the following two conditional probability distributions

$$X = \mathcal{E}_4(x_0^{h'}, k) \mid \mathcal{E}_4(x_0^1, k) = y_1, ..., \mathcal{E}_4(x_0^{h'-1}, k) = y_{h'-1}$$

and

$$Y = \mathcal{E}_4(x_1^{h'}, k) \mid \mathcal{E}_4(x_1^1, k) = y_1, ..., \mathcal{E}_4(x_1^{h'-1}, k) = y_{h'-1},$$

where $y_u = \{(i_j^u, s_j^u) \in \Pi \times Z_3 \mid 1 \leq j \leq m-1\}$, $1 \leq u \leq h'-1$. For oracle queries x_i^u and x_j^v, since $|x_i^u - x_j^v| \leq g_1(\lambda) \leq \frac{g_2(\lambda)-1}{2}$, they are comparable according to Lemma 6.1. So $y_1, ..., y_{h'-1}$ will affect $\mathcal{E}_4(x_0^{h'}, k)$ ($\mathcal{E}_4(x_1^{h'}, k)$). First, for $1 \leq u \leq h'-1$, there exists unique i_j^u (for some j) appears in $\mathcal{E}_4(x_0^{h'}, k)$ ($\mathcal{E}_4(x_1^{h'}, k)$). On the other hand, there exists unique $i_{j'}^u$ (for some j') appears in $y_{u'}$, $1 \leq u' \neq u \leq h'-1$; hence those $i_{j'}^u$ will not appear in $\mathcal{E}_4(x_0^{h'}, k)$ ($\mathcal{E}_4(x_1^{h'}, k)$). Thus, let

$$\bar{\Pi}_u = \{i_j^u \mid i_j^u \text{ appears in } y_u \text{ but does not appear in } y_{u'} \text{ for any } 1 \leq u' \neq u \leq h'-1\},$$

$1 \leq u \leq h'-1$. Then there exists $i_j^u \in \bar{\Pi}_u$ such that $(i_j^u, s_j^u - 1)$ appears in $\mathcal{E}_4(x_0^{h'}, k)$ ($\mathcal{E}_4(x_1^{h'}, k)$), $1 \leq u \leq h'-1$. Note that the elements of a set do not have orders, without loss of generality, for $1 \leq u \leq h'-1$, let (I_u, S_u) be the probability distribution such that $\Pr[(I_u, S_u) = (i_j^u, s_j^u - 1)] = \frac{1}{|\bar{\Pi}_u|}$. The rest probability distributions are similar to those for the situation of $h = 1$. Let

$$\bar{\Pi} = \{(x, x') \in \Pi \mid (x, x') \text{ does not appear in } y_u, 1 \leq u \leq h'-1\}.$$

For $h' \leq u \leq g_2(\lambda) - 1$, let I_u be the probability distribution such that $\Pr[I_{h'} = i_{h'}, ..., I_{g_2(\lambda)-1} = i_{g_2(\lambda)-1}] = \frac{1}{\prod_{j=0}^{g_2(\lambda)-h'-1}(|\bar{\Pi}|-j)}$ for $(i_{h'}, ..., i_{g_2(\lambda)-1}) \in \bar{\Pi}^{g_2(\lambda)-h'}$ and $i_u \neq i_{u'}$ if $u \neq u'$. Let S_u, $h' \leq u \leq g_2(\lambda) - 1$, be the uniform distribution on Z_3. Then

$$X = \{(I_u, S_u) \mid 1 \leq u \leq g_2(\lambda) - 1\} = Y. \tag{7}$$

Consequently,

$$\begin{aligned}
&\Pr[\mathcal{E}_4(x_0^1, k) = y_1, ..., \mathcal{E}_4(x_0^{h'}, k) = y_{h'}] \\
&= \Pr[\mathcal{E}_4(x_0^1, k) = y_1, ..., \mathcal{E}_4(x_0^{h'}, k) = y_{h'}] \cdot \Pr[X = y_{h'}] \\
&= \Pr[\mathcal{E}_4(x_1^1, k) = y_1, ..., \mathcal{E}_4(x_1^{h'}, k) = y_{h'}] \cdot \Pr[Y = y_{h'}] \qquad \text{(induction hypothesis and (7))} \\
&= \Pr[\mathcal{E}_4(x_1^1, k) = y_1, ..., \mathcal{E}_4(x_1^{h'}, k) = y_{h'}].
\end{aligned}$$

Hence it implies that the two probability distributions are identical for $h = h'$, which completes induction.

To achieve efficiency of $\mathcal{SE}_4$, π and π^{-1} can be substituted with deterministic symmetric-key encryption and decryption algorithms, and $r_{xx'}$ can be generated by a pseudorandom number generator. It is obvious that the revision is efficient and remains secure under IND-OLCPA. □

Note that the original $\mathcal{SE}_4$ is given because it is easier to understand its GOPE construction. It is revised to achieve better efficiency. For convenience, from this point onwards, $\mathcal{SE}_4$ refers to the revised version. Next we design $\mathcal{SE}_5 = (\mathcal{K}_5, \mathcal{E}_5, \mathcal{C}_5)$. Since $\mathcal{SE}_4$ supports decryption and "local" comparisons, $\mathcal{SE}_5$ does not need a decryption algorithm but should support "remote" comparisons. In

order to assure security, the ciphertexts should have small statistical distances if the corresponding plaintexts are close to each other. To achieve this, for $1 \leq i \leq l$, $\mathcal{SE}_5(i,k)$ are randomly selected from $[n']$, where n' and l are positive integers. Then the subsequent ciphertexts are gradually increased. The construction of $\mathcal{SE}_5$ is shown as follows.

- $\mathcal{K}_5$: It randomly selects $r_i \in [n']$ for $0 \leq i \leq l-1$ and returns $(r_0, ..., r_{l-1})$;

- $\mathcal{E}_5$: For plaintext $x \in [m]$, we compute a and b, $a \geq 0$ and $0 \leq b < l$, such that $x - 1 = a \cdot l + b$. $\mathcal{E}_5$ returns ciphertext $y = r_b + a$;

- $\mathcal{C}_5$: For ciphertexts y and y', if $y > y'$, it returns $>$; if $y < y'$, it returns $<$.

The correctness of $\mathcal{SE}_5$ is presented in Lemma 6.3.

Lemma 6.3. *For plaintexts $x_1, x_2 \in [m]$, the comparison of $\mathcal{E}_5(x_1,k)$ and $\mathcal{E}_5(x_2,k)$ is correct if $|x_1 - x_2| \geq n'l + l$.*

Proof. Without loss of generality, we assume that $x_1 < x_2$. Then $x_2 - x_1 \geq n'l + l$. Let $x_i - 1 = a_i \cdot l + b_i$ satisfying $a_i \geq 0$ and $0 \leq b_i < l$, $1 \leq i \leq 2$. Then

$$n'l + l \leq x_2 - x_1 = (a_2 - a_1) \cdot l + (b_2 - b_1) < (a_2 - a_1) \cdot l + l \Rightarrow a_2 - a_1 > n'.$$

Hence,

$$\mathcal{E}_5(x_1,k) = r_{b_1} + a_1 < r_{b_1} + (a_2 - n') = (r_{b_1} - n') + a_2 < r_{b_2} + a_2 = \mathcal{E}_5(x_2,k),$$

which implies that the comparison is correct. □

If the queries by the adversary against $\mathcal{SE}_5$ are in the interval $[cl+1, (c+1)l]$, for some $c \geq 0$, then the adversary cannot distinguish the corresponding ciphertexts because they are independent identical random variables generated by $\mathcal{E}_5$. If the queries involve plaintexts in two consecutive intervals $[cl+1, (c+1)l]$ and $[(c+1)l+1, (c+2)l]$, then the advantage of the adversary is not 0, but it can be controlled by l and n'. The security of $\mathcal{SE}_5$ is given in the following Lemma.

Lemma 6.4. *Suppose that the range of oracle queries under IND-OLCPA is bounded by polynomial g_1. For polynomial $g \geq 1$, $\mathbf{Adv}^{IND\text{-}OLCPA}_{\mathcal{SE}_5,\mathcal{A}} \leq \frac{1}{g(\lambda)}$ if $l > g_1(\lambda)$ and $n' \geq g(\lambda) \cdot g_1(\lambda)$.*

Proof. Assume that the adversary queries $\{(x_0^u, x_1^u)\}_{u=1}^h$ under IND-OLCPA. According to the restriction condition (2) under IND-OLCPA, $x_0^u = x_0^v \Leftrightarrow x_0^u = x_0^v$. Since it will not increase the advantage by querying two identical plaintexts pairs, it suffices to consider $x_0^1 < x_0^2 < ... < x_0^h$ and $x_1^1 < x_1^2 < ... < x_1^h$. Let $x_i^u - 1 = a_i^u \cdot l + b_i^u$ satisfying $a_i^u \geq 0$ and $0 \leq b_i^u < l$, then $\mathcal{E}_5(x_i^u, k) = r_{b_i^u} + a_i^u$, $1 \leq u \leq h$ and $0 \leq i \leq 1$. Hence, the adversary views $(r_{b_0^1} + a_0^1, ..., r_{b_0^h} + a_0^h)$ for $b = 0$, and the adversary views $(r_{b_1^1} + a_1^1, ..., r_{b_1^h} + a_1^h)$ otherwise. Let Δ be the statistical distance between $(r_{b_0^1} + a_0^1, ..., r_{b_0^h} + a_0^h)$ and $(r_{b_1^1} + a_1^1, ..., r_{b_1^h} + a_1^h)$. Since $\mathbf{Adv}^{\text{IND-OLCPA}}_{\mathcal{SE}_5,\mathcal{A}} \leq \Delta$, it suffices to prove that $\Delta \leq \frac{1}{g(\lambda)}$.

We study the properties of those probability distributions. Since

$$|(a_i^u - a_j^v) \cdot l + (b_i^u - b_j^v)| = |x_i^u - x_j^v| \leq g_1(\lambda) < l,$$

it implies that $|a_i^u - a_j^v| \leq 1$ and $b_i^u = b_j^v \Rightarrow a_i^u = a_j^v$, $1 \leq u, v \leq h$ and $0 \leq i, j \leq 1$. Furthermore $b_i^u = b_j^v \Rightarrow a_i^u = a_j^v$ and $x_0^u \neq x_0^v$ if $u \neq v$ implies that $b_0^u \neq b_0^v$ if $u \neq v$. Therefore $r_{b_0^1} + a_0^1, ..., r_{b_0^h} + a_0^h$

are independent uniform distributions on $[n'] + a_0^1, ..., [n'] + a_0^h$. Similarly, $r_{b_1^1} + a_1^1, ..., r_{b_1^h} + a_1^h$ are independent uniform distributions on $[n']+a_1^1, ..., [n']+a_1^h$. Hence Δ equals to the statistical distance between independent uniform distributions $X_1, ..., X_h$ on $[n'] + a_0^1, ..., [n'] + a_0^h$ and independent uniform distributions $Y_1, ..., Y_h$ on $[n'] + a_1^1, ..., [n'] + a_1^h$, i.e.

$$\Delta = \frac{1}{2} \sum_{w_u \in ([n']+a_0^u) \cup ([n']+a_1^u), 1 \le u \le h} \left| \Pr[(X_1, ..., X_h) = (w_1, ..., w_h)] - \Pr[(Y_1, ..., Y_h) = (w_1, ..., w_h)] \right|.$$

Since $|a_0^u - a_1^u| \le 1$ for $1 \le u \le h$, $\Delta \le \frac{h \cdot n'^{h-1} + h \cdot n'^{h-1}}{2n'^h} = \frac{h}{n'} \le \frac{g_1(\lambda)}{n'} \le \frac{1}{g(\lambda)}$. □

$\mathcal{SE}_3 = (\mathcal{K}_3, \mathcal{E}_3, \mathcal{D}_3, \mathcal{C}_3)$ is constructed by combining $\mathcal{SE}_4$ and $\mathcal{SE}_5$. In order to achieve full comparison capability, g_2, l, and n' are chosen to satisfy the condition $\frac{g_2-1}{2} \ge n'l + l$ (Lemmas 6.1 and 6.3). In order to achieve security, g_2, l, and n' are chosen to satisfy the conditions $g_2 \ge 2g_1 + 1$, $l > g_1$, and $n' \ge g \cdot g_1$ (Lemmas 6.2 and 6.4). We can solve these inequalities, and get $l > g_1$, $n' \ge g \cdot g_1$, and $g_2 \ge \max\{2(n'l + l) + 1, 2g_1 + 1\} = 2(n'l + l) + 1$. Specifically, we can set $l = g_1 + 1$, $n' = g \cdot g_1$, and $g_2 = 2(n'l + l) + 1$. $\mathcal{SE}_3$ encrypts plaintext x into $(\mathcal{E}_4(x, k), \mathcal{E}_5(x, k))$. Since g and g_1 are polynomials, $\mathcal{SE}_3$ is an efficient encryption scheme. Given two ciphertexts $(\mathcal{E}_4(x_1, k), \mathcal{E}_5(x_1, k))$ and $(\mathcal{E}_4(x_2, k), \mathcal{E}_5(x_2, k))$, $\mathcal{SE}_3$ first compares $\mathcal{E}_4(x_1, k)$ and $\mathcal{E}_4(x_2, k)$ by using $\mathcal{C}_4$; if it fails, $\mathcal{SE}_3$ then compares $\mathcal{E}_5(x_1, k)$ and $\mathcal{E}_5(x_2, k)$ by using $\mathcal{C}_5$. Also, $\mathcal{E}_4(x, k)$ can be decrypted by $\mathcal{D}_4$. We summarize these results in the following theorem.

Theorem 6.5. *Suppose that the range of oracle queries under IND-OLCPA is bounded by polynomial g_1. For any polynomial $g \ge 1$, there exists an efficient GOPE scheme $\mathcal{SE}_3$ such that* $\mathbf{Adv}^{IND\text{-}OLCPA}_{\mathcal{SE}_3,\mathcal{A}} \le \frac{1}{g(\lambda)}$.

Proof. The proof is based on Lemmas 6.1 6.2 6.3 and 6.4. □

7 Conclusion and Future Research

In this paper, we first prove that the ideal OPE object may not be the most secure OPE. To this end, we construct a "real" OPE scheme for the specific plaintext domain and ciphertext range and prove that it is secure under IND-OCPA, and prove that the ideal OPE object for the specific plaintext domain and ciphertext range is not secure under IND-OCPA. The results indicates that we need to use the "ideal" encryption object more cautiously in the security analysis of OPE.

We then study the security of OPE and GOPE schemes under various security notions. First we consider polynomial-sized domains and show that the ideal OPE object is not secure under IND-OCPA even in polynomial-sized domains. Also, we construct a generalized OPE scheme which is secure under IND-OCPA in polynomial-sized domains. Then, we weaken the security notion from IND-OCPA to IND-OLCPA to prevent the big jump attack in superpolynomial-sized domains. Correspondingly, we design a small jump attack under IND-OLCPA and derive the lower bound on the advantage of an adversary against OPE schemes under IND-OLCPA. Based on the small jump attack, it is shown that the ideal OPE object does not achieve the lower bound. So, we construct another generalized OPE schemes to achieve the lower bound. The results are summarized in Tables 1, where $\mathcal{SE}$ represents any OPE scheme, $\mathcal{SE}^*$ represents the ideal OPE object, $\mathcal{SE}_2$ represents the GOPE scheme constructed in Subsection 4.2, $\mathcal{SE}_3$ represents the GOPE scheme constructed in

Section 6; $\mathcal{A}$ represents any PPT adversary, $\mathcal{A}_{BJ}$ represents the big jump attack, $\mathcal{A}_{RSJ}$ represents repeated small jump attacks, $\mathcal{A}_1$ represents the attack against the ideal OPE object defined in Proposition 4.2, $\mathcal{A}_2$ represents the attack against the ideal OPE object defined in Corollary 4.3.

	OPE in the polynomial-sized domain	OPE in the superpolynomial-sized domain
IND-OCPA	$\mathbf{Adv}^{\text{IND-OCPA}}_{\mathcal{SE},\mathcal{A}_{BJ}} \geq 1 - \frac{2\log n}{m-1}$ ([4]) $\mathbf{Adv}^{\text{IND-OCPA}}_{\mathcal{SE}^*,\mathcal{A}_1} = \Omega(1)$ (Proposition 4.2) $\mathbf{Adv}^{\text{IND-OCPA}}_{\mathcal{SE}^*,\mathcal{A}_2} = \Omega(1)$ (Corollary 4.3)	$\mathbf{Adv}^{\text{IND-OCPA}}_{\mathcal{SE},\mathcal{A}_{BJ}} \geq 1 - \nu(\lambda)$ ([4])
IND-OLCPA		$\mathbf{Adv}^{\text{IND-OLCPA}}_{\mathcal{SE},\mathcal{A}_{RSJ}} \geq \frac{1}{g(\lambda)}$ (Proposition 5.4) $\mathbf{Adv}^{\text{IND-OLCPA}}_{\mathcal{SE}^*,\mathcal{A}_1} = \Omega(1)$ $\mathbf{Adv}^{\text{IND-OLCPA}}_{\mathcal{SE}^*,\mathcal{A}_2} = \Omega(1)$

	GOPE in the polynomial-sized domain	GOPE in the superpolynomial-sized domain
IND-OCPA	$\mathbf{Adv}^{\text{IND-OCPA}}_{\mathcal{SE}_2,\mathcal{A}} = 0$ (Theorem 4.6)	
IND-OLCPA		$\mathbf{Adv}^{\text{IND-OLCPA}}_{\mathcal{SE}_3,\mathcal{A}} \leq \frac{1}{g(\lambda)}$ (Theorem 6.5)

Table 1: Security of OPE and GOPE Schemes Under Various Security Notions

There are many unsolved problems that require further research. For instance, we prove that the ideal OPE object is not the most secure OPE for the specific plaintext domain $[m] = \{1, 2\}$. Unfortunately the construction of the "real" OPE scheme cannot be generalized to other plaintext domains. Thus, it is still unknown whether the ideal OPE is the most secure OPE for plaintext domains with size greater than 2. The more important problem is: if the ideal OPE object is not the secure OPE for plaintext domains with size greater than 2, then what is the most secure OPE. For GOPE, we construct GOPE schemes $\mathcal{SE}_2$ and $\mathcal{SE}_3$ to achieve lower bounds on the advantage of an adversary against OPE schemes under IND-OCPA in the polynomial-sized domain and under IND-OLCPA in the superpolynomial-sized domain. A natural question is: can GOPE schemes do better, i.e. can we construct GOPE schemes to exceed those bounds? We plan to conduct research to investigate these problems. Also, we plan to investigate other novel approaches to improve the design of OPE schemes.

References

[1] R. Agrawal, J. Kiernan, R. Stikant, and Y. Xu, *Order-preserving encryption for numeric data*, *ACM SIGMOD International Conference on Management of Data*, pp. 563-574, 2004.

[2] G. Amanatidis, A. Boldyreva, and A. O'Neill, *Provably-secure schemes for basic query support in outsourced databases*, *Working Conference on Data and Applications Security*, pp. 14-30, 2007.

[3] G. Bebek. *Anti-tamper database research: Inference control techniques*, *Technical Report EECS 433 Final Report*, Case Western Reserve University, 2002.

[4] A. Boldyreva, N. Chenette, Y. Lee, and A. O'Neill, *Order-preserving symmetric encryption*, *Eurocrypt*, pp. 224-241, 2009.

[5] A. Boldyreva, N. Chenette, and A. O'Neill, *Order-Preserving Encryption Revisited: Improved Security Analysis and Alternative Solutions*, *Crypto11*, 2011.

[6] D. Boneh and B. Waters, *Conjunctive, subset, and range queries on encrypted data*, *TCC*, pp. 535-554, 2007.

[7] H. Hacigumus, B.R. Iyer, C. Li, and S. Mehrotra, *Executing SQL over encrypted data in the database-service-provider model*, *ACM SIGMOD Conference on Management of Data*, 2002.

[8] M. Halloush and M. Sharif, *Global heuristic search on encrypted data (GHSED)*, *International Journal of Computer Science Issues (IJCSI)*, 1:13-17, 2009.

[9] J. Li and E.R. Omiecinski, *Efficiency and security trade-off in supporting range queries on encrypted databases*, *Data and Applications Security*, pp. 69-83, 2005.

[10] G. Ozsoyoglu, D.A. Singer, and S.S. Chung, *Anti-tamper databases: Querying encrypted databases*, *Conference on Database and Applications Security*, 2003.

[11] E. Shi, J. Bethencourt, T-H.H. Chan, D. Song, and A. Perrig, *Multi-dimensional range query over encrypted data*, *Symposium on Security and Privacy*, pp. 350-364, 2007.

[12] D.X. Song, D. Wagner, and A. Perrig, *Practical techniques for searches on encrypted data*, *IEEE Symposium on Security and Privacy*, pp. 44-55, 2000.