Order-Preserving Encryption Secure Beyond One-Wayness

Tal Malkin[1], Isamu Teranishi[2], and Moti Yung[1,3]

[1] Columbia University
[2] NEC Japan
[3] Google Inc.
tal@cs.columbia.edu, teranisi@ah.jp.nec.com, moti@cs.columbia.edu

Abstract. Semantic-security of individual bits under a ciphertext are fundamental notion in modern cryptography. In this work we present the first results about this fundamental problem for Order-Preserving Encryption (OPE): "what plaintext information can be semantically hidden by OPE encryptions?" While OPE has gained much attention in recent years due to its usefulness in secure databases, any partial-plaintext indistinguishability (semantic security) result for it was open. Here, we propose a new indistinguishability-based security notion for OPE, which can ensure *secrecy of lower bits of a plaintext* (under essentially a random ciphertext probing setting). We then propose a new scheme satisfying this security notion (while earlier schemes do not satisfy it!). We note that the known security notions tell us nothing about the above partial- plaintext indistinguishability because they are limited to being one-way-based. In addition, we show that our security notion with specific parameters implies the known security notion called WOW, and further, our scheme achieves WOW with better parameters than earlier schemes.

Keywords: Order-preserving encryption, secure encryption, security notions, indistinguishability, one-way, foundations.

1 Introduction

1.1 Backgrounds

Securing data of outsourced databases while keeping their functions is critical to the emerging *"cloud computing"* and *"database as a service"* paradigms, describing a setting where users' data is kept in remote servers and is employed when needed. These databases often contain sensitive financial, medical, legal, scientific, and intellectual property data, and security may even be required by law, e.g., HIPPA [26].

The prototypical new problem in this modern setting is that the database manager may be untrusted, in spite of the fact that it holds users' data. This is a fundamentally different concern than merely the secure end-to-end communication of traditional cryptography. Securing cloud database with untrusted cloud servers, therefore, needs to hide information from the database manager itself, and has resulted in new research areas, such as: verifiable computations [13,2], proofs of data possessions [3,28,5], searchable encryptions [43,20,23,21,42], deterministic encryptions [8,10,19], proxy re-encryptions [14,4], prefix-preserving encryptions [50,7], and PIRs [30,22,35].

Order-Preserving Encryption (OPE): This is one of the most promising new primitives in the area of encrypted database processing [1,36,6,16,18,51]. It is a symmetric encryption over the integers such that ciphertexts preserve the numerical orders of the corresponding plaintexts. That is, $\forall m, m' \{m < m' \Rightarrow \mathsf{Enc}_K(m) < \mathsf{Enc}_K(m')\}$. OPE was originally studied heuristically in the database community by Agrawal, Kiernan, Ramakrishnan, Srikant and Xu [1], and seemed like a clever heuristics. Remarkably, its careful foundational study was initiated with surprising formal cryptographic models and proofs by Boldyreva, Chenette, Lee, and O'Neill [16,18]. Overall, it has received much recent attention in the cryptographic community [16,18,51], in the database community [1,36,6], as well as in other applied areas such as sensor networks [46], multimedia content protections [25,31], and so on [50,42,45].

OPE is attractive since it allows one to simultaniously perform very efficiently over encrypted data numerous fundamental database operations: sorting, simple matching (i.e., finding m in a database), range queries (i.e., finding all messages m within a given range $\{i, \ldots, j\}$), and SQL operations [1,40,41,44]. In cotrast, each of the other primitives mentioned above enables at most one of these operations. Furthermore, OPE is more efficient than these other primitives. For instance, the simple matching operation realized by OPE only requires logarithmic time in the database size [1], while the same operation realized by, say, searchable encryption [21,42], needs linear time in the size, which is too costly for a database containing a few millions data items [27].

1.2 Security of OPE

Despite its importance, security of OPE is far from being understood at this time. Even the most fundamental problem: "what plaintext information can be semantically hidden" is open because security of OPE is quite different from that of other primitives. Indeed, a naturally defined indistinguishability notion for OPE, *indistinguishability under ordered CPA attack (IND-O-CPA)* [16], was not only unachievable but it was shown that **any** OPE under this notion is broken with **overwhelming** probability if the OPE scheme has a super-polynomial size message space. In other words, OPE cannot ensure this naturally defined indistinguishability at all or has to have a very small message space (to a point of losing its utility).

OPE is an inherently "leaky" method: The reason that an OPE scheme cannot achieve a natural indistinguishable notion is that an OPE scheme has to reveal something about plaintexts other than their order: i.e., information about the distance between the two plaintexts. By definition (as stated above) an OPE scheme's encryption function Enc_K has to satisfy the monotone increasing property, $m_0 < m_1 \Rightarrow \mathsf{Enc}_K(m_0) < \mathsf{Enc}_K(m_1)$. Hence, the difference $\mathsf{Enc}_K(m_1) - \mathsf{Enc}_K(m_0)$ of two ciphertexts has to become noticeably large if the difference $m_1 - m_0$ of the corresponding plaintexts becomes large. The negative result of [16] mentioned above is, in fact, proved using an attack based on this observation.

To date, no one can tell what exactly OPE has to leak and what it can protect. Our motivation is the fact that the existing security notions (which we list next) are not really helpful in understanding this simple question.

IND-O-CPA [16]: This notion is defined as follows: an adversary selects plaintexts by herself adaptively (her queries have to satisfy some order-preserving property), she gets their corresponding ciphertexts, and then guesses a secret bit of the challenger. However, as we stated above, this notion is not achievable for schemes with a super-polynomial size message space [16].

POPF-CCA [16]: Intuitively, this indistinguishability-based notion says that outputs of an encryption and decryption oracles are indistinguishable from those of an order-preserving random oracle and its inverse, even when an adversary queries the oracle in a CCA fashion. However, this notion does not tell us what information of the plaintext is hidden and what is leaked, as was already pointed out by [18].

$(r, q+1)$-WOW [17] (Window One-Wayness): This is defined as follows[4]: An adversary A gets $q+1$ encryptions $C_*, C_1, \ldots, C_q$ of uniformly randomly selected unknown messages and outputs an interval I such that the length of I is not more than r. A is considered to win iff $\mathsf{Dec}_K(C^*) \in I$.

[4] Boldyreva et.al. [17] gave two definitions of the window one-wayness notion in Section 3 and Appendix B of [17] and the first one can be reduced to the second one and vise versa (due to Lemma B.1 of [17]). We adopt the definition given in Appendix B of [17], which they called *(specified) window one-wayness*, because it is simpler.

This notion is important for OPE since it captures the following natural database setting: Randomly selected $q+1$ elements are stored in a database system in their encrypted form and an adversary A who wants to know one of them breaches the database and gets all ciphertexts in it. On the other hand, this notion does not ensure anything about secrecy of internal plaintext information, since it is "one-way-based" in nature.

$(r, q+1)$-WDOW [17]: $(r, q+1)$-WDOW is another one-way-based notion defined in [17]. But since it is one-way-based, it also does not tell us what information about the plaintexts is hidden.

1.3 Our Indistinguishability Notion — $(\mathcal{X}, \theta, q)$-Indistinguishability

This paper presents the first attempt to give a new perspective to the fundamental open problem: "go beyond one-wayness security and investigate what internal plaintext information OPE can hide." Y Here we propose the first (achievable) indistinguishability notion for OPE regarding partial plaintext information hiding. More specifically: our notion can assure *secrecy of lower bits of a plaintext* in some natural setting.

Our Approach to Defining Indistinguishability: Our starting points are two known security notions for OPE, namely, $(r, q+1)$-WOW [18] and IND-O-CPA [16]. The former is achievable (by the scheme of [16]), is useful in certain database settings, but it is inherently one-way-based. The latter is indistinguishability-based, but is too strong and hence cannot be achieved by any OPE scheme with a super-polynomial size message space.

Our approach defines a new security notion, by considering (in some sense) a "hybrid" of the above two approaches: Our adversary plays an indistinguishability game, although she is in the setting of WOW: That is, she is given $\mathsf{Enc}_K(m_b^*)$, $\mathsf{Enc}_K(m_1)$, ..., $\mathsf{Enc}_K(m_q)$, where messages $m_1, \ldots, m_q$ are selected randomly by the challenger (not the adversary!).

Then, our security notion is of the same natural setting as that of WOW, but is indistinguishability-based and is weaker than IND-O-CPA. This is our initial goal (since we want a realizable notion). Then, we improve the above notion so as to make our definition stronger and to avoid the known attack related to the inherent leaky nature of OPE (reviewed in of Section 1.2). This will be desribed in more details in the sequel. In turn, we will show that our indistinguishability notion can ensure secrecy of the plaintext's lower order bits.

Definition: Our indistinguishability notion, $(\mathcal{X}, \theta, q)$-*indistinguishability*, is as follows: Let $\mathcal{X} = (\mathcal{X}_i)$ be a tuple of message distributions. Consider two polynomial time machines A and Mg, named adversary and message generator and consider the following game: first: $(m_0^*, m_1^*, \mathsf{info}) \leftarrow \mathsf{Mg}$, and then: $m_1 \xleftarrow{\$} \mathcal{X}_1$, ..., and $m_q \xleftarrow{\$} \mathcal{X}_q$. Then a bit b is chosen at random, and A wins iff $\mathsf{A}(\mathsf{Enc}_K(m_b^*), (\mathsf{Enc}_K(m_i))_{i=1..q}, \mathsf{info})$ succeeds in outputting b. Here info is a bit string which intuitively contains some information about messages m_0^* and m_1^*. We say that an OPE scheme is $(\mathcal{X}, \theta, q)$-*indistinguishable* if the advantage of A (guessing the bit beyond probability $1/2$) is negligible for any A and for any Mg whose output satisfies

$$|m_1^* - m_0^*| \leq \theta. \tag{1.1}$$

Note that, when $m_i \in [m_0^* .. m_1^*]$ holds for some i, an OPE scheme under the above notion is broken trivially by an adversary simply checking $\mathsf{Enc}_K(m_b^*) > \mathsf{Enc}_k(m_i)$. However, this is not big issue since we employ this security notion only when the probability that $m_i \in [m_0^* .. m_1^*]$ is negligible.

In the above definition, the restriction (1.1) is imposed in order to avoid the known attack [16] reviewed in Section 1.2. Recall that this known attack applies only when the distance between m_0^* and m_1^* is large. Hence, (1.1) prevents an adversary from executing this attack.

The above definition is improved in three ways. First, we allow an adversary to access the plaintexts $m_1, \ldots, m_q$, secondly, we consider the case where $m_1, \ldots, m_q$ are selected from any given distributions $\mathcal{X}_1, \ldots, \mathcal{X}_q$: these improvements make our security notion stronger and more general, and finally, we impose restriction (1.1) as mentioned before.

Low-order Bits Can be Hidden: Our security notion ensures the secrecy of the least significant $\lfloor \log_2 \theta \rfloor$ bits of a plaintext, due to the following: Let $L = \lfloor \log_2 \theta \rfloor$ and take any (maximal) interval I satisfying the following property: for any two elements of I, all bits of them *except the least significant L bits* are the same. That is, I can be written as

$$I = \{2^L u + x \mid x \in [0..2^L - 1]\}$$

for some u. By definition the length of I is not more than θ.

Next, recall that our security notion ensures the indistinguishability of ciphertexts of m_0^* and m_1^* for any m_0^* and m_1^* satisfying $|m_0^* - m_1^*| \leq \theta$. This fact, in particular, means that a ciphertext of any element m_0^* of I is indistinguishable from that of a uniformly random element m_1^* of I. Since the least significant L bits of uniformly random element m_1^* of I distribute uniformly at random on the L-bit space $[0..2^L - 1]$, the indistinguishability of m_0^* and m_1^* can ensure secrecy of the least significant L bits of m_0^*. Note that, by similar semantic-security-like arguments, we can also ensure the secrecy of other information of m_0^* as well: e.g., the least significant $\lfloor \log_2 \theta \rfloor$ bits of $m_0^* + s$ for any fixed s.

Our security notion does not ensure the secrecy of higher bits of the plaintext, and ,in fact, there is no known scheme which can ensure the secrecy of higher bits of the plaintext, since the scheme of [16] also reveals higher bits of it [18].

Application: Our $(\mathcal{X}, \theta, q)$-indistinguishability notion captures, in fact, the same natural setting as that of WOW mentioned before. That is, consider a database containing $q + 1$ data elements $m_*, m_1, \ldots, m_q$ in their encrypted forms, where these messages are distributed according to given distributions, and an adversary A who wants to know m_* breaches a database system and gets all ciphertexts in it. In this setting, our $(\mathcal{X}, \theta, q)$-indistinguishability notion can ensure a stronger fact that WOW cannot assure, namely: that A cannot learn the least significant $\lfloor \log_2 \theta \rfloor$ bits of m_* even when she has (partial or all) information about the other data elements $m_1, \ldots, m_q$. (We note that our notions as well as previous notions do not deal with guaranteeing anything about messages depending on the known plaintexts, such as $m_1 + 1$, (whether notions which can deal with such strong dependencies are possible is left open)).

1.4 Our Scheme and Its Security

The known scheme [16] is designed based on a remarkable idea but this scheme is irrelevant to our goal since we show that it cannot satisfy our security notion, for good parameters at least (Moreover, there is no evidence that this scheme satisfies our security notion even for bad parameters, since our goal is new). See Section 1.6 and Appendix 4.2 for the details.

Hence, we propose a new OPE scheme $\mathcal{E}^{\beta,t}$ parameterized by two values β and t (to be elaborated on in the sequel) and study its $(\mathcal{X}, \theta, q)$-indistinguishability. Our scheme is constructed based on a pseudo-random function PRF, and its security is shown by reduction to the function's security.

First, we show that our scheme $\mathcal{E}^{\beta,t}$ with $\beta = 1$ satisfies the following theorem in the very important case where $\mathcal{X}$ is the tuple of uniform distributions on the message space. Below, λ is a security parameter.

Theorem 1. *Let $\mathcal{U}^q = (\mathcal{U}, \ldots, \mathcal{U})$ be the tuple of q uniform distributions on the message space $[1..M]$ and $0 \leq t < 1$ be any constant. Then our proposed scheme $\mathcal{E}^{1,t}$ satisfies*

$$\forall \mathsf{Mg} \forall \mathsf{A} \exists \mathsf{B} \quad : \quad \mathsf{Adv.Exp}^{(\mathcal{U}^q, M^t, q)\text{-indis.}}_{\mathcal{E}^{1,t}}(\mathsf{Mg}, \mathsf{A}) \leq O\left(\frac{q}{M^{\frac{1-t}{3}}}\right) + \mathsf{Adv.Exp}_{\mathsf{PRF}}(\mathsf{B}) + \mathsf{neg}(\lambda). \tag{1.2}$$

Here $\mathsf{neg}()$ is some negligible function.

Next, we generalize Theorem 1 as follows. Below, $H_\infty(\mathcal{Y})$ denotes the *min-entropy* of $\mathcal{Y}$, that is, $H_\infty(\mathcal{Y}) = -\log_2 \max_x \Pr[m \xleftarrow{\$} \mathcal{Y} \ : \ m = x]$.

Theorem 2. *Let $\beta > 0$ be any constant and $\mathcal{X} = (\mathcal{X}_1, \ldots, \mathcal{X}_q)$ be any tuple of distributions on the message space $[1..M]$ such that $\mathcal{X}_1, \ldots, \mathcal{X}_q$ are independent from each other (and one can take a sample from $\mathcal{X}_i$ in time polynomial in λ). Suppose that $\mathcal{X} = (\mathcal{X}_1, \ldots, \mathcal{X}_q)$ satisfies*

$$\forall i \in [1..q] \quad : \quad H_\infty(\mathcal{X}_i) \geq \beta \log_2 M. \tag{1.3}$$

Then, for any constant $t < \beta$ our scheme $\mathcal{E}^{\beta,t}$ satisfies

$$\forall \mathsf{Mg} \forall \mathsf{A} \exists \mathsf{B} \quad : \quad \mathsf{Adv.Exp}^{(\mathcal{X}, M^t, q)\text{-indis.}}_{\mathcal{E}^{\beta,t}}(\mathsf{Mg}, \mathsf{A}) \leq O\left(\frac{q}{M^{\frac{\beta-t}{3}}}\right) + \mathsf{Adv.Exp}_{\mathsf{PRF}}(\mathsf{B}) + \mathsf{neg}(\lambda). \tag{1.4}$$

The right hand sides of these theorems become negligible under the condition that the message space size M is super-polynomial of λ: which is exactly the same condition assumed by Boldyreva et.al.[18] to get their results.

Theorems 1 and 2 ensure that we can show stronger security when t is closer to 1 and β respectively, though the advantage bound decrease is slower in this case. The values β and t affect neither the ciphertext length, nor the encryption, nor decryption costs. They only affect the order of the advantage bound and the strength of the security notion.

The value $\log_2 M$ in (1.3) represents the maximum of the min-entropy because any distribution $\mathcal{Y}$ on $[1..M]$ satisfies $H_\infty(\mathcal{Y}) \leq \log_2 M$. Hence, the value β in (1.3) represents a lower bound of the ratio of $H_\infty(\mathcal{X}_i)$ to the maximum $\log_2 M$. Note that the maximum $\log_2 M$ is achieved by the uniform distribution and we therefore can set β to 1 in this case. Hence, Theorem 1 is reduced to Theorem 2 in the case of the uniform distributions.

Any Fraction of Low-order Bits Are Hidden in the Uniform Distribution Case: Recall that the discussion of Section 1.3 shows that $(\mathcal{X}, M^t, q)$-indistinguishability implies secrecy of the least significant $\lfloor \log M^t \rfloor$ bits of a plaintext. Moreover, the maximum bit length of a message is $\lfloor \log_2 M \rfloor + 1$ since the message space is $[1..M]$. Hence, Theorem 1 and 2 show that our scheme with the above parameters can ensure secrecy of the fraction

$$\frac{\lfloor \log M^t \rfloor}{\lfloor \log_2 M \rfloor + 1} \approx t \tag{1.5}$$

of the least significant bits of a plaintext. Since $0 \leq t < 1$ is an arbitrary value in Theorem 1, our scheme can ensure secrecy of any fraction of the least significant bits of the plaintext in this significant case where plaintexts distribute uniformly at random.

Fraction $t < \beta$ of Lower Bits Are Hidden in the General Case: Similarly, Theorem 2 shows that our scheme can ensure secrecy of the fraction $t < \beta$ of the least significant bits of a plaintext, where β is the value given in (1.3).

Importance of Theorem 2 in Database Security: This theorem can ensure some indistinguishability when we are uncertain about values. Namely, $(\mathcal{X}, M^t, q)$-indistinguishability of our scheme $\mathcal{E}^{\beta,t}$ is achieved even when we do not have the complete knowledge of the tuple $\mathcal{X}$ of plaintext distributions (whose min-entropies are larger than β). This fact is very important in the main application of OPE, the database security, because we cannot know the plaintext distributions completely in advance when, for instance, plaintexts are names of new students (with the lexicographic order) or scores of some examination. Our scheme can ensure the secrecy of the least significant $t < \beta$ bits of plaintexts even in this case.

Allowing Decryption Queries: As in [18], we can naturally make our scheme secure even when we allow the adversary to make decryption queries at any time, using the "Encryption-then-Mac" composition (adding MAC data) [11], which, essentially, privately "signs" messages. See Appendix 3.3 for details.

1.5 Stronger Variant of (r, q)-WOW Security [17]

Finally, we define a stronger variant of (r, q)-WOW notion, called (r, q)-WOWM, which is informally discussed in [17]. Then, we show the relationship between our basic results and the known result using this notion [17].

Definition: (r, q)-*WOWM* (WOW viewing Messages) is defined as follows: an adversary A is fed $\mathsf{Enc}_K(m^*)$ and $(m_1, \mathsf{Enc}_K(m_1)), \ldots, (m_q, \mathsf{Enc}_K(m_q))$ and outputs an interval I such that the length of I is not more than r. A is considered to win iff $m^* \in I$. This notion is stronger than (r, q)-WOW because (r, q)-WOWM allows an adversary to watch $m_1, \ldots, m_q$ while (r, q)-WOW prohibits her from doing this.

Relationship with $(\mathcal{X}, \theta, q)$-indistinguishability: Below, $\mathsf{Succ.Exp}^{\mathcal{S}}(\mathsf{A})$ denotes the success probability of A in the experiment of a security notion $\mathcal{S}$.

Theorem 3. *Let $\mathcal{U}^q$ be the tuple of q uniform distributions on the message space $[1..M]$. Then, for any r and θ satisfying $0 \le r < \theta \le M$ and for any OPE $\mathcal{E}$, $(\mathcal{U}^q, \theta, q)$-indistinguishability of $\mathcal{E}$ implies $(r, q+1)$-WOWM of it if r/θ, θ/M, and q/M are negligible. Specifically,*

$$\forall \mathsf{A} \exists \mathsf{Mg} \exists \mathsf{B} \quad : \quad \mathsf{Succ.Exp}_{\mathcal{E}}^{(r,q+1)\text{-WOWM}}(\mathsf{A}) \le \mathsf{Adv.Exp}_{\mathcal{E}}^{(\mathcal{U}^q,\theta,q)\text{-indis}}(\mathsf{Mg}, \mathsf{B}) + O\Big(\frac{r}{\theta}\Big) + O\Big(\frac{\theta}{M}\Big) + O\Big(\frac{q}{M}\Big). \tag{1.6}$$

By definitions, the following property holds as well for any adversary A for $(r, q+1)$-WOW:

$$\mathsf{Succ.Exp}_{\mathcal{E}}^{(r,q+1)\text{-}WOW}(\mathsf{A}) \le \mathsf{Succ.Exp}_{\mathcal{E}}^{(r,q+1)\text{-}WOWM}(\mathsf{A}). \tag{1.7}$$

$(M^\rho, q+1)$-WOWM of our Proposed Scheme for $0 \le \rho < 1$: The following theorem quite immediately follows from Theorems 2 and 3.

Theorem 4. *For any constant*

$$0 \le \rho < 1,$$

our scheme $\mathcal{E}^{\beta,t}$ with $(\beta, t) = (2-\rho, (\rho+1)/2)$ is $(M^\rho, q+1)$-WOWM. Specifically,

$$\forall \mathsf{A} \exists \mathsf{B} \quad : \quad \mathsf{Succ.Exp}_{\mathcal{E}^{t,\theta}}^{(M^\rho,q+1)\text{-WOWM}}(\mathsf{A}) \le O\left(\frac{q}{M^{\frac{1-\rho}{2}}}\right) + \mathsf{Adv.Exp}_{\mathsf{PRF}}(\mathsf{B}) + \mathsf{neg}(\lambda). \tag{1.8}$$

This fact seems interesting since it was proved in [18] that the scheme of [16] was not able to achieve (M^ρ, q)-WOW (and therefore (M^ρ, q)-WOWM) for any $\rho > 1/2$ although $(1, q)$-WOW of this scheme was proved.

1.6 Comparisons with the Known Scheme [16]

The earlier results on OPE are indeed remarkable and opened the door to our investigation, but there are some crucial differences which we would like to point out explicitly.

About Our Security Notion: Our scheme satisfies $(\mathcal{U}^q, M^t, q)$-indistinguishability for any $0 \leq t < 1$, where the tuple $\mathcal{U}^q$ of the uniform distributions on the message space. This means that it can hide *any fraction t of* the least significant bits of a plaintext in our setting with uniformly randomly selection of plaintexts. We also show the secrecy of lower bits of a plaintext for non-uniform distributions based on min-entropy of the distributions.

On the other hand, we can prove that the known scheme [16] *cannot* satisfy $(\mathcal{U}^q, M^t, q)$-indistinguishability for $t > 1/2$, which means that (due to the discussion of Section 1.3) the scheme has to reveal any bit of plaintext other than the lower half bits of it with non-negligible advantage in the same setting. This is because Theorem 3 shows that $(\mathcal{U}^q, M^t, q)$-indistinguishability implies $(M^\rho, q+1)$-WOW for suitable parameter $\rho > 1/2$ and it is proved [18] that the known scheme [16] cannot achieve $(M^\rho, q+1)$-WOW for $\rho > 1/2$. See Appendix 4.2 for the details.

There is no known partial-plaintext indistinguishability result other than ours, in particular not about the earlier scheme [16]. However, even if the scheme can hide something, the proof of this fact seems to us to be very hard, given that even the one-wayness (WOW) proof for that scheme is very long and elaborated.

About WOW [17]: Our scheme satisfies a stronger variant of $(M^\rho, q+1)$-WOW, namely, $(M^\rho, q+1)$-WOWM, for any constant $0 \leq \rho < 1$. On the other hand it is proved that the known scheme [16] *cannot* achieve $(M^\rho, q+1)$-WOW [17] (and therefore $(M^\rho, q+1)$-WOWM) for any $\rho > 1/2$ although it satisfies $(1, q+1)$-WOW (and $(1, q+1)$-WOWM as well, due to the informal discussion in Section 4.1 of [17]).

About POPF-CCA [16]: Our scheme does not satisfy this notion, but, as pointed out above, this notion tells nothing about secrecy of plaintext partial information [18].

1.7 Other Related Works

We note the following results which are beyond the scope of this work.

Other Known Results about OPE [49,51,48]: Yum, Kim, Kim, Lee and Hong [51] propose a more efficient method to compute the encryption and decryption functions of the known scheme [16]. Xiao, Yen, and Huynh [49] study OPE in a multi-user setting. Xiao and Yen [48] estimates the min-entropy of an unknown plaintext encrypted by the known scheme [16], but this is very different from our approach since, first, this does not tell us which bits of a plaintext are hidden; secondly, it is about the known scheme [16]; and thirdly, it is not about indistinguishability.

Property Preserving Encryption [37]: This notion, introduced by Pandey and Rouselakis, is a variant of the OPE whose encryption function preserve some given property. The security notions for this scheme can be considered as those for OPE. However, almost the same attack as that of [16] can break any OPE scheme under these security notions when the scheme has a super-polynomial size message space. See Appendix 4.1 for details.

CEOE and MOPE [18]: These notions achieve stronger security than OPE by sacrificing some functionalities of it: CEOE requires that the set of plaintexts be encrypted is pre-determined and static. MOPE does not tell us the order of plaintexts themselves although it tells us their order modulo the message space size. These notions are of independent interests and beyond our scope.

mOPE and stOPE [38,39]: These notions, introduced by Popa, Li, and Zeldovich, are interactive protocols which allow a user to store plaintexts to a database in their encrypted forms and to compare the sizes of plaintexts in the database. These protocols can be run without revealing any information about plaintexts except their order. However, they are different from and more specialize than the original OPE, since they are interactive and they employ stateful encryption functions.

GOPE[47]: GOPE schemes, introduced by Xiao and Yen, also achieve a stronger security by changing the model of OPE. But their schemes can be used in the very restrictive cases where the message space size is less than polynomial in the security parameter, or all encryption queries are within some polynomial size interval.

2 Preliminaries

2.1 Notations and Terminologies

We next introduce notations and terminology used throughout this paper.

Intervals: For integers a and $b \geq a$, *interval* $[a..b]$ is the set $\{a, \ldots, b\}$. $[b]$, $(a..b]$, $[a..b)$, and $(a..b)$ denote $[1..b]$, $[a+1..b]$, $[a..b-1]$, and $[a+1..b-1]$, respectively.

Order-Preserving Encryption: An *OPE* scheme is a symmetric key encryption $\mathcal{E} = (\mathsf{Kg}, \mathsf{Enc}, \mathsf{Dec})$ whose message space $\mathcal{M}$ and ciphertext space are intervals in $\mathbb{N}$ and which satisfies $m < m' \Rightarrow \mathsf{Enc}_K(m) < \mathsf{Enc}_K(m')$ for any $m, m' \in \mathcal{M}$ and any possible output K of Kg. Here "$<$" represents the numerical order. Throughout this paper, we assume w.l.o.g. that the message space $\mathcal{M}$ of $\mathcal{E}$ can be written as $[1..M]$.

2.2 Notations, Definitions, and Known Facts about Probabilities

Probabilities: We denote by $X \sim \mathcal{P}$ when a random variable X follows a distribution $\mathcal{P}$. For a probability distribution $\mathcal{P}$, a random variable X, and a finite set A, we write $x \xleftarrow{\$} \mathcal{P}$, $x \xleftarrow{\$} X$, and $x \xleftarrow{\$} A$ to denote that x is, respectively, sampled according to $\mathcal{P}$, distribution of X, and the uniform distribution on A.

Binomial and Hypergeometric Distributions: For natural numbers n and real number $0 \leq p \leq 1$, the *binomial distribution* $\mathcal{B}(n,p)$ is the probability distribution of the number of heads in a sequence of n independent coin flips, each of which yields head with probability p. For natural numbers n, and $k, \ell \leq n$, the *hypergeometric distribution* $\mathcal{HG}(n,t,\ell)$ is a probability distribution of the number of drawn black balls in a sequence of ℓ draws without replacement from a bin which contains $n-t$ white balls and t black balls. It is well known that if $X \sim \mathcal{B}(n,p)$ and $Y \sim \mathcal{HG}(n,t,\ell)$ then

$$\Pr[X=i] = \binom{n}{i} p^i (1-p)^{n-i}, \quad \Pr[Y=i] = \binom{\ell}{i}\binom{n-\ell}{t-i} \Big/ \binom{n}{t}. \quad \text{Here } \binom{n}{k} = \frac{n!}{(n-k)!k!}.$$

Statistical Distance: For random variables X and Y, $\mathsf{SD}(X,Y)$ denote the *statistical distance* of X and Y, that is,

$$\mathsf{SD}(X,Y) = \sum_z |\Pr[X=z] - \Pr[Y=z]|$$

2.3 Known Facts About Probabilities

Proposition 5 (Conditional Distribution of Binomial is Hypergeometric) *Let a, b, and $c \le a+b$ be natural numbers. Let α and β be independent random variables which follow $\mathcal{B}(a,q)$ and $\mathcal{B}(b,q)$ respectively. Then the distribution of α given the condition $\alpha+\beta=c$ is $\mathcal{HG}(a+b,c,a)$ (regardless of the value q).*

Proof (Proposition 5). By definitions, $\alpha \sim \mathcal{B}(a,q)$ and $\beta \sim \mathcal{B}(b,q)$ hold and α and β are independent. Since the sum of independent binomial random variables is also binomial random variable, $\alpha+\beta \sim \mathcal{B}(a+b,q)$ holds. From $\alpha \sim \mathcal{B}(a,q)$, $\beta \sim \mathcal{B}(b,q)$, and $\alpha+\beta \sim \mathcal{B}(a+b,q)$, it follows that

$$\Pr[\alpha=i] = \binom{a}{i} q^i (1-q)^{a-i},\ \Pr[\beta=j] = \binom{b}{j} q^j (1-q)^{b-j},\ \Pr[\alpha+\beta=c] = \binom{a+b}{c} q^c (1-q)^{a+b-c}.$$

Set $j=c-i$. Then, it follows that

$$\Pr[\alpha=i \mid \alpha+\beta=c] = \frac{\Pr[\alpha=i \ \wedge\ \alpha+\beta=c]}{\Pr[\alpha+\beta=c]} = \frac{\Pr[\alpha=i]\Pr[\beta=c-i]}{\Pr[\alpha+\beta=c]} = \binom{a}{i}\binom{b}{c-i} \Big/ \binom{a+b}{c}.$$

□

Proposition 6 (Large Uniform Randomness Hides Small Values) *Let θ and $L > 2\theta$ be natural numbers. Then, for all $\alpha,\beta \in [-\theta..\theta]$, the statistical distance between the random variables $\alpha+\delta$ and $\beta+\zeta$ for $\delta,\zeta \xleftarrow{\$} [1..L]$ is less than $O(\theta/L)$.*

Proof (Proposition 6). Take $\delta,\zeta \leftarrow [1..2^\lambda\theta]$ and set $U \leftarrow \alpha+\delta$ and $V \leftarrow \beta+\zeta$. W.l.o.g. we can assume $\alpha \ge \beta$. Then clearly, $(\Pr[U=z],\Pr[V=z])$ is $(0,1/L)$, $(1/L,1/L)$, $(1/L,1/L)$, and $(0,0)$, if $\beta < z \le \alpha$, $\alpha < z \le L+\beta$, $L+\beta < z \le L+\alpha$, and otherwise respectively. Hence, the following equations hold. Below, $\mathsf{SD}(U,V)$ denote the statistical distance and the last equation follows from the assumption $\alpha,\beta \in [-\theta..\theta]$.

$$\mathsf{SD}(U,V) = \sum_z |\Pr[U=z] - \Pr[V=z]| = \sum_{\beta<z\le\alpha} \frac{1}{L} + \sum_{L+\beta<z\le L+\alpha} \frac{1}{L} = \frac{2(\alpha-\beta)}{L} = O(\theta/L) \quad \square$$

3 Security Notions and Their Relationships

Next, we introduce $(\mathcal{X},\theta,q)$-indistinguishability and (r,q)-WOWM, whose intuitive meanings are given in Sections 1.3 and 1.5 respectively. Then we will prove Theorem 3.

3.1 Security Definitions

Below, λ is a security parameter and $\mathcal{E} = (\mathsf{Kg},\mathsf{Enc},\mathsf{Dec})$ is an OPE scheme.

Definition 7 ($(\mathcal{X},\theta,q)$-indistinguishability) For a non-negative real numbers $\theta = \theta(\lambda) \geq 0$, a polynomial $q = q(\lambda)$, and a tuple $\mathcal{X} = (\mathcal{X}_i)_{i\in[1..q]}$ of distributions on the message space of $\mathcal{E}$, $\mathcal{E}$ is said to be $(\mathcal{X},\theta,q)$-*indistinguishable* if for any polynomial time machine Mg whose outputs $(m_0^*, m_1^*, \mathsf{info})$ satisfies $|m_1^* - m_0^*| \leq \theta$ and $m_0^* < m_1^*$, and any polynomial time adversary A, $|\Pr[\mathsf{Exp}_{\mathcal{E}}^{(\mathcal{X},\theta,q)\text{-indis.}}(\mathsf{Mg},\mathsf{A},1) = 1] - \Pr[\mathsf{Exp}_{\mathcal{E}}^{(\mathcal{X},\theta,q)\text{-indis.}}(\mathsf{Mg},\mathsf{A},0) = 1]|$ is negligible. Here $\mathsf{Exp}_{\mathcal{E}}^{(\mathcal{X},\theta,q)\text{-indis.}}(\mathsf{Mg},\mathsf{A},b)$ is defined as follows.

$$K \leftarrow \mathsf{Kg}(1^\lambda), (m_0^*, m_1^*, \mathsf{info}) \leftarrow \mathsf{Mg}(1^\lambda), m_1 \xleftarrow{\$} \mathcal{X}_1, \ldots, m_q \xleftarrow{\$} \mathcal{X}_q,$$
$$d \leftarrow \mathsf{A}(\mathsf{Enc}_K(m_b^*), (m_i, \mathsf{Enc}_K(m_i))_{i\in[1..q]}, \mathsf{info}), \text{Return } d.$$

Above, info is a bit string whose intuitive meaning is any information about messages m_0^* m_1^*. Due to the bit string info, we can re-interpret the above definition such that algorithms Mg and A are the "guess stage" and the "find stage" of a single adversary (Mg,A) and info is her state.

An OPE scheme under this notion is broken trivially when $m_i \in [m_0^*..m_1^*]$ holds for some i but we employ the above security notion only when $m_i \in [m_0^*..m_1^*]$ holds with negligible probability.

Next, we define (r,q)-WOWM notion based on the definition of (specified) $(1,q)$-WOW given in Appendix B of [17], (which can be reduced to $(1,q)$-WOW given in Section 3 of [17] and vise versa).

Definition 8 ((r,q)-WOWM) Let $\mathsf{Comb}_{q+1}(M)$ be the set of $(q+1)$-element subset of the message space $[1..M]$ of $\mathcal{E}$. We say that $\mathcal{E}$ satisfies $(1,q)$-*WOWM (Window One-Wayness viewing Messages)* if for any polynomial time adversary A, $\Pr[\mathsf{Exp}_{\mathcal{E}}^{(r,q)\text{-WOWM}}(\mathsf{A}) = 1]$ is negligible. Here, experiment $\mathsf{Exp}_{\mathcal{E}}^{(r,q)\text{-WOWM}}(\mathsf{A})$ is defined as follows:

$$K \leftarrow \mathsf{Kg}(1^\lambda), \boldsymbol{m} \xleftarrow{\$} \mathsf{Comb}_{q+1}(M), m_* \xleftarrow{\$} \boldsymbol{m}, (m_L, m_R) \leftarrow \mathsf{A}(\mathsf{Enc}_K(m_*), (m, \mathsf{Enc}_K(m))_{m\in\boldsymbol{m}\setminus\{m_*\}}),$$
$$\text{Return 1 iff } m_* \in \mathcal{S}(m_L, m_R), \quad \text{where } \mathcal{S}(m_L, m_R) = \begin{cases} [m_L..m_R] & \text{if } m_L \leq m_R \\ [1..m_R] \cup [m_L..M] & \text{otherwise.} \end{cases}$$

Here "$m_* \xleftarrow{\$} \boldsymbol{m}$" means that "choose a message m_* from the tuple $\boldsymbol{m}$ uniformly at random". The output (m_L, m_R) of A has to satisfy $\#\mathcal{S}(m_L, m_R) \leq r$.

3.2 Proof of Theorem 3

We now prove Theorem 3 presented in Section 1.

Proof. Since (1.7) holds by definition, we show (1.6). Consider game Game, which is the same as the game of $(r, q+1)$-WOWM except that the challenger of it selects $m_1, \ldots, m_q$ not from $\mathsf{Comb}_{q+1}(M)$ but from the uniform distributions. In other words, while the challenger of the $(r, q+1)$-WOWM game draws q different elements of $[1..M]$, that of the Game draws q elements of $[1..M]$ (with repetition). By definitions, the difference between the advantages of an adversary in the $(r, q+1)$-WOWM game and the Game game is only $O(q/M)$.

We next construct Mg and B for $(\mathcal{X},\theta,q)$-indistinguishability by using A for Game as follows: $\mathsf{Mg}(1^\lambda)$ selects $m_0^* \xleftarrow{\$} [1..M]$ and $s \xleftarrow{\$} [0..\theta]$ uniformly at random, sets m_1^* to $m_0^* + s$ and info to the null string, and outputs $(m_0^*, m_1^*, \mathsf{info})$ if $m_1^* \leq M$ holds. Otherwise, Mg outputs fail. Then the challenger selects $m_1, \ldots, m_q$ uniformly at random and compute their ciphertexts $C_1, \ldots, C_q$ and the challenge ciphertext C_*.

B gets $(C_*, (m_i, C_i)_{i\in[1..q]})$ and info as inputs and computes $(m_L, m_R) \leftarrow \mathsf{A}(C_*, (m_i, C_i)_{i\in[1..q]})$. If there exists $d \in \{0,1\}$ such that $m_d^* \in \mathcal{S}(m_L, m_R)$ but $m_{1-d}^* \notin \mathcal{S}(m_L, m_R)$, B outputs d, where $\mathcal{S}(m_L, m_R)$ is the set given in Definition 8; otherwise, B selects $d \xleftarrow{\$} \{0,1\}$ uniformly at random and outputs d.

By simply calculating, $\Pr[m_1^* = m] = 1/M$ holds for any $m \in [\theta, M]$. Hence, m_1^* is uniformly distributed under the condition that $m_1 \in [\theta..M]$ holds. All other data, m_i and m_0^*, are uniformly distributed as well by definition. Therefore, with probability $\Pr[m_1^* \in [\theta..M]] = 1 - \frac{\theta}{M}$, B succeeds in simulating correctly the view of A, where all data have to be selected uniformly at random. (Note that Mg does not output fail either, when $m_1^* \in [\theta..M]$.)

We let Good be the event that $m_1^* \in [\theta..M]$ holds thereafter and let $\varepsilon \leftarrow \Pr[\mathsf{A}\text{ wins} \mid \mathsf{Good}]$. Then, the probability that the output d of B equals b is

$$\Pr[d = b \mid \mathsf{Good}] \geq \Pr[(1) \mid \mathsf{Good}] + \Pr[(2) \mid \mathsf{Good}] = \varepsilon \cdot \left(1 - \frac{r}{\theta}\right) + (1-\varepsilon) \cdot \left(1 - \frac{r}{\theta}\right) \cdot \frac{1}{2} \geq \frac{1+\varepsilon}{2} - \frac{r}{\theta},$$

where (1) and (2) are the following events:

(1) $m_b^* \in \mathcal{S}(m_L, m_R)$ and $m_{1-b}^* \notin \mathcal{S}(m_L, m_R)$.
(2) $m_b^* \notin \mathcal{S}(m_L, m_R)$ and $m_{1-b}^* \notin \mathcal{S}(m_L, m_R)$. But, fortunately, the bit d selected by B is equal to b.

Since $|\Pr[E] - \Pr[E \mid \mathsf{Good}]| \leq \Pr[\neg\mathsf{Good}]$ holds [12] for any event E, the advantage of $\mathsf{B} = |\Pr[d = 1 \mid b = 1] - \Pr[d = 1 \mid b = 0]| = 2|\Pr[d = b] - \frac{1}{2}|$ satisfying (1.6), which completes the proof.

3.3 Encryption-then-Mac Makes Security of Our Scheme Stronger

Next, we show that an OPE scheme be secure even when an adversary can make decryption queries any time, using the "Encryption-then-Mac" composition (adding MAC data) [11], which, essentially, privately "signs" messages.

Definitions:

Definition 9 (Message Authentication Code (MAC)) A *message authentication code* is a tuple $\mathcal{MAC} = (\mathsf{Gen}, \mathsf{Tag}, \mathsf{Ver})$ of algorithms whose inputs and outputs are as follows: Gen takes 1^λ as an input and outputs *key* K. Tag takes K and a *message* m and outputs *tag* σ on m. Ver takes K, m, and a candidate of tag σ' on m and outputs accept or reject. $\mathcal{MAC}$ has to satisfy the *correctness property*, $\mathsf{Ver}_K(m, \mathsf{Tag}_K(m)) = \mathsf{accept}$ for any message m and any possible output K of Gen.

Definition 10 (Strong Unforgeability) We say that $\mathcal{MAC} = (\mathsf{Gen}, \mathsf{Tag}, \mathsf{Ver})$ is *strongly unforgeable (under chosen message attack)* if for any polynomial time adversary A, the probability that the following game outputs 1 is negligible. Below, (m, σ) has to be different from any pair of Tag_K-query of A and its answer from the oracle.

$$K \leftarrow \mathsf{Gen}(1^\lambda), \quad (m, \sigma) \leftarrow \mathsf{A}^{\mathsf{Tag}_K(\cdot), \mathsf{Ver}_K(\cdot,\cdot)}(1^\lambda). \text{ Return 1 iff } \mathsf{Ver}_K(m, \sigma) = \mathsf{accept}.$$

Definition 11 (Encrypt-then-MAC) For OPE scheme $\mathcal{E} = (\mathsf{Kg}, \mathsf{Enc}, \mathsf{Dec})$ and message authentication code $\mathcal{MAC} = (\mathsf{Gen}, \mathsf{Tag}, \mathsf{Ver})$, the "*Encrypt-then-MAC*" composition $\mathcal{E}' = (\mathsf{Kg}', \mathsf{Enc}', \mathsf{Dec}')$ [11] of them is as follows:

– $\mathsf{Kg}'(1^\lambda) : K \leftarrow \mathsf{Kg}(1^\lambda)$, $L \leftarrow \mathsf{Gen}(1^\lambda)$. Output $K' \leftarrow (K, L)$.
– $\mathsf{Enc}'_{K'}(m)$: Parse K' as (K, L). Compute $C \leftarrow \mathsf{Enc}_K(m)$. Output $C' \leftarrow (C, \mathsf{Tag}_L(C))$.
– $\mathsf{Dec}'_{K'}(C')$: Parse K' and C' as (K, L) and (C, σ). If $\mathsf{Ver}_L(C, \sigma) = \mathsf{accept}$, output $\mathsf{Dec}_K(C)$. Otherwise, output $\perp$.

A ciphertext of $\mathcal{E}'$ is a pair of C and σ while the definition of OPE requires that a ciphertext is an integer. However, we loosely say that "$\mathcal{E}'$ is an OPE scheme", because the first element C of a ciphertext satisfies the order preserving property (in addition, we can embed the two elements in a larger domain integer, where the first element determines the high order bits).

Result: Let $(\mathcal{X}, \theta, q)$-indis-dec be the same notion as $(\mathcal{X}, \theta, q)$-indistinguishability except that a message generator and an adversary can make decryption queries any time (if the decryption queries are different from the challenge ciphertext). (Here we consider the case where both a message generator and an adversary can make decryption queries, so as to make our result strong, although the message generator making decryption queries may not be natural.)

Then, we can show that if an OPE scheme $\mathcal{E}$ is $(\mathcal{X}, \theta, q)$-indistinguishable then $\mathcal{E}$ with the "Encryption-Then-MAC" satisfies $(\mathcal{X}, \theta, q)$-indis-dec. The intuition is that successful decryption query breaks the MAC, and formally, we can show the following theorem:

Theorem 12. *Let $\mathcal{E}$ be an OPE scheme, $\mathcal{M}$ be a strongly unforgeable MAC, and $\mathcal{E}'$ be the OPE scheme obtained from them by "Encrypt-then-MAC" composition [11]. Then*

$$\forall \mathsf{Mg}\ \ \forall \mathsf{A}\ \ \exists \mathsf{Mg}^\dagger\ \ \exists \mathsf{A}^\dagger\ \ \exists \mathsf{B}\ \ :$$
$$\mathsf{Adv}.\mathsf{Exp}^{(\mathcal{X},\theta,q)\text{-indis-dec}}_{\mathcal{E}'}(\mathsf{Mg}, \mathsf{A}) \le \mathsf{Adv}.\mathsf{Exp}^{(\mathcal{X},\theta,q)\text{-indis}}_{\mathcal{E}}(\mathsf{Mg}^\dagger, \mathsf{A}^\dagger) + \mathsf{Succ}.\mathsf{Exp}^{\text{strong unforge.}}_{\mathcal{M}}(\mathsf{B})$$

holds. Here Mg *and* $\mathsf{Mg}^\dagger$ *denote message generators for $(\mathcal{X}, \theta, q)$-indis-dec of $\mathcal{E}'$ and $(\mathcal{X}, \theta, q)$-indistinguishability of $\mathcal{E}$ respectively,* A, $\mathsf{A}^\dagger$ *and* B *respectively denote adversaries for $(\mathcal{X}, \theta, q)$-indis-dec of $\mathcal{E}'$, $(\mathcal{X}, \theta, q)$-indistinguishability for $\mathcal{E}$, and strong unforgeability for $\mathcal{M}$, and* $\mathsf{Adv}.\mathsf{Exp}^{\mathcal{S}}$ *and* $\mathsf{Succ}.\mathsf{Exp}^{\mathcal{T}}$ *denote the advantage and the success probability in the experiments for the security notion $\mathcal{S}$ and $\mathcal{T}$.*

Proof (Theorem 12). Take any message generator Mg and any adversary A for $(\mathcal{X}, \theta, q)$-indis-dec. W.l.o.g. we assume that neither Mg nor A send an answered ciphertext C' from the encryption oracle to the decryption oracle, because they already know $\mathsf{Dec}_{K'}(C')$.

Let $\mathsf{Game}^{(0)}(\mathsf{Mg}, \mathsf{A})$ be the game of $(\mathcal{X}, \theta, q)$-indis-dec and $\mathsf{Game}^{(1)}(\mathsf{Mg}, \mathsf{A})$ be the same game as $\mathsf{Game}^{(0)}(\lambda)$, except that A is considered to win if A can make a decryption query whose answer is not $\perp$. We construct an adversary B for $\mathcal{M}$ satisfying

$$\mathsf{Adv}.\mathsf{Game}^{(0)}(\mathsf{Mg}, \mathsf{A}) \le \mathsf{Adv}.\mathsf{Game}^{(1)}(\mathsf{Mg}, \mathsf{A}) + \mathsf{Succ}.\mathsf{Exp}^{\text{strong unforge.}}_{\mathcal{E}}(\mathsf{B}). \tag{3.1}$$

$\mathsf{B}(1^\lambda)$ generates $K \leftarrow \mathsf{Kg}(1^\lambda)$ and executes $\mathsf{Mg}(1^\lambda)$. Each times Mg makes decryption query $C' = (C, \sigma)$, B sends (C, σ) to Ver_L oracle. If the answer from Ver_L is accept, B outputs (C, σ) and terminates. Otherwise, she sends $\perp$ back to Mg.

When Mg outputs the challenge query (m_0^*, m_1^*) and bit string info, B randomly takes $b \in \{0, 1\}$ and $m_1 \xleftarrow{\$} \mathcal{X}_1, \ldots, m_q \xleftarrow{\$} \mathcal{X}_q$, computes $C_* \leftarrow \mathsf{Enc}_K(m_b^*)$, $C_1 \leftarrow \mathsf{Enc}_K(m_1)$, …, $C_q \leftarrow \mathsf{Enc}_K(m_q)$ using K, sends them to Tag_L oracle, gets σ_*, σ_1, …, σ_q as answers, sets $C'_* \leftarrow (C_*, \sigma_*)$, $C'_1 \leftarrow (C_1, \sigma_1)$, …, $C'_q \leftarrow (C_q, \sigma_q)$, and sends C_*, (m_1, C'_1), …, (m_q, C'_q) and info to A. B simulates the decryption oracle in the same way as she did with Mg. If A terminates, the attack of B fails.

The output (C, σ) of B is different from (C_*, σ_*) because A does not send to the decryption oracle a ciphertext answered by the encryption oracle and because A is not allowed to send $C'_* = (C_*, \sigma_*)$ to the decryption oracle. That is, B follows the rule of strong unforgeability. Since the success probability of B is equal to the difference between the advantages of A in $\mathsf{Game}^{(0)}$ and that of $\mathsf{Game}^{(1)}$, (3.1) holds.

Let $\mathsf{Mg}^\dagger$ be the same message generator as Mg except that $\mathsf{Mg}^\dagger$ does not make a decryption query. Specifically, $\mathsf{Mg}^\dagger(1^\lambda)$ executes $\mathsf{Mg}(1^\lambda)$ as a subroutine, answers $\perp$ to Mg each time Mg makes a decryption query, passes all other queries of Mg to oracles, and returns the output of Mg. We define $\mathsf{A}^\dagger$ from A in a similar manner to $\mathsf{Mg}^\dagger$. Clearly, $\mathsf{Mg}^\dagger$ and $\mathsf{A}^\dagger$ are a message generator and an adversary for the game of $(\mathcal{X}, \theta, q)$-indistinguishability and the advantage of $(\mathsf{Mg}^\dagger, \mathsf{A}^\dagger)$ in this game is equal to $\mathsf{Adv.Game}^{(1)}(\mathsf{Mg}, \mathsf{A})$. Hence, the theorem follows from (3.1).□

4 Unachievability of Security Notions

4.1 Unachievability of the Security Notions of [37] for OPE with Super-Polynomial Size Message Space

The subject of this section is to show that all the security notions proposed by Pandey and Rouselakis [37] are unachievable for any OPE scheme with a super-polynomial size message space.

To this end, we review the study of Pandey and Rouselakis [37]. They introduced a generalized variant of the OPE, *property preserving encryption (PPE)*, whose encryption function preserved some given property. Then they proposed several security notions for PPE, such as FTG and sLOR, and showed that for any security notion $\mathcal{S}$ defined by them, either "$\mathcal{S} \Rightarrow$ sLOR" or "$\mathcal{S} \Rightarrow$ FTG" held.

Therefore, we will review the definition of sLOR and FTG and will show that they are unachievable for any OPE scheme with a super-polynomial size message space. Then, from [37] it follows that no security notion defined by them is achievable for such OPE schemes.

About sLOR: Their sLOR notion (for OPE) is a weaker variant of IND-O-CPA [16] such that an adversary has to proclaim which one of the i-th encryption query and the j-th encryption query are larger for any i and j in advance. She then can make encryption queries (m_1^0, m_1^1), ..., (m_q^0, m_q^1) to Left-Or-Right oracle adaptively, gets answers $\mathsf{Enc}_K(m_1^b)$, ..., $\mathsf{Enc}_K(m_q^b)$, and outputs d. She is considered to win iff $b = d$ holds and $(m_1^u, \ldots, m_q^u)$ satisfies the proclaimed rules for any $u \in \{0, 1\}$. It is easy to see that the adversary [16] for IND-O-CPA can be re-interpret as an adversary for sLOR. Hence, sLOR cannot be achieved by any OPE scheme with the super-polynomial size message space either.

About FTG: FTG notion (for OPE) is defined as follows: an adversary makes a challenge query (m_0^*, m_1^*), gets $\mathsf{Enc}_K(m_b^*)$ as an answer, and guess b. She can makes any number of encryption queries at any time but an encryption query has to satisfy the order-preserving property $m_0^* < m \Leftrightarrow m_1^* < m$. We additionally impose that an adversary does not make encryption queries m_0^* and m_1^* because otherwise, she can break the OPE scheme, whose encryption function is deterministic, trivially.

We can show that almost the same attack of [16] for IND-O-CPA breaks an OPE scheme under this notion when the scheme has the super-polynomial size message space. The detailed description of an adversary is as follows. She selects two messages m_1 and m_2 uniformly at random. (W.l.o.g. we assume $m_1 \leq m_2$). Then she makes two encryption queries m_1 and m_2, receives their ciphertexts $C_1 < C_2$ as answers, makes the challenge query $(m_0^*, m_1^*) = (m_1 + 1, m_2 - 1)$, gets $C^* = \mathsf{Enc}_K(m_b^*)$ as an answer, and outputs 1 iff $C_2 - C^* > C^* - C_1$.

An advantage of her is non-negligible. The proof of this fact is as follows. Let Exp be the experiment of FTG, $\mathcal{E} = (\mathsf{Kg}, \mathsf{Enc}, \mathsf{Dec})$ be any OPE scheme on a message space $[1..M]$, and $[1..N]$ be the ciphertext space of $\mathcal{E}$. Let E be the event that $m_1 \leq \lceil M/2 \rceil \leq m_2$ holds. Since the adversary selects m_1 and m_2 uniformly at random from $[1..M]$, E holds with probability $\geq 1/4 - O(1/M)$.

When the secret bit b of the challenger is 0, the challenge ciphertext can be written as $C_* = \mathsf{Enc}_K(m_0^*) = \mathsf{Enc}_K(m_1 + 1)$ Hence, from the definition of A,

$$
\begin{aligned}
&\Pr[\mathsf{A} \text{ outputs } 1 \mid b = 0] \\
&= \Pr[\mathsf{Enc}_K(m_2) - \mathsf{Enc}_K(m_1 + 1) \geq \mathsf{Enc}_K(m_1 + 1) - \mathsf{Enc}_K(m_1) \mid b = 0] \\
&\geq \frac{1}{4} \Pr[\mathsf{Enc}_K(m_2) - \mathsf{Enc}_K(m_1 + 1) \geq \mathsf{Enc}_K(m_1 + 1) - \mathsf{Enc}_K(m_1) \mid b = 0 \wedge E] - O(1/M) \\
&\geq \frac{1}{4} \Pr[\mathsf{Enc}_K(\lceil M/2 \rceil) - \mathsf{Enc}_K(m_1 + 1) \geq \mathsf{Enc}_K(m_1 + 1) - \mathsf{Enc}_K(m_1) \mid b = 0 \wedge E] - O(1/M).
\end{aligned}
$$

We set $f = \mathsf{Enc}_K$. By applying Lemma 3.2 of [16] to the function $f|_{[1..\lceil M/2 \rceil]} : [1..\lceil M/2 \rceil] \to [1..N]$, we can show that the number of m_1 satisfying $\mathsf{Enc}_K(\lceil M/2 \rceil) - \mathsf{Enc}_K(m_1 + 1) < \mathsf{Enc}_K(m_1 + 1) - \mathsf{Enc}_K(m_1)$ is less than $\log N$. From the above discussion,

$$
\Pr[\mathsf{A} \text{ outputs } 1 \mid b = 0] \geq \frac{1}{4} \left(1 - O\left(\frac{\log N}{M} \right) \right)
$$

holds. We can similarly show that

$$
\Pr[\mathsf{A} \text{ outputs } 0 \mid b = 1] \geq \frac{1}{4} \left(1 - O\left(\frac{\log N}{M} \right) \right)
$$

Hence, the advantage of A is more than $(1/4) - O(\log N/M) \geq 1/5$ when $M \to \infty$.

4.2 The Known Scheme [16] Cannot Achieve $(\mathcal{U}^q, M^t, q)$-Indistinguishability for $1/2 < t \leq 1$

Proof. We show that the known scheme [16] cannot satisfy $(\mathcal{U}^q, M^t, q)$-indistinguishability for $1/2 < t < 1$. Then, since $(\mathcal{U}^q, M^1, q)$-indistinguishability is weaker than $(\mathcal{U}^q, M^{2/3}, q)$-indistinguishability, it cannot satisfy $(\mathcal{U}^q, M^t, q)$-indistinguishability for $t = 1$ either.

We prove this fact by contradiction. Suppose that the known scheme [16] is $(\mathcal{U}^q, M^{2/3}, q)$-indistinguishable. By applying (1.6) and (1.7) to $(\theta, r) = (M^t, M^{(2t+1)/4})$, we can conclude that

$$
\begin{aligned}
\forall \mathsf{A} \exists \mathsf{Mg} \exists \mathsf{B} \quad : \quad & \mathsf{Succ.Exp}_{\mathcal{E}}^{(M^{(2t+1)/4}, q+1)\text{-WOW}}(\mathsf{A}) \\
& \leq \mathsf{Adv.Exp}_{\mathcal{E}}^{(\mathcal{U}^q, M^t, q)\text{-indis}}(\mathsf{Mg}, \mathsf{B}) + O\left(\frac{1}{M^{\frac{2t-1}{4}}} \right) + O\left(\frac{1}{M^{1-t}} \right) + O\left(\frac{q}{M} \right). \qquad (4.1)
\end{aligned}
$$

Since the second, the third, and the last terms of the right hand side of (4.1) are negligible for any $1/2 < t < 1$, and since we suppose that the scheme is $(\mathcal{U}^q, M^t, q)$-indistinguishable, the known scheme [16] has to satisfy $(M^{(2t+1)/4}, q+1)$-WOW for $(2t+1)/4 > 1/2$ This fact contradicts with the known fact [17] that $(M^\rho, q+1)$-WOW cannot be achieved by the known scheme [16] for any $\rho > 1/2$.□

Caveat: One may think that the above discussion is strange because the convergence speed of the the third term $1/M^{1-t}$ of (4.1) get worse when t becomes close to 1 (that is, when $(\mathcal{U}^q, M^t, q)$-indistinguishability notion becomes weaker) and in particular, the third term $1/M^{1-t}$ does not converge to 0 in the weakest case, $t = 1$.

But this is because our bound (4.1) becomes very loose when t is close to 1. Hence, when $t \to 1$, the bound (4.1) seems to get worse although two values $\mathsf{Succ}.\mathsf{Exp}_{\mathcal{E}}^{(M^{(2t+1)/4},q+1)\text{-WOW}}(\mathsf{A})$ and $\mathsf{Adv}.\mathsf{Exp}_{\mathcal{E}}^{(\mathcal{U}^q,M^t,q)\text{-indis}}(\mathsf{Mg},\mathsf{B})$ actually become close in this case.

5 Sufficient Condition for $(\mathcal{X},\theta,q)$-Indistinguishability

Next, we introduce a sufficient condition for $(\mathcal{X},\theta,q)$-indistinguishability. It is a new security notion called (k,θ)-*FTG-O-nCPA*, introduced in order to ease the security proof of our scheme, but it may be of independent interest.

(k,θ)-FTG-O-nCPA: Is an indistinguishability notion such that, unlike $(\mathcal{X},\theta,q)$-indistinguishability, an adversary can select, both, the challenge and encryption queries by herself, but she has to select them in nCPA fashion. Here *nCPA (non-adaptive CPA)* [34,32,33] is a type of attacks where an adversary is required to output all her (encryption or challenge) queries simultaneously all together, and gets their answers thereafter. The queries have to satisfy some restrictions (described later in the formal definition) as well. We call this notion (k,θ)-FTG-O-nCPA, because it is Find-Then-Guess-type [9,29] indistinguishability for OPE under nCPA attack. The formal definition is as follows:

Definition 13 ((k,θ)-FTG-O-nCPA) For non-negative real numbers $k = k(\lambda)$ and $\theta = \theta(\lambda)$, $\mathcal{E}$ is said to be (k,θ)-*FTG-O-nCPA secure* if for any polynomial time adversary $\mathsf{A} = (\mathsf{A}_{\mathsf{find}}, \mathsf{A}_{\mathsf{guess}})$, $|\Pr[\mathsf{Exp}_{\mathcal{E}}^{(k,\theta)\text{-FTG-O-nCPA}}(\mathsf{A},1)=1] - \Pr[\mathsf{Exp}_{\mathcal{E}}^{(k,\theta)\text{-FTG-O-nCPA}}(\mathsf{A},0)=1]|$ is negligible.

Here $\mathsf{Exp}_{\mathcal{E}}^{(k,\theta)\text{-FTG-O-nCPA}}(\mathsf{A},b)$ is defined as follows (below, q is an arbitrary number selected by A):

$$K \leftarrow \mathsf{Kg}(1^\lambda),\ ((m_0^*,m_1^*),(m_i)_{i\in[1..q]},\mathsf{st}) \leftarrow \mathsf{A}_{\mathsf{find}}(1^\lambda),$$
$$d \leftarrow \mathsf{A}_{\mathsf{guess}}(\mathsf{Enc}_K(m_b^*),(\mathsf{Enc}_K(m_i))_{i\in[1..q]},\mathsf{st}),\text{Return } d.$$

(m_0^*,m_1^*) and $m_1,\ldots,m_q$ are called a *challenge query* and *encryption queries* respectively. The output of A has to satisfy the following (5.1), (5.2), and (5.3). We also assume (5.4) throughout this paper w.l.o.g.

$$\forall i \ :\ m_i < m_0^* \ \Leftrightarrow\ m_i < m_1^*, \tag{5.1}$$
$$|m_0^* - m_1^*| \le \theta, \tag{5.2}$$
$$\forall d \in \{0,1\}, \forall i \ :\ |m_d^* - m_i| \ge k\theta. \tag{5.3}$$
$$m_0^* < m_1^* \tag{5.4}$$

Condition (5.1) is the ordered-preserving property and condition (5.2) is the same as that of $(\mathcal{X},\theta,q)$-indistinguishability. The new restriction (5.3) requires that the distance between the challenge message m_d^* and the encryption query m_i has to be larger than a pre-determined constant $k\theta$.

Sufficient Condition: Our sufficient condition for $(\mathcal{X},\theta,q)$-indistinguishability is as follows. Below, λ is a security parameter, $\mathcal{E}$ is an OPE scheme on a message space $[1..M]$ and $q = q(\lambda)$ is a polynomial.

Theorem 14. *Let $\mathcal{X} = (\mathcal{X}_1, \ldots, \mathcal{X}_q)$ is a tuple of distributions on $[1..M]$ such that $\mathcal{X}_1, \ldots, \mathcal{X}_q$ are independent from each other and one can sample $\mathcal{X}_i$ in time polynomial in λ. Let $\beta > 0$ be any constant. For $k > 0$ and $\theta > 0$, if*

$$\forall i \in [1..q] \quad : \quad H_\infty(\mathcal{X}_i) \geq \beta \log_2 M \tag{5.5}$$

holds for any large enough λ, then

$$\forall \mathsf{Mg} \forall \mathsf{A} \exists \mathsf{B} \quad : \quad \mathsf{Adv.Exp}_{\mathcal{E}}^{(\mathcal{X},\theta,q)\text{-indis.}}(\mathsf{Mg}, \mathsf{A}) \leq \mathsf{Adv.Exp}_{\mathcal{E}}^{(k,\theta)\text{-FTG-O-nCPA}}(\mathsf{B}) + O\left(\frac{qk\theta}{M^\beta}\right). \tag{5.6}$$

Above, A and Mg denote an adversary and a message generator for $(\mathcal{X}, \theta, q)$-indistinguishability, respectively, and B denotes an adversary for (k, θ)-FTG-O-nCPA.

Hence, (k, θ)-FTG-O-nCPA implies $(\mathcal{X}, \theta, q)$-indistinguishability when the last term of (5.6) is negligible.

Proof (sketch). For Mg and A for $(\mathcal{X}, \theta, q)$-indistinguishability, consider an adversary B for (k, θ)-FTG-O-nCPA which takes $(m_0^*, m_1^*, \mathsf{info}) \leftarrow \mathsf{Mg}(1^\lambda)$ and $m_1 \xleftarrow{\$} \mathcal{X}_1, \ldots, m_q \xleftarrow{\$} \mathcal{X}_q$, makes query $((m_0^*, m_1^*), m_1, \ldots, m_q)$, gives info and an answer to the query to A, and produces the output of A.

The above B will violate the constraint (5.3) if some m_i becomes $|m_i - m_d^*| < k\theta$ for some d. But the probability that B will violate (5.3) is bounded as follows: let I be the interval $(m_0^* - k\theta .. m_1^* + k\theta)$.

$$\begin{aligned}\sum_{i \in [1..q]} \Pr[m_i \leftarrow \mathcal{X}_i \ : \ m_i \in I] &\leq (\text{length of } I) \cdot \sum_{i \in [1..q]} \max_{x \in I} \Pr[m_i \leftarrow \mathcal{X}_i \ : \ m_i = x] \\ &\leq \sum_{i \in [1..q]} \frac{(2k+1)\theta}{2^{H_\infty(\mathcal{X}_i)}} \leq O\left(\frac{qk\theta}{M^\beta}\right).\end{aligned}$$

When $m_i \notin I$ holds, (5.1) is also satisfied. Moreover, (5.2) is always satisfied, due to the corresponding constraints on $(\mathcal{X}, \theta, q)$-indistinguishability. Thus, Theorem 14 follows.

Proving Theorem 2 by reduction: Due to the last result, what is left is to construct a (k, θ)-FTG-O-nCPA secure scheme for suitable k. In the next two sections (in two steps) we will construct an OPE scheme $\mathcal{E}_{k,\theta}$ using a pseudo-random function PRF and will prove the following theorem as our central theorem.

Theorem 15. *For $k \geq 1$, and $\theta \geq 1$,*

$$\forall \mathsf{A} \exists \mathsf{B} \quad : \quad \mathsf{Adv.Exp}_{\mathcal{E}_{k,\theta}}^{(k,\theta)\text{-FTG-O-nCPA}}(\mathsf{A}) \leq O\left(\frac{1}{\sqrt{k}}\right) + \mathsf{Adv.Exp}_{\mathsf{PRF}}(\mathsf{B}) + \mathsf{neg}(\lambda) \tag{5.7}$$

holds. (The value θ does not affect the advantage bound.) Above, λ is a security parameter and $\mathsf{neg}(\cdot)$ is some negligible function which is determined independently of (k, θ).

Our main result, Theorem 2, in turn, follows from Theorems 14 and 15 by setting

$$(k, \theta) = (M^{2(\beta - t)/3}, M^t) \tag{5.8}$$

and Theorem 1 follows, as well, by setting $\beta = 1$ in (5.8).

6 (k, θ)-FTG-O-nCPA Secure Scheme with Polysize Message Space

Next, we propose a (k, θ)-FTG-O-nCPA secure scheme $\mathcal{E}_{k,\theta}$ with the advantage bound given in Theorem 15. Due to the reductions discussed at the end of the previous Section, this scheme with suitable parameters satisfies our main theorems, Theorems 1 and 2.

The scheme $\mathcal{E}_{k,\theta}$, however, has the restriction that the message space size must be bounded by some polynomial in the security parameter λ. This restriction will be removed in Section 7. Note that the scheme of this section does not use a PRF although Theorem 15 refers about it. The PRF will be used to construct the scheme of the next section.

6.1 The Idea Behind Our Constrction

The First Tentative Scheme: The starting point of our idea is the idea spanning any OPE scheme $\mathcal{E} = (\mathsf{Kg}, \mathsf{Enc}, \mathsf{Dec})$, that $\mathsf{Enc}_K(m)$ can be represented by an equation of the form

$$\mathsf{Enc}_K(m) = R + \sum_{i \in (0..m]} \delta_i, \qquad \text{where } R = \mathsf{Enc}_K(0),\ \delta_i = \mathsf{Enc}_K(i) - \mathsf{Enc}_K(i-1). \tag{6.1}$$

We set R to a random quite larger number, much larger than δ_i. Say, $R \xleftarrow{\$} [0..M \cdot 2^\lambda]$ and $\delta_i \leftarrow 1$ where M is the message space size. Then if an adversary makes no encryption query, the secret bit b of the challenge ciphertext $C_b^* = \mathsf{Enc}_K(m_b) = R + \sum_{i \in (0..m_b]} \delta_i$ is hidden from her because the large random value R hides the smaller value $\sum_{i \in (0..m_b]} \delta_i$.

However, if the adversary possesses a pair $(m, \mathsf{Enc}_K(m))$ of an encryption query and its answer, she can compute $C_b^* - \mathsf{Enc}_K(m) = \sum_{i \in (m..m_b]} \delta_i$, which does not contain R and therefore cannot hide the challenge bit b. E.g., if $\delta_i = 1$, she can recover m_b from $C_b^* - \mathsf{Enc}_K(m) = m_b - m$ and m.

The Second Tentative Scheme: Therefore, we choose δ_i satisfying the monotonically decreasing property (MDP):

$$R \gg \delta_1 \gg \delta_2 \gg \delta_3 \gg \cdots$$

where the notation "$x \gg y$" means that the randomness of x within its space of choice, statistically hides y (or, more precisely, it means that $x + y$ is statistically indistinguishable from x). Then, if $m < m_0, m_1$ holds, an adversary cannot detect b from $C_b^* - \mathsf{Enc}_K(m)$ due to the MDP of δ_i. In fact, two values $C_0^* - \mathsf{Enc}_K(m) = \sum_{i \in (m..m_0]} \delta_i$ and $C_1^* - \mathsf{Enc}_K(m) = \sum_{i \in (m..m_1]} \delta_i$ contains much larger δ_i than their difference $\Delta = C_1^* - C_0^* = \sum_{i \in (m_0..m_1]} \delta_i$ and therefore, the difference Δ is hidden by these values in this case. (See Fig. 1.)

However, if $m > m_0, m_1$ holds, $\mathsf{Enc}_K(m) - C_0^* = \sum_{i \in (m_0..m]} \delta_i$ is much larger than $\mathsf{Enc}_K(m) - C_1^* = \sum_{i \in (m_1..m]} \delta_i$ due to the MDP of δ_i. (Here we assume $m_1 > m_0$, w.l.o.g.) Hence, the adversary can distinguish these values easily. If we change $(\delta_i)_i$ of our scheme with a monotonically increasing ones, $\delta_1 \ll \delta_2 \ll \delta_3 \ll \cdots$, the above problem is solved. In this case, however, an adversary can break the scheme when $m < m_0, m_1$ in a symmetric way to the above.

Our Proposed Scheme: We solve the above dilemma by changing the bit length of δ_i randomly. Specifically, we flip a random coin ρ_i which becomes 0 with small probability p and then samples δ_i randomly from a large set $[0..2^\lambda M]$ if $\rho_i = 0$ and set $\delta_i \leftarrow 1$ otherwise. Here M is the message space size. We call δ_i *large number* if it is taken from $[0..2^\lambda M]$.

Since intervals $[m_0 - k\theta..m_0]$ and $[m_1..m_1 + k\theta]$ are k times larger than $[m_0..m_1]$, the probabilities that $[m_0 - k\theta..m_0]$ and $[m_1..m_1 + k\theta]$ will contain a large δ_i is much larger than the probability that $[m_0..m_1]$ will contain a large δ_i. Therefore, if p is taken suitably, we can ensure the three properties

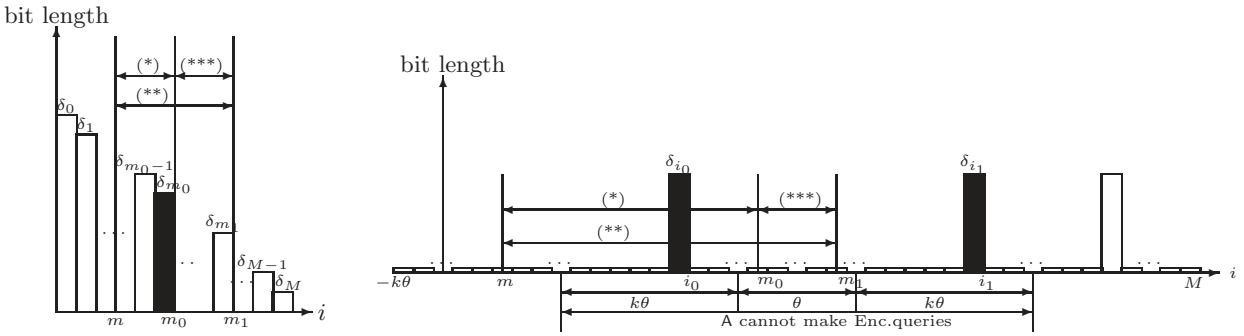

Fig. 1. The Second Tentative Schemes (left) and The Proposed Scheme (right) of Section 6. In both figures, $\mathsf{Enc}_K(m_0) - \mathsf{Enc}_K(m)$, $\mathsf{Enc}_K(m_1) - \mathsf{Enc}_K(m)$, and the difference of them are the sum of δ_i in (*), (**), and (***) respectively. Since both (*) and (**) contain a large randomness (δ_{m_0} of the left figure or δ_{i_0} of the right figure), the difference (***), which is smaller, is hidden by this large randomness. $\mathsf{Enc}_K(m_0) - \mathsf{Enc}_K(m)$ and $\mathsf{Enc}_K(m_1) - \mathsf{Enc}_K(m)$ are therefore indistinguishable.

Message Space $= [1..M]$, $p = 1 - (1 - 1/\sqrt{k})^{1/\theta}$, $A = -k\theta - 1$.

$\mathsf{Kg}(1^\lambda)$	$\mathsf{Enc}_K(m)$	$\mathsf{Dec}_K(C)$
11. For $i \in (A..M]$,	21. Parse K as $(\delta_i)_{i\in(A..M]}$.	31. Parse K as $(\delta_i)_{i\in(A..M]}$.
12. $\quad \rho_i \xleftarrow{\$} \mathsf{Binom}(1, 1-p)$.	22. Output $C \leftarrow \sum_{i\in(A..m]} \delta_i$.	32. For $i \in [0..M]$,
13. $\quad$ If $\rho_i = 0$, then $\delta_i \xleftarrow{\$} \mathcal{X}_\lambda$.		33. $\quad$ If $C = \sum_{i\in(A..m]} \delta_i$, output m.
14. $\quad$ Else $\delta_i \leftarrow 1$		34. Output $\perp$.
15. Output $K \leftarrow (\delta_i)_{i\in(A..M]}$.		

Fig. 2. The Scheme of Section 6 and its Parameters

below with high probability. (See Fig.1). (Below, we say "$\delta_i = \mathsf{Enc}_K(i) - \mathsf{Enc}_K(i-1)$ is in interval I" to mean that both integers $i-1$ and i used to define δ_i are contained in I.)[5] The three properties are:

$$\text{All } \delta_i \text{ in } [m_0..m_1] \text{ are } 1, \tag{6.2}$$

$$\text{Some } \delta_{i_0} \text{ in } [m_0 - k\theta..m_0] \text{ is large}, \tag{6.3}$$

$$\text{Some } \delta_{i_1} \text{ in } [m_1..m_1 + k\theta] \text{ is large}. \tag{6.4}$$

Encryption query m has to satisfy $m \le m_0 - k\theta$ or $m \ge m_1 + k\theta$ due to the constraints (5.3) and (5.4) of (k,θ)-FTG-O-nCPA. In the former case, the difference $\mathsf{Enc}_K(m_b) - \mathsf{Enc}_K(m) = \sum_{i\in(m..m_b]} \delta_i = \sum_{i\in(m..m_0]} \delta_i + \sum_{i\in(m_0..m_b]} \delta_i$ contains the large dominant randomness δ_{i_0} as a summand. Since the term $\sum_{i\in(m_0..m_b]} \delta_i$ depending on b can be hidden by δ_{i_0}, an adversary cannot detect b from $\mathsf{Enc}_K(m_b) - \mathsf{Enc}_K(m)$.

In the latter case, similarly, the sum $\mathsf{Enc}_K(m) - \mathsf{Enc}_K(m_b) = \sum_{i\in(m_b..m]} \delta_i$ contains the other large dominant randomness δ_{i_1}. An adversary therefore cannot detect b from $\mathsf{Enc}_K(m) - \mathsf{Enc}_K(m_b)$ due to a similar argument as above. Thus, our scheme is (k,θ)-FTG-O-nCPA secure.

[5] That is, δ_i is in $I = [a..b]$ iff $i \in (a..b]$. Seemingly asymmetry of the interval, which is a "left-open" one $(a..b]$ but is not "right open" one $[a..b)$, comes from how we number δ_i. If we set δ_i not to $\mathsf{Enc}_K(i) - \mathsf{Enc}_K(i-1)$ but to $\mathsf{Enc}_K(i+1) - \mathsf{Enc}_K(i)$, it becomes a right open one $[a..b)$.

6.2 Description

The formal description of the scheme is given in Fig.2. Here $\mathsf{Binom}(n,p)$ is a binomial distribution. $\mathcal{X}_\lambda$ is a probability distribution such that a random variable taken from it can hide other values. Specifically, $\mathcal{X}_\lambda$ satisfies the following property (where SD denotes statistical distance):

$$\exists \xi \ : \ \text{(negligible func.)}, \quad \forall \alpha, \beta \in [-\theta..\theta], \text{ for } \delta \xleftarrow{\$} \mathcal{X}_\lambda, \ \mathsf{SD}(\alpha+\delta, \beta+\delta) \leq \xi(\lambda). \tag{6.5}$$

We can use the uniform distribution on $[1..2^\lambda\theta]$ as $\mathcal{X}_\lambda$ for example. (See Section 2.3 for details.) But we will use another distribution due to a technical reason when we will improve this scheme in Section 7.

The value $p > 0$ is set as follows (so as to show the advantage bound of (5.7)):

$$p = 1 - \left(1 - \frac{1}{\sqrt{k}}\right)^{\frac{1}{\theta}}.$$

The sum of δ_i in the scheme begins from $i = A+1$, where $A = -k\theta - 1$ since otherwise, $\mathsf{Enc}_K(0)$ is always 0 and the scheme, in turn, becomes insecure.

Message Space Size: The message space size M of this scheme has to satisfy $M = \mathsf{poly}(\lambda)$ because the encryption cost of this scheme is clearly $O(M)$. We will remove this restriction in Section 7.

Upper Bound on Advantage: The advantage is calculated as follows. Let E_1, E_2, and E_3 be, respectively, the events that condition (6.2), (6.3), and (6.4) does *not* hold and Bad be $E_1 \vee E_2 \vee E_3$. Then, the previous discussion showed that the advantage of an adversary for our scheme is less than $\Pr[\mathsf{Bad}] + \mathsf{neg}(\lambda)$.

Recall that nCPA adversary has to make her challenge query (m_0, m_1) and encryption queries at the same time. Hence, she has to determine her challenge query (m_0, m_1) without knowing any information about ciphertexts, in particular, any information about δ_i. Therefore, the distributions of $(\delta_i)_i$ and (m_0, m_1) are independent. Since they are independent, E_1, E_2, E_3 are smaller than $1-(1-p)^\theta = 1/\sqrt{k}$, $(1-p)^{k\theta} = (1-1/\sqrt{k})^k$, and $(1-p)^{k\theta} = (1-1/\sqrt{k})^k$, respectively. Due to the same reason, it follows that

$$\Pr[\mathsf{Bad}] \leq \frac{1}{\sqrt{k}} + 2\left(1-\frac{1}{\sqrt{k}}\right)^k = \frac{1}{\sqrt{k}} + 2\left\{\left(1-\frac{1}{\sqrt{k}}\right)^{\sqrt{k}}\right\}^{\sqrt{k}} = \frac{1}{\sqrt{k}} + O\left(\mathrm{e}^{-\sqrt{k}}\right) = O\left(\frac{1}{\sqrt{k}}\right), \tag{6.6}$$

which is the bound given in Theorem 15 (to formally be proven below).

About CPA Security: The above proof does not work well when we consider the CPA attack because our proof crucially relies on the independence of the distributions of challenge query (m_0^*, m_1^*) and $(\delta_i)_i$ which is ensured in the nCPA setting. In fact, a CPA adversary can choose (m_0^*, m_1^*) in the region $(m_i..m_{i+1}]$ where the length of $(m_i..m_{i+1}]$ is the smallest, where $m_1 < \ldots < m_q$ are the encryption queries. Then, since $[m_i..m_{i+1}]$ is the smallest, $\mathsf{Enc}_K(m_{i+1}) - \mathsf{Enc}_K(m_i) = \sum_{i \in (m_i..m_{i+1}]} \delta_i$ must be small. Therefore, δ_i contained in $(m_i..m_{i+1}]$ must also be small. This means that the probability that conditions (6.3) and (6.4) hold must be smaller than that of the case of nCPA, which is the reason why our proof does not work well in this case.

7 (k, θ)-FTG-O-nCPA Secure Scheme with No Message Space Restrictions

Finally, we improve the scheme of Section 6 and achieve an OPE scheme whose encryption and decryption costs are $O(\log M)$, where M is the number of elements in the message space. This improvement enables the message space size M to become a super-polynomial in the security parameter λ, allowing the encryption and decryption algorithms to stay polynomial time even in this case.

Idea of [15]: The starting point of our improvement is the following idea of Section 6 of [15]. They construct a polynomial time algorithm[6] $\bar{G}$ which takes two pairs (u, C_u) and (v, C_v) of messages and their encryptions, and outputs a data whose distribution is the same as that of ciphertext C_w of w, where w is the "midpoint" $\lceil (u+v)/2 \rceil$ of u and v. They then compute a ciphertext C_m of m using $\bar{G}$ based on a binary search recursion. Specifically, their improved encryption algorithm $\overline{\mathsf{Enc}}(m)$ takes some initial values u, v such that $m \in (u..v]$ holds and C_u and C_v are known. (We denote by Init an algorithm which outputs the encryption C_u and C_v of the initial values.) $\overline{\mathsf{Enc}}(m)$ then computes C_w using $\bar{G}$, replaces interval $(u..v]$ with $(u..w]$ or $(w..v]$ depending on whether $m \leq w$ or not, and recursively executes $\overline{\mathsf{Enc}}$ itself. The computational cost of $\overline{\mathsf{Enc}}$ is $O(\log M)$, where M is the message space size, because the binary search recursion is terminated in time $O(\log M)$. Their decryption algorithm $\overline{\mathsf{Dec}}$ is constructed in a similar fashion.

The Idea Behind Our Scheme: Our efficient encryption and decryption algorithms are constructed based on the above idea, but our innovation is that our binary search recursion $\bar{G}$ and initializing algorithm Init are constructed based not on a ciphertext C_u itself but on I_u defined below. This is so, since our elaborated scheme of Section 6 does not allow construction of $\bar{G}$ to be based simply on C_u. Below, ρ_i, δ_i, and A are as defined in the scheme of Section 6.

$$I_u \leftarrow (C_u^{(0)}, C_u^{(1)}) \leftarrow \Big(\sum_{\substack{i \in (A..u] \\ \rho_i = 0}} \delta_i, \sum_{\substack{i \in (A..u] \\ \rho_i = 1}} \delta_i \Big), \tag{7.1}$$

We will construct Init and $\bar{G}$ satisfying the following properties:

Output Init is indistinguishable from (I_A, I_M). (7.2)

For any $u, v \in (A..M]$ and any I'_u and I'_v, the distribution of an output of $\bar{G}(u, v, I'_u, I'_v)$ is the same as the conditional distribution of I_w when $(I_u, I_v) = (I'_u, I'_v)$ holds. Here $w = \lceil (u+v)/2 \rceil$. (7.3)

Then our efficient encryption algorithm can get I_m in time logarithm $O(\log M)$ in the message space size M by executing a recursion based on Init and $\bar{G}$. It can get the ciphertext of m from $I_m = (C_m^{(0)}, C_m^{(1)})$ because an encryption $\mathsf{Enc}_K(m)$ of Section 6 is $\sum_{i \in (A..u]} \delta_i$, and therefore satisfies

$$\mathsf{Enc}_K(m) = C_m^{(0)} + C_m^{(1)}. \tag{7.4}$$

As in the case of [16,15], the efficient decryption algorithm is also constructed based on a similar idea.

Ideas Behind the Construction of Init and $\bar{G}$: The remaining issue to take care of is the construction of Init and $\bar{G}(I_u, I_v)$. To this end, we set $\mathcal{X}_\lambda$ of (6.5) to a binomial distribution

$$\mathcal{X}_\lambda = \mathcal{B}(2^\lambda \theta^2, 1/2) \tag{7.5}$$

[6] To simplify, here we only consider the case where inputs of $\bar{G}$ are (u, C_u) and (v, C_v), although [15] considers more general case due to some technical reasons.

with suitable parameters. Note that this $\mathcal{X}_\lambda$, in fact, satisfies (6.5), which is the property required to ensure the security of the scheme of Section 6. Formally, the following fact holds (See Section 8.1 for the proof):

Proposition 16 (Binomial Satisfies (6.5)) *There exists a negligible function ξ such that for all $\alpha, \beta \in [-\theta..\theta]$, the statistical distance between the random variables $\alpha + \delta$ and $\beta + \zeta$ for $\delta, \zeta \xleftarrow{\$} \mathcal{B}(2^\lambda\theta^2, 1/2)$ is less than $\xi(\lambda)$.*

(7.5) allows us to write I_u by using two binomial distributions because (7.1) shows that I_u can be written as sums of δ_i, step 13 of Fig.2 and (7.5) show that δ_i is taken from a binomial distribution, and the sum of binomials is also binomial. Since $I_A = (0, 0)$, this means that our algorithm Init satisfying (7.2) can be constructed by using two binomial distributions for generating I_M.

Moreover, it is also known that the conditional distributions of binomials can be written as hypergeometric distributions. (See Section 2.3.) Hence, our algorithm $\bar{G}$ satisfying (7.3) can be constructed by using hypergeometric distributions. Since the values which follow the binomial and hypergeometric distributions can be generated in polynomial time [24], our algorithms Init and $\bar{G}$ can terminate in polynomial time.

The description of our algorithms $\bar{G}$ and Init is given in Fig.3. Here $\mathsf{Binom}(n, p)$ and $\mathsf{HG}(a, b, c)$ are algorithms whose outputs follow binomial distribution and hypergeometric distribution. We can show that our algorithms Init and $\bar{G}$ in fact satisfy (7.2) and (7.3); see Section 8 for the proof.

Proposition 17 (Init and $\bar{G}$ Work Well) *Let A and M be constants given in Fig.3. Let $(\delta_i)_{i\in(A..M]}$ and $(\rho_i)_{i\in(A..M]}$ be tuples which are generated as in $\mathsf{Kg}(1^\lambda)$ of Fig.2. Define I_u as is (7.1). Then (7.2) and (7.3) hold.*

We denote the encryption function given in the above way by $\widetilde{\mathsf{Enc}}$. Then, from (7.2), (7.3), and the construction of $\widetilde{\mathsf{Enc}}$, the following proposition holds. (See Section 8 for the formal proof.)

Proposition 18 *Take A, M, Kg, Enc, $\overline{\mathsf{Kg}}$, and $\overline{\mathsf{Enc}}$ as in Fig.2 and 3. Then for $\bar{K} \leftarrow \overline{\mathsf{Kg}}(1^\lambda)$ and $K \leftarrow \mathsf{Kg}(1^\lambda)$, the distributions of $(\widetilde{\mathsf{Enc}}_{\bar{K}}(i))_{i\in[A..M]}$ and $(\mathsf{Enc}_K(i))_{i\in[A..M]}$ are perfectly indistinguishable.*

Finally, we replace the randomness of $\widetilde{\mathsf{Enc}}$ with a pseudo-random value output by a pseudo-random function, so as to make it deterministic, as in [16,15]. Then our final encryption algorithm $\overline{\mathsf{Enc}}$ is obtained.

Description: The formal description of our scheme is given in Fig.3. Here k and θ are the values which we want to show (k, θ)-FTG-nCPA security for, M is the value such that the message space is $[1..M]$, and p and A are the same values used in the scheme of Section 6. Cph, in turn, is an algorithm which computes a ciphertext C_u from I_u based on (7.4). The notation $\bar{G}(u, v, I_u, I_v; cc)$ means that we compute $\bar{G}(u, v, I_u, I_v)$ using cc as the random tape. PRF is a pseudorandom function.

Order-Preserving Property: The order-preserving property of $\overline{\mathsf{Enc}}$ can be shown from the same property for Enc of Section 6.2 because of Proposition 18.

Correctness: For $C = \overline{\mathsf{Enc}}_{\bar{K}}(m)$, it is shown as follows. The algorithm of $\overline{\mathsf{Dec}}_{\bar{K}}(C)$ is the same as that of $\overline{\mathsf{Enc}}_{\bar{K}}(m)$ except that the conditional branches (52) and (56) of Fig.3 are replaced with (62) and (66) of the same figure. However, the replaced conditions such as "$C \leq \mathsf{Cph}(I_w)$" are essentially the same as those of the original ones such as "$m \leq w$" because

$$m \leq w \Leftrightarrow \overline{\mathsf{Enc}}_{\bar{K}}(m) \leq \overline{\mathsf{Enc}}_{\bar{K}}(w) \Leftrightarrow C \leq \mathsf{Cph}(I_w) \tag{7.6}$$

Message Space $= [1..M]$, $p = 1-(1-1/\sqrt{k})^{1/\theta}$, $A = -k\theta - 1$.

$\overline{\mathsf{Kg}}(1^\lambda)$
41. Randomly take λ bit string K'.
42. $(I_A, I_M) \leftarrow \mathsf{Init}(1^\lambda)$.
43. Return $\bar{K} \leftarrow (K', A, M, I_A, I_M)$.

$\overline{\mathsf{Enc}}_{\bar{K}}(m)$
51. Parse $\bar{K}$ as (K', u, v, I_u, I_v).
52. If $m = v$ holds,
    return $\mathsf{Cph}(I_v)$.
53. $w \leftarrow \lceil (u+v)/2 \rceil$.
54. $cc \leftarrow \mathsf{PRF}_{K'}(u, v)$
55. $I_w \leftarrow \bar{G}(u, v, I_u, I_v; cc)$
56. Return $\begin{cases} \overline{\mathsf{Enc}}_{(K',u,w,I_u,I_w)}(m) & \text{if } m \le w \\ \overline{\mathsf{Enc}}_{(K',w,v,I_w,I_v)}(m) & \text{otherwise} \end{cases}$

$\overline{\mathsf{Dec}}_{\bar{K}}(C)$
61. Parse $\bar{K}$ as (K', u, v, I_u, I_v).
62. If $C = \mathsf{Cph}(I_v)$ or $u = v$ holds,
    return v or $\perp$ respectively.
63. $w \leftarrow \lceil (u+v)/2 \rceil$.
64. $cc \leftarrow \mathsf{PRF}_{K'}(u, v)$
65. $I_w \leftarrow \bar{G}(u, v, I_u, I_v; cc)$
66. Return $\begin{cases} \overline{\mathsf{Dec}}_{(K',u,w,I_u,I_w)}(m) & \text{if } C \le \mathsf{Cph}(I_w) \\ \overline{\mathsf{Dec}}_{(K',w,v,I_w,I_v)}(m) & \text{otherwise} \end{cases}$

$\mathsf{Init}(1^\lambda)$
81. $C_M^{(1)} \leftarrow \mathsf{Binom}(M-A, 1-p)$,
82. $C_M^{(0)} \leftarrow \mathsf{Binom}(2^\lambda\theta^2(M-A-C_M^{(1)}), 1/2)$,
83. $I_A \leftarrow (0,0)$, $I_M \leftarrow (C_M^{(0)}, C_M^{(1)})$.
84. Output (I_A, I_M).

$\mathsf{Cph}(I)$
71. Parse I as $(C^{(0)}, C^{(1)})$.
72. Output $C^{(0)} + C^{(1)}$.

$\bar{G}(u, v, I_u, I_v)$
91. Parse I_u and I_v as $(C_u^{(0)}, C_u^{(1)})$ and $(C_v^{(0)}, C_v^{(1)})$. $w \leftarrow \lceil (u+v)/2 \rceil$.
92. $C_w^{(1)} \leftarrow C_u^{(1)} + \mathsf{HG}(v-u, C_v^{(1)} - C_u^{(1)}, w-u)$,
93. $C_w^{(0)} \leftarrow C_u^{(0)} + \mathsf{HG}(2^\lambda\theta^2((v-u) - (C_v^{(1)} - C_u^{(1)})), C_v^{(0)} - C_u^{(0)}, 2^\lambda\theta^2((w-u) - (C_w^{(1)} - C_u^{(1)})))$,
94. Output $I_w \leftarrow (C_w^{(0)}, C_w^{(1)})$.

Fig. 3. The Scheme of Section 7 and its Parameters and Subroutines

holds due to the order preserving property, (7.1), (7.4), and the definition of Cph. Hence $\overline{\mathsf{Enc}}_{\bar{K}}(m)$ and $\overline{\mathsf{Dec}}_{\bar{K}}(C)$ execute the same recursion and compute the same data. In particular, $m \in (u..v]$ always holds in the recursion of $\overline{\mathsf{Dec}}_{\bar{K}}(C)$ because this property holds in tht of $\overline{\mathsf{Enc}}_{\bar{K}}(m)$ by definition. Since these algorithms cut $(u, v]$ by half in each recursions, they finally assign $m = v$. Due to a similar reason to (7.6), the condition $C = \mathsf{Cph}(I_v)$ of (62) is equivalent to $m = v$, which is now satisfied. Hence, it outputs $v = m$ at this step. (Note that the other condition $u = v$ does not occur at step (62) because $m \in (u..v]$ always holds in the recursion and therefore $u \lneqq m \le v$.)

Security: Theorem 15 follows from Proposition 16, 17, and 18, and the security of the scheme of Section 6. We will prove this fact formally in Section 8.6.

Ciphertext and Key Lengths: We estimate these values when $M \ge k\theta$, $k \ge 1$, and $M \ge 1$ hold, since by (5.3) (k, θ)-FTG-O-nCPA is meaningful only when $k\theta$ is not larger than the message space size M, (5.7) is about $k \to \infty$, and the message space $[1..M]$ has to have at least one element. In this case, the bit lengths of a ciphertext is not more than $\lambda + \lfloor 3 \log M \rfloor + 3$, because a ciphertext can be represented as the sum of positive numbers $\delta_A, \ldots, \delta_M$, δ_i is not more than $2^\lambda\theta^2$ (due to (7.5)), and $A = -k\theta - 1$ holds. Due to similar reasons, the key length is not more than $4\lambda + \lfloor 14 \log M \rfloor + 14$ + (the PRF's key length).

8 Formal Security Proof of Our Scheme of Section 7

The goal of this section is to prove that our scheme of Section 7 is (k, θ)-FTG-O-nCPA secure with advantage bound as given in (5.7). We prove this fact by reducing it to the security of the scheme of Section 6. Hence, security of the scheme of Section 6 is shown as well from the proof of this section.

We use the notations of Section 7 throughout this section. The rest of this section is as follows. We will prove Proposition 16 in Section 8.1. Then, we will give notation and facts for the proof of Proposition 17 in Section 8.2, and will prove (7.2) and (7.3) in Section 8.3 and 8.4, respectively, which, in turn, will imply Proposition 17. We will then prove Proposition 18 in Section 8.5. Finally, we will show in Section 8.6 that our scheme is (k, θ)-FTG-O-nCPA secure whose advantage bound is given in (5.7).

8.1 Proof of Proposition 16

To show Proposition 16, we show the following lemma:

Lemma 19 *Let μ be a natural number and X be a random variable which is distributed according to $\mathcal{B}(\mu, 1/2)$. Then when $\mu \to \infty$, it follows that*

$$\max_x \Pr[X = x] = O(\sqrt{1/\mu}). \tag{8.1}$$

Proof (Lemma 19). For positive valued functions $f(x)$ and $g(x)$, let "$g = \Theta(f)$" denotes the fact that both g/f and f/g converge, each to a finite non-zero value when $x \to \infty$. Let $\mathrm{e} = 2.71828...$. From Stirling series

$$\mu! = \sqrt{2\pi\mu}\left(\frac{\mu}{\mathrm{e}}\right)^{\mu}\left(1 + \frac{1}{12\mu} + \cdots\right) = \sqrt{2\pi\mu}\left(\frac{\mu}{\mathrm{e}}\right)^{\mu}\left(1 + \Theta\left(\frac{1}{\mu}\right)\right),$$

it follows that

$$\max_x \Pr[X = x] = \Pr[X = \lceil \mu/2 \rceil] = \frac{1}{2^{\mu}}\binom{\mu}{\lceil \mu/2 \rceil} = \frac{\mu!}{2^{\mu}(\lceil \mu/2 \rceil !)^2} \le \sqrt{\frac{2}{\pi\mu}} \cdot \frac{1 + \Theta(\frac{1}{\mu})}{1 + \Theta(\frac{1}{\mu})} = O(1/\sqrt{\mu}). \square$$

Proof (Proposition 16). Let

$$\eta = 2^{\lambda}\theta^2. \tag{8.2}$$

Take arbitrary

$$\alpha, \beta \in [-\theta..\theta] \tag{8.3}$$

Let $U = \alpha + \delta$ and $V = \beta + \zeta$ be random variables where $\delta, \zeta \leftarrow \mathcal{B}(\eta, 1/2)$. W.l.o.g. we can assume $\alpha \ge \beta$. Let $q(x)$ be the probability mass function of $\mathcal{B}(\eta, 1/2)$ and $p(x)$ be $2^{\eta}q(x)$. in other words,

$$q(x) = \Pr[\delta \xleftarrow{\$} \mathcal{B}(\eta, 1/2) \ : \ \delta = x] = \frac{1}{2^{\eta}}\binom{\eta}{x}, \qquad p(x) = \binom{\eta}{x}$$

Then the probability mass functions of the distributions of U and V are $q(x - \alpha)$ and $q(x - \beta)$ respectively. Let p_{max} be $\max_x p(x)$. Then,

$$p_{\mathsf{max}} = \max_x p(x) = 2^{\eta} \max_x q(x) \underset{(8.1)}{=} O(2^{\eta}/\sqrt{\eta}) \tag{8.4}$$

Let x_{mid} be $(\eta/2) + (\alpha + \beta)/2$.

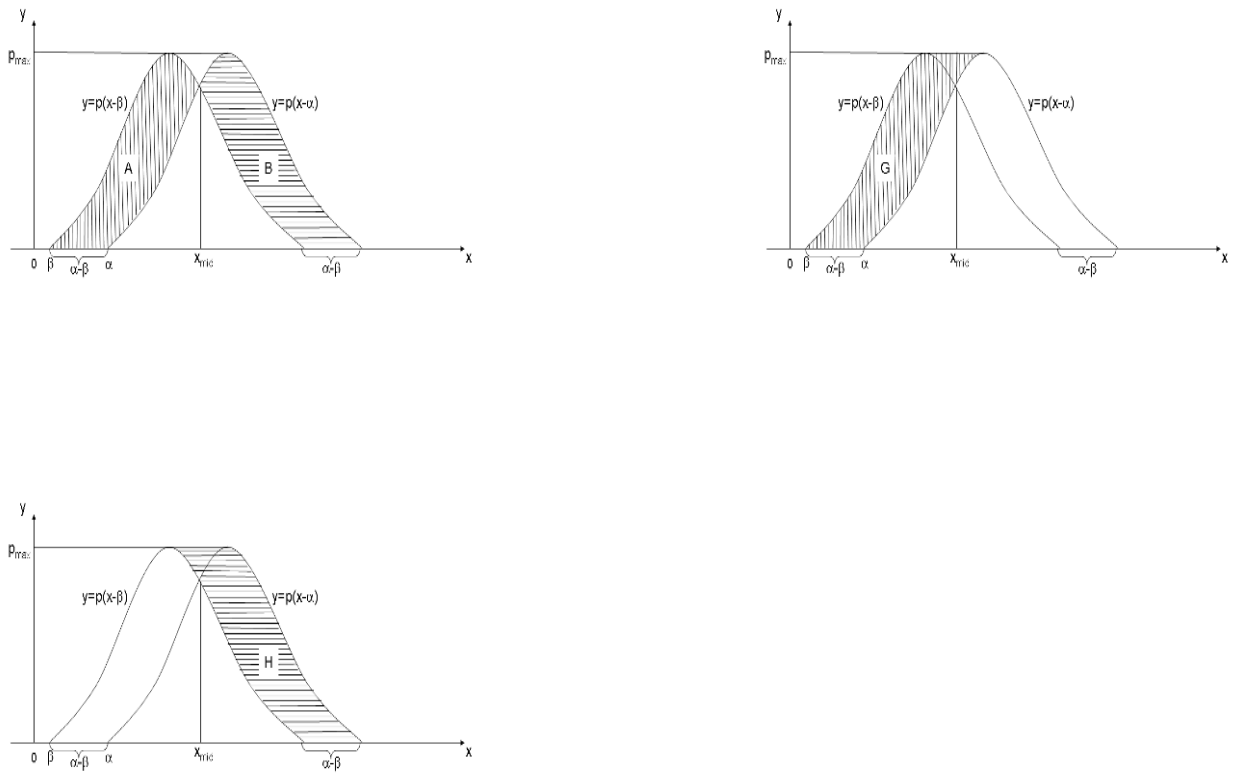

Fig. 4. Sets A, B, G, H.

Define sets A, B, G, $H \subset \mathbb{Z}^2$ as follows. (See Fig.4.)

$$A = \{(x, y) \in \mathbb{Z} \times \mathbb{Z}_{\geq 0} \mid p(x - \alpha) < y \leq p(x - \beta)\},$$
$$B = \{(x, y) \in \mathbb{Z} \times \mathbb{Z}_{\geq 0} \mid p(x - \beta) < y \leq p(x - \alpha)\},$$

$$G = \{(x, y) \in \mathbb{Z} \times \mathbb{Z}_{\geq 0} \mid 0 \leq y \leq p_{\mathsf{max}}, \quad x \in [\rho(y) + \beta..\rho(y) + \alpha]\},$$
$$H = \{(x, y) \in \mathbb{Z} \times \mathbb{Z}_{\geq 0} \mid 0 \leq y \leq p_{\mathsf{max}}, \quad s \in [\tau(y) + \beta..\tau(y) + \alpha]\},$$

where

$$\rho(y) = \min_x \{p(x) \geq y\}$$
$$\tau(y) = \max_x \{p(x) \geq y\}.$$

We will prove the following five facts, where $\#G$ denotes the number of elements of G:

$$A \subset G, \qquad B \subset H. \tag{8.5}$$
$$\#G = (\beta - \alpha + 1)(p_{\mathsf{max}} + 1), \qquad \#H = (\beta - \alpha + 1)(p_{\mathsf{max}} + 1) \tag{8.6}$$

$$\mathsf{SD}(U,V) = \frac{\#A + \#B}{2^\eta}, \tag{8.7}$$

From the above facts, Proposition 16 can be shown as follows:

$$\begin{aligned}\mathsf{SD}(U,V) &\underset{(8.7)}{=} \frac{\#A+\#B}{2^\eta} \underset{(8.5)}{\leq} \frac{\#G+\#H}{2^\eta} \underset{(8.6)}{=} \frac{2(\beta-\alpha+1)(p_{\mathsf{max}}+1)}{2^\eta} \\ &\underset{(8.4)}{=} O\left(\frac{\beta-\alpha+1}{\sqrt{\eta}}\right) \underset{\substack{(8.2)\\(8.3)}}{=} O(1/2^{\lambda/2}).\end{aligned}$$

Before proceeding, we give the intuition behind the proofs of (8.5), (8.6), and (8.7). First, (8.5) and (8.7) clearly follow from Fig.4. The first equation of (8.6) holds because from Fig.4, the number of elements of G (which is almost the same as "the area of G") can be computed as the product of "the length of the base" $(\beta-\alpha+1)$ and the "height" $p_{\mathsf{max}}+1$. The second equation of (8.6) holds by similar reasoning.

We finally prove (8.5), (8.6), and (8.7) formally.

proof of (8.5): Here, we only prove $A \subset G$. The proof for $B \subset H$ is similar. To this end, we show some facts which will be used in showing $A \subset G$. By the definition of ρ, the following fact holds for any y and y'.

$$y \leq y' \Rightarrow \rho(y) = \min_x\{p(x) \geq y\} \leq \min_x\{p(x) \geq y'\} = \rho(y') \tag{8.8}$$

That is, ρ is monotonically increasing.

Since an integer x is clearly an element of the set $\{u \mid p(u) \geq p(x)\}$, it follows that for any x,

$$x \geq \min_u\{p(u) \geq p(x)\} = \rho(p(x)), \tag{8.9}$$

where the last equation holds by the definition of ρ. Hence, it follows that

$$\begin{aligned}(x,y) \in A &\underset{\text{def. of } A}{\Leftrightarrow} p(x-\alpha) < y \leq p(x-\beta) \underset{(8.8)}{\Rightarrow} \rho(p(x-\alpha)) \leq \rho(y) \leq \rho(p(x-\beta)) \\ &\underset{(8.9)}{\Rightarrow} \rho(p(x-\alpha)) \leq \rho(y) \leq x-\beta.\end{aligned} \tag{8.10}$$

It is well known that the function $p(x) = \binom{\eta}{x}$ is monotonically increasing at $x \leq \lfloor \eta/2 \rfloor$ and is monotonically decreasing at $x \geq \lceil \eta/2 \rceil$. Using these facts, we can show that for any x' and x'',

$$x' \leq x'' \ \wedge \ p(x') < p(x'') \quad \Rightarrow \quad x' \leq \lfloor \eta/2 \rfloor \tag{8.11}$$

holds. This fact can be shown by contradiction. If $x' > \lfloor \eta/2 \rfloor$, $x'' \geq x' > \lfloor \eta/2 \rfloor$ holds from the assumption. But since $p(x)$ is monotonically decreasing on $x > \lfloor \eta/2 \rfloor$, $p(x') \geq p(x'')$ has to holds. This contradicts the assumption $p(x') < p(x'')$.

Since $\alpha \geq \beta$,

$$x - \alpha \leq x - \beta \tag{8.12}$$

holds for any x. By applying (8.11) and (8.12) to $x' = x-\alpha$ and $x'' = x-\beta$, it follows that

$$(x,y) \in A \underset{\text{def. of } A}{\Leftrightarrow} p(x-\alpha) < y \leq p(x-\beta) \underset{\substack{(8.11)\\(8.12)}}{\Rightarrow} x - \alpha \leq \lfloor \eta/2 \rfloor \tag{8.13}$$

As mentioned before, the function $p(x) = \binom{\eta}{x}$ is monotonically increasing at $x \le \lfloor \eta/2 \rfloor$ and is monotonically decreasing at $x \ge \lceil \eta/2 \rceil$. Hence, the value $\rho(p(x)) = \min\{u \mid p(u) \ge p(x)\}$ is x itself if $x \le \lfloor \eta/2 \rfloor$. That is,

$$\text{if } x \le \lfloor \eta/2 \rfloor,\ \rho(p(x)) = x \tag{8.14}$$

Hence, it follows that

$$(x,y) \in A \underset{\substack{(8.14)\\(8.13)\\(8.10)}}{\Rightarrow} x - \alpha \le \rho(y) \le x - \beta \Leftrightarrow \rho(y) + \beta \le x \le \rho(y) + \alpha \underset{\text{def. of } G}{\Leftrightarrow} (x,y) \in G.$$

This is what we want to prove.

Proof of (8.6):

$$\#G = \sum_{(x,y)\in G} 1 = \sum_{y\in[0..p_{\mathsf{max}}]} \sum_{x\in[\rho(y)+\beta..\rho(y)+\alpha]} 1 = (p_{\mathsf{max}}+1)(\beta-\alpha+1).$$

The second equation of (8.6) can be shown similarly.

Proof of (8.7):

$$\begin{aligned}
\mathsf{SD}(U,V) =& \sum_x |q(x-\alpha) - q(x-\beta)| \\
=& \frac{1}{2^\eta} \sum_x |p(x-\alpha) - p(x-\beta)| \\
\le& \frac{1}{2^\eta} \sum_{x\le x_{\mathsf{mid}}} (p(x-\beta) - p(x-\alpha)) + \frac{1}{2^\eta} \sum_{x\ge x_{\mathsf{mid}}} (p(x-\alpha) - p(a-\beta)) \\
=& \frac{1}{2^\eta} \sum_{x\le x_{\mathsf{mid}}} \sum_{y\in(p(x-\alpha)..p(x-\beta)]} 1 \quad + \quad \frac{1}{2^\eta} \sum_{x\ge x_{\mathsf{mid}}} \sum_{y=(p(x-\beta)..p(x-\alpha)]} 1 \\
\underset{\text{def. of } A,B}{=}& \frac{\#A + \#B}{2^\eta}
\end{aligned}$$

□

8.2 Notation and Facts for the proof of Proposition 17

Next, we present some notations and show equalities which will be used to prove Propositions 17 and 18. Let δ_i and ρ_i be values given in the statement of Proposition 17 (and therefore in Fig.2.) For any integers s, t, we let

$$I_{s,t} \leftarrow (C^{(0)}_{s,t}, C^{(1)}_{s,t}) \leftarrow \Big(\sum_{\substack{i\in(s..t]\\ \rho_i=0}} \delta_i, \sum_{\substack{i\in(s..t]\\ \rho_i=1}} \delta_i \Big), \tag{8.15}$$

Then the value $I_u = (C^{(0)}_u, C^{(1)}_u)$ given in the statement of Proposition 17 satisfies

$$I_u = (C^{(0)}_u, C^{(1)}_u) = (C^{(0)}_{A,u}, C^{(1)}_{A,u}) = I_{A,u}. \tag{8.16}$$

The following equations clearly follow from (8.15) for any r, s, t :

$$C^{(1)}_{r,s} + C^{(1)}_{s,t} = C^{(1)}_{r,t} \qquad C^{(0)}_{r,s} + C^{(0)}_{s,t} = C^{(0)}_{r,t} \tag{8.17}$$

In particular, the following equation holds, where $I_{r,s} + I_{s,t}$ denote the component-wise sum.

$$I_{r,s} + I_{s,t} = I_{r,t} \tag{8.18}$$

8.3 Proof of (7.2) of Proposition 17

Let δ_i, ρ_i, A, and M be values given in the statement of Proposition 17 (and therefore in Fig.2.) Consider the values I_A and $I_M = (C^{(0)}_M, C^{(1)}_M)$ given by (7.1). We will show that for any s, t,

(x) $I_A = (0, 0)$,
(y) $C^{(1)}_{s,t} \sim \mathcal{B}(t - s, 1 - p)$,
(z) $C^{(0)}_{s,t} \sim \mathcal{B}(2^\lambda \theta^2 (t - s - C^{(1)}_{s,t}), 1/2)$.

(7.2) clearly follows from them due to (7.1), (8.15), (8.16), and the definition of Init given in Fig.3. (x) clearly follows from (7.1). The proof of (y) is as follows.

$$\begin{aligned} C^{(1)}_{s,t} &\underset{(8.15)}{=} \sum_{\substack{i \in (s..t] \\ \rho_i = 1}} \delta_i \\ &\underset{(a)}{=} (\text{the number of } i \in (s..t] \text{ satisfying } \rho_i = 1) \\ &\underset{(b)}{=} \sum_{i \in (s..t]} \rho_i \\ &\underset{(c)}{\sim} \mathcal{B}(t - s, 1 - p), \end{aligned} \tag{8.19}$$

where

- (a) holds because step 14 of Fig.2 shows that $\delta_i = 1$ holds when $\rho_i = 1$.
- (b) holds because step 12 of Fig.2 ensures $\rho_i \in \{0, 1\}$.
- (c) holds because ρ_i follows $\mathcal{B}(1, 1 - p)$ from step 12 of Fig.2 and each ρ_i is independent from each other and because the sum of binomial follows binomial as well.

Finally, the proof of (z) is as follows. Below, V is the number of $i \in (s..t]$ satisfying $\rho_i = 0$.

$$C^{(0)}_{s,t} \underset{(8.15)}{=} \sum_{\substack{i \in (s..t] \\ \rho_i = 0}} \delta_i \underset{(d)}{\sim} \mathcal{B}(2^\lambda \theta^2 V, 1/2) \underset{(e)}{=} \mathcal{B}(2^\lambda \theta^2 (t - s - C^{(1)}_{s,t}), 1/2). \tag{8.20}$$

Above, (d) holds due to the following three reasons.

- When $\rho_i = 0$, $\delta_i \sim \mathcal{B}(2^\lambda \theta^2, 1/2)$ holds because step 13 of Fig.2 shows that δ_i for $\rho_i = 0$ is taken from $\mathcal{X}_\lambda \underset{(7.5)}{=} \mathcal{B}(2^\lambda \theta^2, 1/2)$.
- The number of $i \in (s..t]$ satisfying $\rho_i = 0$ is V by the definition of V.
- The sum of V values, each of which follows $\mathcal{B}(2^\lambda \theta^2, 1/2)$, is $\mathcal{B}(2^\lambda \theta^2 V, 1/2)$.

(e) holds because (8.19), in particular, shows that the number of $i \in (s..t]$ satisfying $\rho_i = 1$ is $C^{(1)}_{s,t}$, which means $V = t - s - C^{(1)}_{s,t}$.

8.4 Proof of (7.3) of Proposition 17

Define I_u and I_v as in (7.1). Let $w = \lceil (u + v)/2 \rceil$. Fix two constants $I'_u = (C'^{(0)}_u, C'^{(1)}_u)$ and $I'_v = (C'^{(0)}_v, C'^{(1)}_v)$ arbitrary. Our goal is to show that the conditional distribution of I_w under the condition $(I_u, I_v) = (I'_u, I'_v)$ is the same as the distribution of the output of $\bar{G}(u, v, I'_u, I'_v)$.

We will achieve our goal in the following three steps. First, we detect the distribution of $(I_{u,w}, I_{w,v})$ (under no condition), where $I_{u,w}$ and $I_{w,v}$ are values defined as in (8.15). Then, from (8.18), if the condition $(I_u, I_v) = (I'_u, I'_v)$ holds, it follows that

$$I'_v = I'_u + I_{u,w} + I_{w,v}. \tag{8.21}$$

Therefore, secondly, we detect the distribution of $(I_{u,w}, I_{w,v})$ under the condition (8.21).

From (8.18), if the condition $(I_u, I_v) = (I'_u, I'_v)$ holds, it follows that

$$I_w = I'_u + I_{u,w} \tag{8.22}$$

Hence, thirdly, we can obtain the conditional distribution of $I_w = I'_u + I_{u,w}$ because we already have detected the conditional distribution of $I_{u,w}$. Then, we can check that this conditional distribution is the same as the distribution of the output of $\bar{G}(u, v, I'_u, I'_v)$. This is what we want to show.

Now, we prove (7.3) based on the above three steps. First, the distributions of $I_{u,w} = (C^{(0)}_{u,w}, C^{(1)}_{u,w})$ and $I_{w,v} = (C^{(0)}_{w,v}, C^{(1)}_{w,v})$ are given as following because of (8.19) and (8.20):

$$C^{(1)}_{u,w} \sim \mathcal{B}(w-u, 1-p), \qquad C^{(1)}_{w,v} \sim \mathcal{B}(v-w, 1-p), \tag{8.23}$$

$$C^{(0)}_{u,w} \sim \mathcal{B}(2^\lambda\theta^2(w-u-C^{(1)}_{u,w}), 1/2) \qquad C^{(0)}_{w,v} \sim \mathcal{B}(2^\lambda\theta^2(v-w-C^{(1)}_{w,v}), 1/2). \tag{8.24}$$

Hence, next, observe that the distribution of $I_{u,w} = (C^{(0)}_{u,w}, C^{(1)}_{u,w})$ under condition (8.21) is as following due to Proposition 5:

$$C^{(1)}_{u,w} \sim \mathcal{HG}(v-u, C'^{(1)}_v - C'^{(1)}_u, w-u).$$
$$C^{(0)}_{u,w} \sim \mathcal{HG}(2^\lambda\theta^2(v-u-(C'^{(1)}_v - C'^{(1)}_u)), C'^{(0)}_v - C'^{(0)}_u, 2^\lambda\theta^2(w-u-(C'^{(1)}_v - C'^{(1)}_u))).$$

Therefore, from (8.17), we can conclude that

$$C^{(1)}_{A,w} \sim C^{(1)}_{A,u} + \mathcal{HG}(v-u, C'^{(1)}_v - C'^{(1)}_u, w-u). \tag{8.25}$$

$$C^{(0)}_{A,w} \sim C^{(0)}_{A,u} + \mathcal{HG}(2^\lambda\theta^2(v-u-(C'^{(1)}_v - C'^{(1)}_u)), C'^{(0)}_v - C'^{(0)}_u, 2^\lambda\theta^2(w-u-(C'^{(1)}_v - C'^{(1)}_u))). \tag{8.26}$$

From (8.16), this means that the conditional distribution of $I_w = (C^{(0)}_w, C^{(1)}_w)$ given by (8.25) and (8.26) is the same as Step 93 and 92 of $\bar{G}(u, v, I'_u, I'_v)$ given in Fig.3. This, in fact, is what we wanted to prove.

8.5 Proof of Proposition 18

We use the notation of Section 7.

8.5.1 Preparations

Before proving Proposition 18, we introduce notations and terminologies, and show facts used in the proof of Proposition 18.

Successive Elements: Let U be any finite set of integers. Then, two elements u and v of U are said to be *successive* elements of U if, when we write $U = \{s_1, \ldots, s_n\}$ $(s_1 < \cdots < s_n)$, there exists $j \in [1..n-1]$ satisfying $(u, v) = (s_j, s_{j+1})$.

Slightly Stronger Variant of (7.3) Next, we introduce the following slightly stronger variant of (7.3), which we will use to prove Proposition 18. Below, I_w and I_s are values defined as in (7.1).

$$\begin{aligned}&\text{Take any set } U \subset [A..M] \text{ and any successive elements } u, v \in U \text{ and set } w = \lceil (u+v)/2 \rceil.\\ &\text{Take any tuple } (s, I'_s)_{s\in U} \text{ as well. Then the distribution of an output of } \bar{G}(u, v, I'_u, I'_v) \text{ is}\\ &\text{the same as the conditional distribution of } I_w \text{ when } (s, I_s)_{s\in U} = (s, I'_s)_{s\in U} \text{ holds.}\end{aligned} \tag{8.27}$$

We can prove (8.27) in a similar fashion to that of (7.3) due to the following reason. By definition, the difference between (8.27) and (7.3) is that the conditions "(s, I_s) = (given value)" for (8.27) are about $s \leq u$ and about $s \geq v$ while those for (7.3) are only for $s = u$ and $s = v$. However, the difference of them is not essential for the proof of (7.3) because the proof of (7.3) is mainly about $I_{u,w}$ and $I_{w,v}$, which are independent from the values I_s for $s < u$ and $s > v$ due to (8.15). Hence, we omit the proof of (8.27).

From (8.27), we can conclude the following fact as well, because if two random variable X and Y have the same distribution under condition $Z = z$ for any z, (Z, X) and (Z, Y) have the same distribution.

$$\begin{aligned}&\text{Take any set } U \subset [A..M] \text{ and any successive elements } u, v \in U \text{ and set } w = \lceil (u+v)/2 \rceil.\\ &\text{Then the distributions of } (s, I_s)_{s\in U} \cup \{(w, \bar{G}(u, v, I_u, I_v))\} \text{ and } (s, I_s)_{s\in U} \cup \{(w, I_w)\}\\ &\text{are the same.}\end{aligned} \tag{8.28}$$

Definition of $\widetilde{\mathsf{Enc}}$: Next, we clarify the details of the definition of $\widetilde{\mathsf{Enc}}_{\bar{K}}$ given in the statement of Proposition 18. This is the algorithm which is obtained from the encryption algorithm $\overline{\mathsf{Enc}}_{\bar{K}}$ of our proposed scheme of Fig.3, by replacing a pseudo-random function with the random oracle. Specifically, $\widetilde{\mathsf{Enc}}_{\bar{K}}$ is the same as $\overline{\mathsf{Enc}}_{\bar{K}}$ except that in the step (54) of Fig.3, it excecutes not "$cc \leftarrow \mathsf{PRF}_{K'}(u, v)$" but "$cc \leftarrow \mathcal{H}(u, v)$", where $\mathcal{H}$ is a random oracle. Hence, although the OPE key $\bar{K} = (K', A, M, I_A, I_M)$ contains a key K' of a pseudo-random function, K' is useless for $\widetilde{\mathsf{Enc}}_{\bar{K}}$ by definition.

We will re-use the step numbers of Fig.3 when we describe the steps of $\widetilde{\mathsf{Enc}}$, since $\widetilde{\mathsf{Enc}}_{\bar{K}}$ is almost the same as $\overline{\mathsf{Enc}}_{\bar{K}}$.

Depth: We will use a notion, *depth* of a message, so as to show some lemmas based on a mathematical induction about it. Let m be any message. We say that the *depth* of m is d if $\widetilde{\mathsf{Enc}}_{\bar{K}}(m)$ is computed by calling $\widetilde{\mathsf{Enc}}_{\bar{K}}$ itself d times. For example, in Fig.3, the depth of w is larger than those of u and v because, at the step (65) of Fig.3, the value I_w is computed by using I_u and I_v, which were computed in the previous recursions. Another example is that the depths of A and M of Fig.3 is clearly 0.

Facts and Terminologies about I_s: Finally, we show lemmas about value I_s given in Fig.3.

Lemma 20 *Let m, m', and w be any messages. Suppose that the value I_w is computed in both recursions of $\widetilde{\mathsf{Enc}}_{\bar{K}}(m)$ and $\widetilde{\mathsf{Enc}}_{\bar{K}}(m')$. Then I_w for $\widetilde{\mathsf{Enc}}_{\bar{K}}(m)$ and that for $\widetilde{\mathsf{Enc}}_{\bar{K}}(m')$ are the same value.*

The proof is easily obtained based on a mathematical induction about the depth of w: the value I_w is set to $\bar{G}(u, v, I_u, I_v; cc)$ at step (65) of Fig.3 in both recursions, which means that I_w is computed from I_u and I_v, and the values I_u and I_v for $\widetilde{\mathsf{Enc}}_{\bar{K}}(m)$ and those for $\widetilde{\mathsf{Enc}}_{\bar{K}}(m')$ are the same value due to the induction hypothesis. Hence, the value I_w for $\widetilde{\mathsf{Enc}}_{\bar{K}}(m)$ and that for $\widetilde{\mathsf{Enc}}_{\bar{K}}(m')$ has to be the same value. The induction bases are the elements of depth 0, that is, A and M and Lemma 20 clearly holds in these cases.

Due to the above lemma, we loosely write "I_w *generated in (the recursion of)* $\widetilde{\mathsf{Enc}}_{\bar{K}}$" if it is generated in the recursion of $\widetilde{\mathsf{Enc}}_{\bar{K}}(m)$ for some m.

Similarly, we can show the following lemma as well.

Lemma 21 *Let m and w be any messages. Suppose that the value I_w is computed in the recursions of $\widetilde{\mathsf{Enc}}_{\bar{K}}(m)$, and let $\mathsf{s}(w)$ and $\mathsf{l}(w)$ be values u and v of step (63). Then, the value $\mathsf{s}(w)$ and $\mathsf{l}(w)$ are independent from m.*

In other words, if the value w is computed in both recursions of $\widetilde{\mathsf{Enc}}_{\bar{K}}(m)$ and $\widetilde{\mathsf{Enc}}_{\bar{K}}(m')$, the values u and v of step (63) in the recursion of $\widetilde{\mathsf{Enc}}_{\bar{K}}(m)$ are the same as those of the same step in the recursion of $\widetilde{\mathsf{Enc}}_{\bar{K}}(m')$.

Moreover, $w = \lceil(\mathsf{s}(w) + \mathsf{l}(w))/2\rceil$ holds by definition.

The proof of the above lemma is again obtained based on a mathematical induction about the depth of w: the value w is set to $w \leftarrow \lceil(u+v)/2\rceil$ in step (63) of Fig.3 and, since the depth of u and v are smaller than that of w, the values u and v can be written as $u = \lceil(\mathsf{s}(u) + \mathsf{l}(u))/2\rceil$ and $v = \lceil(\mathsf{s}(v) + \mathsf{l}(v))/2\rceil$, where $\mathsf{s}(u)$, $\mathsf{l}(u)$, $\mathsf{s}(v)$, and $\mathsf{l}(v)$ do not depend on m due to the induction hypothesis. Hence, u, v, and $w = \lceil(u+v)/2\rceil$ themselves do not depend on m either. The proof for the induction base is trivial.

8.5.2 Proof

Proof (Proposition 18, sketch). For any message u, consider the two values I^*_u and $\tilde{I}_u$ which are generated in the following ways.

(1) Generate $\bar{K} \leftarrow \overline{\mathsf{Kg}}(1^\lambda)$, consider I_u generated in the recursion of $\widetilde{\mathsf{Enc}}_{\bar{K}}$ (in Fig.3), and set $\tilde{I}_u \leftarrow I_u$.
(2) Generate $(\rho_i, \delta_i)_{i\in(A..M]}$ in the same way as $\mathsf{Kg}(1^\lambda)$ (and set $K \leftarrow (\delta_i)_{i\in(A..M]}$ as in $\mathsf{Kg}(1^\lambda)$). Compute I_u based on (7.1) and set $I^*_u \leftarrow I_u$.

Above, (1) is well-defined due to Lemma 20.

We will show the following three facts. Then Proposition 18 clearly follows from them.

$$\forall\ u\ :\ \ \widetilde{\mathsf{Enc}}_{\bar{K}}(u) = \mathsf{Cph}(\tilde{I}_u), \tag{8.29}$$

$$\forall\ u\ :\ \ \mathsf{Enc}_K(u) = \mathsf{Cph}(I^*_u), \tag{8.30}$$

$$(s, \tilde{I}_s)_{s\in[A..M]} \approx (s, I^*_s)_{s\in[A..M]}, \tag{8.31}$$

where $\bar{K}$ and K are keys generated by $\overline{\mathsf{Kg}}(1^\lambda)$ and $\mathsf{Kg}(1^\lambda)$ in the above (1) and (2), Cph is a function given in Fig.3, and "$\approx$" means the perfect indistinguishability.

(8.29) follows because the output of $\widetilde{\mathsf{Enc}}_{\bar{K}}(u)$ is determined at step (62) of Fig.3. (8.30) follows due to (7.1), the definition of $\mathsf{Cph}(\cdot)$ (given in Fig.3), and the definition of the encryption function Enc_K (given in Fig.2).

(8.31) will be shown based on a mathematical induction. Specifically, we will define sets $U_0 \subsetneq U_1 \subsetneq U_2 \subsetneq \ldots$ of messages and will show the following fact based on a mathematical induction about i:

$$(s, \tilde{I}_s)_{s\in U_i} \approx (s, I^*_s)_{s\in U_i}. \tag{8.32}$$

(8.32), in particular, means that

$$(s, \tilde{I}_s)_{s\in U_{\mathsf{all}}} \approx (s, I^*_s)_{s\in U_{\mathsf{all}}},$$

holds where U_{all} is the set of all element of the message space, that is, $U_{\mathsf{all}} = [A..M]$. This is what we want to prove.

The lest of thing we have to do is to define U_i and to show (8.32) based on a mathematical induction about i.

Definition of U_i: U_0 is set to $\{A, M\}$ and when $U_{i-1} \subsetneq [A..M]$, U_i is recursively defined as follows.

$$\begin{array}{l}\text{Take (any) element } w_i \notin U_{i-1} \text{ such that } \mathsf{s}(w_i) \text{ and } \mathsf{l}(w_i) \text{ are in } U_{i-1}.\\ \text{Set } U_i \leftarrow U_{i-1} \cup \{w_i\}.\end{array} \tag{8.33}$$

Above, we can, in fact, take $w_i \notin U_{i-1}$ satisfying $\mathsf{s}(w_i), \mathsf{l}(w_i) \in U_{i-1}$, when $U_{i-1} \subsetneq [A..M]$. E.g. if we take $w_i \in [A..M] \setminus U_{i-1}$ such that the depth of w_i is smallest in $[A..M] \setminus U_{i-1}$. Then, since the depths of $\mathsf{s}(w_i)$ and $\mathsf{l}(w_i)$ are smaller than w_i, they have to be in U_{i-1}.

Before proceeding, we show the following fact, which we will use to prove (8.32):

$$(\mathsf{s}(w_j), \mathsf{l}(w_j)) \neq (\mathsf{s}(w_k), \mathsf{l}(w_k)) \qquad \text{for any } j \neq k. \tag{8.34}$$

The proof of (8.34) is quite simple. Suppose $j > k$ w.l.o.g. Then, $w_j \neq w_k$ has to hold because (8.33) shows that w_k is element of $U_{k+1} \subset U_j$ while w_j is not. But if $(\mathsf{s}(w_j), \mathsf{l}(w_j)) = (\mathsf{s}(w_k), \mathsf{l}(w_k))$, $w_j = \lceil (\mathsf{s}(w_j) + \mathsf{l}(w_j))/2 \rceil = \lceil (\mathsf{s}(w_k) + \mathsf{l}(w_k))/2 \rceil = w_k$ has to hold due to Lemma 21. This means that (8.34) has to hold.

Induction Base for showing (8.32): We show that (8.32) holds for $U_0 = \{A, M\}$. This fact can be shown by tracing the definitions of $\tilde{I}_u$, $\overline{\mathsf{Kg}}(1^\lambda)$, and $\widetilde{\mathsf{Enc}}_{\bar{K}}$ and by using (7.2).

Specifically, by definition, $\tilde{I}_A$ is computed in the following two steps.

(a) generate $\bar{K} = (K', A, M, I_A, I_M)$ using $\overline{\mathsf{Kg}}(1^\lambda)$,
(b) consider the value I_A generated in $\widetilde{\mathsf{Enc}}_{\bar{K}}$ and set $\tilde{I}_A \leftarrow I_A$.

By definition of $\overline{\mathsf{Kg}}(1^\lambda)$, the values I_A and I_M in step (a) are generated using the subroutine $(I_A, I_M) \leftarrow \mathsf{Init}(1^\lambda)$ of $\overline{\mathsf{Kg}}(1^\lambda)$. Hence, $\widetilde{\mathsf{Enc}}_{\bar{K}} = \widetilde{\mathsf{Enc}}_{(K',A,M,I_A,I_M)}$ of step (b) uses the value (I_A, I_M) generated by $\mathsf{Init}(1^\lambda)$. Moreover, Lemma 20 shows that we can replace $\widetilde{\mathsf{Enc}}_{\bar{K}}$ of step (2) with $\widetilde{\mathsf{Enc}}_{\bar{K}}(A)$. Hence, by definition of $\widetilde{\mathsf{Enc}}_{\bar{K}}(A) = \widetilde{\mathsf{Enc}}_{(K',A,M,I_A,I_M)}(A)$, $\tilde{I}_A$ of step (b) is equal to I_A generated by $\mathsf{Init}(1^\lambda)$. Due to a similar reason, value $\tilde{I}_M$ is equal to I_M generated by $\mathsf{Init}(1^\lambda)$. Since (7.2) shows that an output (I_A, I_M) of $\mathsf{Init}(1^\lambda)$ is perfectly indistinguishable from (I^*_A, I^*_M), the above discussion shows that

$$(\tilde{I}_A, \tilde{I}_M) \approx (I^*_A, I^*_M)$$

holds. Since A and M are constants, this means that

$$(s, \tilde{I}_s)_{s\in\{A,M\}} \approx (s, I^*_s)_{s\in\{A,M\}}$$

holds. This is what we want to prove.

Induction Step for Showing (8.32): Fix any $i \geq 0$ and assume the induction hypothesis

$$(s, \tilde{I}_s)_{s\in U_{i-1}} \approx (s, I^*_s)_{s\in U_{i-1}}. \tag{8.35}$$

Let $\mathcal{H}$ be the random oracle used in $\widetilde{\mathsf{Enc}}_{\bar{K}}$, which was replaced with PRF of step (54) of Fig.3.

Let w_i be a value given in (8.33). Then, by definitions of $\tilde{I}_{w_i}$, $\mathsf{s}(\cdot)$, and $\mathsf{l}(\cdot)$, $\tilde{I}_{w_i}$ is set to $\bar{G}(u, v, \tilde{I}_u, \tilde{I}_v; cc)$ at step (55) of Fig.3, where $u = \mathsf{s}(w_i)$, $v = \mathsf{l}(w_i)$, and $cc = \mathcal{H}(u, v)$ (due to step (54) of Fig.3). That is,

$$\tilde{I}_{w_i} = \bar{G}(\mathsf{s}(w_i), \mathsf{l}(w_i), \tilde{I}_{\mathsf{s}(w_i)}, \tilde{I}_{\mathsf{l}(w_i)}; \mathcal{H}(\mathsf{s}(w_i), \mathsf{l}(w_i))) \tag{8.36}$$

holds. Since $U_i = U_{i-1} \cup \{w_i\}$ holds from (8.33), this means that

$$\begin{aligned}(s, \tilde{I}_s)_{s \in U_i} &= (s, \tilde{I}_s)_{s \in U_{i-1}} \cup \{(w_i, \tilde{I}_{w_i})\} \\ &= (s, \tilde{I}_s)_{s \in U_{i-1}} \cup \{(w_i, \bar{G}(\mathsf{s}(w_i), \mathsf{l}(w_i), \tilde{I}_{\mathsf{s}(w_i)}, \tilde{I}_{\mathsf{l}(w_i)}; \mathcal{H}(\mathsf{s}(w_i), \mathsf{l}(w_i))))\}. \end{aligned} \tag{8.37}$$

Due to the definition of the random oracle, (8.34) means that $\mathcal{H}(\mathsf{s}(w_j), \mathsf{l}(w_j))$ and $\mathcal{H}(\mathsf{s}(w_k), \mathsf{l}(w_k))$ are independent randomness for any $j \neq k$. Hence, we can replace $\mathcal{H}(\mathsf{s}(w_i), \mathsf{l}(w_i))$ in (8.37) with a uniformly selected randomness cc_i for any i. We therefore can get

$$(s, \tilde{I}_s)_{s \in U_i} \approx (s, \tilde{I}_s)_{s \in U_{i-1}} \cup \{\bar{G}(\mathsf{s}(w_i), \mathsf{l}(w_i), \tilde{I}_{\mathsf{s}(w_i)}, \tilde{I}_{\mathsf{l}(w_i)}; cc_i)\}. \tag{8.38}$$

From (8.33), $\mathsf{s}(w_i)$ and $\mathsf{l}(w_i)$ are elements of U_{i-1}. Hence, from the induction hypothesis (8.35), we can replace $\tilde{I}_s$, $\tilde{I}_{u_i}$, and $\tilde{I}_{v_i}$ of the right hand side of (8.38) with I^*_s, $I^*_{u_i}$, and $I^*_{v_i}$ respectively. That is,

$$(s, \tilde{I}_s)_{s \in U_i} \approx (s, I^*_s)_{s \in U_{i-1}} \cup \{\bar{G}(\mathsf{s}(w_i), \mathsf{l}(w_i), I^*_{\mathsf{s}(w_i)}, I^*_{\mathsf{l}(w_i)}; cc_i)\} \tag{8.39}$$

holds. Here we use the fact that the distribution of cc_i is independent from those of $(\tilde{I}_s)$ and (I^*_s).

Then by applying (8.28) to the right hand side of (8.39), we can get

$$(s, \tilde{I}_s)_{s \in U_i} \approx (s, I^*_s)_{s \in U_{i-1}} \cup \{(w_i, I^*_{w_i})\}. \tag{8.40}$$

Since $U_i = U_{i-1} \cup \{w_i\}$ holds from (8.33), we can conclude that

$$(s, \tilde{I}_s)_{s \in U_i} \approx (s, I^*_s)_{s \in U_i} \tag{8.41}$$

holds. This is what we want to prove. □

8.6 Security Proof

Finally, we show that our scheme is (k, θ)-FTG-O-nCPA secure with advantage bound given in (5.7). To prove this fact, we use the following proposition.

Proposition 22 For any natural number θ, for any probability distribution $\mathcal{X}_\lambda$ satisfying property (6.5), and for any (even computationally unbounded) adversary $\mathsf{D} = (\mathsf{D}_{\mathsf{find}}, \mathsf{D}_{\mathsf{guess}})$, the advantage of D in the following game is negligible for λ Below, $\mathsf{D}_{\mathsf{find}}$ has to select α and β satisfying $\alpha, \beta \in [-\theta, \theta]$.

$$\begin{aligned}&(\alpha, \beta, \mathsf{st}) \leftarrow \mathsf{D}_{\mathsf{find}}(1^\lambda), \quad \delta, \zeta \xleftarrow{\$} \mathcal{X}_\lambda, \quad (S_b, T_b) \leftarrow \begin{cases}(\delta, \zeta) & \text{If } b = 0 \\ (\delta + \alpha, \zeta + \beta) & \text{If } b = 1\end{cases}, \\ &d \leftarrow \mathsf{D}_{\mathsf{guess}}(S_b, T_b, \mathsf{st}), \quad \text{return } d.\end{aligned}$$

Intuitively, the above proposition holds because the large randomness in δ and ζ hide small values α and β. The formal proof is as follows.

Proof (Proposition 22). Fix any adversary D and take $(\alpha, \beta, \mathsf{st}) \leftarrow \mathsf{D}_{\mathsf{find}}(1^\lambda)$, $\delta, \zeta \xleftarrow{\$} \mathcal{X}_\lambda$, $(S_0, T_0) \leftarrow (\delta, \zeta)$, and $(S_1, T_1) \leftarrow (\delta + \alpha, \zeta + \beta)$, as in the game in the proposition. We will show

$$\mathsf{SD}((S_0, T_0, \mathsf{st}), (S_1, T_0, \mathsf{st})) \le \mathsf{neg}(\lambda). \tag{8.42}$$

holds. In a similar manner, we can also show

$$\mathsf{SD}((S_0, T_1, \mathsf{st}), (S_1, T_1, \mathsf{st})) \le \mathsf{neg}(\lambda). \tag{8.43}$$

Proposition 22 clearly follows from (8.42) and (8.43). (This proposition holds even when D is computationally unbounded because the above proof is based on the statistical distance.)

The proof of (8.42) is as following (Below, (*) holds because the distribution of δ and $(\alpha, \mathsf{st}, \zeta)$ are independent):

$$\begin{aligned}
&\mathsf{SD}((S_0, T_0, \mathsf{st}), (S_1, T_0, \mathsf{st})) \\
&= \sum_{u,v,w} |\Pr[(S_0, T_0, \mathsf{st}) = (u, v, w)] - \Pr[(S_1, T_0, \mathsf{st}) = (u, v, w)]| \\
&= \sum_{u,v,w} |\Pr[(\delta, \zeta, \mathsf{st}) = (u, v, w)] - \Pr[(\delta + \alpha, \zeta, \mathsf{st}) = (u, v, w)]| \\
&\underset{(*)}{=} \sum_{u,v,w} \left| \sum_y \Pr[(\alpha, \mathsf{st}, \zeta) = (y, v, w)](\Pr[\delta = u] - \Pr[\delta + y = u]) \right| \\
&\le \sum_{v,w,y} \Pr[(\alpha, \mathsf{st}, \zeta) = (y, v, w)] \sum_u |\Pr[\delta = u] - \Pr[\delta + y = u]| \\
&= \sum_{v,w,y} \Pr[(\alpha, \mathsf{st}, \zeta) = (y, v, w)] \mathsf{SD}(\delta, \delta + y) \\
&\underset{(6.5)}{\le} \mathsf{neg}(\lambda).
\end{aligned}$$

Proof (Security of Our Scheme). We show here the (k, θ)-FTG-O-nCPA security of the scheme of Fig.3 with the advantage upper-bound given in (5.7).

Definition of Games: Define games $\mathsf{Game}^{(0)}(\lambda)$, ..., $\mathsf{Game}^{(2)}(\lambda)$ as follows:

$\mathsf{Game}^{(0)}(\lambda)$: The same as the game of $(k, \theta)-$FTG-nCPA.

$\mathsf{Game}^{(1)}(\lambda)$: The same as $\mathsf{Game}^{(0)}(\lambda)$ except that when a challenger computes answers to encryption and challenge queries, he uses not $\overline{\mathsf{Enc}}_{\bar{K}}$ but $\widetilde{\mathsf{Enc}}_{\bar{K}}$. Here $\widetilde{\mathsf{Enc}}$ is an algorithm obtained from $\overline{\mathsf{Enc}}$ by replacing $\mathsf{PRF}_{K'}$ of step 54 with the random oracle.

$\mathsf{Game}^{(2)}(\lambda)$: The same as $\mathsf{Game}^{(1)}(\lambda)$ except that the challenger does not generate $\bar{K} \leftarrow \overline{\mathsf{Kg}}(1^\lambda)$ of the scheme of Section 7 but $K \leftarrow \mathsf{Kg}(1^\lambda)$ of the scheme of Section 6 instead, and use not $\widetilde{\mathsf{Enc}}_{\bar{K}}$ but the encryption algorithm Enc_K of the scheme of Section 6 instead, when he computes answers to encryption and challenge queries.

Before proceeding, we give a caveat about the computational cost of $\mathsf{Game}^{(2)}(\lambda)$. The challengers of $\mathsf{Game}^{(2)}(\lambda)$ may have to use exponential time to answer to encryption query because the computational cost of Enc_K of Section 6 is proportional to the message space size M and M can be exponential. However, this fact does not become a problem because the part of the proof in which we will use this game is purely information theoretic one.

Indistinguishability of Games: Let Bad be the event of Section 6.2, $\mathsf{Adv.Game}^{(i)}_{\mathsf{A}}(1^\lambda)$ be the advantage of an adversary A for $\mathsf{Game}^{(i)}$, and $\mathsf{Adv.Good.Game}^{(i)}_{\mathsf{A}}(1^\lambda)$ be the advantage of an adversary A for $\mathsf{Game}^{(i)}$ under the condition that event $\neg\mathsf{Bad}$ holds.

We will show the following facts:

$$\exists \mathsf{B} \quad : \quad |\mathsf{Adv.Game}^{(0)}_{\mathsf{A}}(1^\lambda) - \mathsf{Adv.Game}^{(1)}_{\mathsf{A}}(1^\lambda)| \leq \mathsf{Adv.Exp}_{\mathsf{PRF}}(\mathsf{B}), \tag{8.44}$$

$$\mathsf{Adv.Game}^{(1)}_{\mathsf{A}}(1^\lambda) = \mathsf{Adv.Game}^{(2)}_{\mathsf{A}}(1^\lambda), \tag{8.45}$$

$$\mathsf{Adv.Game}^{(2)}_{\mathsf{A}}(1^\lambda) \leq \mathsf{Adv.Good.Game}^{(2)}_{\mathsf{A}}(1^\lambda) + O(1/\sqrt{k}). \tag{8.46}$$

$$\mathsf{Adv.Good.Game}^{(2)}_{\mathsf{A}}(1^\lambda) \leq \mathsf{neg}(\lambda). \tag{8.47}$$

Then, Theorem 15 clearly follows from the above facts. The proof of (8.44), ..., (8.46) are as follows:

- (8.44) follows because $\mathsf{PRF}_{K'}$ is a pseudorandom function.
- (8.45) follows because the indistinguishability of outputs encryption functions (That is, Proposition 18).
- (8.46) follows from the estimation (6.6) of $\Pr[\mathsf{Bad}]$.

Finally, we prove (8.47).

Proof of (8.47): Let $\mathsf{A} = (\mathsf{A}_{\mathsf{find}}, \mathsf{A}_{\mathsf{guess}})$ be an adversary for $\mathsf{Game}^{(2)}(\lambda)$. By using A as a subroutine, we construct a (possibly computationally unbounded) adversary $\mathsf{D} = (\mathsf{D}_{\mathsf{find}}, \mathsf{D}_{\mathsf{guess}})$ for the game of Proposition 22. We then show that D succeeds in simulating the view of A, under the condition that event $\neg\mathsf{Bad}$ holds, where Bad is the event of Section 6.2.

Note that here we can use Proposition 22 because Proposition 16 ensures that, even when we set $\mathcal{X}_\lambda$ as in (7.5), $\mathcal{X}_\lambda$ satisfies (6.5).

Description of D: $\mathsf{D}_{\mathsf{find}}$ executes $\mathsf{A}_{\mathsf{find}}(1^\lambda)$ and gets the challenge query (m^*_0, m^*_1), encryption queries $m_1, \ldots, m_q$, and the state st as outputs of $\mathsf{A}_{\mathsf{find}}$. $\mathsf{D}_{\mathsf{find}}$ makes query

$$(\alpha, \beta) \leftarrow (m^*_1 - m^*_0, -(m^*_1 - m^*_0)) \tag{8.48}$$

to the challenger of her. The above (α, β), in fact, satisfies the condition $\alpha, \beta \in [-\theta, \theta]$ of Proposition 22, because the definition of (k, θ)-FTG-O-nCPA ensures $|m^*_1 - m^*_0| \leq \theta$.

Then $\mathsf{D}_{\mathsf{guess}}$ gets an answer (S, T) to the query and randomly take[7]

$$i_0 \xleftarrow{\$} (m^*_0 - k\theta..m^*_0], \qquad\qquad i_1 \xleftarrow{\$} (m^*_1..m^*_1 + k\theta].$$

Intuitively, i_0 and i_1 are the values such that δ_{i_0} and δ_{i_1} are large, whose existences are ensured by the conditions (6.3) and (6.4) of $\neg\mathsf{Bad}$ of Section 6.1.

$\mathsf{D}_{\mathsf{guess}}$ sets

$$\delta_i \leftarrow 1 \text{ for } i \in (m^*_0..m^*_1]. \tag{8.49}$$

For $i \notin (m^*_0..m^1_*] \cup \{i_0, i_1\}$, $\mathsf{D}_{\mathsf{guess}}$ takes δ_i in the same way as Kg of Section 6. That is,

$$\rho_i \xleftarrow{\$} \mathcal{B}(1, 1-p), \qquad\qquad \delta_i \leftarrow \begin{cases} 1 & \text{if } \rho_i = 1 \\ \mathcal{X}_\lambda & \text{otherwise.} \end{cases}$$

[7] As described in the footnote of Section 6.1, seeming asymmetry of the interval, which is "left-open" one $(a..b]$ but is not "right open" one $[a..b)$, comes from how we number δ_i. If we set δ_i not to $\mathsf{Enc}_K(i) - \mathsf{Enc}_K(i-1)$ but to $\mathsf{Enc}_K(i+1) - \mathsf{Enc}_K(i)$, it becomes right open one $[a..b)$.

(Here $\mathsf{D}_{\mathsf{guess}}$ has to spend super-polynomial time to compute all δ_i if the message space size is super-polynomial in λ. But it is not a problem because Proposition 22 allows $\mathsf{D}_{\mathsf{guess}}$ to use super-polynomial time.)

Then $\mathsf{D}_{\mathsf{guess}}$ computes

$$C^* \leftarrow S + \sum_{i \in (A..m_0^*] \setminus \{i_0\}} \delta_i. \tag{8.50}$$

The encryption queries m_j of $\mathsf{A}_{\mathsf{find}}$ has to satisfy $m_j \leq m_0^* - k\theta$ or $m_j \geq m_1^* + k\theta$ because of Definition 13 of (k,θ)-FTG-O-nCPA. $\mathsf{D}_{\mathsf{guess}}$ computes for each $j \in [1..q]$,

$$C_j \leftarrow \begin{cases} \sum_{i \in (A..m_j]} \delta_i & \text{if } m_j \leq m_0^* - k\theta \\ S + T + \sum_{i \in (A..m_j] \setminus \{i_0, i_1\}} \delta_i & \text{if } m_j \geq m_1^* + k\theta \end{cases} \tag{8.51}$$

and sends $(C_*, C_1, \ldots, C_q)$ and st to $\mathsf{A}_{\mathsf{guess}}$. If $\mathsf{A}_{\mathsf{guess}}$ outputs a bit d, $\mathsf{D}_{\mathsf{guess}}$ outputs d and terminates.

Next, we show that D succeeds in simulating the view of A, under the condition that $\neg\mathsf{Bad}$ holds. Here Bad is the event given in Section 6.2.

D Succeeds in Generating (δ_i) Correctly: From (8.48) and the statement of Proposition 22, (S,T) can be written as

$$(S,T) = (\delta + \alpha_b, \zeta + \beta_b), \tag{8.52}$$

where

$$(\alpha_b, \beta_b) \leftarrow \begin{cases} (0,0) & \text{if } b = 0, \\ (m_1^* - m_0^*, -(m_1^* - m_0^*)) & \text{otherwise.} \end{cases} \tag{8.53}$$

We set

$$(\delta_{i_0}, \delta_{i_1}) \leftarrow (\delta, \zeta). \tag{8.54}$$

Then, $(\delta_i)_i$ generated in this game has the same distribution as $(\delta_i)_i$ generated by $\mathsf{Kg}(1^\lambda)$ of Fig.2, under the condition that $\neg\mathsf{Bad}$ holds. In fact, the condition (6.2), (6.3), and (6.4) of $\neg\mathsf{Bad}$ of Section 6.1 ensures that

- $\delta_i = 1$ holds for each $i \in (m_0^*..m_1^*]$
- there are $i_0 \in (m_0 - k\theta..m_0]$ and $i_1 \in (m_1..m_1 + k\theta]$ such that δ_{i_0} and δ_{i_1} are selected from $\mathcal{X}_\lambda$.

They clearly correspond with (8.49) and (8.54). Hence $\mathsf{D}_{\mathsf{guess}}$ generated (δ_i) correctly.

D Succeeds in Generating C_* Correctly: Since $\delta_i = 1$ holds for $i \in (m_0^*..m_1^*]$, it follows that if $b = 0$,

$$(\alpha_b, \beta_b) \underset{(8.53)}{=} (m_1^* - m_0^*, -(m_1^* - m_0^*)) \underset{(8.49)}{=} (\sum_{i \in (m_0^*..m_1^*]} \delta_i, -\sum_{i \in (m_0^*..m_1^*]} \delta_i). \tag{8.55}$$

Hence,

$$C^* \underset{(8.50)}{=} S + \sum_{i \in (A..m_0^*] \setminus \{i_0\}} \delta_i \underset{(8.52)}{=} \delta_{i_0} + \alpha_b + \sum_{i \in (A..m_0^*] \setminus \{i_0\}} \delta_i. = \alpha_b + \sum_{i \in (A..m_0^*]} \delta_i.$$

Since

$$\alpha_b \underset{\substack{(8.53)\\(8.55)}}{=} \begin{cases} 0 & \text{if } b = 0 \\ \sum_{i \in (m_0^* .. m_1^*]} \delta_i & \text{if } b = 1, \end{cases}$$

we can conclude that

$$C^* = \sum_{i \in (A..m_b^*]} \delta_i$$

holds. This value is equal to $\mathsf{Enc}_K(m^*)$. Hence, $\mathsf{D}_{\mathsf{guess}}$ generated C^* correctly.

$\mathsf{D}_{\mathsf{guess}}$ Succeeds in Generating Answers C_i to the Encryption Queries m_i Correctly: An encryption query m_j of $\mathsf{A}_{\mathsf{find}}$ has to satisfy $m_j \leq m_0^* - k\theta$ or $m_j \geq m_1^* + k\theta$ because of Definition 13 of (k, θ)-FTG-O-nCPA. In the former case, $\mathsf{D}_{\mathsf{guess}}$ answered $C_j \underset{(8.51)}{=} \sum_{i \in (A..m]} \delta_i$, which is, of course, equal to $\mathsf{Enc}_K(m)$. In the latter case, the answer C of $\mathsf{D}_{\mathsf{guess}}$ satisfies

$$\begin{aligned} C_j &\underset{(8.51)}{=} S + T + \sum_{i \in (A..m_j] \setminus \{i_0, i_1\}} \delta_i \\ &\underset{\substack{(8.52)\\(8.54)}}{=} (\delta_{i_0} + \alpha_b) + (\delta_{i_1} + \beta_b) + \sum_{i \in (A..m_j] \setminus \{i_0, i_1\}} \delta_i \\ &= \alpha_b + \beta_b + \sum_{i \in (A..m_j]} \delta_i. \end{aligned}$$

From (8.53), we can conclude that

$$C_j = \sum_{i \in (A..m_j]} \delta_i$$

holds in both cases where $b = 0$ and $b = 1$. This value is, of course, equal to $\mathsf{Enc}_K(m)$. Hence, $\mathsf{D}_{\mathsf{guess}}$ generated C_j correctly.

Since Proposition 22 shows that D cannot have non-negligible advantage, (8.47) is therefore shown.□

9 Conclusions

Indistinguishability in encryption is a fundamental property. We proposed the first achievable indistinguishability notion for OPE. In fact, Our notion, $(\mathcal{X}, \theta, q)$-indistinguishability, ensures secrecy of the least significant $\lfloor \log_2 \theta \rfloor$ bits of a plaintext m_* under the following setting: the database containing $q + 1$ data $m_*, m_1, \ldots, m_q$ in their encrypted forms, where these messages were distributed according to given distributions, and an adversary who wanted to know m_* breached the database and got all ciphertexts in it. Here she was allowed to have (partial or all) information about the other data elements $m_1, \ldots, m_q$ as well.

Then we proposed an OPE scheme $\mathcal{E}^{\beta,t}$ such that, when the (independent) distributions $\mathcal{X}_1, \ldots, \mathcal{X}_q$ of $m_1, \ldots, m_q$ had min-entropy $\beta \log_2 M$, $\mathcal{E}^{\beta,t}$ satisfied $(\mathcal{X}, M^t, q)$-indistinguishability for any $t < \beta$ when $M \to \infty$ where M was the message space size and $\mathcal{X}$ was $(\mathcal{X}_1, \ldots, \mathcal{X}_q)$, (although the advantage bound decreased somewhat when t became closer to β.) This meant that our scheme was

able to hide the least significant $t < \beta$ bits of the plaintext m_*. In particular, when $m_1, \ldots, m_q$ are distributed uniformly at random, the above fact meant that our scheme was able to hide any fraction of the low order bits of the plaintext m_*.

We then showed that $(\mathcal{X}, \theta, q)$-indistinguishability with suitable parameters implies the known security one-way-ness-type notion, $(r, q+1)$-WOW [17], (and its stronger variant $(r, q+1)$-WOWM which allowed an adversary to watch the plaintext messages other than the target one). We then showed that our scheme satisfies $(r, q+1)$-WOW with better parameter than the known scheme [17] did.

Our investigation is the first to consider indistinguishability notions for OPE and many open questions remain.

References

1. R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu. Order-preserving encryption for numeric data. In *SIGMOD Conference*, pages 563–574, 2004.
2. G. Amanatidis, A. Boldyreva, and A. O'Neill. Provably-secure schemes for basic query support in outsourced databases. In *DBSec*, pages 14–30, 2007.
3. G. Ateniese, R. C. Burns, R. Curtmola, J. Herring, L. Kissner, Z. N. J. Peterson, and D. X. Song. Provable data possession at untrusted stores. In *ACM Conference on Computer and Communications Security*, pages 598–609, 2007.
4. G. Ateniese, K. Fu, M. Green, and S. Hohenberger. Improved proxy re-encryption schemes with applications to secure distributed storage. *ACM Trans. Inf. Syst. Secur.*, 9(1):1–30, 2006.
5. G. Ateniese, S. Kamara, and J. Katz. Proofs of storage from homomorphic identification protocols. In *ASIACRYPT*, pages 319–333, 2009.
6. G. Bebek. *Anti-Tamper Databases: Inference control techniques*. Case Western Reserve University, 2002.
7. M. Bellare, A. Boldyreva, L. R. Knudsen, and C. Namprempre. Online ciphers and the hash-cbc construction. In *CRYPTO*, pages 292–309, 2001.
8. M. Bellare, A. Boldyreva, and A. O'Neill. Deterministic and efficiently searchable encryption. In *CRYPTO*, pages 535–552, 2007.
9. M. Bellare, A. Desai, E. Jokipii, and P. Rogaway. A concrete security treatment of symmetric encryption. In *FOCS*, pages 394–403, 1997.
10. M. Bellare, M. Fischlin, A. O'Neill, and T. Ristenpart. Deterministic encryption: Definitional equivalences and constructions without random oracles. In *CRYPTO*, pages 360–378, 2008.
11. M. Bellare and C. Namprempre. Authenticated encryption: Relations among notions and analysis of the generic composition paradigm. In *ASIACRYPT*, pages 531–545, 2000.
12. M. Bellare and P. Rogaway. The security of triple encryption and a framework for code-based game-playing proofs. In *EUROCRYPT*, pages 409–426, 2006.
13. S. Benabbas, R. Gennaro, and Y. Vahlis. Verifiable delegation of computation over large datasets. In *CRYPTO*, pages 111–131, 2011.
14. M. Blaze, G. Bleumer, and M. Strauss. Divertible protocols and atomic proxy cryptography. In *EUROCRYPT*, pages 127–144, 1998.
15. A. Boldyreva, N. Chenette, Y. Lee, and A. O'Neill. Full paper of [16], 2009. Available at http://www.cc.gatech.edu/~aboldyre/papers/bclo.pdf.
16. A. Boldyreva, N. Chenette, Y. Lee, and A. O'Neill. Order-preserving symmetric encryption. In *EUROCRYPT*, pages 224–241, 2009.
17. A. Boldyreva, N. Chenette, and A. O'Neill. Full paper of [18], 2009. Available at http://www.cc.gatech.edu/~aboldyre/papers/operev.pdf.
18. A. Boldyreva, N. Chenette, and A. O'Neill. Order-preserving encryption revisited: Improved security analysis and alternative solutions. In *CRYPTO*, pages 578–595, 2011.
19. A. Boldyreva, S. Fehr, and A. O'Neill. On notions of security for deterministic encryption, and efficient constructions without random oracles. In *CRYPTO*, pages 335–359, 2008.
20. D. Boneh, G. D. Crescenzo, R. Ostrovsky, and G. Persiano. Public key encryption with keyword search. In *EUROCRYPT*, pages 506–522, 2004.
21. D. Boneh and B. Waters. Conjunctive, subset, and range queries on encrypted data. In *TCC*, pages 535–554, 2007.

22. B. Chor, E. Kushilevitz, O. Goldreich, and M. Sudan. Private information retrieval. *J. ACM*, 45(6):965–981, 1998.
23. R. Curtmola, J. A. Garay, S. Kamara, and R. Ostrovsky. Searchable symmetric encryption: improved definitions and efficient constructions. In *ACM Conference on Computer and Communications Security*, pages 79–88, 2006.
24. L. Devroye. *Non-Uniform Random Variate Generation*. Springer, 1986.
25. Z. Erkin, A. Piva, S. Katzenbeisser, R. L. Lagendijk, J. Shokrollahi, G. Neven, and M. Barni. Protection and retrieval of encrypted multimedia content: When cryptography meets signal processing. *EURASIP J. Information Security*, 2007, 2007.
26. HIPAA — General Information. Available at https://www.cms.gov/bHIPAAGenInfo/.
27. IBM Help—estimating data size. Available at http://publib.boulder.ibm.com/infocenter/wpdoc/v510/topic/com.ibm.wp.ent.doc/pzn/pzn_estimate_db_size.html.
28. A. Juels and B. S. K. Jr. Pors: proofs of retrievability for large files. In *ACM Conference on Computer and Communications Security*, pages 584–597, 2007.
29. J. Katz and M. Yung. Characterization of security notions for probabilistic private-key encryption. *J. Cryptology*, 19(1):67–95, 2006.
30. E. Kushilevitz and R. Ostrovsky. Replication is not needed: Single database, computationally-private information retrieval. In *FOCS*, pages 364–373, 1997.
31. W. Lu, A. L. Varna, and M. Wu. Security analysis for privacy preserving search of multimedia. In *ICIP*, pages 2093–2096, 2010.
32. U. M. Maurer, Y. A. Oswald, K. Pietrzak, and J. Sjödin. Luby-rackoff ciphers from weak round functions? In *EUROCRYPT*, pages 391–408, 2006.
33. U. M. Maurer, K. Pietrzak, and R. Renner. Indistinguishability amplification. In *CRYPTO*, pages 130–149, 2007.
34. K. Minematsu and Y. Tsunoo. Hybrid symmetric encryption using known-plaintext attack-secure components. In *ICISC*, pages 242–260, 2005.
35. R. Ostrovsky and W. E. S. III. A survey of single-database private information retrieval: Techniques and applications. In *Public Key Cryptography*, pages 393–411, 2007.
36. G. Özsoyoglu, D. A. Singer, and S. S. Chung. Anti-tamper databases: Querying encrypted databases. In *DBSec*, pages 133–146, 2003.
37. O. Pandey and Y. Rouselakis. Property preserving symmetric encryption. In *EUROCRYPT*, pages 375–391, 2012.
38. R. A. Popa, F. H. Li, and N. Zeldovich. An ideal-security protocol for order-preserving encoding. *IACR Cryptology ePrint Archive*, 2013:129, 2013.
39. R. A. Popa, F. H. Li, and N. Zeldovich. An ideal-security protocol for order-preserving encoding. In *IEEE Symposium on Secrity an Privacy*, pages ?–?, 2013.
40. R. A. Popa, C. M. S. Redfield, N. Zeldovich, and H. Balakrishnan. Cryptdb: protecting confidentiality with encrypted query processing. In *SOSP*, pages 85–100, 2011.
41. R. A. Popa, C. M. S. Redfield, N. Zeldovich, and H. Balakrishnan. Cryptdb: processing queries on an encrypted database. *Commun. ACM*, 55(9):103–111, 2012.
42. E. Shi, J. Bethencourt, H. T.-H. Chan, D. X. Song, and A. Perrig. Multi-dimensional range query over encrypted data. In *IEEE Symposium on Security and Privacy*, pages 350–364, 2007.
43. D. X. Song, D. Wagner, and A. Perrig. Practical techniques for searches on encrypted data. In *IEEE Symposium on Security and Privacy*, pages 44–55, 2000.
44. S. Tu, M. F. Kaashoek, S. Madden, and N. Zeldovich. Processing analytical queries over encrypted data. *Proceedings of the VLDB Endowment*, Volume 6, No. 5, March 2013:289–300, 2013.
45. C. Wang, N. Cao, J. Li, K. Ren, and W. Lou. Secure ranked keyword search over encrypted cloud data. In *ICDCS*, pages 253–262, 2010.
46. D. Westhoff, J. Girão, and M. Acharya. Concealed data aggregation for reverse multicast traffic in sensor networks: Encryption, key distribution, and routing adaptation. *IEEE Trans. Mob. Comput.*, 5(10):1417–1431, 2006.
47. L. Xiao and I.-L. Yen. A note for the ideal order-preserving encryption object and generalized order-preserving encryption. *IACR Cryptology ePrint Archive*, 2012:350, 2012.
48. L. Xiao and I.-L. Yen. Security analysis for order preserving encryption schemes. In *CISS*, pages 1–6, 2012.
49. L. Xiao, I.-L. Yen, and D. T. Huynh. Extending order preserving encryption for multi-user systems. *IACR Cryptology ePrint Archive*, 2012:192, 2012.
50. J. Xu, J. Fan, M. H. Ammar, and S. B. Moon. Prefix-preserving ip address anonymization: Measurement-based security evaluation and a new cryptography-based scheme. In *ICNP*, pages 280–289, 2002.
51. D. H. Yum, D. S. Kim, J. S. Kim, P. J. Lee, and S. J. Hong. Order-preserving encryption for non-uniformly distributed plaintexts. In *WISA*, 2011.