Order-Preserving Encryption Secure Beyond One-Wayness

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

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

Abstract. Semantic-security of individual plaintext bits given the corresponding ciphertext is a fundamental notion in modern cryptography. We initiate the study of this basic problem for Order-Preserving Encryption (OPE), asking "what plaintext information can be semantically hidden by OPE encryptions?" OPE has gained much attention in recent years due to its usefulness for secure databases, and has received a thorough formal treamtment with innovative and useful security notions. However, all previous notions are one-way based, and tell us nothing about partial-plaintext indistinguishability (semantic security).

In this paper, we propose the first 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 justify the definition, from the theoretical plausibility and practicality aspects. Finally, we propose a new scheme satisfying this security notion (the first one to do so). In order to be clear, we note that the earlier security notions, while innovative and surprising, nevertheless tell us nothing about the above partial- plaintext indistinguishability because they are limited to being one-way-based.

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

1 Introduction

Securing cloud database with untrusted cloud servers needs to hide information from the database manager itself, and has resulted in new research areas.

Order-Preserving Encryption (OPE): This is, perhaps, the most promising new primitives in the area of encrypted database processing [1,17,3,7,8,28]. 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 in an ad-hoc fashion in the database community by Agrawal, Kiernan, Ramakrishnan, Srikant and Xu [1], and seemed like a clever heuristics. However, its careful foundational study was initiated with surprising formal cryptographic models and proofs by

P. Sarkar and T. Iwata (Eds.): ASIACRYPT 2014, PART II, LNCS 8874, pp. 42–61, 2014.

Boldyreva, Chenette, Lee, and O'Neill [7,8]. Overall, it has received much recent attention in the cryptographic community [7,8,28], in the database community [1,17,3], as well as in other applied areas.

OPE is attractive since it allows one to simultaneously 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 [1,20,21,23]. 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 [9,22], needs linear time in the size, which is too costly for a database containing a few millions data items.

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. This is important. Imagine the following "string embedding" problem: we concatenate numerical strings to get a larger number and we have degree of freedom in this concatenation, don't we want to hide the most crucial string by embedding it at a location within the large number which hides it better than otherwise? Hasn't this very issue (partial information security in a ciphertext) been at the heart of cryptographic formalisms of encryption technologies in the last 30 years or so? Indeed, a naturally defined indistinguishability notion for OPE, *indistinguishability under ordered CPA attack (IND-O-CPA)* [7], 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. (And if the message space is only polynomial size, an OPE scheme completely loses its utility, of course.)

OPE Is an Inherently "Leaky" Method: The reason behind the above negative result 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 [7] mentioned above is, in fact, proved using an attack based on this observation.

To date, no one can tell what exactly OPE *must* leak and what it can protect. Our motivation is the fact that the existing security notions are not really helpful in understanding this simple basic question. If we have started to take the formal approach to the problem, why should we stop short of answering such a question? Here are a few notions to date:

IND-O-CPA [7]: It is similar to the LOR-based indistinguishability notion [4] for symmetric key encryption, except that queries of the adversary have to satisfy some order-preserving property. This notion is natural but as we stated above, it is not achievable for schemes with a super-polynomial size message space [7].

POPF-CCA [7]: This is a very important notion which says that a CCA adversary cannot distinguish a pair of an encryption and decryption oracles from a pair of an order-preserving random oracle and its inverse. This notion is natural and therefore should be further studied. But currently, nothing is known about what partial information it can hide and what it cannot hide, as pointed out in [8].

$(r, q+1)$**-WOW [8] (Window One-Wayness):** It says that[1] no adversary, who gets $q+1$ encryptions $C_*, C_1, \ldots, C_q$ of uniformly randomly selected unknown messages, can find an interval I of length $\leq r$ satisfying $\mathsf{Dec}_K(C^*) \in I$. This notion is important since it captures the following natural database setting: Randomly selected $q+1$ elements 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 the ciphertexts in it. This notion, however, does not ensure anything about the secrecy of internal plaintext partial information, since it is "one-way-based" in nature.

$(r, q+1)$**-WDOW [8]:** It is another one-way-based notion defined in [8]. Since it is one-way-based, it also does not tell us what partial information about the plaintexts is hidden.

1.1 Our Contributions

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 partial information OPE can hide." Here (while respecting earlier important works on the subject) we propose the first achievable indistinguishability notion for OPE regarding partial plaintext information hiding. More specifically: we show that our notion can assure *secrecy of lower bits of a plaintext* in the same natural settings as WOW [8].

Our Security Notion — $(\mathcal{X}, \theta, q)$**-indistinguishability:** It is defined based on $(r, q+1)$-WOW [8]. But since WOW is inherently one-way-based, our security notion is defined as a "hybrid" of WOW and indistinguishability as follows. Consider the same database setting as WOW, where an honest entity (not the adversary!) stores $q+1$ his data elements $m^*, m_1, \ldots, m_q$ in their encrypted forms in a database and an adversary A, who wants to get knowledge of m^*, breaches the database system and gets all ciphertexts in it. Above, the messages $m_1, \ldots, m_q$ have been selected according to given distributions $\mathcal{X}_1, \ldots, \mathcal{X}_q$.

The difference from WOW is that m^* has been selected as follows: two messages m_0^* and m_1^* are generated using a polynomial time machine Mg called *message generator*, and set $m^* \leftarrow m_b^*$, where b is a random bit hidden from A.

For $\mathcal{X} = (\mathcal{X}_i)_{i=1,\ldots,q}$, an OPE scheme is called $(\mathcal{X}, \theta, q)$-*indistinguishable* if the advantage of A in the above game (guessing the bit b beyond probability $1/2$) is negligible for any A and for any Mg whose output satisfies

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

[1] Here we adopt the simpler definition of the window one-wayness notion given in Appendix B of the full paper of [8], which can be reduced to the definition of Section 3 of that paper and and vise versa.

Restriction (1.1) enables us to avoid the known attack [7] since it applies only when the distance between m_0^* and m_1^* is large.

Our Results: We will show in Section 2 the following fact:

Fact 1 (informal). *If an OPE scheme satisfies $(\mathcal{X}, \theta, q)$-indistinguishability, the least significant $\lfloor \log_2 \theta \rfloor$ bits of a plaintext are hidden from the adversary in the above database setting.*

We then propose a new OPE scheme $\bar{\mathcal{E}}_{k,\theta}$ based on a pseudo-random function PRF and show the following facts in Section 4. Below, $\mathcal{X}_1, \ldots, \mathcal{X}_q$ are distributions on $[1..M]$ such that they are independent from one another and one can take a sample from $\mathcal{X}_i$ in time polynomial in λ.

Theorem 2 (informal). *Let β and t be constants satisfying $0 < t < \beta \leq 1$. Suppose that the message space size M is super-polynomial in the security parameter λ. Then, for any $\mathcal{X} = (\mathcal{X}_i)_i$ satisfying $\forall i \ : \ H_\infty(\mathcal{X}_i) \geq \beta \log_2 M$, $\bar{\mathcal{E}}_{k,\theta}$ satisfies $(\mathcal{X}, M^t, q)$-indistinguishability under the condition that PRF is secure.*

Remarks: First, 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 their secrecy, since the scheme of [7] also reveals its high order bits [8]. Second, since any distribution $\mathcal{Y}$ on $[1..M]$ satisfies $0 \leq H_\infty(\mathcal{Y}) \leq \log_2 M$, the condition $H_\infty(\mathcal{X}_i) \geq \beta \log_2 M$ means that the ratio of $H_\infty(\mathcal{X}_i)$ to the maximum $\log_2 M$ has to be more than β. Third, Theorem 2 requires that the message space size M is super-polynomial in λ: which is exactly the same condition assumed by Boldyreva et.al.[8] to get their results. Fourth, Theorem 2 shows stronger security when t is closer to β, though the advantage bound decrease is slower in this case.

Due to the above results, we can conclude the following crucial facts:

Knowledge of $\mathcal{X}$: Theorem 2 only requires $\mathcal{X}$ to satisfy the entropy bound. Hence, we can show $(\mathcal{X}, \theta, q)$-indistinguishability *even when we do know the tuple $\mathcal{X}$ of message distributions completely in advance.* This fact is very important because the complete knowledge of $\mathcal{X}$ is not realistic in a central application, like the secure database above, when, for instance, plaintexts are names of new students (with the lexicographic order) or scores of some examination.

Fraction $t < \beta$ of Lower Bits Are Hidden: Due to Fact 1, $(\mathcal{X}, M^t, q)$-indistinguishability implies secrecy of the least significant $\lfloor \log M^t \rfloor$ bits of a plaintext. Since the maximum bit length of a message in the message space $[1..M]$ is $\lfloor \log_2 M \rfloor + 1$, Theorem 2 shows that our scheme with the above parameters can ensure secrecy of the fraction

$$\frac{\lfloor \log M^t \rfloor}{\lfloor \log_2 M \rfloor + 1} \approx t < \beta \tag{1.2}$$

of the least significant bits of a plaintext. The above secrecy can be shown even when we do not have complete knowledge of plaintext distributions.

Any Fraction of Low-Order Bits Are Hidden in the Uniform Distribution Case: In the most significant case where plaintexts distribute uniformly at random, Theorem 2, in particular, shows that our scheme can ensure secrecy of any fraction of the least significant bits of the plaintext because the maximum $\log_2 M$ of the min-entropy is achieved by the uniform distribution and we therefore can set β to 1 in this case.

Allowing Decryption Queries: As in [8], we can naturally make our scheme secure even when we allow the adversary to make decryption queries at any time, using the "Encrypt-then-Mac" composition (adding MAC data) [5].

Open Problem: We can show that our scheme satisfies $\mathsf{Enc}_K(m+1) = \mathsf{Enc}_K(m) + 1$ with high probability. Hence, an adversary can break the scheme if she can get $\mathsf{Enc}_K(m_b^* +$ (small value)), where (m_0^*, m_1^*) is a challenge query of her. (Our proof for Theorem 2 ensures that she can get it only with negligible probability.) Designing a scheme ensuring security for this case is an important open problem.

Finally, we give a note about the construction of our scheme. Since Boldyreva et.al. [7] already gave a natural security notion, POPF-CCA, one important approach to study indistinguishability of OPE is to show that POPF-CCA implies some indistinguishability notion, such as ours. However, we take a different approach in this paper because currently, we do not have much knowledge about the random order-preserving function used in the definition of POPF-CCA, which means that showing our security notion based on POPF-CCA seems to us to be hard. Rather, we define a specific scheme $\bar{\mathcal{E}}_{k,\theta}$ designed for showing our security notion. Showing some indistinguishability results for the more natural security notion, POPF-CCA, is, of course, of independent interest and we leave it as an important open problem.

1.2 Other Security Notions

We also introduce two more security notions for OPE.

(k, θ)-FTG-O-nCPA: This is an (artificial) variant of an indistinguishability notion. We will give the definition of it in Section 3 and show that this notion with suitable parameter implies $(\mathcal{X}, \theta, q)$-indistinguishability for any $\mathcal{X} = (\mathcal{X}_i)_i$ such that $H_\infty(\mathcal{X}_i)$ is larger than the predetermined constant. We then use this fact to show $(\mathcal{X}, \theta, q)$-indistinguishability of our proposed scheme.

WOWM — Stronger Variant of WOW [8]: Informally, $(r, q+1)$-WOWM says that no adversary given $\mathsf{Enc}_K(m^*)$, and $(m_i, \mathsf{Enc}_K(m_i))_{i=1,\ldots,q}$ can find an interval I of length $\leq r$ satisfying $m^* \in I$. This is stronger than $(r, q+1)$-WOW because it allows an adversary to watch $(m_i)_i$ while $(r, q+1)$-WOW prohibits her from doing this.

We will show in Section 5 the following facts. The $(\mathcal{X}, \theta, q)$-indistinguishability notion with suitable parameters implies $(r, q+1)$-WOWM. For any constant $0 \leq \rho < 1$, our scheme with suitable parameters satisfies $(M^\rho, q+1)$-WOWM (and therefore $(M^\rho, q+1)$-WOW).

1.3 Comparison with Known Results [7,8]

First, we clarify what our results owe to [8]: we consider the same natural "database as a service" setting of WOW as described in Section 1.1, and our results are shown under the same condition as WOW [8], that is, the message space size M is super-polynomial in λ. (Note that, technically, our proposed scheme owes the excellent "lazy sampling" of [7] as well.)

Next, we clarify the difference of them. 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: $(\mathcal{X}, \theta, q)$-indistinguishability of the scheme [7] is unknown, because our goal is newly defined. Moreover, we can prove that the known scheme [7] cannot satisfy $(\mathcal{U}^q, M^t, q)$-indistinguishability for $t > 1/2$. (See our full paper for the proof.)

Our scheme achieves $(\mathcal{U}^q, M^t, q)$-indistinguishability for any $0 \leq t < 1$, where $\mathcal{U}^q$ was the tuple of the uniform distributions on the message space. This means that it can hide (in the sense of semantic security) any fraction t of the least significant bits of a plaintext in our setting with uniformly randomly selections of plaintexts. Even when plaintext distributions are not the uniform ones, the scheme can hide fraction $t < \beta$ of lower bits of a plaintext. (β is determined depending on the min-entropy measure of other plaintexts).

About WOW [8]: The known best result [8] is $(1, q)$-WOW security of the scheme of [7]. But it is proved that this scheme cannot achieve $(M^\rho, q+1)$-WOW [8] for any $\rho > 1/2$. In contrast, for any constant $0 \leq \rho < 1$, our scheme with suitable parameters can satisfies $(M^\rho, q+1)$-WOWM (and therefore $(M^\rho, q+1)$-WOW, in particular).

Finally, we describe the POPF-CCA notion given in the seminal work [7].

About POPF-CCA [7]: POPF-CCA is very important notion which can ensure indistinguishability from an ideal object, while our security notion cannot ensure it. Hence, POPF-CCA, as a notion, is more natural and has much potential like other real-vs-ideal definitions and it can ensure security in lots of situations while ours can ensure it in the specific situation described before. E.g. our notion can ensure nothing when an adversary knows $\mathsf{Enc}_K(m)$ and $\mathsf{Enc}_K(m+1)$ while POPF-CCA can ensure something even in this situation. In particular, our notion does not imply POPF-CCA and therefore, POPF-CCA has independent interest.

But currently and unfortunately, nothing is known about what POPF-CCA can hide and what it cannot hide, as pointed out in [8]. This is the motivation behind our entire investigation. Our result is the first positive result in the sense of indistinguishability. Showing some indistinguishability for a more natural notion like, say, POPF-CCA, is an important open issue.

1.4 Other Related Works

Property preserving encryptions [18,2,10] was introduced by Pandey and Rouselakis [18] as a variants of the OPE. Although the security notions for this scheme

can be the same as for OPE, almost the same attack as that of [7] can break any OPE scheme under these security notions when the scheme has a super-polynomial size message space. See our full paper for the details.

CEOE and *MOPE* schemes (introduced by Boldyreva, Chenette, Lee, and O'Neill [8]), *mOPE* and *stOPE* schemes (introduced by Popa, Li, and Zeldovich [19]), and *GOPE* schemes (introduced by Xiao and Yen [25]) achieve stronger security than OPE by sacrificing some of their functionalities, by allowing interactions, or by considering restrictive cases, respectively. *Comparable encryption* schemes (introduced by Furukawa [12,13]) consider an encrypted database where the database manager can search messages m satisfying $m > u$ on behalf of a user if a key K_u depending on u is given from the user as a query. These notions are of independent interests, some may require further formalizations, and are all beyond the scope of this work.

Yum, Kim, Kim, Lee and Hong [28] propose a more efficient method to compute the encryption and decryption functions of the known scheme [7]. Xiao, Yen, and Huynh [27] study OPE in a multi-user setting. Xiao and Yen [26] estimates the min-entropy of a plaintext encrypted by the known scheme [7].

2 $(\mathcal{X}, \theta, q)$-indistinguishability

We introduce notations and terminology and then define our security notion.

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 $\forall m, m' \in \mathcal{M}$ and $\forall K \leftarrow \mathsf{Kg}(1^\lambda)$. Here "$<$" represents the numerical order. Throughout this paper, we assume w.l.o.g. that $\mathcal{M}$ can be written as $[1..M]$.

Definition 3 ($(\mathcal{X}, \theta, q)$-indistinguishability). Let λ, $\mathcal{E} = (\mathsf{Kg}, \mathsf{Enc}, \mathsf{Dec})$, $\theta = \theta(\lambda) > 0$, and $q = q(\lambda) > 0$ be a security parameter, an OPE scheme, a real number, and a polynomial respectively and $\mathcal{X} = (\mathcal{X}_i)_{i \in [1..q]}$ be a tuple 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 (called *message generator*) whose outputs $(m_0^*, m_1^*, \mathsf{info})$ satisfies

$$m_0^* < m_1^*, \qquad |m_1^* - m_0^*| \leq \theta \tag{2.1}$$

and any polynomial time adversary A, $\mathsf{Adv}.\mathsf{Exp}_{\mathcal{E}}^{(\mathcal{X},\theta,q)\text{-indis.}}(\mathsf{Mg}, \mathsf{A}) = |\Pr[\mathsf{Exp}_{\mathcal{E}}^{(\mathcal{X},\theta,q)\text{-indis.}}(\mathsf{Mg}, \mathsf{A}, 1) = 1] - \Pr[\mathsf{Exp}_{\mathcal{E}}^{(\mathcal{X},\theta,q)\text{-indis.}}(\mathsf{Mg}, \mathsf{A}, 0) = 1]|$ is negligible. Here $\mathsf{Exp}_{\mathcal{E}}^{(\mathcal{X},\theta,q)\text{-indis.}}(\mathsf{Mg}, \mathsf{A}, b)$ is defined as follows.

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

Remarks: First, when we consider the above notion, the probability that $m_i \in [m_0^*..m_1^*]$ has to be negligible because otherwise, an OPE scheme under the above notion is broken by an adversary simply by checking $\mathsf{Enc}_K(m_b^*) > \mathsf{Enc}_k(m_i)$. This condition will be automatically satisfied in our theorems due to the selection of parameters of our scheme. Second, due to the bit string info moving between the parties, we can re-interpret the above definition as Mg and A being the "guess and find stages" of a single adversary $(\mathsf{Mg}, \mathsf{A})$ where info is her state.

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 theorem: for any two elements of I, all of their bits 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.

Then, our security notion, in particular, ensures that any element m_0^* of I is indistinguishable from that of a uniformly random element m_1^* of I, because our condition (1.1) is satisfied due to the definition 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^*.

3 (k, θ)-FTG-O-nCPA

In this section, we introduce a security notion, (k, θ)-FTG-O-nCPA, and using it, we give a sufficient condition for $(\mathcal{X}, \theta, q)$-indistinguishability.

(k, θ)-FTG-O-nCPA: It is Find-Then-Guess [4] type indistinguishability for nCPA adversary whose queries satisfy the conditions (3.1), ... (3.4) described later. Here *nCPA (non-adaptive CPA)* [16,14,15] is a type of attack where the adversary is required to output encryption queries $m_1, \ldots, m_q$ and challenge query (m_0^*, m_1^*) together at the same time and gets their answers thereafter.

Definition 4 ((k, θ)-FTG-O-nCPA). For real numbers $k = k(\lambda) > 0$ and $\theta = \theta(\lambda) > 0$, an OPE $\mathcal{E}$ is said to be (k, θ)-*FTG-O-nCPA secure* if for any polynomial time adversary $\mathsf{A} = (\mathsf{A}_{\mathsf{find}}, \mathsf{A}_{\mathsf{guess}})$, the advantage $\mathsf{Adv}.\mathsf{Exp}_{\mathcal{E}}^{(k,\theta)\text{-FTG-O-nCPA}}(\mathsf{A}) = |\Pr[\mathsf{Exp}_{\mathcal{E}}^{(k,\theta)\text{-FTG-O-nCPA}}(\mathsf{A}, 1) = 1] - \Pr[\mathsf{Exp}_{\mathcal{E}}^{(k,\theta)\text{-FTG-O-nCPA}}(\mathsf{A}, 0) = 1]|$ 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 (3.1), (3.2), and (3.3). We also assume (3.4) throughout this paper w.l.o.g.

$$\forall i \ : \ m_i < m_0^* \ \Leftrightarrow \ m_i < m_1^*, \tag{3.1}$$

$$|m_0^* - m_1^*| \leq \theta, \tag{3.2}$$

$$\forall d \in \{0,1\}, \forall i \ : \ |m_d^* - m_i| \geq k\theta. \tag{3.3}$$

$$m_0^* < m_1^* \tag{3.4}$$

Above, (3.1) requires the order preserving property, (3.2) requires the same condition as $(\mathcal{X}, \theta, q)$-indistinguishability, and (3.3) requires the distance $|m_d^* - m_i|$ has to be bigger than the given constant $k\theta$ for any d and i. (3.3) is required because without it, an adversary can take (m_0^*, m_1^*) and m_1 such that $|m_1^* - m_1|$ is much larger than $|m_0^* - m_1|$ (when θ is big. Say, take any m_0^* and set $m_1^* \leftarrow m_0^* + \theta$ and $m_1 \leftarrow m_0^* - 1$). Then, since OPE reveals information about the distance between the two plaintexts, an adversary can know b by checking $|\mathsf{Enc}_K(m_b^*) - \mathsf{Enc}_K(m)|$.

Sufficient Condition: Using (k, θ)-FTG-O-nCPA, we can give the following sufficient condition for $(\mathcal{X}, \theta, q)$-indistinguishability. Below, λ, $\mathcal{E}$, $q = q(\lambda)$ are a security parameter, an OPE scheme on a message space $[1..M]$, and a polynomial respectively. $\mathcal{X}_1, \ldots, \mathcal{X}_q$ are distributions on $[1..M]$ such that they are independent from one another and one can take a sample from $\mathcal{X}_i$ in time polynomial in λ. (M and $\mathcal{X}$ can depend on λ.) 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.

Theorem 5 (Sufficient Condition for $(\mathcal{X}, \theta, q)$-indistinguishability). *Let $\beta > 0$ be any constant. For $k = k(\lambda) > 0$, $\theta = \theta(\lambda) > 0$, if*

$$\forall i \in [1..q] \quad : \quad H_\infty(\mathcal{X}_i) \geq \beta \log_2 M \tag{3.5}$$

holds for any λ, then $\forall \mathsf{Mg} \forall \mathsf{A} \exists \mathsf{B}$:

$$\mathsf{Adv}.\mathsf{Exp}_{\mathcal{E}}^{(\mathcal{X},\theta,q)\text{-indis.}}(\mathsf{Mg}, \mathsf{A}) \leq \mathsf{Adv}.\mathsf{Exp}_{\mathcal{E}}^{(k,\theta)\text{-FTG-O-nCPA}}(\mathsf{B}) + O\left(\frac{qk\theta}{M^\beta}\right). \tag{3.6}$$

We next give two notes reg. Theorem 5. First, as in Theorem 2, condition (3.5) means that the ratio of $H_\infty(\mathcal{X}_i)$ to the maximum $\log_2 M$ has to be more than β. Second, the right hand side of (3.6) is negligible only when $k\theta/M^\beta$ is negligible. We will show that $k\theta/M^\beta$ is, in fact, negligible (for suitable parameters k and θ we will choose) in the proof of Theorem 7, which uses the above theorem.

Proof. 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.

Let I be the interval $(m_0^* - k\theta..m_1^* + k\theta)$. The above B will violate constraint (3.3) if $m_i \in I$ holds for some i. But the probability that $m_i \in I$ holds for some i is $\sum_{i\in[1..q]} \Pr[m_i \leftarrow \mathcal{X}_i \ : \ m_i \in I] \leq (\text{length of } I) \cdot \sum_{i\in[1..q]} \max_{x\in I} \Pr[m_i \leftarrow \mathcal{X}_i \ : \ m_i = x] \leq \sum_{i\in[1..q]} \frac{(2k+1)\theta}{2^{H_\infty(\mathcal{X}_i)}} \leq O\left(\frac{qk\theta}{M^\beta}\right)$. When $m_i \notin I$ holds, (3.1) is also satisfied. Moreover (2.1) implies (3.2). Thus, Theorem 5 follows. □

4 Our Scheme

4.1 Our Goal

This section is devoted to constructing our scheme $\bar{\mathcal{E}}_{k,\theta}$ satisfying the following theorem: Below, A and B are adversaries for (k,θ)-FTG-O-nCPA and PRF respectively, λ is a security parameter, and $\mathsf{neg}(\cdot)$ is some negligible function which is determined independently of (k,θ,A).

Theorem 6 ((k,θ)-FTG-O-nCPA of $\bar{\mathcal{E}}_{k,\theta}$). *For $k,\theta > 0$, $\forall\mathsf{A}\exists\mathsf{B}$:*

$$\mathsf{Adv.Exp}^{(k,\theta)\text{-FTG-O-nCPA}}_{\bar{\mathcal{E}}_{k,\theta}}(\mathsf{A}) \leq O\left(\frac{1}{\sqrt{k}}\right) + \mathsf{Adv.Exp}_{\mathsf{PRF}}(\mathsf{B}) + \mathsf{neg}(\lambda) \tag{4.1}$$

holds when $k \to \infty$. (The value θ does not affect the advantage bound.)

Moreover, the computational costs of algorithms of $\bar{\mathcal{E}}_{k,\theta}$ and the ciphertext length of it are within polynomial of $\log k$, $\log\theta$, $\log M$, and λ, where M is the size of the message space $[1..M]$.

Due to Theorem 5, our scheme satisfies the following theorem as well. Below, M is the size of message space $[1..M]$ of our scheme $\bar{\mathcal{E}}_{k,\theta}$, $q = q(\lambda)$ is a polynomial, and $\mathcal{X}_1,\ldots,\mathcal{X}_q$ are distributions on $[1..M]$ such that they are independent from one another and one can take a sample from $\mathcal{X}_i$ in time polynomial in λ, $\mathsf{neg}(\cdot)$ is some negligible function, A and Mg are an adversary and a message generator for $(\mathcal{X},M^t,q)$-indistinguishability, B is an adversary for PRF, and $\mathsf{Adv.Exp}_{\mathsf{PRF}}(\mathsf{B})$ is an advantage of B in the experiments of PRF.

Theorem 7 ($(\mathcal{X},\theta,q)$-Indistinguishability of Our Scheme, Formal Version of Theorem 2). *Let $0 < \beta \leq 1$ be any constant. Suppose that $\mathcal{X} = (\mathcal{X}_1,\ldots,\mathcal{X}_q)$ satisfies*

$$\forall i \in [1..q] \quad : \quad H_\infty(\mathcal{X}_i) \geq \beta \log_2 M. \tag{4.2}$$

Then, for any constant $0 < t < \beta(\leq 1)$, our scheme $\bar{\mathcal{E}}_{k,\theta}$ with suitable (k,θ) (depending on (M,β,t)) satisfies $\forall\mathsf{Mg}\forall\mathsf{A}\exists\mathsf{B}$:

$$\mathsf{Adv.Exp}^{(\mathcal{X},M^t,q)\text{-indis.}}_{\bar{\mathcal{E}}_{k,\theta}}(\mathsf{Mg},\mathsf{A}) \leq O\left(\frac{q}{M^{\frac{\beta-t}{3}}}\right) + \mathsf{Adv.Exp}_{\mathsf{PRF}}(\mathsf{B}) + \mathsf{neg}(\lambda). \tag{4.3}$$

Moreover, the computational costs of algorithms of $\bar{\mathcal{E}}_{k,\theta}$ and the ciphertext length of it are within polynomial of t, β, $\log M$, and λ.

The right hand sides of (4.3) becomes negligible under the condition that the message space size M is super-polynomial in λ.

Reduction from Theorem 7 to Theorem 5 and 6: Theorem 7 follows if we set parameters (k,θ) of our scheme $\bar{\mathcal{E}}_{k,\theta}$ as

$$(k,\theta) = (M^{2(\beta-t)/3}, M^t) \tag{4.4}$$

because in this case, terms of (3.6) and (4.1) become $O\left(\frac{qk\theta}{M^\beta}\right) = O(\frac{qM^{2(\beta-t)/3} \cdot M^t}{M^\beta})$ $= O(\frac{q}{M^{(\beta-t)/3}})$ and $O(\frac{1}{\sqrt{k}}) = O(\frac{1}{M^{(\beta-t)/3}})$. They are negligible when $M \to \infty$ because the constants t and β satisfy the condition $0 < t < \beta \leq 1$ of Theorem 7. The computational costs of algorithms in our scheme and the ciphertext length of it are polynomial in $\log M$ even when parameters are set as in (4.4), due to the latter part of Theorem 6 and the condition $0 < t < \beta \leq 1$. □

4.2 Scheme $\mathcal{E}_{k,\theta}$ with Polysize Message Space

The goal of this section is designing an OPE scheme $\mathcal{E}_{k,\theta}$ whose advantage bound regarding (k, θ)-FTG-O-nCPA is given in Theorem 6. But the message space size M of $\mathcal{E}_{k,\theta}$ must be bounded by some polynomial in the security parameter λ. (Hence, e.g. the upper bound (4.3) of an advantage for this scheme is not negligible although the bound itself holds even for this scheme.) We stress that $\mathcal{E}_{k,\theta}$ is *not* our proposed scheme.

The scheme $\mathcal{E}_{k,\theta}$ does not use PRF although Theorem 6 refers to it and the discussion in this subsection is purely information theoretic ones. The PRF will be used to design our proposed scheme $\bar{\mathcal{E}}_{k,\theta}$ in the next subsection.

Ideas behind Construction. The scheme $\mathcal{E}_{k,\theta}$ is constructed based mainly on three ideas. Firstly, we write an OPE encryption $\mathsf{Enc}_K(m)$ on a message space $[1..M]$ as $\mathsf{Enc}_K(m) = R + \sum_{i \in [2..m]} \delta_i$, where $R = \mathsf{Enc}_K(1)$ and $\delta_i = \mathsf{Enc}_K(i) - \mathsf{Enc}_K(i-1)$. Then, a design of an OPE encryption can be reduced to the selections of R and (δ_i).

Secondly, we take some values j_0, j_1, ..., and set δ_{j_0}, δ_{j_1}, ... and/or R to random values which are very large compare to other δ_i, so as to hide a (smaller) secret value which the adversary wants to know. A naive way to apply this idea is that we set R to a large random value, while setting all δ_i to 1. Then, the large randomness R seems to hide the secret bit b of a challenge ciphertext $\mathsf{Enc}_K(m_b^*) = R + \sum_{i \in [2..m_b^*]} \delta_i = m_b^* + R - 1$. But, in fact, the adversary can recover b because she can cancel out R by computing $\mathsf{Enc}_K(m_b) - \mathsf{Enc}_K(m') = m_b - m'$, where m' and $\mathsf{Enc}_K(m')$ are her encryption query and its answer.

Therefore, we set some $\delta_{j_0}, \delta_{j_1}, \ldots$, to large random values as well and expect that the set $\{j_0, j_1, \ldots\}$ of indices of them and queries of the adversary to satisfy "good relation" in the sense that, for some j_s, the adversary cannot cancel out δ_{j_s} even when she has encryption queries and their answers. (The precise meaning of this "good relation" will be given later.)

But, the problem is that we cannot know her queries in advance. Therefore, after we fix $j_0, j_1, \ldots$, she may choose her queries such that the queries and $\{j_0, j_1, \ldots\}$ do not satisfy the good relation. So, thirdly, we solve the above problem by introducing another key idea: changing the bit length of δ_i randomly. Specifically, for each i, we flip a random coin ρ_i which becomes 0 with small probability p and then samples δ_i randomly from some given large set if $\rho_i = 0$ and set $\delta_i \leftarrow 1$ otherwise. Then the set $I = \{j_0, j_1, \ldots, \}$ of indices of large δ_i varies randomly and, (due to the definition of nCPA,) we can hide I from the view of the adversary until she determines her queries. Hence, the adversary

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

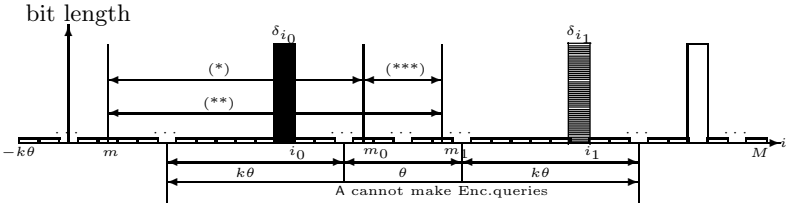

Fig. 1. The Scheme of Section 4.2 (upper) and the Intuition Behind Its Security (lower). In the lower figure, $\mathsf{Enc}_K(m_0) - \mathsf{Enc}_K(m)$, $\mathsf{Enc}_K(m_1) - \mathsf{Enc}_K(m)$, and the difference of them are the sum of δ_i in (*), (**), and (***) respectively. Since both (*) and (**) contain a large randomness δ_{i_0}, 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.

cannot arrange intentionally her queries such that the queries and I do not satisfy the good relation.

Note that this idea has resemblance to the partitioned technique [24] of Waters for an identity based encryption, where one takes some parameters (which determine a "partition") randomly and secretly and expects that queries of an adversary fall into some good places.

Scheme $\mathcal{E}_{k,\theta}$: The formal description of our scheme is given in Fig.1. Here k and θ be the values which we want to show (k, θ)-FTG-nCPA security for, p is a parameter which we will determine in (4.9), and $\mathsf{Binom}(n, p)$ is a binomial distribution.

We set in Fig.1 $\mathsf{Enc}_K(m) = \sum_{i\in[A..m]} \delta_i$ where $A = -k\theta - 1 < 0$ is a fixed value while in the idea described before, we set $\mathsf{Enc}_K(m) = R + \sum_{i\in[2..m]} \delta_i$. (That is, we set $R \leftarrow \sum_{i\in[A..1]} \delta_i$.) Due to this change, we can simplify the security proof for the case where an adversary take as a query a small value m, such as $m = 0$.

$\mathcal{X}_\lambda$ is a probability distribution such that a random variable taken from it can hide other values, specifically,

$$\exists \xi \ : \ (\text{negligible func.}), \quad \forall \alpha, \beta \in [-\theta..\theta], \text{ for } \delta \xleftarrow{\$} \mathcal{X}_\lambda, \ \mathsf{SD}(\alpha + \delta, \beta + \delta) \le \xi(\lambda), \tag{4.5}$$

where SD denotes statistical distance. We can use the uniform distribution on $[1..2^\lambda\theta]$ as $\mathcal{X}_\lambda$ for example. But the scheme in Section 4.3 will use another distribution due to a technical reason.

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 4.3.

(k, θ)-FTG-nCPA Security of $\mathcal{E}_{k,\theta}$: Let k and θ be the values which we want to show (k, θ)-FTG-nCPA security for. Then, 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 below with high probability. (See Fig.1). Bellow, we call δ_i *large number* if it is taken from $[0..2^\lambda M]$ and 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.[2]

$$\text{All } \delta_i \text{ in } [m_0..m_1] \text{ are 1,} \tag{4.6}$$

$$\text{Some } \delta_{i_0} \text{ in } [m_0 - k\theta..m_0] \text{ is large,} \tag{4.7}$$

$$\text{Some } \delta_{i_1} \text{ in } [m_1..m_1 + k\theta] \text{ is large.} \tag{4.8}$$

Note that the precise meaning of "good relation" given in "Ideas behind Construction" is that $(\delta_i)_{i\in(A..M]}$ and queries of an adversary satisfy all of the above three properties.

Here we exploit constraints (3.3) and (3.4) of (k, θ)-FTG-nCPA. Due to them, encryption query m has to satisfy $m \le m_0 - k\theta$ or $m \ge m_1 + k\theta$. 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.

The above discussion shows that the secret bit b is hidden by "barriers" δ_{i_0} and δ_{i_1}. Based on the same idea, we can show, more generally, that the distribution of the secret bit b is independent from the view of an adversary even when she knows encryption queries and their answers, under the assumption that (4.6), (4.7), and (4.8) hold. (See the full paper for the formal proof.)

Upper Bound on Advantage: The rest of thing we have to do is to show the advantage bound of (4.1) by estimating the probabilities that (4.6), (4.7), and (4.8) hold. To this end, we set the parameter p of the scheme $\mathcal{E}_{k,\theta}$ as follows:

$$p = 1 - \left(1 - \frac{1}{\sqrt{k}}\right)^{\frac{1}{\theta}}. \tag{4.9}$$

[2] 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 $[a..b)$.

Then the advantage bound is calculated as follows. Let E_1, E_2, and E_3 be, respectively, the events that condition (4.6), (4.7), and (4.8) 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

$$\begin{aligned}\Pr[\mathsf{Bad}] &\le \frac{1}{\sqrt{k}} + 2\left(1-\frac{1}{\sqrt{k}}\right)^k = \frac{1}{\sqrt{k}} + 2\left\{\left(1-\frac{1}{\sqrt{k}}\right)^{\sqrt{k}}\right\}^{\sqrt{k}} \\ &= \frac{1}{\sqrt{k}} + O\left(\mathrm{e}^{-\sqrt{k}}\right) = O\left(\frac{1}{\sqrt{k}}\right), \end{aligned} \tag{4.10}$$

which is the bound given in Theorem 6.

About CPA Security: The above 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. However, a CPA adversary can choose (m_0^*, m_1^*) in the region $(m_i..m_{i+1}]$ where $\mathsf{Enc}_K(m_{i+1}) - \mathsf{Enc}_K(m_i)$ is the smallest, where $m_1 < \ldots < m_q$ are the encryption queries and $(\mathsf{Enc}_K(m_i))_i$ are their answers. Then all δ_i contained in the sum $\mathsf{Enc}_K(m_{i+1}) - \mathsf{Enc}_K(m_i) = \sum_{i \in (m_i..m_{i+1}]} \delta_i$ must be small as well. This means that the probabilities that conditions (4.7) and (4.8) hold must be smaller than those of the case of nCPA. Hence, our proof does not work well in the CPA setting.

4.3 The Proposed Scheme

By improving the scheme $\mathcal{E}_{k,\theta}$ of Section 4.2, we achieve our proposed OPE scheme $\bar{\mathcal{E}}_{k,\theta}$. The encryption and decryption algorithms of it stay polynomial time in the logarithm in the message space M, which enables M to become a super-polynomial in the security parameter λ.

Idea of the Full Paper of [7]: The starting point of our improvement is the following excellent new "lazy sampling" [6] technique of Section 6 of the full paper of [7]: They construct a polynomial time algorithm[3] $\bar{G}$ 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. Using $\bar{G}$, their improved encryption algorithm $\overline{\mathsf{Enc}}(m)$ computes a ciphertext C_m of m the following binary search recursion:

[3] To simplify, here we only consider the case where inputs of $\bar{G}$ are (u, C_u) and (v, C_v), although the full paper of [7] considers more general case due to some technical reasons.

It 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 algorithms $\bar{G}$ and 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 4.2 does not allow construction of $\bar{G}$ to be based simply on C_u. Below, ρ_i, δ_i, and $A = -k\theta - 1$ are as defined in the scheme of Section 4.2.

$$I_u \leftarrow (C_u^{(0)}, C_u^{(1)}) \leftarrow \Big(\sum_{\substack{i \in (A..u] \\ \rho_i = 0}} \delta_i, \sum_{\substack{i \in (A..u] \\ \rho_i = 1}} \delta_i \Big). \tag{4.11}$$

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

Output Init is indistinguishable from (I_A, I_M). (4.12)

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 $w = \lceil (u+v)/2 \rceil$. (4.13)

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 4.2 is $\sum_{i \in (A..u]} \delta_i$, and therefore satisfies

$$\mathsf{Enc}_K(m) = C_m^{(0)} + C_m^{(1)}. \tag{4.14}$$

As in the case of [7], 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 (4.5) to a binomial distribution

$$\mathcal{X}_\lambda = \mathcal{B}(2^\lambda \theta^2, 1/2) \tag{4.15}$$

with suitable parameters. Note that this $\mathcal{X}_\lambda$, in fact, satisfies (4.5), which is the property required to ensure the security of the scheme of Section 4.2. Formally, the following fact holds (See the full paper for the proof):

Proposition 8 (Binomial Satisfies (4.5)). *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 $\xi(\lambda)$.*

(4.15) allows us to write I_u by using two binomial distributions because (4.11) shows that I_u can be written as sums of δ_i, step 13 of Fig.1 and (4.15) 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 (4.12) 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. Hence, our algorithm $\bar{G}$ satisfying (4.13) can be constructed by using hypergeometric distributions. Since the values which follow the binomial and hypergeometric distributions can be generated in polynomial time [11], our algorithms Init and $\bar{G}$ can terminate in polynomial time.

The description of our algorithms $\bar{G}$ and Init is given in Fig.2. 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 (4.12) and (4.13); see the full paper for the proof.

Proposition 9 (Init and $\bar{G}$ Work Well). *For constants A and M be given in Fig.1, tuples $(\delta_i)_{i\in(A..M]}$ and $(\rho_i)_{i\in(A..M]}$ generated as in $\mathsf{Kg}(1^\lambda)$ of Fig.1, and I_u defined as is (4.11), (4.12) and (4.13) hold.*

We denote the encryption function given in the above way by $\widetilde{\mathsf{Enc}}$. Then, from (4.12), (4.13), and the construction of $\widetilde{\mathsf{Enc}}$, the following proposition holds. (See the full paper for the formal proof.)

Proposition 10. *Take A, M, Kg, Enc as in Fig.1 and take $\overline{\mathsf{Kg}}$ as in Fig.2. 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 [7]. Then our final encryption algorithm $\overline{\mathsf{Enc}}$ is obtained.

Formal Description of Our Scheme: It is given in Fig.2. Here k and θ are the values which we want to show (k, θ)-FTG-O-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 1. Cph, in turn, is an algorithm which computes a ciphertext C_u from I_u based on (4.14). 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 pseudo-random function.

(k, θ)-FTG-O-nCPA: Theorem 6 follows from Proposition 8, 9, and 10, and the security of the scheme of Section 4.2. See the full paper for the formal proof of Proposition 8, 9, and 10 and Theorem 6.

Efficiency: The algorithms of our scheme can terminate within polynomial time in $\log M$, $\log k$, $\log \theta$, and security parameter λ due to our binary recursion search and polynomial time algorithms [11] of binomial and hypergeometric distributions. The ciphertext bit length is not more than $\lambda + 2\log_2 \theta + \log_2(M + k\theta + 1)$ because, due to Proposition 10, a ciphertext can be written as $\sum_{i\in[A..m)} \delta_i$

Parameters:
· 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. $(I_A, I_M) \leftarrow \mathsf{Init}$.
43. Return $\bar{K} \leftarrow (K', A, M, I_A, I_M)$.

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

Init
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).

$\overline{\mathsf{Enc}}_{\bar{K}}(m)$
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. $\bar{K} \leftarrow \begin{cases} (K', u, w, I_u, I_w) & \text{if } m \le w \\ (K', w, v, I_w, I_v) & \text{otherwise} \end{cases}$
57. Return $\overline{\mathsf{Enc}}_{\bar{K}}(m)$.

$\overline{\mathsf{Dec}}_{\bar{K}}(C)$
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. $\bar{K} \leftarrow \begin{cases} (K', u, w, I_u, I_w) & \text{if } C \le \mathsf{Cph}(I_w) \\ (K', w, v, I_w, I_v) & \text{otherwise} \end{cases}$
67. Return $\overline{\mathsf{Dec}}_{\bar{K}}(C)$

$\bar{G}(u, v, I_u, I_v)$
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. 2. The Schemes of Section 4.3, its Parameters, and its Subroutines

for some $m \in [1..M]$ and each δ_i is not more than $2^\lambda\theta^2$ due to (4.15). When we set $(k, \theta) = (M^{2(\beta-t)/3}, M^t)$ as in (4.4), the ciphertext bit length becomes $\lambda + 3\log_2 M +$ (lower terms) due to $0 < t < \beta < 1$.

On the other hand, the known scheme [7] can ensure $(1, q + 1)$-WOW if the ciphertext length is more than $(\log_2 M) + 1$ when M is super-polynomial of λ.

5 Stronger Window-OneWayness of Our Scheme

Finally, we study a stronger variant of (r, q)-WOW notion, called (r, q)-WOWM (studied in [8] intuitively as well). Our definition of WOWM is based on the simpler definition of WOW given in Appendix B of the full paper of [8] which can be reduced to the original WOW given in Section 3 of that paper and vise versa.

Definition 11 ((r, q)-WOWM). An OPE scheme $\mathcal{E}$ on the message space $[1..M]$ is said to be (r, q)-*WOWM (Window One-Way viewing Messages)* if for any polynomial time adversary A, $\mathsf{Succ}.\mathsf{Exp}_{\mathcal{E}}^{(r,q)\text{-WOWM}}(\mathsf{A}) = \Pr[\mathsf{Exp}^{(r,q)\text{-WOWM}}_{\mathcal{E}}(\mathsf{A}) = 1]$ is negligible for the message space size M. Here, experiment

$\mathsf{Exp}_{\mathcal{E}}^{(r,q)\text{-WOWM}}(\mathsf{A})$ is defined as follows. Below, $\mathsf{Comb}_q(M)$ be the set of q-element subset of $[1..M]$.

$$
\begin{aligned}
&K \leftarrow \mathsf{Kg}(1^\lambda), \boldsymbol{m} \xleftarrow{\$} \mathsf{Comb}_q(M), m_* \xleftarrow{\$} \boldsymbol{m},\\
&(m_L, m_R) \leftarrow \mathsf{A}(\mathsf{Enc}_K(m_*), (m, \mathsf{Enc}_K(m))_{m \in \boldsymbol{m} \setminus \{m_*\}}),\\
&\text{Return 1 iff } m_* \in \mathcal{S}(m_L, m_R),
\end{aligned}
$$

$$
\text{where } \mathcal{S}(m_L, m_R) = \begin{cases} [m_L..m_R] & \text{if } m_L \le m_R \\ [1..m_R] \cup [m_L..M] & \text{otherwise.,} \end{cases}
$$

"$m_* \xleftarrow{\$} \boldsymbol{m}$" 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 $\#\mathcal{S}(m_L, m_R) \le r$.

The following property holds for WOWM and WOW of Appendix B of the full paper of [8] because they are the same except that A can view $\boldsymbol{m} \setminus \{m_*\}$.

$$
\forall \mathsf{A} \ : \ \mathsf{Adv}.\mathsf{Exp}_{\mathcal{E}}^{(r,q)\text{-WOW}}(\mathsf{A}) \le \mathsf{Adv}.\mathsf{Exp}_{\mathcal{E}}^{(r,q)\text{-WOWM}}(\mathsf{A}). \tag{5.1}
$$

Lemma 12 (Relationship between $(\mathcal{U}^q, \theta, q)$-indis. and WOWM). *Let $q = q(\lambda)$ be a polynomial of security parameter λ, $\mathcal{E}$ be an OPE scheme with a message space $[1..M]$, $\mathcal{U}^q$ be the tuple of q uniform distributions on $[1..M]$, and $0 < t < 1$ be a constant. Suppose that $\mathcal{E}$ is $(\mathcal{U}^q, M^t, q)$-indistinguishable. Then for any constant ρ satisfying*

$$
0 \le \rho < t (< 1), \tag{5.2}
$$

$\mathcal{E}$ is $(M^\rho, q+1)$-WOWM when M is super-polynomial of λ. Specifically,

$$
\begin{aligned}
&\forall \mathsf{A} \exists \mathsf{Mg} \exists \mathsf{B} \ : \ \mathsf{Succ}.\mathsf{Exp}_{\mathcal{E}}^{(M^\rho, q+1)\text{-WOWM}}(\mathsf{A})\\
&\le \mathsf{Adv}.\mathsf{Exp}_{\mathcal{E}}^{(\mathcal{U}^q, M^t, q)\text{-indis}}(\mathsf{Mg}, \mathsf{B}) + O\Big(\frac{1}{M^{t-\rho}}\Big) + O\Big(\frac{1}{M^{1-t}}\Big) + O\Big(\frac{q}{M}\Big).
\end{aligned} \tag{5.3}
$$

The right hand side of (5.3) is negligible when M is super-polynomial to λ because of (5.2). See the full paper for the formal proof of the above lemma. Lemma 12 and Theorem 7 show the following theorem.

Theorem 13 (WOWM of Our Scheme). *For a polynomial $q = q(\lambda)$ and for any constant*

$$
0 \le \rho < 1, \tag{5.4}
$$

our scheme $\bar{\mathcal{E}}_{k,\theta}$ with suitable parameter (depending on (M, ρ)) is $(M^\rho, q+1)$-WOWM under security of PRF (although the advantage bound decreases slower when ρ becomes closer to 1). Specifically,

$$
\forall \mathsf{A} \exists \mathsf{B} \ : \ \mathsf{Succ}.\mathsf{Exp}_{\bar{\mathcal{E}}_{k,\theta}}^{(M^\rho, q+1)\text{-WOWM}}(\mathsf{A}) \le O\left(\frac{q}{M^{\frac{1-\rho}{4}}}\right) + \mathsf{Adv}.\mathsf{Exp}_{\mathsf{PRF}}(\mathsf{B}) + \mathsf{neg}(\lambda). \tag{5.5}
$$

It achieve better ρ than [8]. See Section 1.3 for details.

Proof (Theorem 13). Take any ρ satisfying (5.4) and set

$$(\beta, t) = (1, (3\rho + 1)/4). \tag{5.6}$$

Let $\mathcal{U}$ be the tuple uniform distributions on the message space $[1..M]$ and let $\mathcal{X} = \mathcal{U}^q$. Then two conditions of Theorem 7, (4.2) and $\beta > t$, are satisfied due to $H_\infty(\mathcal{U}) = \log_2 M$, (5.6), and (5.4). Hence, our scheme $\bar{\mathcal{E}}_{k,\theta}$ with suitable parameter (k, θ) is $(\mathcal{U}^q, M^t, q)$-indistinguishable and satisfies (4.3). (Due to (4.4), the parameters are $(k, \theta) = (M^{2(\beta-t)/3}, M^t) = (M^{(1-\rho)/2}, M^{(3\rho+1)/4})$). The condition (5.2) of Lemma 12 follows from (5.6) and (5.4). Hence, our scheme with the above parameters is $(M^\rho, q+1)$-WOWM and satisfies (5.3). The bound (5.5) comes from (5.4) and (5.6) because in (4.3) and (5.3), $O(\frac{q}{M^{(\beta-t)/3}}) = O(\frac{q}{M^{\frac{1}{3}\cdot(1-(3\rho+1)/4)}}) = O(\frac{q}{M^{(1-\rho)/4}})$, $O(\frac{1}{M^{t-\rho}}) = O(\frac{1}{M^{(1-\rho)/4}}) \leq O(\frac{q}{M^{(1-\rho)/4}})$, $O(\frac{1}{M^{1-t}}) = O(\frac{1}{M^{3(1-\rho)/4}}) \leq O(\frac{1}{M^{(1-\rho)/4}})$, and $O(\frac{q}{M}) \leq O(\frac{q}{M^{(1-\rho)/4}})$. □

Acknowledgements. We thank the anonymous reviewers of ASIACRYPT 2014 for useful comments.

References

1. Agrawal, R., Kiernan, J., Srikant, R., Xu, Y.: Order-preserving encryption for numeric data. In: SIGMOD, pp. 563–574 (2004)
2. Agrawal, S., Agrawal, S., Badrinarayanan, S., Kumarasubramanian, A., Prabhakaran, M., Sahai, A.: Function private functional encryption and property preserving encryption: New definitions and positive results. Cryptology ePrint Archive, Report 2013/744 (2013)
3. Bebek, G.: Anti-Tamper Databases: Inference control techniques. Case Western Reserve University (2002)
4. Bellare, M., Desai, A., Jokipii, E., Rogaway, P.: A concrete security treatment of symmetric encryption. In: FOCS, pp. 394–403 (1997)
5. Bellare, M., Namprempre, C.: Authenticated encryption: Relations among notions and analysis of the generic composition paradigm. In: Okamoto, T. (ed.) ASIACRYPT 2000. LNCS, vol. 1976, p. 531. Springer, Heidelberg (2000)
6. Bellare, M., Rogaway, P.: The security of triple encryption and a framework for code-based game-playing proofs. In: Vaudenay, S. (ed.) EUROCRYPT 2006. LNCS, vol. 4004, pp. 409–426. Springer, Heidelberg (2006)
7. Boldyreva, A., Chenette, N., Lee, Y., O'Neill, A.: Order-preserving symmetric encryption. In: Joux, A. (ed.) EUROCRYPT 2009. LNCS, vol. 5479, pp. 224–241. Springer, Heidelberg (2009)
8. Boldyreva, A., Chenette, N., O'Neill, A.: Order-preserving encryption revisited: Improved security analysis and alternative solutions. In: Rogaway, P. (ed.) CRYPTO 2011. LNCS, vol. 6841, pp. 578–595. Springer, Heidelberg (2011)
9. Boneh, D., Waters, B.: Conjunctive, subset, and range queries on encrypted data. In: Vadhan, S.P. (ed.) TCC 2007. LNCS, vol. 4392, pp. 535–554. Springer, Heidelberg (2007)

10. Chatterjee, S., Das, M.P.L.: Property preserving symmetric encryption revisited. Cryptology ePrint Archive, Report 2013/830 (2013)
11. Devroye, L.: Non-Uniform Random Variate Generation. Springer (1986)
12. Furukawa, J.: Request-based comparable encryption. In: Crampton, J., Jajodia, S., Mayes, K. (eds.) ESORICS 2013. LNCS, vol. 8134, pp. 129–146. Springer, Heidelberg (2013)
13. Furukawa, J.: Short comparable encryption. In: CANS (2014)
14. Maurer, U.M., Oswald, Y.A., Pietrzak, K., Sjödin, J.: Luby-rackoff ciphers from weak round functions? In: Vaudenay, S. (ed.) EUROCRYPT 2006. LNCS, vol. 4004, pp. 391–408. Springer, Heidelberg (2006)
15. Maurer, U.M., Pietrzak, K., Renner, R.S.: Indistinguishability amplification. In: Menezes, A. (ed.) CRYPTO 2007. LNCS, vol. 4622, pp. 130–149. Springer, Heidelberg (2007)
16. Minematsu, K., Tsunoo, Y.: Hybrid symmetric encryption using known-plaintext attack-secure components. In: Won, D.H., Kim, S. (eds.) ICISC 2005. LNCS, vol. 3935, pp. 242–260. Springer, Heidelberg (2006)
17. Özsoyoglu, G., Singer, D.A., Chung, S.S.: Anti-tamper databases: Querying encrypted databases. In: De Capitani di Vimercati, S.I., Ray, I., Ray, I. (eds.) Data and Applications Security XVII. IFIP, vol. 142, pp. 133–146. Springer, Boston (2003)
18. Pandey, O., Rouselakis, Y.: Property preserving symmetric encryption. In: Pointcheval, D., Johansson, T. (eds.) EUROCRYPT 2012. LNCS, vol. 7237, pp. 375–391. Springer, Heidelberg (2012)
19. Popa, R.A., Li, F.H., Zeldovich, N.: An ideal-security protocol for order-preserving encoding. In: IEEE Symposium on S&P, pp. 463–477 (2013)
20. Popa, R.A., Redfield, C.M.S., Zeldovich, N., Balakrishnan, H.: CryptDB: protecting confidentiality with encrypted query processing. In: SOSP, pp. 85–100 (2011)
21. Popa, R.A., Redfield, C.M.S., Zeldovich, N., Balakrishnan, H.: CryptDB: processing queries on an encrypted database. Commun. ACM 55(9), 103–111 (2012)
22. Shi, E., Bethencourt, J., Chan, H.T.-H., Song, D.X., Perrig, A.: Multi-dimensional range query over encrypted data. In: IEEE Symposium on S&P, pp. 350–364 (2007)
23. Tu, S., Kaashoek, M.F., Madden, S., Zeldovich, N.: Processing analytical queries over encrypted data. VLDB Endowment 6(5), 289–300 (2013)
24. Waters, B.: Efficient identity-based encryption without random oracles. In: Cramer, R. (ed.) EUROCRYPT 2005. LNCS, vol. 3494, pp. 114–127. Springer, Heidelberg (2005)
25. Xiao, L., Yen, I.-L.: A note for the ideal order-preserving encryption object and generalized order-preserving encryption. IACR Cryptology ePrint Archive 2012, 350 (2012)
26. Xiao, L., Yen, I.-L.: Security analysis for order preserving encryption schemes. In: CISS, pp. 1–6 (2012)
27. Xiao, L., Yen, I.-L., Huynh, D.T.: Extending order preserving encryption for multi-user systems. IACR Cryptology ePrint Archive 2012, 192 (2012)
28. Yum, D.H., Kim, D.S., Kim, J.S., Lee, P.J., Hong, S.J.: Order-preserving encryption for non-uniformly distributed plaintexts. In: Jung, S., Yung, M. (eds.) WISA 2011. LNCS, vol. 7115, pp. 84–97. Springer, Heidelberg (2012)