Order-Preserving Encryption Revisited: Improved Security Analysis and Alternative Solutions

Alexandra Boldyreva[1], Nathan Chenette[1], and Adam O'Neill[2,*]

[1] Georgia Institute of Technology
{sasha,nchenette}@gatech.edu
[2] University of Texas at Austin
adamo@cs.utexas.edu

Abstract. We further the study of order-preserving symmetric encryption (OPE), a primitive for allowing efficient range queries on encrypted data, recently initiated (from a cryptographic perspective) by Boldyreva et al. (Eurocrypt '09). First, we address the open problem of characterizing what encryption via a random order-preserving function (ROPF) leaks about underlying data (ROPF being the "ideal object" in the security definition, POPF, satisfied by their scheme.) In particular, we show that, for a database of randomly distributed plaintexts and appropriate choice of parameters, ROPF encryption leaks neither the precise value of any plaintext nor the precise distance between any two of them. The analysis here is quite technically non-trivial and introduces useful new techniques. On the other hand, we also show that ROPF encryption does leak both the value of any plaintext as well as the distance between any two plaintexts to within a *range* of possibilities roughly the square root of the domain size. We then study schemes that are not order-preserving, but which nevertheless allow efficient range queries and achieve security notions stronger than POPF. In a setting where the entire database is known in advance of key-generation (considered in several prior works), we show that recent constructions of "monotone minimal perfect hash functions" allow to efficiently achieve (an adaptation of) the notion of IND-O(rdered) CPA also considered by Boldyreva et al., which asks that *only* the order relations among the plaintexts is leaked. Finally, we introduce *modular* order-preserving encryption (MOPE), in which the scheme of Boldyreva et al. is prepended with a shift cipher. MOPE improves the security of OPE in a sense, as it does not leak any information about plaintext location. We clarify that our work should not be interpreted as saying the original scheme of Boldyreva et al., or the variants that we introduce, are "secure" or "insecure." Rather, the goal of this line of research is to help practitioners decide whether the options provide a suitable security-functionality tradeoff for a given application.

Keywords: Searchable encryption, symmetric encryption, hypergeometric distribution, range queries.

[*] Part of the work done while at the Georgia Institute of Technology.

P. Rogaway (Ed.): CRYPTO 2011, LNCS 6841, pp. 578–595, 2011.

1 Introduction

Background and Motivation. An order-preserving symmetric encryption (or OPE) scheme is a deterministic symmetric encryption scheme whose encryption algorithm produces ciphertexts that preserve numerical ordering of the plaintexts. OPE was proposed in the database community by Agrawal et al. [1] in 2004 as a tool to support efficient range queries on encrypted data. (When encryption is done using an OPE scheme, a range query simply consists of the encryptions of the two end-points.) However, the first formal cryptographic treatment of OPE did not appear until recently, in the paper by Boldyreva et al. [8]. The authors formalized a security requirement for OPE and proposed an efficient blockcipher-based scheme provably meeting their security definition.

Yet despite having an OPE scheme that provably satisfies their security notion, the authors warn against its practical use before further studies of its security are performed. To explain this, consider the security notion (or "ideal object") from [8], called a pseudorandom order-preserving function (POPF).

Informally, the POPF notion calls an OPE scheme secure if oracle access to its encryption algorithm is indistinguishable from that to a *random* order-preserving function (ROPF), i.e., a random element of the set of all strictly-increasing functions on the same domain and range. This is a rather straight-forward adaptation of the classical notion of pseudorandom function (PRF)—which asks that oracle access to a function be indistinguishable from that to a truly random function on the same domain and range—to the order-preserving context, and it captures some intuition of what should be the "best possible" OPE scheme. However, the POPF definition is somewhat deceiving and confusing in terms of giving an idea of what kind of security it describes. A random function's behavior is well understood: on a new input the output is a random point in the range. Hence, an adversary seeing a function value learns absolutely no information about the pre-image, unless the former happens to coincide with one it has previously seen. But the situation with a random OPF is much harder to describe. It is clear that a random OPF cannot provide such strong security, but what exactly is leaked about the data and what is protected? The distribution of ciphertexts is known and it is not immediately clear if encryption is even one-way.

Despite its authors' warning of lingering unanswered questions, the OPE scheme from [8] immediately received attention from the applied community [21,20,18,17,14]. We agree that a secure OPE is better than no encryption at all and understand why the idea of its implementation may sound appealing. But practical use without a clear security understanding can be very dangerous and thus it is very important to clarify the security questions as soon as possible.

In this work we first address this open problem. We revisit the security of the "ideal object" ROPF introduced by [8] and provide results that help characterize what it leaks and what it protects about the underlying data. We then observe that it may be possible to achieve stronger security notions than POPF using schemes that fall outside the OPE class but nevertheless allow efficient range queries on encrypted data, and propose two such schemes. We now discuss our contributions in more detail.

New Definitions for Studying ROPF Security. As (perhaps surprisingly) pointed out by [8], a random order-preserving function—the ideal object in the POPF definition from that paper—itself requires a cryptographic treatment.

In order to better understand the strengths and limitations of encryption with an ROPF we first propose several security notions. One captures a basic one-wayness security and measures the probability that an adversary, given a set of ciphertexts of random messages, decrypts one of them. (The fact that messages are chosen uniformly at random we call the "uniformity assumption," and it will be discussed later.) We give the adversary multiple challenge ciphertexts because this corresponds to practical settings and because the ciphertexts are not independent from each other: learning more points of the OPE function may give the adversary additional information. We actually consider a more general security notion that asks the adversary given same inputs to guess an interval (window) within which the underlying challenge plaintext lies. This definition helps us get a better sense of how accurately the adversary can identify the location of a data point. The size of the window and the number of challenge ciphertexts are parameters of the definition. When the window size is one, the notion collapses to the case of simple one-wayness.

Our subsequent definitions address leakage of information not about the *location* of the data points but rather the *distances* between them, which seems crucial in other applications (e.g., a database of salaries). Indeed, [8] showed that an ROPF with a practical range size does not hide distances between plaintexts. We attempt to clarify this intuition. We consider a definition measuring the adversary's success in (precisely) guessing the distance[1] between the plaintexts corresponding to any two out of the set of ciphertexts of random messages given to the adversary. Again, we also consider a more general definition where the adversary is allowed to specify a window in which the distance falls.

We analyze security of an ROPF under these definitions as we believe this helps to understand secure pseudorandom OPE schemes' security guarantees and limitations, and also to evaluate the risk of their usage in various applications. (Indeed, we believe they capture the information about the data, namely location and relative distances, that practitioners are most likely to care about in applications.) However, especially in light of the uniformity assumption (which is unlikely to be satisfied in practice), we view our results as providing important steps in the direction of this understanding (as even under this assumption our results are challenging to prove) but still warn against practical usage of OPE based on current knowledge.

Analysis of an ROPF. We first give an upper bound on the one-wayness advantage of any adversary attacking an ROPF. The proof is quite involved (and is explained in detail in the full version [9]), but the result is a very concise, understandable bound that, under reasonable assumptions, does not even

[1] Technically, for purposes that will become clear in the paper, "distance" actually refers to "directed modular distance," i.e. the distance from one point "up" to the other point, possibly wrapping around the space. As such, distance in our context is non-commutative.

depend on the size of the ciphertext space. (Intuitively, an ROPF's one-wayness comes from the function's probability to deviate from points on the linear OPF $m \mapsto (N/M)m$. Increasing the ciphertext space size beyond a certain amount has little to no effect on these deviations.) We evaluate the bound for several parameters to get an idea of its quality. Our evaluation demonstrates that on practical parameters ROPF and POPF-secure OPEs significantly resist one-wayness attacks, i.e. the maximum one-wayness advantage of any adversary is quite low.

On the other hand, our ROPF analysis under the window one-wayness definition shows that a very efficient adversary can successfully break window one-wayness if the size of the window is not very small. In particular, for message space size M and arbitrary constant b, if the window size is approximately $b\sqrt{M}$, there exists an adversary A whose window one-wayness is at least $1 - 2e^{-b^2/2}$. Thus, for b large enough (say, $b \geq 8$), there exists an adversary with window one-wayness advantage very close to one.

We then extend our analysis of an ROPF to the distance one-wayness and window distance one-wayness definitions. Using similar techniques we show entirely analogous results, namely that the former is very small but the latter becomes large when the adversary is allowed to specify a window of size approximately $b\sqrt{M}$.

We conclude our ROPF analysis with several important supplemental remarks regarding the effect of known-plaintext attacks in the schemes, choosing an appropriate ciphertext space size, and the need to satisfy the uniformity assumption in practical implementations.

Achieving Stronger Security. We next consider the question of whether different types of schemes that support efficient range queries can achieve stronger security than POPF. To capture such schemes we introduce a general notion of *efficiently orderable encryption* (EOE), that covers all schemes supporting standard range queries by requiring a publicly computable function that determines order of the underlying plaintexts given any two ciphertexts. Since EOE leaks order of ciphertexts, IND-OCPA (which [8] showed is unachievable by OPE) is an ideal level of security for EOE schemes (although what information about the data can be inferred from it is outside the scope of the current paper).

An Optimally Secure Committed EOE Scheme. We focus on a scenario where we can show something like IND-OCPA security is possible. We define "committed" versions of EOE and IND-OCPA, called CEOE and IND-CCPA, corresponding to a setting where the database is static and completely known to the user in advance of encryption. Such a scenario is apparently important as it was considered in the first paper to propose an order-preserving scheme [1], and was also studied in several works including [13] for the case of exact-match queries. We observe that the more restrictive functionality in this setting allows one to achieve IND-CCPA. We propose a new scheme that uses a monotone minimal perfect hash function (MMPHF) directly as an "order preserving tagging algorithm" for the given message set, together with a secure encryption. The construction allows for easy implementation of range queries while also achieving the strongest security. Moreover, while MMPHFs are known to require long keys [4], recent constructions [4] are close to being space-optimal. Thus, this application of MMPHFs for tagging seems to be a novel, nearly efficient-as-possible way to support range queries, leaking nothing but the order of ciphertexts, when the database is fixed in advance.

A New Modular OPE Scheme and its Analysis. Finally, we propose a technique that improves on the security of any OPE scheme without sacrificing efficiency. Recall that our ROPF analysis reveals information leakage in OPE not alluded to by [8], namely about the *locations* of the data points rather than just the distances between them. We suggest a modification to (that can be viewed as a generalization of) an OPE scheme that overcomes this. The resulting scheme is not order-preserving per se, but still permits range queries—in this case, modular range queries. (When the left end of the queried range is greater than the right end, a modular range query returns the "wrap-around range," i.e. everything greater than the left end or less than the right end.) The modification to the scheme is simple and generic: the encryption algorithm just adds (modulo the size of the message space) a secret offset to the message before encryption. (The secret offset is the same for all messages.) We call a scheme obtained this way a modular OPE scheme, and generalize the security notion: the ideal object is now a random modular OPF (RMOPF), i.e. a random OPF applied to messages with a randomly picked offset. It is easy to see that any secure OPE scheme yields a secure modular OPE scheme using the above transformation.

We show that a random modular OPF, unlike a random OPF, completely hides the locations of the data points (but has the same leakage with respect to distance and window-distance one-wayness). On the other hand, if the adversary is able to recover a single known plaintext-ciphertext pair, security falls back to that of a random OPF.

We also note that the technique with a secret offset can be applied to the CEOE scheme to enhance its security even beyond IND-CCPA when support for modular range queries is sufficient.

Related Work. Efficient (sub-linear time) search on encrypted data for the case of simple exact-match queries has been addressed by [2] in the symmetric-key setting and [6,10,7] in the public-key setting. The work of [16] suggested enabling efficient range queries on encrypted data not by using OPE but so-called *prefix-preserving encryption* (PPE) [22,5]. But as discussed in [16,2], PPE schemes are subject to certain attacks. Allowing range queries on encrypted data in the public-key setting was studied in [11,19], but the solutions are not suitable for large databases, requiring to scan the whole database on every query. As we mentioned, order preserving encryption as an efficient solution for range queries has been proposed in [1], however, they do not provide any formal security analysis.

2 Preliminaries

Notation. If M is an integer, then $[M]$ denotes the set $\{1, \ldots, M\}$. For a set S and $n \leq |S|$, let Comb_n^S denote the set of n-element subsets of S. If $\mathcal{E}nc$ is an encryption function with key K, $\mathbf{x} = (x_1, \ldots, x_\ell)$ is a vector, and $X = \{x_1, \ldots, x_\ell\}$ is a set, then $\mathcal{E}nc(K, \mathbf{x})$ is shorthand for $(\mathcal{E}nc(K, x_1), \ldots, \mathcal{E}nc(K, x_\ell))$ and $\mathcal{E}nc(K, X)$ is shorthand for $\{\mathcal{E}nc(K, x_1), \ldots, \mathcal{E}nc(K, x_\ell)\}$. The same holds for decryption $\mathcal{D}ec$.

A Convention. For simplicity, in many cases we will assume a domain/plaintext space $[M]$ and range/ciphertext space $[N]$, for $N \geq M$. Naturally, all results for arbitrary spaces $\mathcal{D}$, $\mathcal{R}$ can be derived from those of $[|\mathcal{D}|]$, $[|\mathcal{R}|]$.

Range Queries. For fixed plaintext and ciphertext spaces $[M]$ and $[N]$, a range query *target* is a pair of plaintexts (m_L, m_R) that comes in two varieties: *standard* if $m_L \leq m_R$, or *wrap-around* if $m_L > m_R$. If (m_L, m_R) is a target, its associated *range* is $[m_L, m_R]$ in the standard case and $[m_L, M] \cup [1, m_R]$ in the wrap-around case.

To model the intended application, suppose a server has a database encrypted under a scheme $(\mathcal{K}, \mathcal{E}nc, \mathcal{D}ec)$ with key $K \overset{\$}{\leftarrow} \mathcal{K}$. In a *standard range query*, the user submits two unordered ciphertexts $\{c_1, c_2\}$ to the server. Let $(m_1, m_2) = \mathcal{D}ec(K, (c_1, c_2))$. Then the target is $(\min\{m_1, m_2\}, \max\{m_1, m_2\})$, and the server must return the set of ciphertexts in the database whose decryptions fall into the associated range. Notice that these targets are always standard.

In a *modular range query*, the user submits two ordered ciphertexts (c_L, c_R). Let $(m_L, m_R) = \mathcal{D}ec(K, (c_L, c_R))$. Then the range query target is (m_L, m_R), and the server must return the set of ciphertexts in the database whose decryptions fall into the associated range. Notice that these targets can be standard or wrap-around.

Order-Preserving Encryption (OPE). Following [8] we say that $\mathcal{SE}_{\mathcal{D},\mathcal{R}} = (\mathcal{K}, \mathcal{E}nc, \mathcal{D}ec)$ with associated *plaintext-space* $\mathcal{D}$ and *ciphertext-space* $\mathcal{R}$ is *deterministic* if the encryption algorithm $\mathcal{E}nc$ is deterministic. For $A, B \subseteq \mathbb{N}$ with $|A| \leq |B|$, a function $f\colon A \to B$ is *order-preserving* if for all $i, j \in A$, $f(i) > f(j)$ iff $i > j$. We say that deterministic encryption scheme $\mathcal{SE}_{\mathcal{D},\mathcal{R}} = (\mathcal{K}, \mathcal{E}nc, \mathcal{D}ec)$ is *order-preserving* if $\mathcal{E}nc(K, \cdot)$ is an order-preserving function from $\mathcal{D}$ to $\mathcal{R}$ for all K output by $\mathcal{K}$ (with elements of $\mathcal{D}, \mathcal{R}$ interpreted as numbers, encoded as strings).

Security of OPE. We recall the security definition for OPE from [8].[2] Informally (refer to [8] for the formal definition), it says that an OPE scheme is secure if oracle access to its encryption function is indistinguishable from oracle access to a random order-preserving function (ROPF) on the same domain and range. Any secure OPE scheme (including the only currently known blockcipher-based scheme from [8]) should "closely" imitate the behavior of an ROPF. Accordingly we focus in this paper on analyzing the ideal object, an ROPF.

[2] For simplicity, we do not discuss chosen-ciphertext attacks in detail. Note that symmetric schemes such as these can be made resistant to chosen-ciphertext attacks by implementing Encrypt-then-MAC with a MAC having strong unforgeability, preventing adversaries from even constructing valid ciphertexts.

An "Ideal" Scheme ROPF. We define the "ideal" ROPF scheme as follows. Let $\mathsf{OPF}_{\mathcal{D},\mathcal{R}}$ denote the set of all order-preserving functions from $\mathcal{D}$ to $\mathcal{R}$. Define $\mathsf{ROPF}_{\mathcal{D},\mathcal{R}} = (\mathcal{K}_\mathrm{r}, \mathcal{E}nc_\mathrm{r}, \mathcal{D}ec_\mathrm{r})$ as the following encryption scheme:

- $\mathcal{K}_\mathrm{r}$ returns a random element g of $\mathsf{OPF}_{\mathcal{D},\mathcal{R}}$.
- $\mathcal{E}nc_\mathrm{r}$ takes the key and a plaintext m to return $g(m)$.
- $\mathcal{D}ec_\mathrm{r}$ takes the key and a ciphertext c to return $g^{-1}(c)$.

Of course the above scheme is not computationally efficient, but our goal is its security analysis for the purpose of clarifying security of all POPF-secure constructions.

Most Likely Plaintext. Fix a symmetric encryption scheme $\mathcal{SE}_{\mathcal{D},\mathcal{R}} = (\mathcal{K}, \mathcal{E}nc, \mathcal{D}ec)$. For given $c \in \mathcal{R}$, if $m_c \in \mathcal{D}$ is a message such that

$$\Pr\left[\, K \xleftarrow{\$} \mathcal{K} \;:\; \mathcal{E}nc(K, m) = c \,\right]$$

achieves a maximum at $m = m_c$, then we call m_c a (if unique, "the") *most likely plaintext* for c.

Most Likely Plaintext Distance. Fix a symmetric encryption scheme $\mathcal{SE}_{[M],[N]} = (\mathcal{K}, \mathcal{E}nc, \mathcal{D}ec)$. For given $c_1, c_2 \in \mathcal{R}$, if $d_{c_1,c_2} \in \{0, 1, \ldots, M-1\}$ such that

$$\Pr\left[\, K \xleftarrow{\$} \mathcal{K} \;:\; (c_1, c_2) = \mathcal{E}nc(K, (m_1, m_2))\;;\; m_2 - m_1 \bmod M = d \,\right]$$

achieves a maximum at $d = d_{c_1,c_2}$, then we call d_{c_1,c_2} a (if unique, "the") *most likely plaintext distance* from c_1 to c_2.

3 New Security Definitions

As explained in the introduction, the "ideal" ROPF scheme defined in Section 2 itself requires a cryptographic treatment. Toward this end, we propose several generalized security definitions that help us understand its security.

Let $\mathcal{SE}_{[M],[N]} = (\mathcal{K}, \mathcal{E}nc, \mathcal{D}ec)$ be a deterministic symmetric encryption scheme.

Window One-Wayness. The most basic question left unanswered by [8] is whether a POPF-secure scheme is even one-way. Towards this end we start with the one-wayness definition. Our definition is a stronger and more general version of the standard notion of one-wayness. For $1 \leq r \leq M$ and $z \geq 1$, the adversary is given a set of z ciphertexts of (uniformly) random messages and is asked to come up with an interval of size r within which one of the underlying plaintexts lies. We call our notion r, z-window one-wayness (or r, z-WOW). Note that when $r = 1$, the definition collapses to the standard one-wayness definition (for multiple ciphertexts), and we will call it one-wayness for simplicity.

The *r, z-window one-wayness (r, z-WOW) advantage* of adversary A against $\mathcal{SE}_{[M],[N]}$ is

$$\mathbf{Adv}^{r,z\text{-wow}}_{[M],[N]}(A) \;=\; \Pr\left[\, \mathbf{Exp}^{r,z\text{-wow}}_{\mathcal{SE}_{[M],[N]}}(A) = 1 \,\right] \;,$$

where

```
Experiment $\mathbf{Exp}^{r,z\text{-wow}}_{\mathcal{SE}_{[M],[N]}}(A)$
    $K \xleftarrow{\$} \mathcal{K}$ ; $\mathbf{m} \xleftarrow{\$} \mathrm{Comb}^{[M]}_z$ ; $\mathbf{c} \leftarrow \mathcal{E}nc(K, \mathbf{m})$
    $(m_L, m_R) \xleftarrow{\$} A(\mathbf{c})$
    Return 1 if $(m_R - m_L) \bmod M + 1 \le r$ and there exists $m \in \mathbf{m}$ so that
        either $m \in [m_L, m_R]$ or $(m_L > m_R$ and $m \in [m_L, M] \cup [1, m_R])$
    Return 0 otherwise
```

Notice that the latter success condition allows the adversary to specify a window that "wraps around" the message space. Granting this extra power to the adversary will be useful in analyzing the MOPE scheme of Section 5.2.

Window Distance One-Wayness. To identify the extent to which an OPE scheme leaks distance between plaintexts, we also provide a definition in which the adversary attempts to guess the interval of size r in which the distance between any two out of z random plaintexts lies, for $1 \le r \le M$ and $z \ge 2$. We call the notion r, z-window distance one-wayness (r, z-WDOW). When $r = 1$, the adversary has to guess the exact distance between any two of z ciphertexts.

The *r, z-window distance one-way (r, z-WDOW) advantage* of adversary A against scheme $\mathcal{SE}_{[M],[N]}$ is

$$\mathbf{Adv}^{r,z\text{-wdow}}_{[M],[N]}(A) \;=\; \Pr\left[\, \mathbf{Exp}^{r,z\text{-wdow}}_{\mathcal{SE}_{[M],[N]}}(A) = 1 \,\right] \;,$$

where

```
Experiment $\mathbf{Exp}^{r,z\text{-wdow}}_{\mathcal{SE}_{[M],[N]}}(A)$
    $K \xleftarrow{\$} \mathcal{K}$ ; $\mathbf{m} \xleftarrow{\$} \mathrm{Comb}^{[M]}_z$ ; $\mathbf{c} \leftarrow \mathcal{E}nc(K, \mathbf{m})$
    $(d_1, d_2) \xleftarrow{\$} A(\mathbf{c})$
    Return 1 if $d_2 - d_1 + 1 \le r$ and there exist distinct $m_i, m_j \in \mathbf{m}$
        with $m_j - m_i \bmod M \in [d_1, d_2]$
    Return 0 otherwise
```

4 One-Wayness of a Random OPF

This section is devoted to analyzing the "ideal" scheme $\mathsf{ROPF}_{[M],[N]}$ under the security definitions given in the previous section. The first result shows an upper bound on $1, z$-WOW advantage against the scheme. This demonstrates that on practical parameters, ROPF and POPF-secure OPEs significantly resist (size-1-window) one-wayness attacks. In contrast, the second result shows the ideal ROPF scheme is susceptible to an efficient large-window (a constant times $\sqrt{M}$) one-wayness attack, by constructing an adversary and lower-bounding its r, z-WOW advantage.

The analysis then proceeds similarly for window distance one-wayness definitions: we will show analogous contrasting results for small- versus large-window experiments. We now turn to the details of the analysis.

An Upper Bound on the $1, z$-WOW Advantage. The following theorem states an upper bound on the $1, z$-WOW advantage of any adversary against $\mathsf{ROPF}_{[M],[N]}$.

Theorem 1. *For any challenge set of size z and adversary A, if $N \geq 2M$ and $M \geq 15 + z$ then*

$$\mathbf{Adv}^{1,z\text{-wow}}_{\mathsf{ROPF}_{[M],[N]}}(A) < \frac{9z}{\sqrt{M - z + 1}} .$$

The formal proof is quite involved and is in the full version [9]. The idea is to first bound $1, z$-WOW security in terms of $1, 1$-WOW security; because ciphertexts are correlated, a simple hybrid argument does *not* work and our reduction uses new ideas. Then, to bound $1, 1$-WOW security, we take a combinatorial strategy, as follows. We define a ciphertext's most likely plaintext (m.l.p.) and recall the negative hypergeometric distribution (NHGD). We first relate the middle ciphertext's m.l.p.'s NHGD probability for a given plaintext/ciphertext space to that of a space twice the size; iterating this result produces a formula for the middle ciphertext's m.l.p.'s NHGD probability in a large space given the analogous value in a small space. We then relate *any* ciphertext's m.l.p.'s NHGD probability to that of the middle ciphertext in the space. Finally, we approximate the sum of m.l.p. NHGD probabilities over the ciphertext space in terms of that of the middle ciphertext, and hence to that of the middle ciphertext in a smaller space. Plugging in a value for the m.l.p. NHGD probability on the small space and simplifying yields the bound.

Evaluating the Bound. The bound of Theorem 1 is quite succinct—it does not even rely on N (as long as $N \geq 2M$). The result in essence shows that as long as the challenge set size z is small compared to M, the bound is a small constant times $z/\sqrt{M}$. This in turn is small as long as z is small compared to $\sqrt{M}$.

Plugging in some parameters, we can see some numerical bounds. (In all the following, we assume $N \geq 2M$.) For $M = 2^{80}$ and $z = 1$, the bound is $1.2 \cdot 2^{-37}$. For $M = 2^{80}$ and $z = 2^{20}$, the bound is $1.2 \cdot 2^{-17}$. For $M = 2^{80}$ and $z = 2^{38}$, the bound is no longer useful at 1.2.

We see that $\mathsf{ROPF}_{[M],[N]}$ has very good one-wayness security for reasonably-sized parameters. Given the results of [8] our bound for ROPF can be easily adjusted for their POPF construction, by taking into account pseudorandomness of an underlying blockcipher. But as we discussed in the introduction, standard one-wayness may not be sufficient in all applications and we have to also analyze the schemes under other security notions. Thus, we turn to the next result.

A Lower Bound on Large Window One-Wayness. Here we show that there exists a very efficient adversary attacking the window one-wayness of an ROPF for a sufficiently large window size. A more intuitive explanation of the result follows the theorem.

Theorem 2. *For any window size r and challenge set size z, there exists an adversary A such that*

$$\mathbf{Adv}^{r,z\text{-wow}}_{\mathsf{ROPF}_{[M],[N]}}(A) \geq \mathbf{Adv}^{r,1\text{-wow}}_{\mathsf{ROPF}_{[M],[N]}}(A) \geq 1 - 2e^{-\frac{(r-1)^2}{2}\frac{(M-1)}{M^2}} .$$

The proof is in the full version [9]. There, we construct a straightforward attack and demonstrate that it has the above probability of success, using some bounds by Chvátal on the tail probabilities of the hypergeometric distribution.

Intuitively, Theorem 2 implies that for $r \approx b\sqrt{M}$, where b is a large enough constant (say $b \geq 8$), there exists an adversary A whose r-window one-wayness is very close to 1. More precisely, let $r = b\frac{M}{\sqrt{M-1}} + 1$, and the theorem implies there exists an A such that

$$\mathbf{Adv}^{r,z\text{-wow}}_{\mathsf{ROPF}_{[M],[N]}}(A) \geq 1 - 2e^{-b^2/2} .$$

An Upper Bound on the $1, z$-WDOW Advantage. The following theorem, with the proof in the full version [9] , states an upper bound on the $1, z$-distance one-wayness of a random OPF that is very similar to the bound in Theorem 1.

Theorem 3. *For any challenge set size z and adversary A, if $N \geq 2M$ and $M \geq 16 + z$ then*

$$\mathbf{Adv}^{1,z\text{-wdow}}_{\mathsf{ROPF}_{[M],[N]}}(A) \leq \frac{9z(z-1)}{\sqrt{M-z+1}} .$$

Naturally, as this result looks very much like that of Theorem 1, the proof follows the same strategy and achieves similar results. The only differences are that the initial reduction relates r, z-WDOW security to $r, 2$-WDOW security, incurring a factor $z(z-1)$ advantage increase as opposed to just z, and the initial (tight) bound formula replaces parameters N, M with $N-1$, $M-1$. for proof details.

Thus, the $1, z$-window distance one-wayness of a random OPF is upper-bounded in a similar fashion as the $1, z$-window one-wayness, and we conclude that random OPFs have good $1, z$-WDOW security. Again, though, that is not the whole story, as we see next.

A Lower Bound on Window Distance One-Wayness of ROPF. Here, we derive a result similar to that of Theorem 2, but for the window distance one-wayness of a random OPF.

Theorem 4. *For any window size r and challenge set size z, there exists an efficient adversary A such that*

$$\mathbf{Adv}^{r,z\text{-wdow}}_{\mathsf{ROPF}_{[M],[N]}}(A) \geq \mathbf{Adv}^{r,1\text{-wdow}}_{\mathsf{ROPF}_{[M],[N]}}(A) \geq 1 - 2e^{-\frac{(r-1)^2}{2}\frac{(M-2)}{(M-1)^2}} .$$

The proof uses directly a result from the proof of Theorem 2 and appears in the full version [9].

Intuitively, Theorem 4 implies that for $r \approx b\sqrt{M}$, where b is a large enough constant (say, $b \geq 8$), there exists an efficient adversary A whose r-window distance one-wayness advantage is very close to 1. More precisely, let $r = b\frac{M-1}{\sqrt{M-2}}+1$, and the theorem implies there exists an A such that

$$\mathbf{Adv}^{r,z\text{-wow}}_{\mathsf{ROPF}_{[M],[N]}}(A) \geq 1 - 2e^{-b^2/2}.$$

4.1 Further Security Considerations for ROPFs

In this section, we explore several important questions regarding our ROPF security analysis.

Effect of Known-Plaintext Attacks. It is a natural question to ask what happens to the security of an ROPF scheme when the adversary knows a certain number of plaintext-ciphertext pairs. In general, we can answer this question for each definition of one-wayness using a simple extension of the arguments above.

In the scheme $\mathsf{ROPF}_{\mathcal{D},\mathcal{R}}$, known plaintext-ciphertext pairs split the plaintext and ciphertext spaces into subspaces. On each subspace, the analysis under each one-wayness definition reduces to that of an ROPF on the domain and range of the subspace. For instance, if (m_1, c_1) and (m_2, c_2) are known for $m_1 < m_2$, and no other known plaintext-ciphertext pairs occur between these two, then for $\mathcal{D}' = \{m \in \mathcal{D} \mid m_1 < m < m_2\}$ and $\mathcal{R}' = \{c \in \mathcal{R} \mid c_1 < c < c_2\}$, we analyze the behavior of the function on this subspace by considering the one-wayness bounds on $\mathsf{ROPF}_{\mathcal{D}',\mathcal{R}'}$.

This brings up an important issue. For much of our analysis to apply to a scheme, it must be the case that the ciphertext space is at least twice the size of the message space. Therefore, in order to make sure that our analysis will still apply to most subspaces once several plaintext-ciphertext pairs are discovered by the adversary, we would like to choose the initial parameters in such a way that subspaces are unlikely to violate this condition.

Choosing the Ciphertext Space Size. This brings us to the question posed in [8]: given a plaintext space of size M, what should be the size N of the ciphertext space? The recommendation and justification given in [8] was ad-hoc, necessarily so because the paper lacked a notion of security that would in any way depend on the size of N compared to M. Indeed, the choice of N has to do with the nature of the ideal object, an ROPF, while [8] was focused only on pseudorandomly sampling that ideal object, not analyzing it. Now that we have ways of characterizing the security of an ROPF using our one-wayness definitions, we can more justifiably discuss the question of what to choose for N.

For $g \in \mathsf{OPF}_{[M],[N]}$, if $m_1 < m_2 \in [M]$ exist such that $g(m_2) - g(m_1) < 2(m_2 - m_1)$, then we say that g is *shallow* on the ciphertext interval $[g(m_1), g(m_2)]$. The bounds found in the previous sections assume that $N \geq 2M$. Thus, any non-shallow interval can be analyzed through our theorems about one-wayness, and as a result we would like to choose N to avoid shallow intervals, both in the original space and in potential subspaces.

In particular, consider the following result, which bounds the probability that an interval between encryptions of two random plaintexts is shallow.

Proposition 1. *Let* $t = (N-1)/(M-1)$*, and assume* $t \geq 7$*. Let* $m_1 \xleftarrow{\$} [M]$*,* $m_2 \xleftarrow{\$} [M] \setminus \{m_1\}$*,* $K \xleftarrow{\$} \mathcal{K}_{\mathrm{r}}$*,* $\mathcal{E}nc_{\mathrm{r}}(K, (m_1, m_2)) = (c_1, c_2)$*,* $w = c_2 - c_1 \bmod M$*, and* $d = m_2 - m_1 \bmod M$*. Then over the choice of* m_1, m_2, K*,*

$$\Pr[\,2d > w\,] < \frac{3}{t}\frac{1}{\sqrt{(M-1)/\ln M}} \,.$$

The proof can be found in [9] and is mostly algebraic fiddling.

This bound gives us an idea of good values for $t \approx N/M$. In particular, it seems that choosing a constant for $t \geq 7$, that is, taking N to be a constant multiple of M, is sufficient in order to make the above probability negligible. Whether the constant should be large or small depends on one's tolerance for random intervals to be shallow.

On Implementing a Scheme to Support Range Queries using POPF. We stress that most of our analysis relies on the uniformity assumption assumption, namely that challenge messages come from a uniform distribution. (Intuitively, the we need this in our analysis so that the ciphertexts fall into a range subset of the range.) It is an open problem to extend our analysis to other input distributions, and until that is accomplished, we do not recommend practitioners draw any conclusions from the analysis.

5 Achieving Stronger Security

We study new ways to achieve better security than the OPE scheme of [8] while still allowing for efficient range queries on encrypted data. But first, we define a general primitive, Efficiently Orderable Encryption (EOE), that includes all schemes that support efficient standard range queries, including OPE. We show that IND-OCPA, defined and shown to be unachievable by OPE in [8], is the ideal security definition for such schemes.

We define "committed" analogues of EOE and IND-OCPA, namely CEOE and IND-CCPA, that apply to the practical scenario where the database to encrypt is pre-determined and static. Such a setting has been studied in several works on searchable encryption, including the first paper to propose an order-preserving scheme [1,13]. We then propose a new CEOE scheme that is CCPA-secure.

Finally, we develop a generic modification of an OPE that supports modular range queries (but not standard range queries) and overcomes some of the security weaknesses of any OPE that we studied in Section 4. The scheme is not EOE because it does not leak order; rather, it leaks only "modular" order.

Efficiently Orderable Encryption. We say that $\mathcal{EOE} = (\mathcal{K}, \mathcal{E}nc, \mathcal{D}ec, W)$ is an *efficiently-orderable encryption* (EOE) scheme if $\mathcal{K}, \mathcal{E}nc, \mathcal{D}ec$ are the algorithms of a symmetric encryption scheme, W is an efficient algorithm that takes two ciphertexts as input, and defining $C_K = \{\mathcal{E}nc(K,m) \mid m \in \mathcal{M}\}$ as the set of valid ciphertexts for key K,

$$W(c_0, c_1) = \begin{cases} 1 & \text{if } \mathcal{D}ec(K,c_0) < \mathcal{D}ec(K,c_1) \\ 0 & \text{if } \mathcal{D}ec(K,c_0) = \mathcal{D}ec(K,c_1) \\ -1 & \text{if } \mathcal{D}ec(K,c_0) > \mathcal{D}ec(K,c_1) \end{cases}$$

for any key K and all $c_0, c_1 \in C_K$. It is easy to see that such a scheme permits efficient standard range queries, as the server can keep the encrypted database sorted using W.

It is also clear that any OPE scheme $(\mathcal{K}, \mathcal{E}nc, \mathcal{D}ec)$ corresponds to an EOE scheme with the same key generation, encryption, and decryption algorithms, and $W(c_0, c_1)$ outputting 1, 0, or -1 if the relation between c_0 and c_1 is $<$, $=$, or $>$, respectively. But in general an EOE scheme does not have to be deterministic.

5.1 Committed Efficiently-Orderable Encryption

Range Queries on a Predetermined Static Database. Now we consider schemes for the settings when it is possible for the user to preprocess the whole data before encrypting and sending it to the server. For that we allow the key generation of an EOE scheme to take the message set as input, which we rename a *committed* EOE scheme.

Committed efficiently-orderable encryption. A *committed efficiently-orderable encryption* (CEOE) scheme on domain $\mathcal{D}$ is a tuple $(\mathcal{K}, \mathcal{E}nc, \mathcal{D}ec, W)$ satisfying the following.

- The randomized key generation algorithm $\mathcal{K}$ takes a message space $\mathcal{M} \subset \mathcal{D}$ (called the *committed* message space) as input and outputs a secret key K.
- For any committed message space $\mathcal{M} \subset \mathcal{D}$, $(\mathcal{K}(\mathcal{M}), \mathcal{E}nc, \mathcal{D}ec, W)$ is an EOE scheme on $\mathcal{M}$.

We will show that a CEOE scheme can achieve very strong security. In particular, it can achieve the "committed" adaptation of the IND-OCPA notion from [8], where the adversary outputs two vectors of plaintexts with the same order and equality pattern and is asked to guess whether it is given encryptions of the first or second vector. We define *indistinguishability under committed chosen plaintext attacks* (IND-CCPA). The definition mimics IND-OCPA except that the adversary chooses the challenge vectors (now viewed as message spaces) before key generation, and the scheme's key generation algorithm takes the appropriate message space as input.

IND-CCPA. Let $\mathcal{CEOE} = (\mathcal{K}, \mathcal{E}nc, \mathcal{D}ec, W)$ be a CEOE scheme on message space $\mathcal{M}$.

For an adversary $A = (A_1, A_2)$ and $b \in \{0,1\}$ consider the following experiment. (σ denotes a state.)

```
Experiment $\mathbf{Exp}_{\mathcal{CEOE}}^{\text{ind-ccpa-b}}(A)$
    $(\mathcal{M}_0, \mathcal{M}_1, \sigma) \overset{\$}{\leftarrow} A_1$ ; If $|\mathcal{M}_0| \neq |\mathcal{M}_1|$ then output $\perp$.
    Otherwise, let $l = |\mathcal{M}_0| = |\mathcal{M}_1|$
    Let $m_1^j < m_2^j < \ldots < m_l^j$ be the elements of $\mathcal{M}_j$, for $j = 0, 1$
    If there exist $1 \leq i \leq l$ so that $|m_i^0| \neq |m_i^1|$ then output $\perp$
    $K \overset{\$}{\leftarrow} \mathcal{K}(\mathcal{M}_b)$ ; $c_j \leftarrow \mathcal{E}nc(K, m_j^b)$ for $j = 1, \ldots, l$
    $d \overset{\$}{\leftarrow} A_2(\sigma, c_1, c_1, \ldots, c_l)$. Return $d$
```

For an adversary A, define its *ind-ccpa advantage* against $\mathcal{SE}$ as

$$\mathbf{Adv}_{\mathcal{CEOE}}^{\text{ind-ccpa}}(A) \;=\; \Pr\left[\mathbf{Exp}_{\mathcal{CEOE}}^{\text{ind-ccpa-1}}(A) = 1\right] - \Pr\left[\mathbf{Exp}_{\mathcal{CEOE}}^{\text{ind-ccpa-0}}(A) = 1\right] .$$

We say that $\mathcal{CEOE}$ is IND-CCPA secure if the ind-ccpa advantage of any adversary against $\mathcal{CEOE}$ is small.

Our CEOE construction and its security. We now propose a CEOE scheme that will achieve IND-CCPA security. A ciphertext in our scheme consists of a semantically-secure ciphertext of the message concatenated with the tag, which indicates the order of the message in the ordered message list. As a building block for our scheme we use monotone minimal perfect hash functions, defined as follows.

Let $\mathcal{M}$ be a set with a total (lexicographical) order. h is a *monotone minimal perfect hash function* [4] (MMPHF) on $\mathcal{M}$ if h sends the ith largest element of $\mathcal{M}$ to i, for $i = 0, 1, \ldots, |\mathcal{M}| - 1$. Notice that the MMPHF on any given domain $\mathcal{M}$ is unique. So that we can use MMPHFs in the upcoming construction, let an *index tagging scheme* $(\mathcal{K}, \tau)$ be a pair of algorithms such that $\mathcal{K}$ takes a domain $\mathcal{M}$ and outputs a secret key $K_{\mathcal{M}}$ so that $\tau(K_{\mathcal{M}}, \cdot)$ is the (unique) MMPHF for $\mathcal{M}$, while $\tau(K, m) = \perp$ for any $m \notin \mathcal{M}$.

Our CEOE construction is based on two building blocks: MMPHF tagging and any symmetric encryption scheme.

Let $(\mathcal{K}_t, \tau)$ be an index tagging scheme. Fix a universe $\mathcal{D}$, and let $\mathcal{SE} = (\mathcal{K}', \mathcal{E}nc', \mathcal{D}ec')$ be any symmetric encryption scheme on $\mathcal{D}$. We construct a CEOE scheme $(\mathcal{K}, \mathcal{E}nc, \mathcal{D}ec, W)$ as follows.

- $\mathcal{K}$ takes $\mathcal{M} \subset \mathcal{D}$ as input, runs $K_t \leftarrow \mathcal{K}_t(\mathcal{M})$ and $K_e \leftarrow \mathcal{K}'$, and returns $K = K_t \| K_e$.
- $\mathcal{E}nc$ takes key $K = K_t \| K_e$ and message m as input, and computes $i = \tau(K_t, m)$. If $i = \perp$ then $\mathcal{E}nc$ returns $\perp$, otherwise it returns $i \| \mathcal{E}nc'(K_e, m)$.
- $\mathcal{D}ec$ takes key $K = K_t \| K_e$ and ciphertext $c = i \| c'$ as input, and returns $\mathcal{D}ec'(K_e, c')$.
- W takes ciphertexts $c_0 = i_0 \| c_0'$ and $c_1 = i_1 \| c_1'$ as input, and returns 1 if $i_0 < i_1$, 0 if $i_0 = i_1$, and -1 if $i_0 > i_1$.

We note that unlike the scheme with pre-processing for exact-match queries [13], when using the above scheme the server does indexing and query processing as for unencrypted data, which is a practical advantage. Also, as the following result shows, the scheme is secure under IND-CCPA.

Theorem 5. *The CEOE scheme defined above is IND-CCPA-secure provided the underlying symmetric encryption scheme is IND-CPA secure.*

The proof is in the full version [9].

Note that our secure CEOE construction relies on an efficient MMHPF implementation. Luckily, MMHPFs were studied recently by [4]. They showed that for a universe of size 2^w and for $n \geq \log w$, the shortest possible description of an MMPHF function (and thus, best possible key length for a tagging scheme) on n elements is unfortunately quite large at $\Omega(n)$ bits. This is somewhat disheartening, as a naive solution, in which the MMPHF key consists of an n-entry array whose ith entry is the ith largest element in the domain, has a key length of $O(nw)$. Nevertheless, the authors of [4] were able to generate MMPHF descriptions that are closer to the optimal bound: one construction uses $O(n \log\log w)$ bits and has query time $O(\log w)$, and the other uses $O(n \log w)$ bits and has constant query time. This is still large, but may be practical depending on the parameters involved.

5.2 Modular OPE and Analysis of an Ideal MOPE Scheme

Modular OPE. We propose a modification to (that can be viewed as a generalization of) an OPE scheme that improves the security performance of any OPE. The resulting scheme is no longer strictly order-preserving, but it still permits range queries. However, now the queries must be *modular* range queries. Standard range queries are not supported, as only "modular order" rather than order is leaked. The modification from OPE is simple, generic, and basically free computation-wise.

Let $\mathcal{SE}_{[M],[N]} = (\mathcal{K}, \mathcal{E}nc, \mathcal{D}ec)$ be an order-preserving encryption scheme. Define a *modular order-preserving encryption scheme* (MOPE) $\mathcal{SE}_{[M],[N]} = (\mathcal{K}_{\mathrm{m}}$, $\mathcal{E}nc_{\mathrm{m}}$, $\mathcal{D}ec_{\mathrm{m}})$ as follows.

- $\mathcal{K}_{\mathrm{m}}$ runs $\mathcal{K}$ to get K, picks $j \overset{\$}{\leftarrow} [M]$ and returns (K, j).
- $\mathcal{E}nc_{\mathrm{m}}$ on input (K, j) and m returns $\mathcal{E}nc(K, m - j \bmod M)$.
- $\mathcal{D}ec_{\mathrm{m}}$ on inputs (K, j) and c returns $\mathcal{D}ec(K, c) + j \bmod M$.

Notice that a MOPE is suitable for modular range query support as follows. To request the ciphertexts of the messages in the range $[m_1, m_2]$ (if $m_1 \leq m_2$), or $[m_1, M] \cup [1, m_2]$ (if $m_1 > m_2$), the user computes $c_1 \leftarrow \mathcal{E}nc_{\mathrm{m}}(K, m_1), c_2 \leftarrow \mathcal{E}nc_{\mathrm{m}}(K, m_2)$ and submits ciphertexts (c_1, c_2) as the query. The server returns the ciphertexts in the interval $[c_1, c_2]$ (if $c_1 \leq c_2$) or $[c_1, N] \cup [1, c_2]$ (if $c_1 > c_2$).

MOPE Security and Random MOPF. In order to define the security of an MOPE scheme, we introduce a generalization of OPFs. For $j \in [M]$, let $\phi_j : [M] \to [M]$ be the cyclic transformation $\phi_j(x) = (x - j - 1) \bmod M + 1$. We define the set of *modular order preserving functions* from $[M]$ to $[N]$ as

$$\mathsf{MOPF}_{[M],[N]} = \{ f \circ \phi_j \mid f \in \mathsf{OPF}_{[M],[N]}, j \in [M] \} \ .$$

Note that all OPFs are MOPFs; on the other hand, most MOPFs are not OPFs. However, a MOPF g is "modular order-preserving" in that the function $g - g(0) \bmod N$ is order-preserving.

Now, define $\mathsf{RMOPF}_{[M],[N]} = (\mathcal{K}_{\mathrm{rm}}, \mathcal{E}nc_{\mathrm{rm}}, \mathcal{D}ec_{\mathrm{rm}})$, the *random modular order-preserving function* scheme, as the following (inefficient) encryption scheme:

- $\mathcal{K}_{\mathrm{rm}}$ returns a random instance g of $\mathsf{MOPF}_{[M][N]}$.
- $\mathcal{E}nc_{\mathrm{rm}}$ takes the key g and a plaintext m to return $g(m)$.
- $\mathcal{D}ec_{\mathrm{rm}}$ takes the key g and a ciphertext c to return $g^{-1}(c)$.

Note that an MOPF could alternatively be defined with a random ciphertext shift following the OPF rather than a random plaintext shift preceding it. The advantage of the above definition is that the map from (OPF, ciphertext offset) pairs to MOPFs is bijective whereas in the alternative it is not one-to-one.

We now are ready to define MOPE security. Fix an MOPE scheme $\mathcal{SE}_{[M],[N]} = (\mathcal{K}_{\mathrm{m}}, \mathcal{E}nc_{\mathrm{m}}, \mathcal{D}ec_{\mathrm{m}})$. Let $\mathsf{RMOPF}_{[M],[N]} = (\mathcal{K}_{\mathrm{rm}}, \mathcal{E}nc_{\mathrm{rm}}, \mathcal{D}ec_{\mathrm{rm}})$ be as defined above. For an adversary A, define its $\mathbf{Adv}^{\mathrm{pmopf}}_{\mathcal{SE}}(A)$, *pmopf-advantage* (or *pseudorandom modular order-preserving function advantage*) against $\mathcal{SE}$ as

$$\Pr\left[\, K \xleftarrow{\$} \mathcal{K}_{\mathrm{m}} \;:\; A^{\mathcal{E}nc_{\mathrm{m}}(K,\cdot)} = 1 \,\right] - \Pr\left[\, g \xleftarrow{\$} \mathsf{RMOPF}_{[M],[N]} \;:\; A^{g(\cdot)} = 1 \,\right] .$$

It is straightforward to show that the MOPE scheme obtained from any POPF-secure OPE scheme via the transformation defined in the beginning of Section 5.2 is PMOPF-secure, under the same assumption as the base scheme. We omit the details.

We now analyze the ideal object, RMOPF, under the one-wayness definitions.

Window One-Wayness of RMOPF. The following proposition, proved in [9], establishes that RMOPF is optimally r, z-window one-way (and hence optimally one-way, taking $r = 1$) in the sense that an adversary cannot do better than an adversary that outputs a random window independent of the challenge set. (Reminder: "window" includes windows that wrap around the edge of the space.)

Proposition 2. *Fix any window size r and challenge set size z. Let $A_{\mathrm{rand}}(r)$ be an r, z-WOW adversary that, on any input, outputs a random r-window from $[M]$. Then for any adversary A,*

$$\mathbf{Adv}^{r,z\text{-wow}}_{\mathsf{RMOPF}_{[M],[N]}}(A) \le \mathbf{Adv}^{r,z\text{-wow}}_{\mathsf{RMOPF}_{[M],[N]}}(A_{\mathrm{rand}}(r)) \le rz/M .$$

As one might surmise, the above "optimal" characterization of the one-wayness of a random MOPF fails to show a complete picture of the information a random MOPF leaks. To investigate further, we turn to distance one-wayness.

WDOW Advantage Bounds for RMOPF. We claim that the distance one-wayness analysis for RMOPF is exactly the same as for ROPF. To see this, consider the following proposition, whose (short) proof is in [9].

Proposition 3. *Let $c_1, c_2 \in [N]$. Then for any $d \in \{0, \ldots, M-1\}$,*

$$\begin{aligned}&\Pr\left[\, \mathcal{D}ec_{\mathrm{r}}(K_1, c_2) - \mathcal{D}ec_{\mathrm{r}}(K_1, c_1) = d \,\right] \\ &= \Pr\left[\, \mathcal{D}ec_{\mathrm{rm}}(K_2, c_2) - \mathcal{D}ec_{\mathrm{rm}}(K_2, c_1) = d \,\right] ,\end{aligned}$$

where the probabilities are over, respectively, $K_1 \xleftarrow{\$} \mathcal{K}_{\mathrm{r}}$ and $K_2 \xleftarrow{\$} \mathcal{K}_{\mathrm{rm}}$.

Therefore, the $1, z$-WDOW advantage upper bound of Theorem 3 and the r, z-WDOW advantage lower bound of Theorem 4 against ROPF schemes also apply to RMOPF schemes on the same parameters.

So, while an RMOPF has similar security to that of an ROPF for distance and window distance one-wayness, it is better in terms of one-wayness and window one-wayness. The analysis easily transfers to any secure MOPE scheme. We now discuss a few supplemental security considerations for RMOPF schemes.

Effect of a Known-Plaintext Attack on RMOPF. In the $\mathsf{RMOPF}_{[M],[N]}$ scheme, if the adversary learns a single plaintext-ciphertext pair, then the one-wayness analysis reduces to that of $\mathsf{ROPF}_{[M-1],[N-1]}$. To see this, note that if g is a random function in $\mathsf{MOPF}_{[M],[N]}$, and it is revealed that $g(m_0) = c_0$, then $f(m) = g(m + m_0 \bmod M) - c_0 \bmod N$ is a random function in $\mathsf{OPF}_{[M-1],[N-1]}$.

On Implementing a Scheme to Support Range Queries using PMOPF. We note that when a pseudorandom MOPF scheme is used to implement a range-query-supporting database, even wrap-around target range queries must be made, for otherwise an adversary may infer the secret offset of the MOPF scheme after observing many non-wrap-around target queries.

Remark. We finally note that the tagging scheme defined in Section 5.1 could be similarly modified so that its tag receives a secret offset. The resulting scheme would support modular range queries in predetermined static database scenario, and satisfy a stronger version of IND-CCPA, leaking only "modular" order.

Acknowledgements. We thank Nigel Smart, Abdullatif Shikfa and the anonymous reviewers for useful comments. We also thank Adam Smith and Brent Waters for useful discussions, and in particular Adam Smith for pointing out that ROPF encryption leaks the high-order bits of the plaintexts. Alexandra Boldyreva and Nathan Chenette are supported in part by Alexandra's NSF CAREER award 0545659 and NSF Cyber Trust award 0831184. Adam O'Neill was supported in part by Brent Waters grants NSF CNS-0915361 and CNS-0952692.

References

1. Agrawal, R., Kiernan, J., Srikant, R., Xu, Y.: Order-preserving encryption for numeric data. In: SIGMOD 2004, pp. 563–574. ACM, New York (2004)
2. Amanatidis, G., Boldyreva, A., O'Neill, A.: Provably-secure schemes for basic query support in outsourced databases. In: DBSec 2007, pp. 14–30. Springer, Heidelberg (2007)
3. Bauer, F.: Decrypted Secrets: Methods and Maxims of Cryptology. Springer, Heidelberg (2006)
4. Belazzougui, D., Boldi, P., Pagh, R., Vigna, S.: Monotone minimal perfect hashing: searching a sorted table with $o(1)$ accesses. In: SODA 2009, pp. 785–794. SIAM, Philadelphia (2009)
5. Bellare, M., Boldyreva, A., Knudsen, L.R., Namprempre, C.: Online ciphers and the hash-CBC construction. In: Kilian, J. (ed.) CRYPTO 2001. LNCS, vol. 2139, pp. 292–309. Springer, Heidelberg (2001)

6. Bellare, M., Boldyreva, A., O'Neill, A.: Deterministic and efficiently searchable encryption. In: Menezes, A. (ed.) CRYPTO 2007. LNCS, vol. 4622, pp. 535–552. Springer, Heidelberg (2007)
7. Bellare, M., Fischlin, M., O'Neill, A., Ristenpart, T.: Deterministic encryption: Definitional equivalences and constructions without random oracles. In: Wagner, D. (ed.) CRYPTO 2008. LNCS, vol. 5157, pp. 360–378. Springer, Heidelberg (2008)
8. 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)
9. Boldyreva, A., Chenette, N., O'Neill, A.: Order-Preserving Encryption Revisited: Improved Security Analysis and Alternative Solutions (2011) Full version of this paper, `http://www.cc.gatech.edu/~aboldyre/publications.html`
10. Boldyreva, A., Fehr, S., O'Neill, A.: On notions of security for deterministic encryption, and efficient constructions without random oracles. In: Wagner, D. (ed.) CRYPTO 2008. LNCS, vol. 5157, pp. 335–359. Springer, Heidelberg (2008)
11. 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)
12. Chvátal, V.: The tail of the hypergeometric distribution. Discrete Mathematics 25(3), 285–287 (1979)
13. Curtmola, R., Garay, J., Kamara, S., Ostrovsky, R.: Searchable symmetric encryption: Improved denitions and efficient constructions. In: CCS 2006, pp. 79–88. ACM, New York (2006)
14. Ding, Y., Klein, K.: Model-Driven Application-Level Encryption for the Privacy of E-health Data. In: International Conference on Availability, Reliability and Security, pp. 341–346 (2010)
15. Kershaw, D.: Some extensions of W. Gautschi's inequalities for the gamma function. Mathematics of Computation 41(164), 607–611 (1983)
16. Li, J., Omiecinski, E.: Efficiency and security trade-off in supporting range queries on encrypted databases. In: DBSec 2005, pp. 69–83. Springer, Heidelberg (2005)
17. Liu, H., Wang, H., Chen, Y.: Ensuring Data Storage Security against Frequency-Based Attacks in Wireless Networks. In: Rajaraman, R., Moscibroda, T., Dunkels, A., Scaglione, A. (eds.) DCOSS 2010. LNCS, vol. 6131, pp. 201–215. Springer, Heidelberg (2010)
18. Lu, W., Varna, A.L., Wu, M.: Security analysis for privacy preserving search of multimedia. In: Image Processing (ICIP), 2010, pp. 26–29 (2010)
19. Shi, E., Bethencourt, J., Chan, T.-H.H., Song, D., Perrig, A.: Multi-dimensional range query over encrypted data. In: Symposium on Security and Privacy 2007, pp. 350–364. IEEE, Los Alamitos (2007)
20. Tang, Q.: Privacy preserving mapping schemes supporting comparison. In: Proceedings of the ACM Workshop on Cloud Computing Security Workshop (CCSW 2010). ACM, New York (2010)
21. Wang, C., Cao, N., Li, J., Ren, K., Lou, W.: Secure Ranked Keyword Search over Encrypted Cloud Data. In: ICDCS 2010, pp. 253–262. IEEE, Los Alamitos (2010)
22. Xu, J., Fan, J., Ammar, M.H., Moon, S.B.: Prefix-preserving IP address anonymization: Measurement-based security evaluation and a new cryptography-based scheme. In: ICNP 2002, pp. 280–289. IEEE, Los Alamitos (2002)