A preliminary version of this paper appears in *Advances in Cryptology - CRYPTO 2011, 31st Annual International Cryptology Conference,* P. Rogaway ed., LNCS, Springer, 2011.

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

Alexandra Boldyreva* Nathan Chenette* Adam O'Neill†

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 introduces useful new techniques. On the other hand, we show that ROPF encryption leaks approximate value of any plaintext as well as approximate distance between any two plaintexts, each to an accuracy of about 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 random 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.

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

* College of Computing, Georgia Institute of Technology, 266 Ferst Drive, Atlanta, GA 30332, USA. E-mail: {sasha,nchenette}@gatech.edu.

† University of Texas at Austin, 1616 Guadalupe, Austin TX 78701, USA E-mail: adamo@cs.utexas.edu.

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. [9]. 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 from [9], 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 straightforward 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 [9] immediately received attention from the applied community [22, 21, 19, 18, 15]. 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 [9] 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 pointed out by [9], a random order-preserving function—the ideal object in the POPF definition from that paper— itself (perhaps surprisingly) 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, [9] 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 Appendix), but the result is a very concise, understandable bound that, under reasonable assumptions, does not even 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, the indistinguishability under ordered chosen plaintext attacks (IND-OCPA) definition, introduced in [9] and showed is unachievable by OPE, is an ideal level of security for EOE schemes.

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,

[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.

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 [14] 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 [9], 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, 11, 7] in the public-key setting. The work of [17] suggested enabling efficient range queries on encrypted data not by using OPE but so-called *prefix-preserving encryption* (PPE) [23, 5]. But as discussed in [17, 2], PPE schemes are subject to certain attacks. Allowing range queries on encrypted data in the public-key setting was studied in [12, 20], 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 $K \xleftarrow{\$} \mathcal{K}$. 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.

If S is a set then $X \xleftarrow{\$} S$ denotes that X is selected uniformly at random from S. For convenience, for any $k \in \mathbb{N}$ we write $X_1, X_2, \ldots, X_k \xleftarrow{\$} S$ as shorthand for $X_1 \xleftarrow{\$} S, X_2 \xleftarrow{\$} S, \ldots, X_n \xleftarrow{\$} S$. If A is a randomized (resp. deterministic) algorithm then $A(x, y, \ldots)$ denotes the result of running A on inputs $x, y, \ldots$ and $a \xleftarrow{\$} A(x, y, \ldots)$ (resp. $a \leftarrow A(x, y, \ldots)$) means that we let $a = A(x, y, \ldots; R)$.

Order-Preserving Encryption (OPE). Following [9] 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 [9]. (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 using Encrypt-then-MAC [8] generic constructions that prevent adversaries from constructing valid ciphertexts.) Informally (refer to [9] 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 [9]) should "closely" imitate the behavior of an ROPF. Accordingly we focus in this paper on analyzing the ideal object, an ROPF.

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 deterministic order-preserving 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 [9] 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 $1 \le r \le M$ and $z \ge 1$, 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 the experiment $\mathbf{Exp}^{r,z\text{-wow}}_{\mathcal{SE}_{[M],[N]}}(A)$ above is defined as follows.

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 the experiment $\mathbf{Exp}^{r,z\text{-wdow}}_{\mathcal{SE}_{[M],[N]}}(A)$ above is defined as follows.

Experiment $\mathbf{Exp}^{r,z\text{-wdow}}_{\mathcal{SE}_{[M],[N]}}(A)$

$K \xleftarrow{\$} \mathcal{K}$; $\mathbf{m} \xleftarrow{\$} \mathrm{Comb}_z^{[M]}$; $\mathbf{c} \leftarrow \mathcal{E}nc(K, \mathbf{m})$

$(d_1, d_2) \xleftarrow{\$} A(\mathbf{c})$

Return 1 if $d_2 - d_1 + 1 \leq r$ 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 4.1 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 Appendix A. 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 4.1 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 [9] 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 4.2 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 Appendix F. 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 4.2 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 Appendix G, states an upper bound on the $1,z$-distance one-wayness of a random OPF that is very similar to the bound in Theorem 4.1.

Theorem 4.3 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 4.1, 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 $N-1$, $M-1$. See Appendix G 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 4.2, but for the window distance one-wayness of a random OPF.

Theorem 4.4 For any window size r and challenge set size z, there exists an efficient adversary A such that

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

Proof: As in Theorem 4.2, the first inequality is trivially true. It is left to prove the second inequality, which we do by constructing an $r,2$-WDOW adversary A as follows.

Adversary $A(\{c_1, c_2\})$
$w \leftarrow c_2 - c_1 \bmod N$
$d_w \leftarrow \lceil \frac{(M-1)w}{N} \rceil$
$\delta \leftarrow \frac{r-1}{2(M-1)}$
$d_L \leftarrow \max\{d_w - \lfloor \delta(M-1) \rfloor, 1\}$
$d_R \leftarrow \min\{d_w + \lfloor \delta(M-1) \rfloor, M-1\}$
Return (d_L, d_R).

(d_L, d_R) is a legal response in the $r, 2$-WDOW experiment since the associated window has size $d_R - d_L + 1 \leq 2\delta(M-1) + 1 \leq r$.

Note that $d_w = \lceil \frac{(M-1)w}{N} \rceil$ is the most likely plaintext distance between c_1 and c_2 by Corollary I.2. The probability that the adversary succeeds in the $r, 2$-WDOW experiment is the probability that $m_2 - m_1 \bmod M = d \in [d_L, d_R]$, or

$$\sum_{d=d_L}^{d_R} P_*(N-1, M-1, w, d) \geq 1 - 2e^{-\frac{(r-1)^2}{2}\frac{(M-2)}{(M-1)^2}},$$

by Lemma F.1. Since A only performs efficient operations, the result follows. ∎

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 [9]: given a plaintext space of size M, what should be the size N of the ciphertext space? The recommendation and justification given in [9] 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 [9] 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 4.5 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 Appendix J. Besides using Lemma F.1, the proof 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 [9] 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 [9], 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, 14]. 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 [9], 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 $A = (A_1, A_2)$, define its *ind-ccpa advantage* against $\mathcal{SE}$ as

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

where for $b \in \{0,1\}$ the experiments $\mathbf{Exp}^{\text{ind-ccpa-b}}_{\mathcal{CEOE}}(A)$ are define as follows.

Experiment $\mathbf{Exp}^{\text{ind-ccpa-b}}_{\mathcal{CEOE}}(A)$
$(\mathcal{M}_0, \mathcal{M}_1, \sigma) \xleftarrow{\$} A_1$; 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 $\perp$
$K \xleftarrow{\$} \mathcal{K}(\mathcal{M}_b)$; $c_j \leftarrow \mathcal{E}nc(K, m_j^b)$ for $j = 1, \ldots, l$
$d \xleftarrow{\$} A_2(\sigma, c_1, c_1, \ldots, c_l)$. Return d

Above σ denotes a state the adversary can preserve. 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.

Construction 5.1 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 [14], 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 5.2 The CEOE scheme of Construction 5.1 is IND-CCPA-secure provided the underlying symmetric encryption scheme is IND-CPA secure.

Proof: Let $\mathcal{CEOE}$ be the scheme from Construction 5.1 , $\mathcal{D}$ be the domain, and suppose $A = (A_1, A_2)$ is an adversary with nontrivial IND-CCPA advantage against $\mathcal{CEOE}$. We construct an IND-CPA adversary B against $\mathcal{SE}$. B has access to O, a left-right encryption oracle for $\mathcal{SE}$ under a random secret key.

B runs A_1 to receive $\mathcal{M}_0, \mathcal{M}_1, \sigma$. Let l be the lengths of $|\mathcal{M}_0|, |\mathcal{M}_1|$. After sorting (separately) the elements of $\mathcal{M}_0$ and $\mathcal{M}_1$, B assigns label m_i^b to the ith smallest element of $\mathcal{M}_b$, for $i = 1, \ldots, l$ and $b = 0, 1$. B queries its left-right $\mathcal{SE}$-encryption oracle with matched pairs of these messages: $c_i' \leftarrow O(m_i^1, m_i^2)$ for $i = 1, \ldots, l$. Note that each pair consists of messages of equal length. Then, B prepends indices $c_i = i \| c_i'$ for $i = 1, \ldots, l$. Finally, it runs $A_2(\sigma, c_1, \ldots, c_l)$ to receive d, and outputs d.

It is clear that B's communication with A perfectly mimics the IND-OCPA experiment, and thus the IND-CPA advantage of B is equal to the IND-CCPA advantage of A. Clearly, B is efficient, since it only needs to sort the elements of $|\mathcal{M}_0|, |\mathcal{M}_1|$. ∎

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 $j \xleftarrow{\$} [M]$ 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 $m_1 \le m_2$), 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 $c_1 \le c_2$) 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 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 5.3 Fix any window size r and challenge set size z. Let $A_{\text{rand}}(r)$ be an r, z-WOW adversary that, on any input, outputs a random r-window from $[M]$. Then for any adversary A,

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

Proof: Let $V_{\mathbf{m}}$ be the set of r-windows in $[M]$ that contain an element of $\mathbf{m}$. Notice that $|V_{\mathbf{m}}| \leq rz$, as each element of the challenge set is contained in at most r windows. Also, the total number of r-windows in $[M]$ is M. An adversary wins if it outputs an element in $V_{\mathbf{m}}$. Since A_{rand} outputs a random r-window, it is clear that $\mathbf{Adv}^{r,z\text{-wow}}_{\mathsf{RMOPF}_{[M],[N]}}(A_{\text{rand}}) \leq rz/M$.

Fix a function $f \in \mathsf{OPF}_{[M],[N]}$ and challenge set $\mathbf{c}$. Let $f^{-1}(\mathbf{c}) = \{x \in [M] \mid f(x) \in \mathbf{c}\}$. Let S be the set of modular intervals $I' \subseteq [M]$ such that $I' \cap f^{-1}(\mathbf{c}) \neq \emptyset$, and let $n = |S|$. For offset j, an adversary wins if it picks $I = (m_L, m_R)$ such that the interval $I + j \ = \ (m_L + j \bmod M, m_R + j \bmod M)$ is in S. For each I, note that there are precisely n values for $j \in [M]$ for which $I + j \in S$, and precisely $M - n$ for which $I + j \notin S$. Thus, over the choice of j, each interval I has the same probability of winning (namely, n/M.) Hence, a random choice of interval has the same probability of success as any other choice of interval. This is true for any function f and challenge set $\mathbf{c}$, so the result follows. ▌

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.

Proposition 5.4 Let $c_1, c_2 \in [N]$. Then for any $d \in \{0, \ldots, M-1\}$,

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

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

Proof: Let $w = c_2 - c_1 \bmod N$. Note that among the $\binom{N-2}{M-2}$ OPFs f with $c_1, c_2 \in f([M])$, there are $\binom{w-1}{d-1}\binom{N-w-1}{M-d-1}$ such that $f^{-1}(c_2) - f^{-1}(c_1) \bmod M = d$. On the other hand, among the $\binom{N-2}{M-2} \cdot M$ MOPFs g with $c_1, c_2 \in g([M])$, there are $\binom{w-1}{d-1}\binom{N-w-1}{M-d-1} \cdot M$ such that $g^{-1}(c_2) - g^{-1}(c_1) \bmod M = d$. The result follows. ▌

Therefore, the $1, z$-WDOW advantage upper bound of Theorem 4.3 and the r, z-WDOW advantage lower bound of Theorem 4.4 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 of Construction 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.

6 Conclusions

We revisited security of symmetric order-preserving schemes defined in [9]. We formally clarify the strengths and limitations of any OPE scheme proven to be a pseudorandom order-preserving function (POPF), and in particular, the efficient OPE scheme proposed in [9]. Namely, for any POPF-secure OPE our analysis together with the result of [9] provides upper bounds on the advantages of any adversaries attacking the one-wayness and distance one-wayness, (2) lower bounds on the window one-wayness and window distance one-wayness advantages. We hope our results help practitioners to estimate the risks and security guarantees of using a secure OPE in their applications. Our analysis also gives directions in selecting the size of the ciphertext space. Finally we propose a simple and efficient transformation that can be applied to any OPE scheme. Our analysis shows that the transformation yields a scheme with improved security in that the scheme resists the one-wayness and window one-wayness attacks.

7 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. Part of the work done while Adam was at the Georgia Institute of Technology.

References

[1] R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu. Order-preserving encryption for numeric data. In *SIGMOD '04*, pp. 563–574. ACM, 2004.

[2] G. Amanatidis, A. Boldyreva, and A. O'Neill. Provably-secure schemes for basic query support in outsourced databases. In *DBSec '07*, pp. 14–30. Springer, 2007.

[3] F. Bauer. *Decrypted Secrets: Methods and Maxims of Cryptology.* Springer, 2006.

[4] D. Belazzougui, P. Boldi, R. Pagh, and S. Vigna. Monotone minimal perfect hashing: searching a sorted table with $o(1)$ accesses. In *SODA '09*, pp. 785–794, SIAM, 2009.

[5] M. Bellare, A. Boldyreva, L. R. Knudsen, and C. Namprempre. Online ciphers and the Hash-CBC construction. In *CRYPTO '01*, pp. 292–309. Springer, 2001.

[6] M. Bellare, A. Boldyreva, and A. O'Neill. Deterministic and efficiently searchable encryption. In *CRYPTO '07*, pp. 535–552. Springer, 2007.

[7] M. Bellare, M. Fischlin, A. O'Neill, and T. Ristenpart. Deterministic encryption: Definitional equivalences and constructions without random oracles. In *CRYPTO '08*, pp. 360–378. Springer, 2008.

[8] M. Bellare and C. Namprempre. *Authenticated encryption: Relations among notions and analysis of the generic composition paradigm.* In *ASIACRYPT 2000*, volume 1976 of *LNCS*. Springer, 2000.

[9] A. Boldyreva, N. Chenette, Y. Lee and A. O'Neill. Order-preserving symmetric encryption. In *Eurocrypt '09*, pp. 224–241. Springer, 2009.

[10] A. Boldyreva, S. Fehr, and A. O'Neill. On notions of security for deterministic encryption, and efficient constructions without random oracles. In *CRYPTO '08*, pp. 335–359. Springer, 2008.

[11] D. Boneh and B. Waters. Conjunctive, subset, and range queries on encrypted data. In *TCC '07*, pp. 535–554. Springer, 2007.

[12] V. Chvátal. The tail of the hypergeometric distribution. *Discrete Mathematics*, 25(3):285–287, 1979.

[13] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky. Searchable symmetric encryption: Improved denitions and efficient constructions. In *CCS 06*, pp. 79–88. ACM, 2006.

[14] Y. Ding and K. Klein, 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] D. Kershaw. Some extensions of W. Gautschi's inequalities for the gamma function. *Mathematics of Computation*, 41(164):607–611, 1983.

[16] J. Li and E. Omiecinski. Efficiency and security trade-off in supporting range queries on encrypted databases. In *DBSec '05*, pp. 69–83. Springer, 2005.

[17] H. Liu, H. Wang, and Y. Chen. Ensuring Data Storage Security against Frequency-based Attacks in Wireless Networks. In *Proceedings of the International Conference on Distributed Computing in Sensor Systems (DCOSS 2010)*. 2010.

[18] W. Lu, A.L. Varna, and M. Wu; Security analysis for privacy preserving search of multimedia. In *Image Processing (ICIP), 2010*, pp. 26–29. 2010.

[19] E. Shi, J. Bethencourt, T-H. H. Chan, D. Song, and A. Perrig. Multi-dimensional range query over encrypted data. In *Symposium on Security and Privacy '07*, pp. 350–364. IEEE, 2007.

[20] Q. Tang. Privacy preserving mapping schemes supporting comparison. In *Proceedings of the 2010 ACM workshop on Cloud computing security workshop (CCSW '10)*. ACM, 2010.

[21] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou. Secure Ranked Keyword Search over Encrypted Cloud Data. In *ICDCS '10*, pp. 253–262. IEEE, 2010.

[22] J. Xu, J. Fan, M. H. Ammar, and S. B. Moon. Prefix-preserving IP address anonymization: Measurement-based security evaluation and a new cryptography-based scheme. In *ICNP '02*, pp. 280–289. IEEE, 2002.

A Proving Theorem 4.1

Before proceeding, we define probabilities relating to the hypergeometric distribution, and note their connection to random OPFs, which was demonstrated in [9]. These probabilities will show up at several points in the analysis.

Let $N \geq M$, $0 \leq y \leq N$, $0 \leq x \leq M$. We define hypergeometric and negative hypergeometric probabilities, respectively, as follows, for $(x, y \neq 0)$

$$P_{HGD}(N, M, y, x) = \frac{\binom{y}{x}\binom{N-y}{M-x}}{\binom{N}{M}} \; ; \quad P_{NHGD}(N, M, y, x) = \frac{\binom{y-1}{x-1}\binom{N-y}{M-x}}{\binom{N}{M}} \; .$$

For convenience, we also define a third, related probability:

$$P_*(N, M, y, x) \;\; = \;\; \frac{\binom{y-1}{x-1}\binom{N-y}{M-x}}{\binom{N-1}{M-1}} \qquad (x, y \neq 0) \; .$$

As shown in [9], random OPFs are naturally linked to negative hypergeometric probabilities. We will use the fact stated in the main body that for $(\mathcal{K}_{\mathrm{r}}, \mathcal{E}nc_{\mathrm{r}}, \mathcal{D}ec_{\mathrm{r}}) = \mathsf{ROPF}_{[M],[N]}$ and $m \in [M]$, $c \in [N]$,

$$\Pr\left[\, K \leftarrow \mathcal{K}_{\mathrm{r}} \; : \; \mathcal{E}nc_{\mathrm{r}}(K, m) = c \,\right] = P_{NHGD}(N, M, c, m) \; .$$

Now, we turn to the proof. The proof relies on two lemmas and a corollary to a third lemma, as follows.

Lemma A.1 For window size r, challenge set size z, and any adversary A, there exists a OW-adversary A' such that

$$\mathbf{Adv}^{r,z\text{-wow}}_{\mathsf{ROPF}_{[M],[N]}}(A) \leq z\mathbf{Adv}^{r,1\text{-wow}}_{\mathsf{ROPF}_{[M-z+1],[N-z+1]}}(A').$$

The proof is in Appendix B.

Lemma A.2 For any adversary A,

$$\mathbf{Adv}^{1,1\text{-wow}}_{\mathsf{ROPF}_{[M],[N]}}(A) \;\; \leq \;\; \frac{1}{N}\sum_{c=1}^{N} P_*(N, M, c, m_c) \; ,$$

where $m_c = \lceil \frac{Mc}{N+1} \rceil$ for any $c \in [N]$.

The proof is in Appendix C.

Lemma A.3 Let $N_0 \geq 2M_0$ be (positive) multiples of 2 and let $M = 2^q M_0$ and $N = 2^q N_0$ for integer $q \geq 1$. Define $\alpha_0 = P_*(N_0, M_0, N_0/2, m_{N_0/2})$. Then

$$\frac{1}{N}\sum_{c=1}^{N} P_*(N, M, c, m_c) < \frac{2}{M} + \frac{\pi\alpha_0}{2^{q/2+1}} \cdot e^{1/M_0+3/2} \; .$$

The proof is in Appendix D.

Corollary A.4 If $N \geq 2M \geq 32$ and $m_c = \lceil \frac{Mc}{N+1} \rceil$ for any $c \in [N]$, then

$$\frac{1}{N}\sum_{c=1}^{N} P_*(N, M, c, m_c) < \frac{9}{\sqrt{M}} \; .$$

Proof: Let $M_0 = 16$. Then $N_0 \geq 32$, and we have

$$\begin{aligned}
\alpha_0 &= P_*(N_0, M_0, N_0/2, M_0/2) \\
&= P_*(N_0, 16, N_0/2, 8) \\
&= \frac{\binom{N_0/2-1}{7}\binom{N_0/2}{8}}{\binom{N_0-1}{15}} \\
&= \frac{(N_0/2-1)\cdots(N_0/2-7)(N_0/2)\cdots(N_0/2-7)15!}{(N_0-1)\cdots(N_0-15)7!8!} \\
&= \frac{N_0(N_0-2)^2(N_0-4)^2\cdots(N_0/2-14)^2 15!}{(N_0-1)\cdots(N_0-15)2^{15}7!8!} \\
&= \frac{N_0(N_0-2)(N_0-4)\cdots(N_0/2-14)15!}{(N_0-1)(N_0-3)\cdots(N_0-15)2^{15}7!8!} \\
&= \left(1+\frac{1}{N_0-1}\right)\left(1+\frac{1}{N_0-3}\right)\cdots\left(1+\frac{1}{N_0-15}\right)\frac{15!}{2^{15}7!8!} \\
&\leq \left(1+\frac{1}{31}\right)\left(1+\frac{1}{29}\right)\cdots\left(1+\frac{1}{17}\right)\frac{15!}{2^{15}7!8!} \\
&< 0.278.
\end{aligned}$$

Since $M = 2^q M_0 = 2^{q+4}$, we have $2^{q/2+1} = \frac{\sqrt{M}}{2}$. Thus,

$$\frac{2}{M} + \frac{\pi\alpha_0 e^{1/M_0+3/2}}{2^{q/2+1}} < \frac{1/\sqrt{16}}{\sqrt{M}} + \frac{2\pi(0.278)e^{25/16}}{\sqrt{M}} < \frac{9}{\sqrt{M}}\,.$$

The result then follows from Lemma A.3. ▌ Now, we are ready to prove the main, generalized, result.

Proof of Theorem 4.1. Let $M' = M - z + 1$, $N' = N - z + 1$.

$$\begin{aligned}
\mathbf{Adv}^{1,z\text{-wow}}_{\mathsf{ROPF}_{[M],[N]}}(A) &\leq z\mathbf{Adv}^{1,1\text{-wow}}_{\mathsf{ROPF}_{[M'],[N']}}(A) && \text{(Lemma A.1)} \\
&\leq z\tfrac{1}{N'}\textstyle\sum_{c=1}^{N'} P_*(N', M', c, m_c) && \text{(Lemma A.2)} \\
&< z\tfrac{9}{\sqrt{M'}}. && \text{(Corollary A.4)}
\end{aligned}$$

In the final step, note that $N \geq 2M$ and $M \geq 15 + z$ imply $N - z + 1 \geq 2(M - z + 1) \geq 32$. ▌

B Proving Lemma A.1

We first introduce a concept related to r, z-WOW security called *specified* r, z-WOW security. The proof then proceeds in two steps. First, we construct an adversary A' whose specified r, z-WOW advantage is at least a factor $1/z$ of the r, z-WOW advantage of A (which, in fact, works for general schemes). In the second step, we exhibit a bijection between OPFs on the space $[M], [N]$ that hit a fixed set $\mathbf{c} \subseteq [N]$ of size $z - 1$, and OPFs on the space $[M - z + 1], [N - z + 1]$. This allows us to construct an efficient r, 1-WOW adversary against $\mathsf{ROPF}_{[M-z+1],[N-z+1]}$ using an efficient specified r, z-WOW adversary against $\mathsf{ROPF}_{[M],[N]}$, with the same advantage. Putting these constructions together yields the result.

An Intermediate Security Definition. The *specified* r, z*-window-one-wayness advantage* of adversary A with respect to scheme $\mathcal{SE}_{\mathcal{D},\mathcal{R}} = (\mathcal{K}, \mathcal{E}nc, \mathcal{D}ec)$ is

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

where the security experiment is as follows.

Experiment $\mathbf{Exp}^{s\text{-}r,z\text{-wow}}_{\mathcal{SE}_{\mathcal{D},\mathcal{R}}}(A)$

$K \xleftarrow{\$} \mathcal{K}$; $\mathbf{m} \xleftarrow{\$} \mathrm{Comb}_z^{[M]}$

$m_0 \xleftarrow{\$} \mathbf{m}$; $\mathbf{c} \leftarrow \mathcal{E}nc(K, \mathbf{m})$; $c_0 \leftarrow \mathcal{E}nc(K, m_0)$

$(m_L, m_R) \xleftarrow{\$} A(\mathbf{c}, c_0)$

Return 1 if $(m_R - m_L + 1 \bmod M) \le r$ and either

$m_0 \in [m_L, m_R]$ or $(m_L > m_R$ and $m_0 \in [m_L, M] \cup [1, m_R])$;

Return 0 otherwise.

The only difference between this experiment and the standard r, z-WOW one is that here, the experiment demands that the adversary return an r-window containing the pre-image of the *specified* ciphertext $c_0 \in \mathbf{c}$ (rather than any ciphertext from $\mathbf{c}$.)

Reducing r, z-WOW Security to Specified r, z-WOW Security for Any Scheme. As our first step, we show that for any efficient r, z-WOW adversary against a general scheme $\mathcal{SE}$, there exists an efficient specified r, z-WOW adversary A' whose success probability is at least a factor of $1/z$ of that of A.

Lemma B.1 For any scheme $\mathcal{SE}_{\mathcal{D},\mathcal{R}}$ and r, z, and any r, z-WOW adversary A, there exists an equally efficient specified r, z-WOW adversary A' such that

$$\mathbf{Adv}^{r,z\text{-wow}}_{\mathcal{SE}_{\mathcal{D},\mathcal{R}}}(A) \le z\mathbf{Adv}^{s\text{-}r,z\text{-wow}}_{\mathcal{SE}_{\mathcal{D},\mathcal{R}}}(A') \; .$$

Proof: Given A, let A' on input $(\mathbf{c}, c)$ simply run $(m_L, m_R) \xleftarrow{\$} A(\mathbf{c})$ and return (m_L, m_R). Whenever A outputs (m_L, m_R) such that $\exists m \in \mathbf{m}$ with $m \in [m_L, m_R]$ or $(m_L > m_R$ and $m \in [m_L, M] \cup [1, m_R])$, then A' wins if $m = m_0$. Since m_0 is random from $\mathbf{m}$, independent of the rest of the experiment, we conclude that A' wins the specified experiment at least $1/z$ of the times that A wins the standard experiment. The result follows. ∎

Reducing Specified r, z-WOW Security to $r, 1$-WOW Security for ROPFs. Now, fix scheme $\mathsf{ROPF}_{[M],[N]} = (\mathcal{K}_{\mathrm{r}}, \mathcal{E}nc_{\mathrm{r}}, \mathcal{D}ec_{\mathrm{r}}) =$ and r, z. It is left to reduce the success probability of a specified r, z-adversary A against this scheme to that of an $r, 1$-WOW adversary against $\mathsf{ROPF}_{[M-z+1],[N-z+1]}$.

We first introduce a number of notations that will be useful in the proof. Let $z' = z - 1$. For orderable sets $\mathcal{D}, \mathcal{R}$, and $H \subset \mathcal{R}$, let $\mathsf{OPF}_{\mathcal{D},\mathcal{R}}(H)$ denote $\{f \in \mathsf{OPF}_{\mathcal{D},\mathcal{R}} \mid H \subset f(\mathcal{D})\}$, i.e., the set of OPFs from $\mathcal{D}$ to $\mathcal{R}$ with all elements of H in their range. Similarly, for a set U, $n \le |U|$, and $H \subset U$ with $|H| \le n$, let $\mathrm{Comb}_n^U(H)$ denote the set of n-element subsets of U that contain H. For set S with elements $x_1 < x_2 < \ldots x_{|S|}$, and $x \in S$, $H \subseteq S$, $i \in [|S|]$, $I \subseteq [|S|]$, let

$$\begin{aligned} \mathrm{Index}_x^S &= j \text{ such that } x = x_j \; , \\ \mathrm{Indices}_H^S &= \{j \mid x_j \in H\} \; , \\ \mathrm{Element}_i^S &= x_i \; , \\ \mathrm{Elements}_I^S &= \{x_i \mid i \in I\} \; . \end{aligned}$$

Finally, for equal-sized orderable sets S_1, S_2, let $\mathrm{UniqueOPF}(S_1, S_2)$ be the unique OPF from S_1 to S_2.

The next lemma demonstrates the connection between OPFs in space $[M], [N]$ that hit a certain z'-element subset of $[N]$, and general OPFs in space $[M - z'], [N - z']$.

Lemma B.2 Fix $\mathbf{c} \subseteq [M]$ with $|\mathbf{c}| = z'$. There is a chain of natural bijections between the following sets.

$$\mathsf{OPF}_{[M],[N]}(\mathbf{c}) \overset{\beta_1}{\rightleftarrows} \mathrm{Comb}_M^{[N]}(\mathbf{c}) \overset{\beta_2}{\rightleftarrows} \mathrm{Comb}_{M-z'}^{[N]\setminus\mathbf{c}} \overset{\beta_3}{\rightleftarrows} \mathrm{Comb}_{[M-z']}^{[N-z']} \overset{\beta_4}{\rightleftarrows} \mathsf{OPF}_{[M-z'],[N-z']}$$

Proof: The bijective functions and their inverses can be defined as follows:

- $\beta_1 : f \mapsto f([M])$; $\quad \beta_1^{-1} : S \mapsto \text{UniqueOPF}([M], S)$
- $\beta_2 : S \mapsto S \setminus \mathbf{c}$; $\quad \beta_2^{-1} : S \mapsto S \cup \mathbf{c}$
- $\beta_3 : S \mapsto \text{Indices}_S^{[N]\setminus\mathbf{c}}$; $\quad \beta_3^{-1} : I \mapsto \text{Elements}_I^{[N]\setminus\mathbf{c}}$
- $\beta_4 : S \mapsto \text{UniqueOPF}([M - z'], S)$; $\quad \beta_4^{-1} : f \mapsto f([M - z'])$

Since all functions are well-defined, the bijections are clear.
See Figure 1 for a visual depiction of elements associated through the bijections. ∎

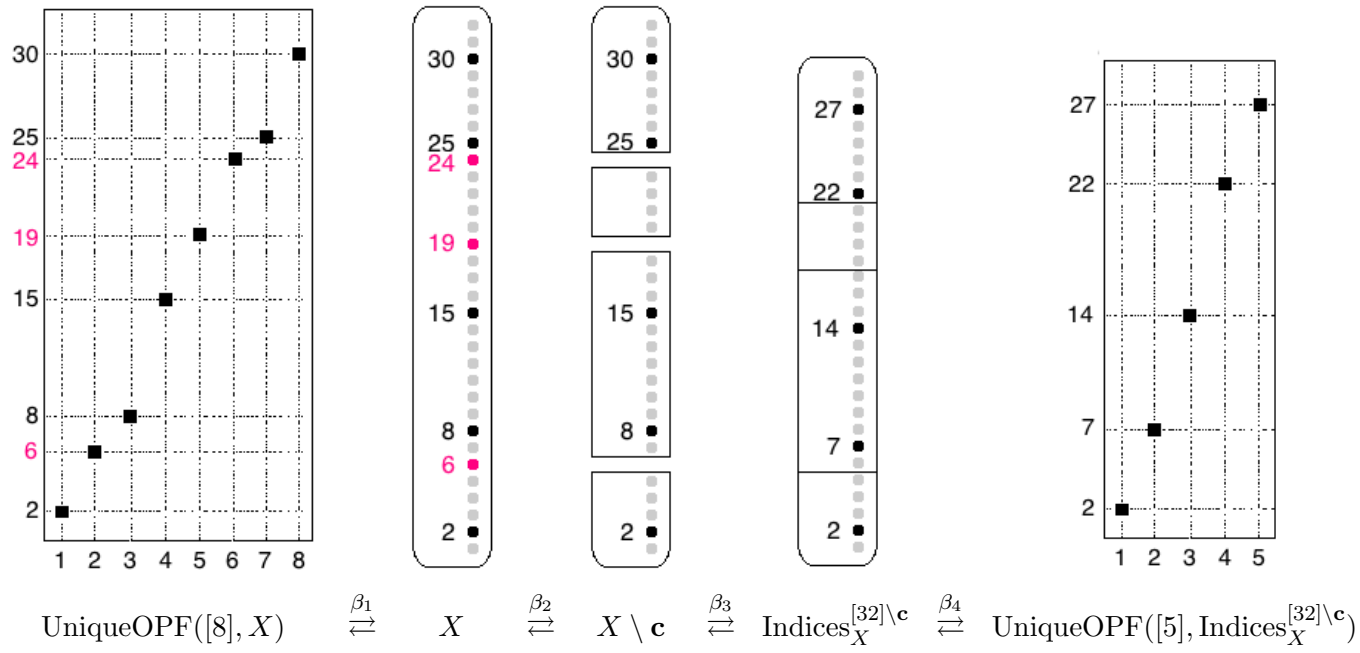

Figure 1: Example of associated elements in the chain of bijections from Lemma B.2. In the example, $N = 32$, $M = 8$, $\mathbf{c} = \{6, 19, 24\}$, and we are looking at the particular OPF $f \in \mathsf{OPF}_{[8],[32]}(\mathbf{c})$ with range $X = \{2, 6, 8, 15, 19, 24, 25, 30\}$.

Before we show the final reduction, we state and prove a small lemma.

Lemma B.3 Let $\mathsf{ROPF}_{[M],[N]} = (\mathcal{K}_{\mathrm{r}}, \mathcal{E}nc_{\mathrm{r}}, \mathcal{D}ec_{\mathrm{r}})$, and $z \geq 1$. Then for any $\mathbf{c} \in \text{Comb}_z^{[N]}$,

$$\Pr\left[K \xleftarrow{\$} \mathcal{K}_{\mathrm{r}}, \mathbf{m} \xleftarrow{\$} \text{Comb}_z^{[M]} \ : \ \mathbf{c} = \{\mathcal{E}nc_{\mathrm{r}}(K, m) \mid m \in \mathbf{m}\} \right] = 1/\binom{N}{z}.$$

Proof: The probability that some $\mathbf{c} \subseteq [N]$ is chosen as the encryptions of the elements of $\mathbf{m}$ is equal to the probability that $\mathcal{E}nc_{\mathrm{r}}(K, \cdot)$ sends *some* z plaintexts $\mathbf{m}' \subseteq [M]$ to $\mathbf{c}$, times the probability that the appropriate $\mathbf{m}$ was picked from $[M]$. The former probability is equal to the likelihood that $\mathbf{c}$ is a

subset of a random M-element subset of N, or $\binom{N-z}{M-z}/\binom{N}{M}$. The latter probability is $1/\binom{M}{z}$. Hence, the desired probability is

$$\frac{\binom{N-z}{M-z}}{\binom{N}{M}}\frac{1}{\binom{M}{z}} = \frac{(N-z)!M!(N-M)!z!(M-z)!}{(M-z)!(N-M)!N!M!} = \frac{(N-z)!z!}{N!} = 1/\binom{N}{z} .$$

∎

Now, here is the second reduction.

Lemma B.4 Fix r, z, M, and N. Let $z' = z - 1$. For any efficient specified r, z-WOW adversary A to scheme $\mathsf{ROPF}_{[M],[N]}$, there exists an efficient $r, 1$-WOW adversary A' to scheme $\mathsf{ROPF}_{[M-z'],[N-z']}$ such that

$$\mathbf{Adv}^{s\text{-}r,z\text{-wow}}_{\mathsf{ROPF}_{[M],[N]}}(A) \le \mathbf{Adv}^{r,1\text{-wow}}_{\mathsf{ROPF}_{[M-z'],[N-z']}}(A') .$$

Proof: Let A be an adversary in experiment $\mathbf{Exp}^{s\text{-}r,z\text{-wow}}_{\mathsf{ROPF}_{[M],[N]}}(A)$. We construct a similarly efficient adversary A' to experiment $\mathbf{Exp}^{r,1\text{-wow}}_{\mathsf{ROPF}_{[M-z'],[N-z']}}(A')$ using A as follows.

Adversary $A'(\{c'\})$
$\mathbf{c}' \xleftarrow{\$} \mathrm{Comb}^{[N]}_{z'}$; $c \leftarrow \mathrm{Element}^{[N]\setminus\mathbf{c}'}_{c'}$; $\mathbf{c} \leftarrow \mathbf{c}' \cup \{c\}$
$(m_L, m_R) \xleftarrow{\$} A(\mathbf{c}, c)$
$z'_- \leftarrow |\{y \in \mathbf{c} \mid y < c\}|$
$m'_L \leftarrow m_L - z'_-$ $\quad m'_R \leftarrow m_R - z'_-$
Return (m'_L, m'_R).

Assume that c' is a random ciphertext in $[N - z']$ (as it is in the experiment $\mathbf{Exp}^{r,1\text{-wow}}_{\mathsf{ROPF}_{[M-z'],[N-z']}}(A')$, by Lemma B.3). Then we must show that the input $(\mathbf{c}, c)$ to A accurately mimics the experiment $\mathbf{Exp}^{s\text{-}r,z\text{-wow}}_{\mathsf{ROPF}_{[M],[N]}}(A)$. That is, it must be that $\mathbf{c}$ looks random from $\mathrm{Comb}^{[N]}_z$ (recalling Lemma B.3 applied to the experiment's challenge sets), and c' looks random from $\mathbf{c}$. Note that c looks uniformly random among $[N] \setminus \mathbf{c}'$ because c' is a random index in $[N - z']$ and c is chosen as the (c')th largest element of $[N] \setminus \mathbf{c}$. Hence, A' accurately simulates the experiment $\mathbf{Exp}^{s\text{-}r,z\text{-wow}}_{\mathsf{ROPF}_{[M],[N]}}(A)$.

Let $\beta_1, \ldots, \beta_4$ be as defined in Lemma B.2. For any OPF f from $[M]$ to $[N]$ with $\mathbf{c}$ in its range, let

$$\beta_f = (\beta_4 \circ \beta_3 \circ \beta_2 \circ \beta_1)(f)$$

be the associated (unique) OPF from $[M - z']$ to $[N - z']$. For fixed $\mathbf{c}$ and c, let $z'_- = |\{c' \in \mathbf{c} \mid c' < c\}|$. Then note that for any $m \in [M]$, if $f(m) = c \notin \mathbf{c}$ then $\beta_f(m - z'_-) = c - z'_-$ and vice versa.

Thus, if A correctly guesses a window m_L, m_R that would succeed in $\mathbf{Exp}^{s\text{-}r,z\text{-wow}}_{\mathsf{ROPF}_{[M],[N]}}(A)$ when f is the OPF picked, then the output m'_L, m'_R of A' succeeds in $\mathbf{Exp}^{r,1\text{-wow}}_{\mathsf{ROPF}_{[M-z'],[N-z']}}(A')$ when β_f is picked; and the converse is also true. Hence, A and A' have the same advantage in their respective experiments.

We also note that A' is efficient if A is efficient, as the extra steps of sampling an element of $\mathrm{Comb}^{[N]}_{z'}$ and re-indexing c, m'_L, and m'_R are all efficient operations. ∎

We are now ready to prove the main lemma of this section.

Proof of Lemma A.1. For any r, z, and any efficient r,z-WOW adversary A, there exist efficient algorithms A'', A' such that

$$\begin{aligned}\mathbf{Adv}^{r,z\text{-wow}}_{\mathsf{ROPF}_{[M],[N]}}(A) &\leq z\mathbf{Adv}^{s\text{-}r,z\text{-wow}}_{\mathsf{ROPF}_{[M],[N]}}(A'') && \text{(Lemma B.1)}\\ &\leq z\mathbf{Adv}^{r,1\text{-wow}}_{\mathsf{ROPF}_{[M-z+1],[N-z+1]}}(A') && \text{(Lemma B.4)}\end{aligned}$$

The result follows. ∎

C Proving Lemma A.2

The proof will use two supporting lemmas. One has already been proved, as Lemma B.3 in the special case $z = 1$ establishes that the uniform choice of plaintext in the experiment ensures a uniformly distributed challenge ciphertext. The second lemma, stated next, allows us to calculate the most likely plaintext for a given ciphertext.

Lemma C.1 For fixed $N, M, c \in \mathbb{N}$, $P_{NHGD}(N, M, c, \cdot)$ achieves its maximum over $[M]$ at some

$$m_0 \in \left[\frac{Mc}{N+1}, \frac{Mc}{N+1} + 1\right] .$$

In particular, if $N = tM$ for some positive integer t, then $P_{NHGD}(N, M, c, \cdot)$ achieves its maximum over $[M]$ at the unique point

$$m_0 = \lceil Mc/N \rceil = \lceil c/t \rceil.$$

Proof: Suppose that $P_{NHGD}(N, M, c, \cdot)$ achieves its maximum over $[M]$ at m_0. Then the function must have a local maximum there; that is,

$$\begin{aligned}P_{NHGD}(N, M, c, m_0 - 1) &\leq P_{NHGD}(N, M, c, m_0) ,\\ P_{NHGD}(N, M, c, m_0) &\geq P_{NHGD}(N, M, c, m_0 + 1) .\end{aligned}$$

Notice that for $m \geq 2$,

$$\begin{aligned}\frac{P_{NHGD}(N, M, c, m)}{P_{NHGD}(N, M, c, m-1)} &= \frac{\binom{c-1}{m-1}\binom{N-c}{M-m}}{\binom{c-1}{(m-1)-1}\binom{N-c}{M-(m-1)}}\\ &= \frac{(M-(m-1))(c-(m-1))}{(m-1)(N-M-c+m)}\\ &= \frac{\frac{Mc}{m-1} - M - c + m - 1}{N-M-c+m}\end{aligned}$$

is at least 1 if and only if $\frac{Mc}{m-1} - 1 \geq N$, or $m \leq \frac{Mc}{N+1} + 1$; it is at most 1 if and only if $\frac{Mc}{m-1} - 1 \leq N$, or $m - 1 \geq \frac{Mc}{N+1}$. The former implies $m_0 \leq \frac{Mc}{N+1} + 1$; the latter implies $m_0 \geq \frac{Mc}{N+1}$.

So, the maximum value of $P_{NHGD}(N, M, c, m)$ occurs at either a unique point, or two adjacent points in $[M]$. Thus, these local maxima are global maxima, and the necessary condition $\frac{Mc}{N+1} \leq m_0 \leq \frac{Mc}{N+1} + 1$ is also sufficient for m_0 to be a global maximum.

To see the second property, note that $N = tM$ implies

$$\left\lceil \frac{Mc}{N+1} \right\rceil = \left\lceil \frac{c}{t}\left(\frac{N}{N+1}\right) \right\rceil$$

$$= \left\lceil \frac{c}{t} - \frac{c}{t(N+1)} \right\rceil$$
$$= \left\lceil \frac{c}{t} \right\rceil ,$$

where in the last step is implied by the following: note that the fractional part of c/t is either 0 or at least $1/t$. In either case, subtracting $c/(t(N+1)) < 1/t$ from c/t will not change the value of its ceiling. Also note that in this case, $Mc/(N+1) = Mc/(tM+1)$ is non-integer and thus m_0 is unique. ▌

Corollary C.2 Fix encryption scheme $(\mathcal{K}_{\mathrm{r}}, \mathcal{E}nc_{\mathrm{r}}, \mathcal{D}ec_{\mathrm{r}}) = \mathsf{ROPF}_{[M],[N]}$, and let $c \in [N]$. Then m_c is a most likely plaintext for c if and only if

$$\frac{Mc}{N+1} \leq m_c \leq \frac{Mc}{N+1} + 1 .$$

In particular, if $N = tM$ for some positive integer t, then m_c is unique for each c and

$$m_c = \lceil Mc/N \rceil = \lceil c/t \rceil .$$

Proof: For any m, c, the probability that $\mathcal{E}nc_{\mathrm{r}}(K, m) = c$ over random $K \in \mathcal{K}_{\mathrm{r}}$ is $P_{NHGD}(N, M, c, m)$. Thus, the result follows directly from Lemma C.1. ▌

We are now ready to prove the lemma.

Proof of Lemma A.2. In the one-wayness experiment, notice that an adversary A is not allowed any oracle access, and in fact the only information A receives is the ciphertext c. Thus, given c, the adversary's best recourse is to output the most likely plaintext for c. By Lemma B.3, the c given to A is uniform from $[N]$, so the OW advantage of A is bounded above by the average probability (over all $c \in [N]$) that c is the image of its most likely plaintext m_c under random $f \in \mathsf{OPF}_{[M],[N]}$, knowing that c is the image of *some* plaintext under f.

Fix $c \in [N]$. Given that $c \in \{f(m) \mid m \in [M]\}$, the probability that $f(m_c) = c$ is equal to the number of OPFs going through (m_c, c), over the number of OPFs that have a point (x, c) for some $x \in [M]$, or

$$\frac{\binom{c-1}{m_c-1}\binom{N-c}{M-c}}{\binom{N-1}{M-1}} = P_*(N, M, c, m_c) .$$

Thus, in the one-wayness experiment, the probability that the (randomly determined) challenge ciphertext is the image of its most likely plaintext is the average of the above quantity for each value of c, and the result follows. ▌

D Proving Lemma A.3

The proof proceeds in several steps. Here is an outline:

- Lemma D.1 relates the middle ciphertext's most likely plaintext's NHGD probability for a given plaintext/ciphertext space to that of a space twice the size, using an algebraic argument.
- Corollary D.2 iterates this result, producing a formula for the middle ciphertext's most likely plaintext's NHGD probability in a large space given the analogous value α_0 in a small space.

- Lemma D.3 and Lemma D.4 together relate any ciphertext's most likely plaintext's NHGD probability to that of the middle ciphertext in the space, using Stirling's approximation and certain bounds on the gamma function.
- Finally, the proof of Lemma A.3 ties these results together, approximating the sum of most likely plaintext 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.

For readability, we introduce the following notation. For a, b, t positive integers such that $a > b$ and $t < a/b$, let

$$(a)_{[b]} = a(a-1)(a-2)\cdots(a-(b-1))\,; \quad (a)_{[b;t]} = a(a-t)(a-2t)\cdots(a-(b-1)t)\,.$$

Approximating most likely NHGD probabilities for the middle ciphertext. Set a domain size M and range size N, larger domain size $M^* = 2M$ and range size $N^* = 2N$, and consider "middle ciphertexts" $c = N/2$ and $c^* = N^*/2 = N$. We show that if M and $N - M$ are large, then the relative most likely NHGD probabilities for c and c^* (knowing that the ciphertexts are hit) in their respective spaces is approximately equal to the constant $1/\sqrt{2}$.

Lemma D.1 Let N, M be multiples of 2 such that $N \geq 2M$, let $M^* = 2M$ and $N^* = 2N$, and let $c = N/2$ and $c^* = N^*/2 = N$. For any $c \in [N]$, let $m_c = \lceil \frac{Mc}{N+1} \rceil$. If M is large, then

$$\frac{P_*(N^*, M^*, c^*, m_{c^*})}{P_*(N, M, c, m_c)} \approx \frac{1}{\sqrt{2}}\,.$$

In particular,

$$\frac{1}{\sqrt{2}} \lessapprox \frac{P_*(N^*, M^*, c^*, m_{c^*})}{P_*(N, M, c, m_c)} < \frac{1}{\sqrt{2}} \cdot e^{1/(2M)}\,.$$

Proof: Set $M' = N - M$. Observe that

$$\begin{aligned}
&\frac{P_*(N^*, M^*, c^*, m_{c^*})}{P_*(N, M, c, m_c)} \\
=\ &\frac{\binom{N-1}{M-1}^2 \binom{N}{M}}{\binom{2N-1}{2M-1}\binom{N/2-1}{M/2-1}\binom{N/2}{M/2}} \\
=\ &\frac{\binom{N}{M}^3}{\binom{2N}{2M}\binom{N/2}{M/2}^2} \\
=\ &\frac{N!^3(2M)!(2M')!(M/2)!^2(M'/2)!^2}{M!^3(M')!^3(2N)!(N/2)!^2} \\
=\ &\frac{N!^2}{(N/2)!^2}\frac{N!}{(2N)!}\frac{(M/2)!^2}{M!^2}\frac{(2M)!}{M!}\frac{(M'/2)!^2}{(M')!^2}\frac{(2M')!}{(M')!} \\
=\ &\frac{((N)_{[N/2]})^2}{(2N)_{[N]}}\frac{(2M)_{[M]}}{((M)_{[M/2]})^2}\frac{(2M')_{[M']}}{((M')_{[M'/2]})^2} \\
=\ &\frac{2^N((N)_{[N/2]})^2}{(2N)_{[N]}}\frac{(2M)_{[M]}}{2^M((M)_{[M/2]})^2}\frac{(2M')_{[M']}}{2^{M'}((M')_{[M'/2]})^2} \\
=\ &\frac{((2N)_{[N/2;2]})^2}{(2N)_{[N]}}\frac{(2M)_{[M]}}{((2M)_{[M/2;2]})^2}\frac{(2M')_{[M']}}{((2M')_{[M'/2;2]})^2} \\
=\ &\frac{(2N)_{[N/2;2]}}{(2N-1)_{[N/2;2]}}\frac{(2M-1)_{[M/2;2]}}{(2M)_{[M/2;2]}}\frac{(2M'-1)_{[M'/2;2]}}{(2M')_{[M'/2;2]}}
\end{aligned}$$

Define the above quantity to be α. Also, let

$$\begin{aligned}\beta &= \frac{(2N-1)_{[N/2;2]}}{(2N-2)_{[N/2;2]}}\frac{(2M-2)_{[M/2;2]}}{(2M-1)_{[M/2;2]}}\frac{(2M'-2)_{[M'/2;2]}}{(2M'-1)_{[M'/2;2]}}\\ \beta' &= \frac{(2N+1)_{[N/2;2]}}{(2N)_{[N/2;2]}}\frac{(2M)_{[M/2;2]}}{(2M+1)_{[M/2;2]}}\frac{(2M')_{[M'/2;2]}}{(2M'+1)_{[M'/2;2]}}\end{aligned}$$

and notice that for large M and N,

$$\beta \lessapprox \alpha \lessapprox \beta'.$$

On the other hand,

$$\begin{aligned}\alpha\beta &= \frac{2N}{N}\frac{M}{2M}\frac{(N-M)}{(2N-2M)}\\ &= 1/2\ ,\end{aligned}$$

and

$$\begin{aligned}\alpha\beta' &= \frac{2N+1}{N+1}\frac{M+1}{2M+1}\frac{(N-M+1)}{2N-2M+1}\\ &< 2\left(\frac{1}{2}+\frac{1/2}{2M+1}\right)\left(\frac{1}{2}+\frac{1/2}{2N-2M+1}\right)\\ &< \frac{1}{2}\left(1+\frac{1}{2M}\right)\left(1+\frac{1}{2(N-M)}\right)\\ &< (1/2)\cdot e^{1/(2M)+1/(2(N-M))}\\ &< (1/2)\cdot e^{1/M}\ .\end{aligned}$$

Hence,

$$\alpha \gtrapprox (\alpha\beta)^{1/2} = \frac{1}{\sqrt{2}}\ ;\ \alpha \lessapprox (\alpha\beta')^{1/2} < \frac{1}{\sqrt{2}}\cdot e^{1/(2M)}\ .$$

∎

Now, we can easily approximate most likely NHGD probabilities for middle ciphertexts in large spaces, in the following manner.

Claim D.2 Let $N_0 \geq 2M_0$ be multiples of 2, let $M = 2^q M_0$ and $N = 2^q N_0$, and let $c = N/2$ and $c_0 = N_0/2$. Define

$$\alpha = P_*(N, M, c, m_c)\ ;\ \alpha_0 = P_*(N_0, M_0, c_0, m_{c_0})\ .$$

Then in particular,

$$\frac{\alpha_0}{2^{q/2}} \lessapprox \alpha < \frac{\alpha_0}{2^{q/2}}\cdot e^{1/M_0}.$$

Proof: The left side of the statement directly follows from repeated application of Lemma D.1. Similarly, by the lemma,

$$\alpha \ < \ \alpha_0 \cdot 2^{-q/2}\prod_{i=1}^{q} e^{1/(2^i M_0)} = \alpha_0 \cdot 2^{-q/2} e^{(1/M_0)\sum_{i=1}^{q} 2^{-i}} < \alpha_0 \cdot 2^{-q/2} e^{1/M_0}\ .$$

Relating general most likely NHGD probabilities to that of the middle ciphertext. In this section we show how to approximate most likely NHGD probabilities for any ciphertext in a large space using the probability corresponding to the middle ciphertext.

Recall the definition of the gamma function: for x a real number,

$$\Gamma(x) = \int_0^\infty r^{x-1} e^{-r}\, dr.$$

The gamma function satisfies the following properties, for x real.

$$\Gamma(x+1) = x\Gamma(x)\ ;\ \Gamma(1) = 1\ .$$

For notational convenience, we will let $\hat{\Gamma}(x) = \Gamma(x+1)$. The above properties imply that $\hat{\Gamma}(x)$ is an extension of the factorial function to real numbers. In particular, for positive integer n,

$$\hat{\Gamma}(n-1) = \Gamma(n) = (n-1)!$$

Also, Stirling's approximation applies to Γ: for real $x > 0$,

$$\hat{\Gamma}(x) = \Gamma(x+1) = \sqrt{2\pi x}(x/e)^x e^{\lambda_x}\ ,$$

where

$$\frac{1}{12x+1} < \lambda_x < \frac{1}{12x}\ .$$

We first prove a short lemma, that will be used in the next proof.

Lemma D.3 Let M, N be multiples of 2 and $N \geq 2M$. Let $k \in (0,1)$, and $k' = 1-k$. Let $M' = N - M$. Then

$$\frac{\Gamma(kN)\,\hat{\Gamma}(k'N)\,\Gamma\left(\frac{M}{2}\right)\hat{\Gamma}\left(\frac{M}{2}\right)\hat{\Gamma}^2\left(\frac{M'}{2}\right)}{\Gamma(kM)\,\hat{\Gamma}(kM')\,\hat{\Gamma}(k'M)\,\hat{\Gamma}(k'M')\,\Gamma\left(\frac{N}{2}\right)\hat{\Gamma}\left(\frac{N}{2}\right)} \leq \frac{1}{2\sqrt{kk'}}.$$

Proof: Using Stirling's approximation,

$$\begin{aligned}
& \frac{\Gamma(kN)\,\hat{\Gamma}(k'N)\,\Gamma\left(\frac{M}{2}\right)\hat{\Gamma}\left(\frac{M}{2}\right)\hat{\Gamma}^2\left(\frac{M'}{2}\right)}{\Gamma(kM)\,\hat{\Gamma}(kM')\,\hat{\Gamma}(k'M)\,\hat{\Gamma}(k'M')\,\Gamma\left(\frac{N}{2}\right)\hat{\Gamma}\left(\frac{N}{2}\right)} \\
= & \frac{kM\frac{N}{2}}{kN\frac{M}{2}} \frac{\hat{\Gamma}(kN)\,\hat{\Gamma}(k'N)\,\hat{\Gamma}\left(\frac{M}{2}\right)\hat{\Gamma}\left(\frac{M}{2}\right)\hat{\Gamma}^2\left(\frac{M'}{2}\right)}{\hat{\Gamma}(kM)\,\hat{\Gamma}(kM')\,\hat{\Gamma}(k'M)\,\hat{\Gamma}(k'M')\,\hat{\Gamma}^2\left(\frac{N}{2}\right)} \\
= & e^{\lambda} \cdot \sqrt{\frac{kNk'N\left(\frac{M}{2}\right)^2\left(\frac{N-M}{2}\right)^2}{kMkM'k'Mk'M'\left(\frac{N}{2}\right)^2}} \cdot \\
& \frac{(kN)^{kN}(k'N)^{k'N}\left(\frac{M}{2}\right)^{2M/2}\left(\frac{M'}{2}\right)^{2M'/2}}{(kM)^{kM}(kM')^{kM'}(k'M)^{k'M}(k'M')^{k'M'}\left(\frac{N}{2}\right)^{2N/2}} \\
= & e^{\lambda} \cdot \frac{1}{2\sqrt{kk'}}\ ,
\end{aligned}$$

where

$$\lambda = \lambda_{kN} + \lambda_{k'N} + 2\lambda_{M/2} + 2\lambda_{M'/2}$$

$$
\begin{aligned}
& -\lambda_{kM} - \lambda_{kM'} - \lambda_{k'M} - \lambda_{k'M'} - 2\lambda_{N/2}) \\
\approx\ & \frac{1}{12}\left(\frac{1}{kN} + \frac{1}{k'N} + \frac{2}{M/2} + \frac{2}{M'/2}\right) \\
& -\frac{1}{12}\left(\frac{1}{kM} + \frac{1}{kM'} + \frac{1}{k'M} + \frac{1}{k'M'} + \frac{2}{N/2}\right) \\
=\ & \frac{1}{12}\left(\frac{1}{M} + \frac{1}{M'} - \frac{1}{N}\right)\left(4 - \frac{1}{k} - \frac{1}{k'}\right) \\
<\ & \frac{1}{6M}\left(4 - \frac{1}{kk'}\right) \\
\leq\ & 0\,,
\end{aligned}
$$

since the maximum value of kk' on $(0,1)$ is $1/4$. ∎

Now, we provide a bound on the ratio between the most likely plaintext probability of a ciphertext c, with $1/M \leq c \leq (M-1)/M$, versus that of the middle ciphertext.

Lemma D.4 Let M be a multiple of 2 and let $N = tM$, where $t \geq 2$ is an integer. Let k be a multiple of $1/N$ such that $1/M \leq k \leq M - 1/M$. Let $k' = 1 - k$, and $M' = N - M$. Then

$$
\frac{P_*(N, M, kN, m_{kN})}{P_*(N, M, N/2, m_{N/2})} \leq 2\sqrt{kk'} \cdot e^{3/2}\,.
$$

Proof: We will use the following bounds of D. Kershaw [16]: for $x > 0$ and $0 < s < 1$,

$$
\left(x + \frac{s}{2}\right)^{1-s} < \frac{\Gamma(x+1)}{\Gamma(x+s)} < \left(x - \frac{1}{2} + \left(s + \frac{1}{4}\right)^{1/2}\right)^{1-s}
$$

Rewriting the bounds, for $y > 1$ and $0 < \delta < 1$, we have

$$
\frac{\Gamma(y)}{\Gamma(y-\delta)} < \left(y - \frac{3}{2} + \left(\frac{5}{4} - \delta\right)^{1/2}\right)^{\delta}\ ;\ \frac{\Gamma(y)}{\Gamma(y+\delta)} < \left(y + \frac{\delta - 1}{2}\right)^{-\delta}\,.
$$

Let $\epsilon = \lceil kM \rceil - kM$. By Lemma C.2, $m_{kN} = \lceil kM \rceil = kM + \epsilon$. Then using Lemma D.3,

$$
\begin{aligned}
& \frac{P_*(N, M, kN, m_{kN})}{P_*(N, M, N/2, m_{N/2})} \\
=\ & \frac{\binom{kN-1}{kM+\epsilon-1}\binom{k'N}{k'M-\epsilon}}{\binom{N/2-1}{M/2-1}\binom{N/2}{M/2}} \\
=\ & \frac{\Gamma(kN)\,\hat{\Gamma}(k'N)\,\Gamma\left(\frac{M}{2}\right)\hat{\Gamma}\left(\frac{M}{2}\right)\hat{\Gamma}^2\left(\frac{M'}{2}\right)}{\Gamma(kM+\epsilon)\,\hat{\Gamma}(kM'-\epsilon)\,\hat{\Gamma}(k'M-\epsilon)\,\hat{\Gamma}(k'M'+\epsilon)\,\Gamma\left(\frac{N}{2}\right)\hat{\Gamma}\left(\frac{N}{2}\right)} \\
\leq\ & \frac{1}{2\sqrt{kk'}}\frac{\Gamma(kM)}{\Gamma(kM+\epsilon)}\frac{\hat{\Gamma}(kM')}{\hat{\Gamma}(kM'-\epsilon)}\frac{\hat{\Gamma}(k'M)}{\hat{\Gamma}(k'M-\epsilon)}\frac{\hat{\Gamma}(k'M')}{\hat{\Gamma}(k'M'+\epsilon)} \\
<\ & \frac{1}{2\sqrt{kk'}}\left(kM + \frac{\epsilon - 1}{2}\right)^{-\epsilon}\left(kM' - \frac{1}{2} + \left(\frac{5}{4} - \epsilon\right)^{1/2}\right)^{\epsilon} \\
& \left(k'M - \frac{1}{2} + \left(\frac{5}{4} - \epsilon\right)^{1/2}\right)^{\epsilon}\left(k'M' + \frac{\epsilon + 1}{2}\right)^{-\epsilon}
\end{aligned}
$$

$$
\begin{aligned}
&< \frac{1}{2\sqrt{kk'}} \left(kM + \frac{\epsilon-1}{2}\right)^{-\epsilon} \left(kM' + \frac{\sqrt{5}-1}{2}\right)^{\epsilon} \\
&\quad \left(k'M + \frac{\sqrt{5}-1}{2}\right)^{\epsilon} \left(k'M' + \frac{\epsilon+1}{2}\right)^{-\epsilon} \\
&= \frac{1}{2\sqrt{kk'}} (kM)^{-\epsilon}(kM')^{\epsilon}(k'M)^{\epsilon}(k'M')^{-\epsilon} \left(1 + \frac{\epsilon-1}{2kM}\right)^{-\epsilon} \\
&\quad \left(1 + \frac{\sqrt{5}-1}{2kM'}\right)^{\epsilon} \left(1 + \frac{\sqrt{5}-1}{2k'M}\right)^{\epsilon} \left(1 + \frac{\epsilon+1}{2k'M'}\right)^{-\epsilon} \\
&< \frac{1}{2\sqrt{kk'}} \exp\left(-\frac{\epsilon(\epsilon-1)}{2kM} + \frac{\epsilon(\sqrt{5}-1)}{2kM'} + \frac{\epsilon(\sqrt{5}-1)}{2k'M}\right) \\
&< \frac{1}{2\sqrt{kk'}} \exp\left(\frac{\epsilon}{2M}\left(\frac{1}{k} + \frac{\sqrt{5}-1}{k(t-1)} + \frac{\sqrt{5}-1}{k'}\right)\right) \\
&< \frac{1}{2\sqrt{kk'}} \exp\left(\frac{\epsilon\sqrt{5}}{2M}\left(\frac{1}{k} + \frac{1}{k'}\right)\right) \\
&< \frac{1}{2\sqrt{kk'}} \exp\left(\frac{\epsilon\sqrt{5}}{2M}\frac{3}{\sqrt{5}}M\right) \\
&< \frac{1}{2\sqrt{kk'}} e^{3/2} \,.
\end{aligned}
$$

∎

The preceding results can now be put together to prove the main lemma statement.

Proof of Lemma A.3. By Lemma D.4 and Corollary D.2,

$$
\begin{aligned}
&\frac{1}{N}\sum_{c=1}^{N} P_*(N, M, c, m_c) \\
&< \frac{2}{M} + \frac{1}{N}\sum_{c=N/M}^{N-N/M} P_*(N, M, c, m_c) \\
&< \frac{2}{M} + \frac{P_*(N, M, N/2, m_{N/2})}{N}\sum_{j=1}^{N} \frac{e^{3/2}}{2\sqrt{(j/N)(1-j/N)}} \\
&\approx \frac{2}{M} + e^{3/2} P_*(N, M, N/2, m_{N/2}) \cdot \int_0^1 \frac{1}{2\sqrt{x(1-x)}}\, dx \\
&= \frac{2}{M} + e^{3/2} P_*(N, M, N/2, m_{N/2}) \frac{\arcsin(2x-1)}{2}\Big|_0^1 \\
&= \frac{2}{M} + P_*(N, M, N/2, m_{N/2}) \cdot e^{3/2} \cdot \frac{\pi}{2} \\
&< \frac{2}{M} + \alpha_0 \cdot \frac{1}{2^{q/2}} \cdot \frac{\pi}{2} \cdot e^{1/M_0 + 3/2} \,.
\end{aligned}
$$

∎

E Comparing tight and simple bounds.

In Table 1, we compare the tight bound of Lemma A.2 and the simple bound of Lemma A.3 for several values of M and N and see that the results are close. We separate the factor of $e^{3/2}$ in for the simple

bound values to show that the bounds would be very close without this factor. In fact, our numerical calculations lead us to informally conjecture that perhaps through some more careful analysis, the factor of $e^{3/2}$ can be removed from the simple bound.

M	N	Tight	Simple
2^8	2^{16}	0.077	$0.087e^{3/2}$
2^9	2^{17}	0.055	$0.060e^{3/2}$
2^{10}	2^{18}	0.039	$0.042e^{3/2}$

Table 1: Sample evaluation of tight vs. simple bounds. For the simple bounds, $M_0 = 2^6$.

F Proving Theorem 4.2

The first half of the result,

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

is obvious, as giving the adversary more challenge ciphertexts can only help it win. It is left to prove the bound on $\mathbf{Adv}^{r,1\text{-wow}}_{\mathsf{ROPF}_{[M],[N]}}(A)$.

We use the following notation for the tail probabilities of the hypergeometric distribution.

$$\begin{aligned} H_+(c,N,M,m_0) &= \sum_{m=m_0}^{M} P_{HGD}(N,M,c,m), \\ H_-(c,N,M,m_0) &= \sum_{m=0}^{m_0} P_{HGD}(N,M,c,m). \end{aligned}$$

The proof of the theorem appears after a lemma.

Lemma F.1 Let M, N, $c \in [N]$, and $r \in [M]$ be given. Let $\delta = \frac{r-1}{2M}$, and let $m_L, m_R \in [M]$ be defined as

$$\begin{aligned} m_L &= \max\{m_c - \lfloor \delta M \rfloor, 1\}, \\ m_R &= \min\{m_c + \lfloor \delta M \rfloor, M\}, \end{aligned}$$

where $m_c = \lceil \frac{Mc}{N+1} \rceil$. Then

$$\sum_{m=1}^{m_L-1} P_*(N,M,c,m) \leq e^{-2\delta^2(M-1)}$$

and

$$\sum_{m=m_R+1}^{M} P_*(N,M,c,m) \leq e^{-2\delta^2(M-1)}.$$

Proof: We will use a bound by Chvátal [13] on the upper tail of the hypergeometric distribution:

$$H_+\left(c,N,M,\left(\frac{c}{N}+d\right)M\right) \leq e^{-2d^2M}.$$

Chvátal's upper tail bound implies a similar lower tail bound:

$$\begin{aligned}
& H_-\left(c, N, M, \left(\frac{c}{N} - d\right) M\right) \\
= & \sum_{i=0}^{(c/N-d)M} P_{HGD}(N, M, c, i) \\
= & \sum_{i=0}^{(c/N-d)M} P_{HGD}(N, M, N-c, M-i) \\
= & \sum_{j=M-(c/N-d)M}^{M} P_{HGD}(N, M, N-c, j) \\
= & H_+\left(N-c, N, M, \left(\frac{N-c}{N} + d\right) M\right) \\
\leq & e^{-2d^2 M}.
\end{aligned}$$

Notice that $m_R \geq \frac{cM}{N-1} + \delta M \geq \left(\frac{c-1}{N-1} + \delta\right)(M-1)$. So

$$\begin{aligned}
& \sum_{m=m_R+1}^{M} P_*(N, M, c, m) \\
= & \sum_{m=m_R+1}^{M} \frac{\binom{c-1}{m-1}\binom{N-c}{M-m}}{\binom{N-1}{M-1}} \\
= & \sum_{m=m_R}^{M-1} \frac{\binom{c-1}{m}\binom{N-c}{M-1-m}}{\binom{N}{M}} \\
= & H_+(c-1, N-1, M-1, m_R) \\
\leq & H_+\left(c-1, N-1, M-1, \left(\frac{c-1}{N-1} + \delta\right)(M-1)\right) \\
\leq & e^{-2\delta^2(M-1)}.
\end{aligned}$$

Similarly, $m_L - 2 \leq \frac{cM}{N-1} - \delta M - 1 \leq \left(\frac{c-1}{N-1} - \delta\right)(M-1)$. So

$$\begin{aligned}
& \sum_{m=1}^{m_L-1} P_*(N, M, c, m) \\
= & \sum_{m=1}^{m_L-1} \frac{\binom{c-1}{m-1}\binom{N-c}{M-m}}{\binom{N-1}{M-1}} \\
= & \sum_{m=0}^{m_L-2} \frac{\binom{c-1}{m}\binom{N-c}{M-1-m}}{\binom{N}{M}} \\
= & H_-(c-1, N-1, M-1, m_L-2) \\
\leq & H_-\left(c-1, N-1, M-1, \left(\frac{c-1}{N-1} - \delta\right)(M-1)\right) \\
\leq & e^{-2\delta^2(M-1)}.
\end{aligned}$$

∎

We now prove the theorem.

Proof of Theorem 4.2. As already mentioned, the first inequality of the theorem is trivially true. It is left to prove the second inequality.

Consider the following $r,1$-WOW adversary A.

Adversary $A(\{c\})$
$m_c \leftarrow \lceil \frac{Mc}{N+1} \rceil$
$\delta \leftarrow \frac{r-1}{2M}$
$m_L \leftarrow \max\{m_c - \lfloor \delta M \rfloor, 1\}$
$m_R \leftarrow \min\{m_c + \lfloor \delta M \rfloor, M\}$
Return (m_L, m_R)

(m_L, m_R) is a legal response in the $r,1$-WOW experiment since the associated window has size $m_R - m_L + 1 \leq 2\delta M + 1 \leq r$. The probability that the adversary succeeds is the probability that $c \in [m_L, m_R]$, or

$$\sum_{m=m_L}^{m_R} P_*(N, M, c, m) \geq 1 - 2e^{-\frac{(r-1)^2}{2}\frac{(M-1)}{M^2}},$$

where the inequality follows from Lemma F.1. Since A only performs efficient operations, the result follows. ∎

G Proving Theorem 4.3

The proof of the theorem parallels that of Theorem 4.1. As such, it requires several intermediate results that are now stated.

Lemma G.1 For window size r, challenge set size z, and any adversary A, there exists a OW-adversary A' such that

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

The proof is in Appendix H.

Lemma G.2 For any adversary A,

$$\mathbf{Adv}^{1,2\text{-wdow}}_{\mathsf{ROPF}_{[M],[N]}}(A) \leq \frac{1}{N-1}\sum_{w=1}^{N-1} P_*(N-1, M-1, w, d_w) ,$$

where $d_w = \lceil \frac{(M-1)w}{N} \rceil$.

The proof is in Appendix I.

Notice that the bound given in Lemma G.2 is precisely the bound in Lemma A.2, only with parameters $M-1$, $N-1$ instead of M, N. Thus, we will be able to use the simple bound from Corollary A.4.

The proof of the theorem now easily follows.

Proof of Theorem 4.3. Let $M' = M - z + 2$, $N' = N - z + 2$.

$$\begin{aligned}
\mathbf{Adv}^{1,z\text{-wow}}_{\mathsf{ROPF}_{[M],[N]}}(A) &\leq z(z-1)\mathbf{Adv}^{1,2\text{-wow}}_{\mathsf{ROPF}_{[M'],[N']}}(A) && \text{(Lemma G.1)}\\
&\leq z(z-1)\tfrac{1}{N'-1}\textstyle\sum_{c=1}^{N'-1} P_*(N'-1, M'-1, w, d_w) && \text{(Lemma G.2)}\\
&< z(z-1)\tfrac{9}{\sqrt{M'-1}} && \text{(Corollary A.4)}\\
&= z(z-1)\tfrac{9}{\sqrt{M-z+1}}.
\end{aligned}$$

In the third step, note that $N \geq 2M$ and $M \geq 15 + z$ imply $N' - 1 \geq 2(M'-1) \geq 32$. ∎

H Proving Lemma G.1

Define *specified r, z-window-distance-one-wayness advantage* of adversary A with respect to scheme $\mathcal{SE}_{\mathcal{D},\mathcal{R}} = (\mathcal{K}, \mathcal{E}nc, \mathcal{D}ec)$ as

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

where the security experiment is as follows.

Experiment $\mathbf{Exp}^{s\text{-}r,z\text{-wow}}_{\mathcal{SE}_{\mathcal{D},\mathcal{R}}}(A)$
$K \xleftarrow{\$} \mathcal{K}$; $\mathbf{m} \xleftarrow{\$} \mathrm{Comb}_z^{[M]}$
$m_0 \xleftarrow{\$} \mathbf{m}$; $m_1 \xleftarrow{\$} \mathbf{m} \setminus \{m_0\}$
$\mathbf{c} \leftarrow \mathcal{E}nc(K, \mathbf{m})$; $(c_0, c_1) \leftarrow \mathcal{E}nc(K, (m_0, m_1))$
$(d_L, d_R) \xleftarrow{\$} A(\mathbf{c}, c_0, c_1)$
Return 1 if $d_2 - d_1 + 1 \le r$ and $m_1 - m_0 \bmod M \in [d_1, d_2]$;
Return 0 otherwise.

Lemma H.1 For any scheme $\mathcal{SE}_{\mathcal{D},\mathcal{R}}$ and r, z, and any r, z-WDOW adversary A, there exists an equally efficient specified r, z-WDOW adversary A' such that

$$\mathbf{Adv}^{r,z\text{-wdow}}_{\mathcal{SE}_{\mathcal{D},\mathcal{R}}}(A) \le z(z-1)\mathbf{Adv}^{s\text{-}r,z\text{-wdow}}_{\mathcal{SE}_{\mathcal{D},\mathcal{R}}}(A').$$

Proof: Given A, let A' on input $(\mathbf{c}, c_0, c_1)$ simply run $(d_L, d_R) \xleftarrow{\$} A(\mathbf{c})$ and return (d_L, d_R). Whenever A outputs legal (d_L, d_R) such that $\exists m_0', m_1' \in [M]$ with $m_1' - m_0' \bmod M \in [d_1, d_2]$, then A' wins if $m_0 = m_0'$ and $m_1 = m_1'$. Since m_0 is random in $[M]$ and m_1 is random in $[M] \setminus \{m_0\}$, independent of the rest of the experiment, we conclude that A' wins the specified experiment at least $\frac{1}{z(z-1)}$ of the times that A wins the standard experiment. The lemma follows. ∎

Lemma H.2 Fix r, z, M, and N. Let $z' = z - 2$. For any efficient specified r, z-WDOW adversary A to scheme $\mathsf{ROPF}_{[M],[N]}$, there exists an efficient $r, 2$-WDOW adversary A' to scheme $\mathsf{ROPF}_{[M-z'],[N-z']}$ such that

$$\mathbf{Adv}^{s\text{-}r,z\text{-wdow}}_{\mathsf{ROPF}_{[M],[N]}}(A) \le \mathbf{Adv}^{r,2\text{-wdow}}_{\mathsf{ROPF}_{[M-z'],[N-z']}}(A').$$

Proof: Let A be an adversary to experiment $\mathbf{Exp}^{s\text{-}r,z\text{-wdow}}_{\mathsf{ROPF}_{[M],[N]}}(A)$. We construct an adversary A' to experiment $\mathbf{Exp}^{r,2\text{-wdow}}_{\mathsf{ROPF}_{[M-z'],[N-z']}}(A')$ using A as follows.

Adversary $A'(\{c_0', c_1'\})$
$\mathbf{c}' \xleftarrow{\$} \mathrm{Comb}_{z'}^{[N]}$
$c_i \leftarrow \mathrm{Element}_{c_i'}^{[N]\setminus\mathbf{c}'}$ for $i = 0, 1$
$\mathbf{c} \leftarrow \mathbf{c}' \cup \{c_0, c_1\}$
$(d_L, d_R) \xleftarrow{\$} A(\mathbf{c}, c_0, c_1)$
$z'_{\mathrm{bt}} \leftarrow |\{y \in \mathbf{c} \mid c_0 < y < c_1\}|$
$d_L' \leftarrow d_L - z'_{\mathrm{bt}}$ $\quad d_R' \leftarrow d_R - z'_{\mathrm{bt}}$
Return (d_L', d_R').

Assume that c'_0, c'_1 are random (distinct) ciphertexts in $[N-z']$ (as it is in the experiment $\mathbf{Exp}^{r,2\text{-wdow}}_{\mathsf{ROPF}_{[M-z'],[N-z']}}(A')$, by Lemma B.3). Then we must show that the input $(\mathbf{c}, c_0, c_1)$ to A accurately mimics the experiment $\mathbf{Exp}^{s\text{-}r,z\text{-wdow}}_{\mathsf{ROPF}_{[M],[N]}}(A)$. That is, it must be that $\mathbf{c}$ looks random from $\mathrm{Comb}^{[N]}_z$ (recalling Lemma B.3 applied to the experiment's challenge sets), and $\{c_0, c_1\}$ looks random from $\mathrm{Comb}^{\mathbf{c}}_2$. Note that c_0, c_1 are uniformly random distinct elements of $[N] \setminus \mathbf{c}$ because c'_0, c'_1 are random distinct indices in $[N - z']$ and c_i is chosen as the (c'_i)th largest element of $[N] \setminus \mathbf{c}$ for $i = 0, 1$. Hence, $\mathbf{c}$ looks random from $\mathrm{Comb}^{[N]}_z$ and $\{c_0, c_1\}$ looks random from $\mathrm{Comb}^{\mathbf{c}}_2$. Thus, A' accurately simulates the experiment $\mathbf{Exp}^{s\text{-}r,z\text{-wdow}}_{\mathsf{ROPF}_{[M],[N]}}(A)$.

Let $\beta_1, \ldots, \beta_4$ be as defined in Lemma B.2. For any OPF f from $[M]$ to $[N]$ with $\mathbf{c}$ in its range, let

$$\beta_f = (\beta_4 \circ \beta_3 \circ \beta_2 \circ \beta_1)(f)$$

be the associated (unique) OPF from $[M - z']$ to $[N - z']$. Let $z'_{\mathrm{bt}} = |\{c' \in \mathbf{c} \mid c_0 < c' < c_1\}|$. Then note that for any $m_0, m_1 \in [M]$, if $f(m_0) = c_0 \notin \mathbf{c}$ and $f(m_1) = c_1 \notin \mathbf{c}$ then $m_1 - m_0 = \beta_f(m_1) - \beta_f(m_0) + z'_{\mathrm{bt}}$ and vice versa.

Thus, if A correctly guesses a window d_L, d_R that would succeed in $\mathbf{Exp}^{s\text{-}r,z\text{-wdow}}_{\mathsf{ROPF}_{[M],[N]}}(A)$ when f is the OPF picked, then the output d'_L, d'_R of A' succeeds in $\mathbf{Exp}^{r,2\text{-wdow}}_{\mathsf{ROPF}_{[M-z'],[N-z']}}(A')$ when β_f is picked; and the converse is also true. Hence, A and A' have the same advantage in their respective experiments.

We also note that A' is efficient if A is efficient, as the extra steps of sampling an element of $\mathrm{Comb}^{[N]}_{z'}$ and re-indexing c_0, c_1, d'_L, and d'_R are all efficient operations. ∎

We are now ready to prove the main lemma of this section.

Proof of Lemma G.1. For any r, z, and any efficient r, z-WDOW adversary A, there exist efficient algorithms A'', A' such that

$$\begin{aligned}
\mathbf{Adv}^{r,z\text{-wdow}}_{\mathsf{ROPF}_{[M],[N]}}(A) &\leq z\mathbf{Adv}^{s\text{-}r,z\text{-wow}}_{\mathsf{ROPF}_{[M],[N]}}(A'') && \text{(Lemma H.1)} \\
&\leq z\mathbf{Adv}^{r,1\text{-wow}}_{\mathsf{ROPF}_{[M-z+1],[N-z+1]}}(A') && \text{(Lemma H.2)}
\end{aligned}$$

The result follows. ∎

I Proving Lemma G.2

A formula for the most likely plaintext distance. In Corollary I.2 below, we derive a formula for the most likely plaintext distance between two given ciphertexts. But first, the following lemma determines the probability that a given ciphertext pair corresponds to a given plaintext distance, which is used to prove the corollary.

Lemma I.1 For any $c_1, c_2 \in [N]$, let

$$\mathsf{OPF}_{[M],[N]}{}^* = \{f \in \mathsf{OPF}_{[M],[N]} : c_1, c_2 \in f([M])\}.$$

Then for any $d \in [M-1]$,

$$\Pr\left[\, f^{-1}(c_2) - f^{-1}(c_1) \bmod M = d \,\right] = P_*(N-1, M-1, w, d)\ ,$$

where $w = c_2 - c_1 \bmod N$ and the probability is over the random choice of f from $\mathsf{OPF}_{[M],[N]}{}^*$.

Proof: Let $c_1, c_2 \in [N]$. If $c_1 = c_2$, the result easily follows, so suppose $c_1 \neq c_2$. c_1 and c_2 partition the rest of the ciphertext space into two sets S and S':

$$\begin{aligned} S &= \begin{cases} c_1+1, c_1+2, \ldots, c_2-1 & c_1 \leq c_2 \\ c_1+1, c_1+2, \ldots, N, 1, 2, \ldots, c_2-1 & c_1 > c_2 \end{cases} \\ S' &= [M] \setminus (S \cup \{c_1, c_2\}) \ . \end{aligned}$$

Let $w = c_2 - c_1 \bmod N$. Then $1 \leq w \leq N-1$, and note that $|S| = w-1$ and $|S'| = N-w-1$.

The probability, over random $f \in \mathsf{OPF}_{[M],[N]*}$, that $f^{-1}(c_2) - f^{-1}(c_1) \bmod M = d$ is equal to the number of OPFs g on $[M], [N]$ such that

$$\begin{aligned} c_1, c_2 &\in g([M]) \\ |g([M]) \cap S| &= d-1 \\ |g([M]) \cap S'| &= M-d-1, \end{aligned}$$

over the number of OPFs g such that $c_1, c_2 \in g([M])$, or

$$\frac{\binom{w-1}{d-1}\binom{N-w-1}{M-d-1}}{\binom{N-2}{M-2}} = P_*(N-1, M-1, w, d).$$

∎

In particular, for fixed c_1, c_2, d and $w = c_2 - c_1 \bmod N$ the lemma says that

$$\Pr[\, \mathcal{D}ec_{\mathrm{r}}(K, c_2) - \mathcal{D}ec_{\mathrm{r}}(K, c_1) \bmod M = d \,] = P_*(N-1, M-1, w, d).$$

where the probability is over K a random key output by $\mathcal{K}_{\mathrm{r}}$ such that $c_1, c_2 \in \mathcal{E}nc_{\mathrm{r}}(K, [M])$.

Now, we can locate the most likely plaintext distance for c_1, c_2.

Corollary I.2 Let $c_1, c_2 \in [N]$ with $c_1 < c_2$, and $w = c_2 - c_1 \bmod N$. Then in $\mathsf{ROPF}_{[M],[N]}$, d_{c_1,c_2} is a most likely plaintext distance from c_1 to c_2 if and only if

$$\frac{(M-1)w}{N} \leq d_{c_1,c_2} \leq \frac{(M-1)w}{N} + 1.$$

Proof: By Lemma C.1, for N, M, w fixed, $P_{NHGD}(N-1, M-1, w, \cdot)$ has a maximum at $d_0 \in [M-1]$ where

$$\frac{(M-1)w}{N} \leq d_0 \leq \frac{(M-1)w}{N} + 1.$$

Therefore, $P_*(N-1, M-1, w, \cdot)$ also has a maximum at d_0, so the result follows from Lemma I.1. ∎

Note in particular that d_{c_1,c_2} depends only on the difference $w = c_2 - c_1 \bmod N$. Thus, for $w \in [N-1]$, we define d_w to be the *most likely plaintext distance for* w and $d_w = d_{c_1,c_2}$ for all $c_1, c_2 \in [N]$ with $w = c_2 - c_1 \bmod N$.

The plaintext distance is uniformly random. Here we establish that no plaintext distance (from 1 to $M-1$) is more or less likely than any other, if the challenge plaintexts are uniformly random and distinct.

Lemma I.3 For any $w \in [N-1]$,

$$\Pr[\, \mathcal{E}nc_{\mathrm{r}}(K, m_2) - \mathcal{E}nc_{\mathrm{r}}(K, m_1) \bmod N = w \,] = \frac{1}{N-1} \ ,$$

where the probability is over the following random choices: $K \xleftarrow{\$} \mathcal{K}_{\mathrm{r}}$, $m_1 \xleftarrow{\$} [M]$, and $m_2 \xleftarrow{\$} [M] \setminus \{m_1\}$.

Proof: In the following, we consider addition and subtraction of ciphertexts to be taken mod N.

$$\begin{aligned}
&\Pr_{K,m_1,m_2}[\,\mathcal{E}nc_{\mathrm{r}}(K,m_2)-\mathcal{E}nc_{\mathrm{r}}(K,m_1)=w\,]\\
=\;&\sum_{c\in[N]}\Pr_{\substack{K,m_1,\\ m_2}}[\,\mathcal{E}nc_{\mathrm{r}}(K,m_1)=c,\mathcal{E}nc_{\mathrm{r}}(K,m_2)=c+w\,]\\
=\;&\sum_{c\in[N]}\Pr_{K}[\,c,c+w\in\mathcal{E}nc_{\mathrm{r}}(K,[M])\,]\cdot\\
&\Pr_{m_1,m_2}\Big[\,c,c+w\in\mathcal{E}nc_{\mathrm{r}}(K,[M])\;:\;\mathcal{E}nc_{\mathrm{r}}(K,m_1)=c,\mathcal{E}nc_{\mathrm{r}}(K,m_2)=c+w\,\Big]\\
=\;&\sum_{c\in[N]}\frac{\binom{N-2}{M-2}}{\binom{N}{M}}\frac{1}{M}\frac{1}{M-1}\\
=\;&\frac{1}{N-1}.
\end{aligned}$$

∎

We are now ready to prove the lemma.

Proof of Lemma G.2. In the DOW experiment, since the adversary A is given only the challenge ciphertexts c_1, c_2, the adversary will have highest probability to win the game if it outputs the most likely plaintext distance for c_1, c_2. By Lemma I.3, $w = c_2 - c_1 \bmod N$ is uniform from $[N-1]$, so the DOW advantage of A is bounded above by the average probability (over all $w \in [N-1]$) that $d_w = \mathcal{D}ec_{\mathrm{r}}(K,c_2) - \mathcal{D}ec_{\mathrm{r}}(K,c_2)$, where K is a random key output by $\mathcal{K}_{\mathrm{r}}$ such that $c_1, c_2 \in \mathcal{E}nc_{\mathrm{r}}(K,[M])$. Thus, the result follows from Lemma I.1. ∎

J Proof of Proposition 4.5

Let $t = (N-1)/(M-1)$. Let b be a fixed value (less than $\sqrt{M-1}$) to be determined later. Define $\beta = \frac{2tb\sqrt{M-1}}{t-2}$.

By Lemma I.3, w is uniformly random in $[N-1]$, so we see that over the choice of m_1, m_2, K,

$$\Pr[\,w < \beta+1\,] \le \frac{\beta+1}{N-1}.$$

Recall that $d_w = \lceil \frac{(M-1)w}{N} \rceil$ is the most likely plaintext distance of w, by Corollary I.2. Let $\delta = \frac{b}{\sqrt{M-1}}$, and define

$$d_R = \min\{d_w + \lfloor \delta(M-1) \rfloor, M-1\}.$$

Then note that whenever $w \ge \beta+1$,

$$\begin{aligned}
d_R &\le d_w + \lfloor \delta(M-1) \rfloor\\
&\le \frac{w}{t} + 1 + \delta(M-1)\\
&= \frac{w}{t} + 1 + b\sqrt{M-1}\\
&\le \frac{2b\sqrt{M-1}}{t-2} + b\sqrt{M-1} + 1 \quad (\text{Since } w > \beta)\\
&= \frac{b\sqrt{M-1}(2+t-2)}{t-2} + 1\\
&= \frac{tb\sqrt{M-1}}{t-2} + 1\\
&= \beta/2 + 1
\end{aligned}$$

$$< \quad w/2 \quad \text{(Since } w \geq \beta + 1.\text{)}$$

Hence,

$$\begin{aligned}
\Pr[\, 2d > w \,:\, w \geq \beta + 1 \,] &\leq \Pr[\, d > d_R \,:\, w \geq \beta + 1 \,] \\
&= \sum_{d=d_R+1}^{M-1} P_*(N-1, M-1, w, d) \\
&\leq e^{-2\delta^2(M-2)} \quad \text{(by Lemma F.1)} \\
&= e^{-2b^2 \frac{M-2}{M-1}} \\
&< e^{-b^2} \,.
\end{aligned}$$

Now we have that, over the choice of m_1, m_2, K,

$$\begin{aligned}
\Pr[\, 2d > w \,] &\leq \Pr[\, 2d > w \,:\, w \geq \beta + 1 \,] + \Pr[\, w < \beta + 1 \,] \\
&\leq e^{-b^2} + \frac{\beta + 1}{N - 1} \\
&= e^{-b^2} + \frac{2tb\sqrt{M-1}}{(t-2)(N-1)} + \frac{1}{N-1} \\
&= e^{-b^2} + \frac{2b}{(t-2)\sqrt{M-1}} + \frac{1}{N-1}.
\end{aligned}$$

We may now select a value for b, say $b = \sqrt{\ln M}$. Then this bound becomes

$$\begin{aligned}
\Pr[\, 2d > w \,] &\leq 1/M + \frac{2\sqrt{\log M}}{(t-2)\sqrt{M-1}} + 1/(N-1) \\
&< 2/M + \frac{2}{t-2} \frac{1}{\sqrt{(M-1)/\ln M}} \\
&< \frac{3}{t} \frac{1}{\sqrt{(M-1)/\ln M}} \,,
\end{aligned}$$

assuming $t \geq 7$.