Order-Revealing Encryption: New Constructions, Applications, and Lower Bounds

(Extended Version)

Kevin Lewi
Stanford University
klewi@cs.stanford.edu

David J. Wu
Stanford University
dwu4@cs.stanford.edu

Abstract

In the last few years, there has been significant interest in developing methods to search over encrypted data. In the case of range queries, a simple solution is to encrypt the contents of the database using an order-preserving encryption (OPE) scheme (i.e., an encryption scheme that supports comparisons over encrypted values). However, Naveed et al. (CCS 2015) recently showed that OPE-encrypted databases are extremely vulnerable to "inference attacks."

In this work, we consider a related primitive called order-revealing encryption (ORE), which is a generalization of OPE that allows for stronger security. We begin by constructing a new ORE scheme for small message spaces which achieves the "best-possible" notion of security for ORE. Next, we introduce a "domain-extension" technique and apply it to our small-message-space ORE. While our domain-extension technique does incur a loss in security, the resulting ORE scheme we obtain is more secure than all existing (stateless and non-interactive) OPE and ORE schemes which are practical. All of our constructions rely only on symmetric primitives. As part of our analysis, we also give a tight lower bound for OPE and show that no efficient OPE scheme can satisfy best-possible security if the message space contains just three messages. Thus, achieving strong notions of security for even small message spaces requires moving beyond OPE.

Finally, we examine the properties of our new ORE scheme and show how to use it to construct an efficient range query protocol that is robust against the inference attacks of Naveed et al. We also give a full implementation of our new ORE scheme, and show that not only is our scheme more secure than existing OPE schemes, it is also faster: encrypting a 32-bit integer requires just 55 microseconds, which is more than 65 times faster than existing OPE schemes.

1 Introduction

Today, large corporations and governments collect and store more personal information about us than ever before. And as high-profile data breaches on companies and organizations (such as Anthem [AC15], eBay [Kel14], and the U.S. Voter Database [FV15]) become startlingly common, it is imperative that we develop practical means for securing our personal data in the cloud.

One way to mitigate the damage caused by a database breach is to encrypt the data before storing it in the cloud. This, however, comes at the price of functionality: once data is encrypted, it is more difficult to execute searches over the data without first decrypting the data. As a result,

This is the extended version of a paper by the same name that appeared in *ACM Conference on Computer and Communications Security* in October, 2016.

security researchers have turned to developing methods that both protect the contents of the database, as well as support efficient operations, such as search, over the encrypted data.

Property-preserving encryption. One way to support searching over an encrypted database is through property-preserving encryption (PPE) [BCLO09, PR12, CD15]. A PPE scheme is an encryption scheme where the ciphertexts reveal a particular property on their underlying plaintexts. Examples include deterministic encryption, where the ciphertexts reveal equality between messages, and order-preserving encryption (OPE) [AKSX04, BCLO09], where the ciphertexts reveal the ordering of messages. Deterministic and order-preserving encryption schemes have been used in CryptDB [PRZB11], and also commercially by SkyHigh Networks, CipherCloud, Google Encrypted BigQuery, and others. One of the main appeals of PPE for encrypting relational databases is that they are lightweight, and hence, can be deployed with minimal changes to existing databases. For instance, in an OPE scheme, the ciphertexts themselves are numeric and the order of the ciphertexts precisely coincides with the order of the plaintexts. Thus, searching over a column encrypted using OPE is *identical* to searching over an unencrypted column.

Limitations of PPE and OPE. While PPE, and in particular, OPE, provides a practical solution for searching on encrypted data, these schemes also leak significant amounts of information about their underlying plaintexts. For instance, Boldyreva et al. [BCO11] showed that a *single* OPE ciphertext leaks half of the most significant bits of its underlying plaintext!

More recently, Naveed et al. [NKW15] described a series of inference attacks on relational databases encrypted using deterministic and order-preserving encryption schemes. They show that, given just a data dump of an encrypted database along with auxiliary information from a public database, an attacker can successfully recover nearly all of the underlying plaintext values from their respective ciphertexts.

Our goals. Motivated by the limited security of existing OPE schemes and the emerging threat of inference attacks on databases encrypted using PPE, our goal in this work is to construct a practical property-preserving encryption for comparisons that achieves stronger security guarantees compared to existing OPE schemes while at the same time providing robustness against offline inference attacks, such as those considered by Naveed et al.

Order-revealing encryption. To address the limitations of OPE, we rely on a closely-related, but more flexible, notion called order-revealing encryption (ORE) [BLR$^+$15, CLWW16] (also called efficiently-orderable encryption (EOE) [BCO11, §5]). In this work, we focus exclusively on non-interactive and stateless schemes—these are the only schemes we know of that are deployed on a large scale. We survey the work on alternative solutions in Section 8.

In an OPE scheme, both the plaintext and ciphertext spaces must be numeric and well-ordered. Moreover, the ciphertexts themselves preserve the order of the underlying plaintexts. While this property makes OPE suitable for performing range queries on encrypted data, it also limits the achievable security of OPE schemes. In their original work, Boldyreva et al. [BCLO09] introduced the notion of "best-possible" semantic security for OPE, which states that the ciphertexts do not leak any information beyond the ordering of the plaintexts. Unfortunately, in the same work and a follow-up work [BCO11], they show that any OPE scheme with best-possible security must have ciphertexts whose length grows exponentially in the length of the plaintexts. Popa et al. [PLZ13] further extended this lower bound to apply to stateful, interactive OPE schemes. These lower bounds rule out any hope of constructing efficient OPE schemes for large message spaces. As a

compromise, Boldyreva et al. [BCLO09] introduced a weaker notion of security (POPF-CCA) for OPE schemes, but it is difficult to quantify the leakage of schemes which are POPF-CCA secure.

Recently, Boneh et al. [BLR+15] studied the more general notion of ORE, which does *not* place any restrictions on the structure of the ciphertext space. An ORE scheme simply requires that there exists a *publicly* computable function that compares two ciphertexts. By relaxing the constraint on the ciphertext space, the Boneh et al. scheme is the first (non-interactive and stateless) scheme to achieve best-possible semantic security. However, their construction relies on multilinear maps [BS03, GGH13a, CLT13], and is extremely far from being practically viable. More recently, Chenette et al. [CLWW16] introduced a new security model for ORE that explicitly models the information leakage of an ORE scheme. They also give the first efficiently-implementable ORE scheme. However their scheme also reveals the index of the first bit that differs between two encrypted values.

1.1 Extending ORE: The Left/Right Framework

Before describing our main contributions, we first highlight the "left/right" framework for order-revealing encryption that we use in this work. Our notions are adapted from similar definitions for multi-input functional encryption [GGG+14, BLR+15], where the encryption function operates on different "input slots." In a multi-input functional encryption scheme (of which ORE is a special case), information about plaintexts is only revealed when one has a ciphertext for *every* slot.

We now describe how this notion of encrypting to different input slots applies to order-revealing encryption. In a vanilla ORE scheme, there is a single encryption algorithm that takes a message and outputs a ciphertext. The comparison algorithm then takes two ciphertexts and outputs the comparison relation on the two underlying messages. In the left/right framework, we modify this interface and decompose the encryption function into two separate functions: a "left" encryption function and a "right" encryption function. Each of these encryption functions takes a message and the secret key, and outputs either a "left" or a "right" ciphertext, respectively. Next, instead of taking two ciphertexts, the comparison function takes a left ciphertext and a right ciphertext, and outputs the comparison relation between the two underlying messages (encrypted by the left and right ciphertexts). We note that any ORE scheme in the left/right framework can be converted to an ORE scheme in the usual sense by simply having the ORE encryption function output both the left and right ciphertexts for a given message.

This left/right notion is a strict generalization of the usual notion of order-revealing encryption, and thus, can be used to strengthen the security guarantees provided by an ORE scheme. In particular, a key advantage of working in this framework is that we can now define additional security requirements on collections of left or right ciphertexts taken in isolation. For example, in both of the ORE constructions we introduce in this work (Sections 3 and 4), a collection of right ciphertexts taken individually is semantically secure—that is, no information about the underlying plaintexts (including their order relations) is revealed given only a collection of right ciphertexts. In Section 5, we describe precisely how semantic security of the right ciphertexts can be leveraged to obtain a range query protocol that is robust against offline inference attacks. We also note that the schemes presented in this work are the first practical ORE constructions in the left/right framework where one side (the right ciphertexts) achieves semantic security.[1]

[1]Concurrent with the publication of this work, Joye and Passelégue [JP16] along with Cash, Liu, O'Neill, and Zhang [CLOZ16] independently gave constructions of ORE based on bilinear groups where the ciphertexts are decomposable into left and right components where one side has semantic security.

Finally, we note that the left/right framework extends naturally to property-preserving encryption schemes, and thus, opens up many new avenues of developing more secure cryptographic primitives for searching on encrypted data.

1.2 Our Contributions

In this work, we describe a new ORE scheme that achieves stronger security compared to existing practical OPE and ORE schemes, as well as a method to leverage our new ORE scheme to efficiently perform range queries while providing robustness against inference attacks. We now highlight our main contributions.

An efficient small-domain ORE. We begin by giving the first construction of a practical, *small-domain* ORE scheme with *best-possible* semantic security that only relies on pseudorandom functions (PRFs).[2] The restriction to "small" domains is due to the fact that the ciphertext length in our scheme grows linearly in the size of the plaintext space. All existing constructions of ORE that achieve best-possible security in the small-domain setting rely on pairings [KLM$^+$16], general-purpose functional encryption [AJ15, BKS15], or multilinear maps [BLR$^+$15], and thus, are not yet practical. Our particular construction is inspired by the "brute-force" construction of functional encryption by Boneh et al. [BSW11, §4.1]. They show that functional encryption with respect to a "small" (i.e., polynomially-sized) class of functions can be constructed using only symmetric primitives. We adapt these methods to show how best-possible ORE (and more generally, functional encryption) can be efficiently constructed from symmetric primitives when the *message space* is small. Our construction is described in Section 3.

Domain extension for ORE. Of course, a small-domain ORE by itself is not very useful for range queries. Our second contribution is a recasting of the Chenette et al. [CLWW16] ORE construction as a general technique of constructing a large-domain ORE from a small-domain ORE. The transformation is not perfect and incurs some leakage. Applying this domain-extension technique to our new small-domain ORE, we obtain an ORE scheme whose leakage profile is significantly better than that of the Chenette et al. construction. In particular, our new ORE scheme operates on blocks (where a block is a sequence of bits) and the additional leakage in our scheme is the position of the first block in which two messages differ. For instance, if blocks are byte-sized (8 bits), then our ORE scheme only reveals the index of the first byte that differs between the two messages (and nothing more). In contrast, the Chenette et al. construction always reveals the index of the first *bit* that differs.[3] Thus, our new ORE construction provides significantly stronger security, at the cost of somewhat longer ciphertexts.

Encrypted range queries. While our new ORE scheme can almost[4] be used as a drop-in replacement for OPE to enable searching over an encrypted database, the scheme remains susceptible to an offline inference attack. To carry out their inference attacks, Naveed et al. [NKW15] rely on the fact that OPE-encrypted ciphertexts enable equality tests and comparisons (by design). In our

[2] We prove security in the random oracle model, but it is possible to replace the random oracle with a PRF to show security under a slightly weaker indistinguishability-based notion of security.

[3] While Chenette et al. also describe a multi-bit generalization of their scheme, the generalized version leaks *more* information, namely the difference of the values in the first differing block. In our construction, only the index and nothing else is revealed.

[4] We say "almost," since using ORE in place of OPE would require writing a custom comparator for database elements.

setting, we take advantage of the special structure of the ciphertexts in our ORE scheme to obtain a way of supporting range queries on encrypted data while protecting against offline inference attacks.

Our range query protocol critically relies on the fact that our ORE scheme is a left/right ORE scheme (Section 1.1). More precisely, a ciphertext ct in our ORE scheme naturally decomposes into a left component ct_L and a right component ct_R. To compare two ciphertexts, the comparison function only requires the left component of one ciphertext and the right component of the other. More importantly, the right components have the property that they are *semantically-secure* encryptions of their messages. To build an encrypted database system with robustness against range queries, the database server only stores the "right" ciphertexts (in sorted order). To perform a range query, the client provides the "left" ciphertexts corresponding to its range. The server can respond to the range query as usual since comparisons are possible between left and right ciphertexts. Robustness against offline inference attacks is ensured since the database dump only contains the right ciphertexts stored on the server, which are semantically-secure encryptions of their underlying messages. We describe our method in greater detail in Section 5.

New lower bounds for OPE. The core building block in our new ORE construction is a small-domain ORE with best-possible security. This raises the natural question of whether we could construct a small-domain OPE that also achieves best-possible security. Previously, Boldyreva et al. [BCLO09, BCO11] and Popa et al. [PLZ13] gave lower bounds that ruled out schemes where the ciphertext space is subexponential in the size of the plaintext space. But when the plaintext has size $\mathrm{poly}(\lambda)$ for a security parameter λ, there could conceivably exist an efficient OPE scheme with best-possible security. In this work, we show that this is in fact impossible. Using a very different set of techniques compared to [BCLO09, BCO11, PLZ13], we show (Section 6) that *no efficient* (stateless and non-interactive) OPE scheme can satisfy best-possible security, even when the message space contains only 3 elements! Thus, to achieve strong security even in the small-domain setting, it is necessary to consider relaxations of OPE, such as ORE.

Experimental evaluation. Finally, we implement and compare our new ORE scheme to the ORE scheme by Chenette et al. [CLWW16] and the OPE scheme by Boldyreva et al. [BCLO09]. For typical parameters, our new ORE scheme is over 65 times faster than the Boldyreva et al. scheme, but has longer ciphertexts. For example, when working with byte-size blocks, encrypting a 32-bit integer requires just 55 μs and produces a ciphertext that is 224 bytes. Typically, range queries are not performed over extremely long fields, so the extra space overhead of our scheme is not unreasonable. Given the superior security conferred by our scheme (in both the online and offline settings), and faster throughputs, our ORE scheme is a very compelling replacement for existing OPE schemes.

Applying ORE. To conclude, we make a cautionary note that because of the leakage associated with any ORE scheme, the primitive is not always suitable for applications that demand a high level of security. Our hope, however, is that by giving precise, concrete characterization of the leakage profile of our construction (in both the online and offline settings when used to support encrypted database queries), practitioners are able to make better-informed decisions on the suitability of our construction for a specific application.

2 Preliminaries

For $n \in \mathbb{N}$, we write $[n]$ to denote the set of integers $\{1, \ldots, n\}$. If $\mathcal{P}$ is a predicate on x, we write $\mathbf{1}(\mathcal{P}(x))$ to denote the indicator function for $\mathcal{P}$: that is, $\mathbf{1}(\mathcal{P}(x)) = 1$ if and only if $\mathcal{P}(x) = 1$, and 0 otherwise. For a distribution $\mathcal{D}$, we write $x \leftarrow \mathcal{D}$ to denote a draw from $\mathcal{D}$. For a finite set S, we write $x \xleftarrow{R} S$ to denote a uniformly random draw from S. In this work, we write λ to denote a security parameter. We say a function $f(\lambda)$ is negligible in λ if $f = o(1/\lambda^c)$ for all $c \in \mathbb{N}$. We write $\mathrm{negl}(\lambda)$ to denote a negligible function in λ and $\mathrm{poly}(\lambda)$ to denote a polynomial in λ. We say that an event occurs with negligible probability if the probability of the event occurring is $\mathrm{negl}(\lambda)$, and that it occurs with overwhelming probability if the complement of the event occurs with negligible probability. For two bit strings $x, y \in \{0,1\}^*$, we write $x \| y$ to denote the concatenation of x and y.

For two distributions $\mathcal{D}_1, \mathcal{D}_2$, we write $\mathcal{D}_1 \overset{c}{\approx} \mathcal{D}_2$ to denote that $\mathcal{D}_1$ and $\mathcal{D}_2$ are computationally indistinguishable (i.e., no efficient adversary can distinguish $\mathcal{D}_1$ from $\mathcal{D}_2$ except with negligible probability). We write $\mathcal{D}_1 \overset{s}{\approx} \mathcal{D}_2$ if $\mathcal{D}_1$ and $\mathcal{D}_2$ are statistically indistinguishable (i.e., the statistical distance between $\mathcal{D}_1$ and $\mathcal{D}_2$ is negligible). Finally, we write $\mathcal{D}_1 \equiv \mathcal{D}_2$ to denote that $\mathcal{D}_1$ and $\mathcal{D}_2$ are identical distributions.

We also review the standard definition of pseudorandom functions (PRFs) [GGM86]. A function $F : \mathcal{K} \times \mathcal{X} \to \mathcal{Y}$ is a secure PRF if no efficient adversary can distinguish (except perhaps with negligible probability) the outputs (on arbitrary points chosen adaptively by the adversary) of $F(k, \cdot)$ for a randomly chosen $k \xleftarrow{R} \mathcal{K}$ from that of a truly random function $f(\cdot)$ from $\mathcal{X}$ to $\mathcal{Y}$. Similarly, a function $F : \mathcal{K} \times \mathcal{X} \to \mathcal{X}$ is a secure pseudorandom permutation (PRP) if for all $k \in \mathcal{K}$, $F(k, \cdot)$ is a permutation on $\mathcal{X}$ and no efficient adversary can distinguish the outputs of $F(k, \cdot)$ where $k \xleftarrow{R} \mathcal{K}$ from the outputs of $\pi(\cdot)$ where π is a random permutation on $\mathcal{X}$.

2.1 Order-Revealing Encryption

An order-revealing encryption (ORE) scheme [BLR+15, CLWW16] is a tuple of three algorithms $\Pi = (\mathsf{ORE.Setup}, \mathsf{ORE.Encrypt}, \mathsf{ORE.Compare})$ defined over a well-ordered domain $\mathcal{D}$ with the following properties:

- $\mathsf{ORE.Setup}(1^\lambda) \to \mathsf{sk}$: On input a security parameter λ, the setup algorithm outputs a secret key sk.
- $\mathsf{ORE.Encrypt}(\mathsf{sk}, m) \to \mathsf{ct}$: On input a secret key sk and a message $m \in \mathcal{D}$, the encryption algorithm outputs a ciphertext ct.
- $\mathsf{ORE.Compare}(\mathsf{ct}_1, \mathsf{ct}_2) \to b$: On input two ciphertexts $\mathsf{ct}_1, \mathsf{ct}_2$, the compare algorithm outputs a bit $b \in \{0,1\}$.

Correctness. We say an ORE scheme over a well-ordered domain $\mathcal{D}$ is correct if for $\mathsf{sk} \leftarrow \mathsf{ORE.Setup}(1^\lambda)$ and all messages $m_1, m_2 \in \mathcal{D}$,

$$\Pr[\mathsf{ORE.Compare}(\mathsf{ct}_1, \mathsf{ct}_2) = \mathbf{1}(m_1 < m_2)] = 1 - \mathrm{negl}(\lambda).$$

Remark 2.1 (ORE Decryption)**.** Our schema for ORE does not include a decryption function, but as noted by Chenette et al. [CLWW16, Remark 2.3], this is without loss of generality. In particular, we can construct a decryption algorithm $\mathsf{ORE.Decrypt}$ using the $\mathsf{ORE.Encrypt}$ and $\mathsf{ORE.Compare}$ algorithms (by performing a binary search).

Security. The "best-possible" notion of security for order-revealing encryption is the notion of indistinguishability under an ordered chosen plaintext attack (IND-OCPA) introduced by Boldyreva et al. [BCLO09]. The IND-OCPA notion of security is a generalization of semantic security [GM84], and states that no efficient adversary can distinguish between the encryptions of any two sequences of messages, provided that the ordering of the messages in the two sequences is identical. We give the formal definition in Section 6 (Definition 6.1).

Due to the apparent difficulty in constructing efficient schemes that satisfy IND-OCPA security, Chenette et al. [CLWW16] introduced a weaker simulation-based notion of security for ORE schemes that allows for some leakage beyond just the ordering of the plaintexts. We recall their definition here.

Definition 2.2 (ORE with Leakage [CLWW16])**.** Let $\Pi = (\mathsf{ORE.Setup}, \mathsf{ORE.Encrypt}, \mathsf{ORE.Compare})$ be an ORE scheme, and let $\mathcal{A} = (\mathcal{A}_1, \ldots, \mathcal{A}_q)$ be an adversary for some $q = \text{poly}(\lambda)$. Let $\mathcal{S} = (\mathcal{S}_0, \mathcal{S}_1, \ldots, \mathcal{S}_q)$ be a simulator, and let $\mathcal{L}(\cdot)$ be a leakage function. We define the experiments $\mathsf{REAL}^{\text{ORE}}_{\mathcal{A}}(\lambda)$ and $\mathsf{SIM}^{\text{ORE}}_{\mathcal{A},\mathcal{S},\mathcal{L}}(\lambda)$ as follows:

$\mathsf{REAL}^{\text{ORE}}_{\mathcal{A}}(\lambda)$:	$\mathsf{SIM}^{\text{ORE}}_{\mathcal{A},\mathcal{S},\mathcal{L}}(\lambda)$:
1. $\mathsf{sk} \leftarrow \mathsf{ORE.Setup}(1^\lambda)$	1. $\mathsf{st}_{\mathcal{S}} \leftarrow \mathcal{S}_0(1^\lambda)$
2. $(m_1, \mathsf{st}_{\mathcal{A}}) \leftarrow \mathcal{A}_1(1^\lambda)$	2. $(m_1, \mathsf{st}_{\mathcal{A}}) \leftarrow \mathcal{A}_1(1^\lambda)$
3. $c_1 \leftarrow \mathsf{ORE.Encrypt}(\mathsf{sk}, m_1)$	3. $(c_1, \mathsf{st}_{\mathcal{S}}) \leftarrow \mathcal{S}_1(\mathsf{st}_{\mathcal{S}}, \mathcal{L}(m_1))$
4. for $2 \le i \le q$:	4. for $2 \le i \le q$:
(a) $(m_i, \mathsf{st}_{\mathcal{A}}) \leftarrow \mathcal{A}_i(\mathsf{st}_{\mathcal{A}}, c_1, \ldots, c_{i-1})$	(a) $(m_i, \mathsf{st}_{\mathcal{A}}) \leftarrow \mathcal{A}_i(\mathsf{st}_{\mathcal{A}}, c_1, \ldots, c_{i-1})$
(b) $c_i \leftarrow \mathsf{ORE.Encrypt}(\mathsf{sk}, m_i)$	(b) $(c_i, \mathsf{st}_{\mathcal{S}}) \leftarrow \mathcal{S}_i(\mathsf{st}_{\mathcal{S}}, \mathcal{L}(m_1, \ldots, m_i))$
5. output $(c_1, \ldots, c_q)$ and $\mathsf{st}_{\mathcal{A}}$	5. output $(c_1, \ldots, c_q)$ and $\mathsf{st}_{\mathcal{A}}$

We say that Π is a secure ORE scheme with leakage function $\mathcal{L}(\cdot)$ if for all polynomial-size adversaries $\mathcal{A} = (\mathcal{A}_1, \ldots, \mathcal{A}_q)$, there exists a polynomial-size simulator $\mathcal{S} = (\mathcal{S}_0, \mathcal{S}_1, \ldots, \mathcal{S}_q)$ such that the outputs of the two distributions $\mathsf{REAL}^{\text{ORE}}_{\mathcal{A}}(\lambda)$ and $\mathsf{SIM}^{\text{ORE}}_{\mathcal{A},\mathcal{S},\mathcal{L}}(\lambda)$ are computationally indistinguishable.

Remark 2.3 (Best-Possible Security)**.** The best-possible notion of simulation-security is security with respect to the leakage function that only reveals the ordering of the plaintexts. This is the minimal leakage possible from an order-revealing encryption scheme. In particular, we define $\mathcal{L}_{\text{CMP}}$ as follows:

$$\mathcal{L}_{\text{CMP}}(m_1, \ldots, m_t) = \{(i, j, \text{CMP}(m_i, m_j) \mid 1 \le i < j \le t)\},$$

where $\text{CMP}(m_i, m_j)$ is the comparison function that outputs -1 if $m_i < m_j$, 0 if $m_i = m_j$ and 1 if $m_i > m_j$.

3 ORE for Small Domains

The order-revealing encryption in [CLWW16] reveals a significant amount of information, namely, the index of the first bit position that differs between two encrypted plaintexts. In this work, we show how to construct an ORE scheme that only leaks the first *block* that differs, where a block is a collection of one or more bits. For instance, we can construct an ORE scheme that only reveals the first *byte* that differs between two encrypted plaintexts, and nothing more.

The starting point for our construction is a "small-domain" ORE scheme with best-possible simulation security. The limitation is that the length of the ciphertexts in our ORE scheme grows

linearly with the size of the message space, hence the restriction to small (polynomially-sized) domains. We show in Section 4 how to extend our small-domain ORE to obtain an order-revealing encryption scheme over large domains (i.e., exponentially-sized) that leaks strictly less information compared to the scheme by Chenette et al. [CLWW16].

As described in Section 1.1, we give our ORE construction in the left/right framework where we decompose the $\mathsf{ORE.Encrypt}$ function into two separate functions: $\mathsf{ORE.Encrypt_L}$ and $\mathsf{ORE.Encrypt_R}$. We refer to them as the "left encryption" and "right encryption" functions, respectively. Our particular construction has the property that only "left ciphertexts" can be compared with "right ciphertexts." Note that this is without loss of generality and we can recover the usual notion of ORE by simply defining the output of $\mathsf{ORE.Encrypt}(\mathsf{sk}, m)$ to be the tuple $(\mathsf{ORE.Encrypt_L}(\mathsf{sk}, m), \mathsf{ORE.Encrypt_R}(\mathsf{sk}, m))$.

3.1 Small-Domain ORE Construction

We begin with a high-level overview of our construction. Our scheme is defined with respect to a plaintext space $[N]$ where $N = \text{poly}(\lambda)$. First, we associate each element $x \in [N]$ in the domain with an encryption key k_x. A (right) ciphertext for a value $y \in [N]$ consists of N encryptions of the comparison output $\text{CMP}(x, y)$ between y and every element $x \in [N]$ in the domain, where the value $\text{CMP}(x, y)$ is encrypted under k_x. The left encryption of a value x is simply the encryption key k_x. Given k_x and an encryption of $\text{CMP}(x, y)$ under k_x, the evaluator can decrypt and learn the comparison bit $\text{CMP}(x, y)$. The values of the other comparison bits are hidden by semantic security of the encryption scheme. Note, however, that we still need a way for the evaluator to determine *which* of the N ciphertexts is encrypted under k_x without learning the value of x. To ensure this, we sample a random permutation π on the domain $[N]$ during setup. The components in the right ciphertexts are then permuted according to π and the left encryption of x includes the permuted position $\pi(x)$. Given $\pi(x)$, the evaluator learns which component in the right ciphertext to decrypt, but learns nothing about x. Finally, to show simulation security, we require a "non-committing" encryption scheme, and for this, we rely on a random oracle [BR93].[5]

Construction. Let $[N]$ be the message space. Let $F : \{0,1\}^\lambda \times \{0,1\}^\lambda \to \{0,1\}^\lambda$ be a secure PRF and $H : \{0,1\}^\lambda \times \{0,1\}^\lambda \to \mathbb{Z}_3$ be a hash function (modeled as a random oracle in the security proof). Let CMP be the comparison function from Remark 2.3. Our ORE scheme $\Pi^{(\mathsf{s})}_{\mathsf{ore}}$ is defined as follows:

- $\mathsf{ORE.Setup}(1^\lambda)$. The setup algorithm samples a PRF key $k \xleftarrow{R} \{0,1\}^\lambda$ for F, and a uniformly random permutation $\pi : [N] \to [N]$. The secret key sk is the pair (k, π).
- $\mathsf{ORE.Encrypt_L}(\mathsf{sk}, x)$. Write sk as (k, π). The left encryption algorithm computes and returns the tuple $\mathsf{ct_L} = (F(k, \pi(x)), \pi(x))$.
- $\mathsf{ORE.Encrypt_R}(\mathsf{sk}, y)$. Write sk as (k, π). First, the right encryption algorithm samples a random nonce $r \xleftarrow{R} \{0,1\}^\lambda$. Then, for each $i \in [N]$, it computes the value

$$v_i = \text{CMP}(\pi^{-1}(i), y) + H(F(k, i), r) \pmod 3.$$

  Finally, it outputs the ciphertext $\mathsf{ct_R} = (r, v_1, v_2 \ldots, v_N)$.

[5] We believe we can replace the random oracle with a PRF if we aim to prove an indistinguishability notion of security for our construction. For simplicity of presentation in this paper, we work with a simulation-based definition and prove security in the random oracle model.

- $\mathsf{ORE.Compare}(\mathsf{ct_L}, \mathsf{ct_R})$. The compare algorithm first parses

$$\mathsf{ct_L} = (k', h) \quad \text{and} \quad \mathsf{ct_R} = (r, v_1, v_2, \ldots, v_N),$$

and then outputs the result $v_h - H(k', r) \pmod 3$.

Correctness. Let $\mathsf{sk} = (k, \pi) \leftarrow \mathsf{ORE.Setup}(1^\lambda)$, and take any $x, y \in [N]$. Let $\mathsf{ct_L}^{(x)} = (k', h) \leftarrow \mathsf{ORE.Encrypt_L}(\mathsf{sk}, x)$ and and $\mathsf{ct_R}^{(y)} = (r, v_1, \ldots, v_N) \leftarrow \mathsf{ORE.Encrypt_R}(\mathsf{sk}, y)$, Then, we have the following:

$$\begin{aligned}
\mathsf{ORE.Compare}(\mathsf{ct_L}^{(x)}, \mathsf{ct_R}^{(y)}) &= v_h - H(k', r) \\
&= \textsc{cmp}(\pi^{-1}(h), y) + H(F(k, h), r) - H(k', r) \\
&= \textsc{cmp}(\pi^{-1}(\pi(x)), y) + H(F(k, \pi(x)), r) - H(F(k, \pi(x)), r) \\
&= \textsc{cmp}(x, y) \in \mathbb{Z}_3,
\end{aligned}$$

Note that $\textsc{cmp}(x, y)$ provides the same amount of information as $\mathbf{1}(x < y)$ and $\mathbf{1}(y < x)$, so correctness follows.

Space usage. Before we give our formal security analysis, we first characterize the length of the ciphertexts in our ORE scheme for a message space of size N. The left ciphertexts $\mathsf{ct_L}$ in our scheme consists of a PRF key and an index, which are $\lambda + \lceil \log N \rceil$ bits long. The right ciphertexts $\mathsf{ct_R}$ consists of a nonce, together with N elements in $\mathbb{Z}_3$, which can be represented using $\lambda + \lceil N \log_2 3 \rceil$ bits. Thus, a complete ciphertext consists of $2\lambda + \lceil \log N \rceil + \lceil N \log_2 3 \rceil$ bits. However, as we note in the following remark, it is possible to obtain shorter ciphertexts if we allow the comparison algorithm to take the full ciphertext $(\mathsf{ct_L}, \mathsf{ct_R})$ as opposed to only the left half of the first ciphertext and the right half of the second ciphertext. Thus, when using the construction as a pure ORE scheme, we can obtain shorter ciphertexts. When leveraging our ORE scheme to build a range query system (Section 5), we will exploit the fact that comparisons can be performed given just the left component of one ciphertext and the right component of the other.

Remark 3.1 (Shorter Ciphertexts)**.** For a domain of size N, the right ciphertexts in our ORE construction contain N elements of $\mathbb{Z}_3$. Suppose instead we replaced the comparison function $\textsc{cmp}$ with a function $\textsc{cmp}'$ where $\textsc{cmp}'(x, y) = 1$ if $x \leq y$ and 0 otherwise. Then, a left encryption $\mathsf{ct_L}^{(x)}$ of x and a right encryption $\mathsf{ct_R}^{(y)}$ of y can be used to compute $\textsc{cmp}'(x, y)$, or equivalently, whether $x \leq y$. If the comparison algorithm takes as input $\mathsf{ct}^{(x)} = \left(\mathsf{ct_L}^{(x)}, \mathsf{ct_R}^{(x)}\right)$ and $\mathsf{ct}^{(y)} = \left(\mathsf{ct_L}^{(y)}, \mathsf{ct_R}^{(y)}\right)$, then it can compute both $\textsc{cmp}'(x, y)$ and $\textsc{cmp}'(y, x)$. This means that given $\mathsf{ct}^{(x)}$ and $\mathsf{ct}^{(y)}$, the comparison algorithm can still determine if $x < y$, $x = y$, or $x > y$. With this modification, the right ciphertexts in our scheme have length N rather than $\lceil N \log_2 3 \rceil$.

Remark 3.2 (Beyond Comparisons)**.** By substituting an arbitrary bivariate function $f(x, y)$ for the comparison function $\textsc{cmp}$ in our construction, we obtain an encryption scheme where any two ciphertexts $\mathsf{ct}^{(x)}$ and $\mathsf{ct}^{(y)}$ encrypting messages x and y, respectively, reveal $f(x, y)$. Moreover, by the security argument given in the proof of Theorem 3.3, we have that $\mathsf{ct}^{(x)}$ and $\mathsf{ct}^{(y)}$ reveal nothing more than the function value $f(x, y)$ and the equality predicate $x \stackrel{?}{=} y$. Note that equality is revealed in our construction since the left ciphertexts are deterministic. Our construction can thus be viewed as a general-purpose property-preserving encryption for two-input functionalities [PR12]

or a two-input functional encryption scheme [GGG+14, BLR+15] that leaks equality. Because the length of the ciphertexts in our construction grow linearly in the size of the domain, our construction is limited to functions over a polynomially-sized domain. However, in contrast to other schemes that rely on primitives such as indistinguishability obfuscation [GGG+14], multilinear maps [BLR+15], or pairings [KLM+16], our construction has the appealing property that it relies only on symmetric primitives, namely, PRFs.

Security. We now state our main security theorem for this section. We give the proof in Appendix A.

Theorem 3.3. *The ORE scheme $\Pi_{\mathsf{ore}}^{(\mathsf{s})}$ is secure with the best-possible leakage function $\mathcal{L}_{\text{CMP}}$ from Remark 2.3 assuming that F is a secure PRF and H is modeled as a random oracle.*

4 Domain Extension: A Large-Domain ORE

Although our small-domain ORE construction from Section 3 achieves the strongest possible notion of security for ORE, it is limited to polynomially-sized message spaces. In this section, we show how to construct an efficient ORE scheme for large domains which achieves provably stronger security guarantees than all existing efficient ORE constructions for large domains. Our construction can be viewed as a composition of our small-domain ORE construction together with the ORE scheme by Chenette et al. [CLWW16].

Intuitively, we can view the techniques used in the Chenette et al. construction as a domain-extension mechanism for ORE. In particular, their construction can be viewed as a general transformation that takes as input a k-bit ORE scheme and outputs an kn-bit ORE scheme, with ciphertext expansion that grows linearly in n and a slight reduction in security (that degrades with n). Under this lens, the Chenette et al. construction can be viewed as taking a 1-bit ORE scheme (with best-possible security) and extending it to an n-bit ORE scheme. In this work, we apply this general domain-extension technique to our small-domain ORE from Section 3, and show how we can start with a d-bit ORE and extend it to a dn-bit ORE. By varying the parameters n and d, we obtain a performance-security tradeoff. At a high level, our composed construction implements encryption via several parallel (prefix-dependent) instances of the small-domain ORE scheme $\Pi_{\mathsf{ore}}^{(\mathsf{s})}$ from Section 3, one for each block of the plaintext. Using the techniques of Chenette et al. [CLWW16], a blinding factor is derived from the prefix of each block and used to mask the $\Pi_{\mathsf{ore}}^{(\mathsf{s})}$ ciphertexts for that block. We give the precise leakage of our construction in Theorem 4.1.

Construction. Fix a security parameter $\lambda \in \mathbb{N}$, a message space size $N > 0$, and integers $d, n > 0$ such that $d^n \geq N$. Let $F : \{0,1\}^\lambda \times [N] \to \{0,1\}^\lambda$ be a secure PRF on variable-length inputs,[6] $H : \{0,1\}^\lambda \times \{0,1\}^\lambda \to \mathbb{Z}_3$ be a hash function (modeled as a random oracle), and $\pi : \{0,1\}^\lambda \times [d] \to [d]$ be a secure PRP. For a d-ary string $x = x_1 x_2 \cdots x_n$, let $x_{|i} = x_1 x_2 \cdots x_i$ denote the d-ary string representing the first i digits of x (i.e., the length-i prefix of x), and let $x_{|0}$ be the empty prefix. We define our ORE scheme $\Pi_{\mathsf{ore}} = (\mathsf{ORE.Setup}, \mathsf{ORE.Encrypt}_{\mathsf{L}}, \mathsf{ORE.Encrypt}_{\mathsf{R}}, \mathsf{ORE.Compare})$ as follows.

- $\mathsf{ORE.Setup}(1^\lambda)$. The setup algorithm samples PRF keys $k_1, k_2 \xleftarrow{\text{R}} \{0,1\}^\lambda$. The master secret key is $\mathsf{sk} = (k_1, k_2)$.

[6] The Chenette et al. ORE construction also used a PRF on variable-length inputs. We refer to their construction [CLWW16, §3] for one possible way of constructing a PRF on variable-length inputs from a standard PRF.

- $\mathsf{ORE.Encrypt_L}(\mathsf{sk}, x)$. Let $\mathsf{sk} = (k_1, k_2)$. For each $i \in [n]$, the left encryption algorithm first computes $\tilde{x} = \pi(F(k_2, x_{|i-1}), x_i)$ and then sets $u_i = (F(k_1, x_{|i-1} \| \tilde{x}), \tilde{x})$. It returns the ciphertext $\mathsf{ct_L} = (u_1, \ldots, u_n)$.

- $\mathsf{ORE.Encrypt_R}(\mathsf{sk}, y)$. Let $\mathsf{sk} = (k_1, k_2)$. First, the right encryption algorithm uniformly samples a nonce $r \xleftarrow{R} \{0,1\}^\lambda$. Then, for each $i \in [n]$ and $j \in [d]$, letting $j^* = \pi^{-1}(F(k_2, y_{|i-1}), j)$, it computes
$$z_{i,j} = \text{CMP}(j^*, y_i) + H(F(k_1, y_{|i-1} \| j), r) \pmod 3.$$
It then defines the tuple $v_i = (z_{i,1}, \ldots, z_{i,d})$ and outputs the ciphertext $\mathsf{ct_R} = (r, v_1, v_2 \ldots, v_n)$.

- $\mathsf{ORE.Compare}(\mathsf{ct_L}, \mathsf{ct_R})$. The compare algorithm first parses
$$\mathsf{ct_L} = (u_1, \ldots, u_n) \quad \text{and} \quad \mathsf{ct_R} = (r, v_1, v_2, \ldots, v_n),$$
where for each $i \in [n]$, we write $u_i = (k'_i, h_i)$ and $v_i = (z_{i,1}, \ldots, z_{i,d})$. Then, let ℓ be the smallest index i for which $z_{i,h_i} - H(k'_i, r) \neq 0 \pmod 3$. If no such ℓ exists, output 0. Otherwise, output $z_{\ell,h_\ell} - H(k'_\ell, r) \pmod 3$.

Correctness. Let $\mathsf{sk} = (k_1, k_2) \leftarrow \mathsf{ORE.Setup}(1^\lambda)$ and take any $x, y \in [N]$. Let $\mathsf{ct_L^{(x)}} \leftarrow \mathsf{ORE.Encrypt_L}(\mathsf{sk}, x)$ and $\mathsf{ct_R^{(y)}} \leftarrow \mathsf{ORE.Encrypt_R}(\mathsf{sk}, y)$. We show that with overwhelming probability, $\mathsf{ORE.Compare}(\mathsf{ct_L^{(x)}}, \mathsf{ct_R^{(y)}}) = \text{CMP}(x, y)$.

Let $x = x_1 \cdots x_n$ and $y = y_1 \cdots y_n$. Let $\mathsf{ct_L^{(x)}} = (u_1, \ldots, u_n)$ and $\mathsf{ct_R^{(y)}} = (r, v_1, \ldots, v_n)$, $u_i = (k'_i, h_i)$ and $v_i = (z_{i,1}, \ldots, z_{i,d})$ for all $i \in [n]$. Next, let $i^* \in [n]$ be the first index i where $x_i \neq y_i$. If $x = y$, set $i^* = n + 1$. Then, if $x \neq y$, we have that $\text{CMP}(x, y) = \text{CMP}(x_{i^*}, y_{i^*})$ By definition, for all $\ell < i^*$, $x_{|\ell} = y_{|\ell}$, and so setting $\kappa_\ell = F(k_2, x_{|\ell}) = F(k_2, y_{|\ell})$, we have that
$$\pi^{-1}(\kappa_{\ell-1}, h_\ell) = \pi^{-1}(\kappa_{\ell-1}, \pi(\kappa_{\ell-1}, x_\ell)) = x_\ell.$$
By definition of $z_{i,j}$, we have that for all $\ell \leq i^*$
$$\begin{aligned} z_{\ell,h_\ell} &= \text{CMP}(\pi^{-1}(\kappa_{\ell-1}, h_\ell), y_\ell) + H(F(k_1, y_{|\ell-1} \| h_\ell), r) \\ &= \text{CMP}(x_\ell, y_\ell) + H(F(k_1, x_{|\ell-1} \| h_\ell), r) \\ &= \text{CMP}(x_\ell, y_\ell) + H(k'_\ell, r). \end{aligned}$$
Thus, for all $\ell < i^*$, $z_{\ell,h_\ell} - H(k'_\ell, r) = \text{CMP}(x_\ell, y_\ell) = 0$, and for $\ell = i^*$, $z_{\ell,h_\ell} - H(k'_\ell, r) = \text{CMP}(x_{i^*}, y_{i^*})$. If $x = y$, then $i^* = n+1$ and for all $\ell \in [n]$, $z_{\ell,h_\ell} - H(k'_\ell, r) = 0$, in which case the comparison algorithm correctly outputs 0. Otherwise, the comparison algorithm outputs $\text{CMP}(x_{i^*}, y_{i^*}) = \text{CMP}(x, y)$.

Security. Before stating our security theorem, we first specify our leakage function $\mathcal{L}^{(d)}_{\text{BLK}}$. Each ciphertext block in our ORE scheme is essentially a ciphertext for the underlying small-domain ORE, and the comparison operation proceeds block-by-block. Intuitively then, since our small-domain ORE scheme leaks nothing except the ordering (Theorem 3.3), the additional leakage of our new ORE scheme is the index of the first block that differs between two ciphertexts. In particular, for messages $x = x_1 x_2 \cdots x_n$ and $y = y_1 y_2 \cdots y_n$ written in base d, we define the first differing block function $\mathsf{ind}^{(d)}_{\mathsf{diff}}(x, y)$ to be the first index $i \in [n]$ such that $x_j = y_j$ for all $j < i$ and $x_i \neq y_i$. If $x = y$,

we define $\mathsf{ind}^{(d)}_{\mathsf{diff}}(x, y)$ to be $n + 1$. Then, our leakage function $\mathcal{L}^{(d)}_{\text{BLK}}$ for our extended ORE scheme is given by

$$\mathcal{L}^{(d)}_{\text{BLK}}(m_1, \ldots, m_t) = \{(i, j, \text{BLK}(m_i, m_j)) \mid 1 \leq i < j \leq t\},$$

where $\text{BLK}(m_i, m_j) = (\text{CMP}(m_i, m_j), \mathsf{ind}^{(d)}_{\mathsf{diff}}(m_i, m_j))$. In general, we refer to the parameter d as the arity (or base) of the plaintext space, which grows exponentially in the length (in bits) of the block. We now state our main security theorem.

Theorem 4.1. *The ORE scheme Π_{ore} is secure with leakage function $\mathcal{L}_{\text{BLK}}$ assuming that F is a secure PRF and H is modeled as a random oracle.*

The proof of Theorem 4.1 can be viewed as a composition of the security proof for our underlying small-domain ORE (Theorem 3.3) and the security proof of the Chenette et al. scheme [CLWW16, Theorem 3.2]. We give the proof in Appendix B.

Space usage. Ciphertexts in our new ORE scheme consist essentially of n ciphertexts for our small-domain ORE scheme (with domain size d). More concretely, a left ciphertext in our new scheme consists of $n(\lambda + \lceil \log d \rceil)$ bits and a right ciphertext consists of $\lambda + n \lceil d \log_2 3 \rceil$ bits. Since the size of the plaintext space N satisfies $N \leq d^n$, ciphertext size in our new ORE scheme grow as $O((\lambda + d) \log_d N)$.

Non-uniform block sizes. In practice, some bits of the plaintext may be more sensitive than others. Leaking information about these bits is less desirable than leaking information about less sensitive bits. To accommodate the different sensitivities, we can use different input bases (e.g., use larger blocks for more sensitive bits) for the different blocks of the ciphertext. The leakage in the resulting scheme is still the index of the first (variable-sized) block that differs between two messages. Correctness is unchanged.

5 Encrypted Range Queries

In this section, we formally define the properties of a client-server protocol for range queries over an encrypted database. In our model, a client stores an encrypted database on the server. The client can update the database (e.g., by adding or removing records) and issue range queries against the database. In a range query, the client specifies a numeric interval and the server responds by returning all ciphertexts whose underlying messages fall within that interval.

Although our definitions are stated in terms of numeric intervals, our methods are broadly applicable to more general settings—in particular, to any well-ordered domain such as English names. For example, when the database consists of encrypted alphanumeric strings, range queries can be used for both exact-keyword as well as prefixed-based search.

Our security definitions are adapted from existing definitions for searchable symmetric encryption (SSE) [CGKO06, CK10]. We survey some of the work on SSE in Section 8. In our definitions we consider both the online and offline settings. In the online setting, the adversary sits on the server and sees both the encrypted database as well as the client's queries, while in the offline setting, the adversary just obtains a dump of the server's encrypted database. By showing that in the offline setting, the server's encrypted database provides semantic security, we can argue that our new range query scheme provides robustness against the kinds of offline inference attacks considered by Naveed et al. [NKW15].

After formally defining the security requirements for a range query protocol, we give a construction based on our ORE scheme Π_{ore} from Section 4. Our protocol not only satisfies our security properties, but also has several additional appealing properties such as sublinear query time (in the size of the database) and optimal round complexity.

Our proposed protocol is easily extensible to the multi-client setting where many clients are interacting with the server. Each authorized client is simply given the secret key needed to query and update the database.

5.1 Range Query Schemes

We begin with a formal definition of a range query scheme, followed by our notions of online and offline security. We describe a range query scheme in terms of a set of algorithms, where each algorithm is a single-round protocol between the client and the server. In each protocol, the client is always stateless, but the server is stateful—in particular, the server's state represents the information stored on the server needed to efficiently respond to the client's queries, including the encrypted database itself.

Initially, the client runs a setup procedure that takes as input a plaintext database $\boldsymbol{D}$ of values and outputs a secret key sk and some token t representing the encrypted database. The token t is given to the server, and the server outputs some initial state st. Then, for each query (range query, insert query, delete query), the client uses the secret key sk to derive a token t representing its query, and sends t to the server. This token contains a masked version of the client's input for the query. On input a query token t, the server processes the query and updates its internal state. In a range query, the server also returns a response r, which the client uses to learn the answer to the range query.

More formally, let $\boldsymbol{D} \in [N]^M$ represent a (possibly empty) database consisting of $M \geq 0$ values, each in the range $[N]$. A range query scheme $\Pi_{\mathsf{rq}} = (\mathsf{RQ.Setup}, \mathsf{RQ.Range}, \mathsf{RQ.Insert}, \mathsf{RQ.Delete})$ consists of a tuple of algorithms defined as follows:

- $\mathsf{RQ.Setup}(1^\lambda, \boldsymbol{D}) \to (\mathsf{t}, \mathsf{st})$. The setup algorithm between the client and server proceeds as follows:
  - $\mathsf{Client}(1^\lambda, \boldsymbol{D}) \to (\mathsf{sk}, \mathsf{t})$. The client, on input the security parameter λ and database $\boldsymbol{D}$, produces a key sk which is kept secret, and a token t which is sent to the server.
  - $\mathsf{Server}(\mathsf{t}) \to \mathsf{st}$. The server takes as input the token t and outputs an initial state st.
- $\mathsf{RQ.Range}(\mathsf{sk}, \mathsf{q}, \mathsf{st}) \to (\mathsf{t}, \mathsf{st}')$. The range query algorithm between the client and server proceeds as follows:
  - $\mathsf{Client}(\mathsf{sk}, \mathsf{q} = (x, y)) \to \mathsf{t}$. The client, on input the secret key sk and a query q for the range $[x, y]$, produces a token t which is sent to the server.
  - $\mathsf{Server}(\mathsf{st}, \mathsf{t}) \to (\mathsf{st}', \mathsf{r})$. The server takes as input its current state st and the token t and produces an updated state st', along with a response r, which is sent to the client.
  - $\mathsf{Client}(\mathsf{sk}, \mathsf{r}) \to \mathsf{S}$. The client, on input the secret key sk and the response r from the server, obtains a subset S of entries which represent the answer to the range query.
- $\mathsf{RQ.Insert}(\mathsf{sk}, \mathsf{q}, \mathsf{st}) \to (\mathsf{t}, \mathsf{st}')$. The insert algorithm between the client and server proceeds as follows:

– $\mathsf{Client}(\mathsf{sk}, \mathsf{q} = x) \to \mathsf{t}$. The client, on input the secret key sk and a query q representing an insertion of the value x, produces a token t which is sent to the server.

– $\mathsf{Server}(\mathsf{st}, \mathsf{t}) \to (\mathsf{st}', \mathsf{r})$. The server takes as input its current state st and the token t and produces an updated state st'.

- $\mathsf{RQ.Delete}(\mathsf{sk}, \mathsf{q}, \mathsf{st}) \to (\mathsf{t}, \mathsf{st}')$. The delete algorithm between the client and server proceeds as follows:

  – $\mathsf{Client}(\mathsf{sk}, \mathsf{q} = x) \to \mathsf{t}$. The client, on input the secret key sk and a query q representing a deletion of the value x, produces a token t which is sent to the server.

  – $\mathsf{Server}(\mathsf{st}, \mathsf{t}) \to (\mathsf{st}', \mathsf{r})$. The server takes as input its current state st and the token t and produces an updated state st'.

We now define the correctness and security properties of a range query scheme. At a high level, we say that a range query scheme is correct if for all range queries (x, y) the client makes, it obtains the set of entries in the database $\boldsymbol{D}$ (taking into account any insertion and deletion queries occurring before the range query) that lie in the interval $[x, y]$.

Correctness. Fix a security parameter λ, positive integers x, y, N, M where $x \leq y \in [N]$, a database $\boldsymbol{D} \in [N]^M$ and a sequence of ℓ insertion, deletion, and range queries $\mathsf{q}_1, \ldots, \mathsf{q}_{\ell-1}$. Let $\mathsf{q}_\ell = (x, y)$ be a range query. Let $(\mathsf{st}_\ell, \mathsf{r}) \leftarrow \mathsf{Server}(\mathsf{st}_{\ell-1}, \mathsf{Client}(\mathsf{sk}, \mathsf{q}_\ell))$ and $\mathsf{S} \leftarrow \mathsf{Client}(\mathsf{sk}, \mathsf{r})$, where $\mathsf{st}_{\ell-1}$ is the server's state after processing queries $\mathsf{q}_1, \ldots, \mathsf{q}_{\ell-1}$. Let $\boldsymbol{D}_0 = \boldsymbol{D}, \boldsymbol{D}_1, \ldots, \boldsymbol{D}_\ell$ be the effective database elements after each query—that is, for all $i \in [\ell]$, $\boldsymbol{D}_i = \boldsymbol{D}_{i-1}$ if q_i is a range query, $\boldsymbol{D}_i = \boldsymbol{D}_{i-1} \cup \{x\}$ if q_{i-1} is an insertion query for x, and $\boldsymbol{D}_i = \boldsymbol{D}_{i-1} \setminus \{x\}$ if q_{i-1} is a deletion query for x. We say a range query scheme $\Pi_{\mathsf{rq}} = (\mathsf{RQ.Setup}, \mathsf{RQ.Range}, \mathsf{RQ.Insert}, \mathsf{RQ.Delete})$ is correct if for all security parameters λ, integers N, M, x, y, databases $\boldsymbol{D} \in [N]^M$ and sequence of queries $\mathsf{q}_1, \ldots, \mathsf{q}_\ell$, we have that the client's response S satisfies $\mathsf{S} = \boldsymbol{D}_\ell \cap [x, y]$.

Security. Our first notion of security is online security, which models the information revealed to a malicious server in the range query protocol. Here, the adversary sees both the contents of the server's state (i.e., the encrypted database) as well as the client's queries. We give a simulation-based definition with respect to a concrete leakage function that operates over the plaintext values in the database and the queries. Our definition is adapted from the standard paradigm used to define security in searchable symmetric encryption schemes [CGKO06, CK10].

Definition 5.1 (Online Security)**.** For all databases $\boldsymbol{D}$ and sequences of ℓ queries $\mathsf{q}_1, \ldots, \mathsf{q}_\ell$, define the sequence of states $\mathsf{st}_0, \ldots, \mathsf{st}_\ell$ and tokens $\mathsf{t}_0, \ldots, \mathsf{t}_\ell$ where $(\mathsf{t}_0, \mathsf{st}_0) \leftarrow \mathsf{RQ.Setup}(1^\lambda, \boldsymbol{D})$, and for each $i \in [\ell]$, $(\mathsf{t}_i, \mathsf{st}_i)$ is the output of the i^{th} query on input sk, q_i, and st_{i-1}. A range query scheme is online secure with respect to a leakage function $\mathcal{L}$ if for every efficient adversary $\mathcal{A}$, there exists a simulator $\mathcal{S}$ where

$$\left| \Pr[\mathcal{A}(1^\lambda, \mathsf{st}_0, \ldots, \mathsf{st}_\ell, \mathsf{t}_0, \ldots, \mathsf{t}_\ell) = 1] - \Pr[\mathcal{S}(1^\lambda, \mathcal{L}(\boldsymbol{D}, \mathsf{q}_1, \ldots, \mathsf{q}_\ell)) = 1] \right| = \mathrm{negl}(\lambda).$$

We also define an "offline" notion of security for a range query scheme. The offline setting models scenarios where the adversary obtains a dump of the contents of the server (i.e., the server's state), but does not observe any queries made by the client. Against offline adversaries, we require the much stronger property that the only thing leaked by the encrypted database is the size of the encrypted database. This is the best-possible leakage.

Definition 5.2 (Offline Security)**.** For all databases $\boldsymbol{D}$ and sequences of ℓ queries $\mathsf{q}_1, \ldots, \mathsf{q}_\ell$, define the sequence of states $\mathsf{st}_0, \ldots, \mathsf{st}_\ell$ and tokens $\mathsf{t}_0, \ldots, \mathsf{t}_\ell$ as in Definition 5.1 Let $|\mathsf{st}_\ell|$ be the bit-length of st_ℓ. A range query scheme is offline secure if for all efficient adversaries $\mathcal{A}$, there exists an efficient simulator $\mathcal{S}$ where

$$\Big|\Pr[\mathcal{A}(1^\lambda, \mathsf{st}_\ell) = 1] - \Pr[\mathcal{S}(1^\lambda, |\mathsf{st}_\ell|) = 1]\Big| = \mathrm{negl}(\lambda).$$

The importance of offline security. Although offline security is strictly weaker than online security, it captures the real-world scenario where an attacker breaks into a server and exfiltrates any data the server has stored on disk. While companies are often able to detect and protect against active online corruption of their servers, the question remains what happens after the fact when the attacker has also exfiltrated the database for offline analysis. Of course, the ideal solution to this problem is an encrypted database system that provides strong online security guarantees. However, existing systems with strong online security typically require redesigning the database management system and implementing elaborate cryptographic protocols for querying [CJJ+14, FJK+15], or leverage heavy, less practical tools such as fully homomorphic encryption [Gen09] or oblivious RAMs [GO96]. On the flip side, an OPE-based solution yields a scheme that does not provide offline security in our model; this is one reason why OPE and other PPE-based encrypted database schemes are vulnerable to inference attacks. This is true even if we use an (interactive) OPE scheme with best-possible security; the ability to directly compare ciphertexts is sufficient to carry out the inference attacks. Thus, there is an interesting intermediate ground where we build systems that achieve decent online security, while still providing strong offline security guarantees to be robust against inference attacks.

5.2 An Efficient Range Query Scheme

We now describe how to build an efficient range query scheme using our ORE construction from Section 4. At a high level, the server's encrypted database consists of right ciphertexts for each value, stored in sorted order. The tokens t for each query consist of a left encryption of the query value. This allows the server to use the ORE comparison algorithm to perform binary search over the encrypted ciphertexts in the database. Thus, the server is able to answer queries efficiently and maintain the database in sorted order (during updates). To answer a range query, the server performs binary search to find the lower and upper boundaries in the encrypted database corresponding to its query and returns all ciphertexts lying within those bounds. The client then decrypts the ciphertexts to learn the response.

More formally, we define our range query scheme $\Pi_{\mathsf{rq}} = (\mathsf{RQ.Setup}, \mathsf{RQ.Range}, \mathsf{RQ.Insert}, \mathsf{RQ.Delete})$ as follows:

- $\mathsf{RQ.Setup}(1^\lambda, \boldsymbol{D}) \to (\mathsf{t}, \mathsf{st})$. The setup algorithm between the client and server proceeds as follows:
  - $\mathsf{Client}(1^\lambda, \boldsymbol{D}) \to (\mathsf{sk}, \mathsf{t})$. The client, on input the security parameter λ and database $\boldsymbol{D}$, generates a secret key $\mathsf{sk} \leftarrow \mathsf{ORE.Setup}(1^\lambda)$. Then, the client *sorts* the database $\boldsymbol{D}$, and for each sequential element $x_i \in \boldsymbol{D}$, the client computes $\mathsf{ct}_i \leftarrow \mathsf{ORE.Encrypt}_{\mathsf{R}}(\mathsf{sk}, x_i)$, and sends the token $\mathsf{t} = (\mathsf{ct}_1, \ldots, \mathsf{ct}_M)$ to the server.
  - $\mathsf{Server}(\mathsf{t}) \to \mathsf{st}$. The server simply sets $\mathsf{st} = \mathsf{t}$.

- $\mathsf{RQ.Range}(\mathsf{sk}, \mathsf{q}, \mathsf{st}) \to (\mathsf{t}, \mathsf{st}')$. The range query algorithm between the client and server proceeds as follows:
  - $\mathsf{Client}(\mathsf{sk}, \mathsf{q} = (x, y)) \to \mathsf{t}$. The client, on input the secret key sk and a query representing a range query for the range $[x, y]$, produces the token $\mathsf{t} = (\mathsf{ORE.Encrypt}_{\mathsf{L}}(\mathsf{sk}, x), \mathsf{ORE.Encrypt}_{\mathsf{L}}(\mathsf{sk}, y))$ which is sent to the server.
  - $\mathsf{Server}(\mathsf{st}, \mathsf{t}) \to (\mathsf{st}', \mathsf{r})$. The server takes as input its current state $\mathsf{st} = (\mathsf{ct}_1, \ldots, \mathsf{ct}_{M'})$ for some integer M', and the token $\mathsf{t} = (\mathsf{ct}_x, \mathsf{ct}_y)$. Using $\mathsf{ORE.Compare}$, it performs a binary search to find the ciphertexts in st that are "at least" ct_x and "at most" ct_y. Let r be the set of ciphertexts lying in this interval. The server outputs the response r and an updated state $\mathsf{st}' = \mathsf{st}$.
  - $\mathsf{Client}(\mathsf{sk}, \mathsf{r}) \to \mathsf{S}$. The client, on input the secret key sk and the response $\mathsf{r} = (\mathsf{ct}_1, \ldots, \mathsf{ct}_m)$ for some integer m, outputs the tuple $\mathsf{S} = (\mathsf{ORE.Decrypt}(\mathsf{sk}, \mathsf{ct}_1), \ldots, \mathsf{ORE.Decrypt}(\mathsf{sk}, \mathsf{ct}_m))$. (Recall from Remark 2.1 that any ORE scheme can be augmented with a decryption algorithm.)
- $\mathsf{RQ.Insert}(\mathsf{sk}, \mathsf{q}, \mathsf{st}) \to (\mathsf{t}, \mathsf{st}')$. The insert algorithm between the client and server proceeds as follows:
  - $\mathsf{Client}(\mathsf{sk}, \mathsf{q} = x) \to \mathsf{t}$. The client, on input the secret key sk and a query representing an insertion of the value x, produces a token $\mathsf{t} = (\mathsf{ORE.Encrypt}_{\mathsf{L}}(\mathsf{sk}, x), \mathsf{ORE.Encrypt}_{\mathsf{R}}(\mathsf{sk}, x))$ which is sent to the server.
  - $\mathsf{Server}(\mathsf{st}, \mathsf{t}) \to (\mathsf{st}', \mathsf{r})$. The server takes as input its current state st and the token $\mathsf{t} = (\mathsf{ct}_1, \mathsf{ct}_2)$. Using $\mathsf{ORE.Compare}(\mathsf{ct}_1, \cdot)$, it performs a binary search over the contents of its database st to find the index at which to insert the new value. The server inserts ct_2 at that position and outputs the updated database st'.
- $\mathsf{RQ.Delete}(\mathsf{sk}, \mathsf{q}, \mathsf{st}) \to (\mathsf{t}, \mathsf{st}')$. The delete algorithm between the client and server proceeds as follows:
  - $\mathsf{Client}(\mathsf{sk}, \mathsf{q} = x) \to \mathsf{t}$. The client, on input the secret key sk and a query representing a deletion of the value x, produces a token $\mathsf{t} = (\mathsf{ORE.Encrypt}_{\mathsf{L}}(\mathsf{sk}, x), \mathsf{ORE.Encrypt}_{\mathsf{R}}(\mathsf{sk}, x))$ which is sent to the server.
  - $\mathsf{Server}(\mathsf{st}, \mathsf{t}) \to (\mathsf{st}', \mathsf{r})$. The server takes as input its current state st and the token $\mathsf{t} = (\mathsf{ct}_1, \mathsf{ct}_2)$. Using $\mathsf{ORE.Compare}(\mathsf{ct}_1, \cdot)$, it performs a binary search over the contents of its database st to find the indices of the elements in st equal to ct_1. It removes the entries at the matching indices and outputs the updated database st'.

Correctness. By correctness of the ORE scheme, the state st maintained by the server after each query is a (sorted) list of right encryptions (under sk) of the values in the database $\boldsymbol{D}$ after the corresponding insertions and deletions. Thus, the response r returned by the server to the client in a range query for the range $[x, y]$ is precisely the subset of ciphertexts whose plaintext values fall in the range $[x, y]$. Correctness follows by correctness of ORE decryption (which in turn follows from correctness of the ORE scheme).

Additional properties. In addition to the core security and correctness properties that we want from a symmetric range query scheme, we also note several useful properties that our construction Π_{rq} achieves for handling efficient range queries in our client-server model.

- **Stateless client and single-round protocols.** The client does not need to maintain state between queries, and each query is a single round trip between the client and the server. Our protocol achieves optimal round complexity.
- **Short query tokens**. The size of each query token t is asymptotically optimal. They are approximately the same length as the inputs used to generate the query, and *independent* of the size of the database.
- **Fast responses**. The running time of the server's algorithms is *sublinear* (logarithmic) in the total number of elements in the database.

Databases with multiple columns. In our description so far, we have modeled the database as containing a single column of values. To apply our methods to databases containing multiple columns, we would instead construct a sorted index for each column that needs to support range queries. Each of these indices are then encrypted with an independent ORE scheme. To support a range query over a particular column, the client would query the index for that column. The server responds with the set of (encrypted) record identifiers that fall within the requested range. The client then decrypts the response to obtain the record identifiers, and finally, retrieves the associated records.

Dual-encryption leakage functions. To define the security of our range query scheme, we first introduce a slight modification to the security notions achieved by our ORE scheme from Section 4. Recall from Definition 2.2 that an ORE scheme is secure with respect to a leakage function $\mathcal{L}(\cdot)$ if for any adversarially-chosen sequence of messages $m_1, \ldots, m_\ell$, there is an efficient simulator $\mathcal{S}$ that can simulate the real ORE ciphertexts given only the leakage $\mathcal{L}(m_1, \ldots, m_i)$.

Here, we consider "dual-encryption" leakage functions $\mathcal{L}'(\cdot, \cdot)$ which take two collections of plaintext values: one associated with "left" values, and the other associated with "right" values. Now, we say that an ORE scheme with separate left and right encryption functions $\mathsf{ORE.Encrypt_L}$ and $\mathsf{ORE.Encrypt_R}$, is secure with respect to the dual-encryption leakage function $\mathcal{L}'(\cdot, \cdot)$ if there exists an efficient simulator such that for any two (adversarially-chosen) collections of plaintexts $x_1, \ldots, x_\ell$ and $y_1, \ldots, y_\kappa$, and $\mathsf{sk} \leftarrow \mathsf{ORE.Setup}(1^\lambda)$, the simulator can simulate the outputs $\mathsf{ORE.Encrypt_L}(\mathsf{sk}, x_i)$ and $\mathsf{ORE.Encrypt_R}(\mathsf{sk}, y_j)$ for all $i \in [\ell]$ and $j \in [\kappa]$, given only the output of $\mathcal{L}'((x_1, \ldots, x_\ell), (y_1, \ldots, y_\kappa))$.

We note that the proof of Theorem 4.1 can be rewritten to prove security with respect to the dual-encryption leakage function $\mathcal{L}_{\text{DUAL}}$ as defined in the following lemma.

Lemma 5.3. *Let $\mathcal{L}_{\text{DUAL}}$ be the following dual-leakage function*

$$\mathcal{L}_{\text{DUAL}}((x_1, \ldots, x_\ell), (y_1, \ldots, y_\kappa)) = \left\{ \left(i, i', j, \text{BLK}(x_i, y_j), \mathsf{ind}^{(d)}_{\mathsf{diff}}(x_i, x_{i'}) \right) \mid i, i' \in [\ell], j \in [\kappa] \right\},$$

where $\text{BLK}(x_i, y_j) = (\text{CMP}(x_i, y_j), \mathsf{ind}^{(d)}_{\mathsf{diff}}(x_i, y_j))$ as defined in Section 4 and used in Theorem 4.1. The ORE scheme Π_{ore} from Section 4 is secure with respect to the dual-encryption leakage function $\mathcal{L}_{\text{DUAL}}$.

Proof. Follows by inspection of the proof of Theorem 4.1. □

Representing the leakage of our ORE scheme in terms of a dual-encryption leakage function allows us to easily reason about the online and offline security properties of our scheme. At a high level, the online leakage of our range query scheme is simply the output of the dual leakage function

on the sets of left ciphertexts appearing in the queries and the set of right ciphertexts appearing in the database. We now describe the leakage more precisely.

In our description below, we refer to the leakage function $\mathcal{L}^{(d)}_{\text{BLK}}(m_1, m_2)$ as the "ORE leakage" between two equal-length values m_1 and m_2. Informally, the "ORE leakage" in our setting is the ordering of m_1 and m_2 and the index of the first differing digit in the d-ary representation of m_1 and m_2. Our range query leakage function $\mathcal{L}_{\text{RQ}}$ then takes as input the database $\boldsymbol{D} = (d_1, \ldots, d_M)$, and a sequence of ℓ queries $\mathsf{q}_1, \ldots, \mathsf{q}_\ell$ and outputs:

- For each $i \in [M]$ and $j \in [\ell]$, the ORE leakage between each database value d_i and query q_j. For a range query of the form $\mathsf{q} = (x, y)$, this includes the ORE leakage between both pairs (d_i, x) and (d_i, y) for $i \in [M]$.
- For each query q_i, and each insertion or deletion query q'_j, the ORE leakage between q_i and q'_j. Similarly, for a range query of the form $\mathsf{q}_i = (x_i, y_i)$, this include the ORE leakage between both pairs (x_i, q'_j) and (y_i, q'_j).

Roughly speaking, our range query scheme reveals the ordering and the index of the first differing digit between every query and every message in the database. We also leak some information between range queries and insertion/deletion queries. We now formalize our security claims.

Online security. For a database $\boldsymbol{D} \in [N]^M$ and sequence of ℓ queries $\mathsf{q}_1, \ldots, \mathsf{q}_\ell$, let $\mathcal{R}, \mathcal{I}, \mathcal{D}$ denote the sequence of values appearing in the range queries, the insert queries, and the delete queries, respectively. Note that the two values in each range query are expanded as separate elements in $\mathcal{R}$. Finally, let $\mathcal{Q} = \mathcal{R} \| \mathcal{I} \| \mathcal{D}$.

Theorem 5.4. *Let $\mathcal{L}_{\text{RQ}}$ be the following leakage function:*

$$\mathcal{L}_{\text{RQ}}(\boldsymbol{D}, \mathsf{q}_1, \ldots, \mathsf{q}_\ell) = (\mathcal{L}_{\text{DUAL}}(\mathcal{Q}, \boldsymbol{D}), \mathcal{L}_{\text{DUAL}}(\mathcal{Q}, \mathcal{I} \| \mathcal{D}))$$

Then, the range query scheme Π_{rq} achieves online security with respect to the leakage function $\mathcal{L}_{\text{RQ}}$.

Proof. The proof follows immediately from observing that in Π_{rq}, the values that are encrypted using the left encryption algorithm are the values appearing in the queries $\mathcal{Q}$, and the values encrypted using the right encryption algorithm are the database elements, along with the respective components appearing in the insertion and deletion queries. Hence, we can directly invoke Lemma 5.3, which proves the theorem. □

Offline security. Offline security (Definition 5.2) of our range query scheme Π_{rq} follows directly from the fact that the encrypted database stored on the server only contains a collection of right ciphertexts, which are simulatable given just the size of the collection (that is, the right ciphertexts are *semantically secure* encryptions of their values).

Theorem 5.5. *The range query scheme Π_{rq} is offline secure.*

Proof. The contents of the server's state after each query in the range query protocol Π_{rq} is always a collection of ORE right ciphertexts. Hence, for any sequence of states $\mathsf{st}_0, \ldots, \mathsf{st}_\ell$ induced by a database $\boldsymbol{D}$ and any sequence of queries $\mathsf{q}_1, \ldots, \mathsf{q}_\ell$, we just need to invoke the simulator (for constructing right ciphertexts) in the proof of Theorem 4.1 a total of $|\mathsf{st}_\ell|$ times to simulate the right encryptions in st_ℓ. This completes the proof. □

Robustness against offline inference attacks. Offline security for our protocol implies that the contents of the server's database are *always* semantically secure. Consequently, ciphertext-only inference attacks, such as those studied by Naveed et al. [NKW15], do not directly apply.

In their model [NKW15, §4.2], an attacker is able to obtain access to the "steady state" of an encrypted database, which describes the database in a state that includes all auxiliary information that is needed to perform encrypted searches efficiently. In our scheme, no such auxiliary information is needed on top of the ORE scheme, and yet we are still able to achieve offline security. In contrast, in other existing PPE-based schemes, comparisons are enabled by a underlying layer of OPE encryption, which is vulnerable to inference attacks. Thus, even though these schemes can be modified to satisfy our notion of offline security, their "steady-state" representation is in the form of OPE ciphertexts which are vulnerable to inference attacks. Our scheme achieves robustness against these ciphertext-only inference attacks because our steady-state representation is precisely our offline representation. Finally, we note that we can always add additional layers of encryption (e.g., onion encryption [PRZB11]) without compromising the security of our range query scheme, which can serve as a useful countermeasure against general adversaries.

Existing schemes and the left/right framework. The key ingredient in our work that enables us to construct an efficient, inference-robust range query protocol is the fact that ciphertexts in our scheme naturally split into left and right components such that the right components, when taken in isolation, are semantically secure. To our knowledge, our scheme is the first practical ORE scheme where the ciphertexts split naturally into left and right components such that one side is semantically secure.[7] In contrast, no OPE scheme can satisfy this property—this is due to the restriction that the comparison operation must be a numeric comparison on the ciphertexts. Since comparisons are transitive, this means that if comparisons are possible between left and right ciphertexts, they are necessarily possible between left ciphertexts or right ciphertexts in isolation. Thus, neither side can be semantically secure.

Ciphertexts in the Chenette et al. [CLWW16] ORE scheme also do not decompose naturally into left and right ciphertexts where one side is semantically secure. In fact, ciphertexts in their scheme are deterministic, and thus, cannot provide semantic security. We note though that the semantically-secure ORE constructions from multilinear maps [BLR$^+$15] or indistinguishability obfuscation [GGG$^+$14] are naturally defined in the left/right framework (specifically, the encryption function in these constructions also take an "input slot," which directly corresponds to our notions of left and right). Thus, these ORE constructions can also be leveraged to obtain a range query scheme with sublinear query complexity and robustness against offline inference attacks. Due to their reliance on extremely powerful tools, however, they are very far from being practically viable.

6 Impossibility Result for OPE

Our ORE construction from Section 4 uses a small-domain ORE scheme with best-possible security as a core building block. A natural question to ask then is whether we could have applied the same kind of transformation starting from a small-domain OPE scheme with best-possible security. While Boldyreva et al. [BCLO09, BCO11] and Popa et al. [PLZ13] have previously ruled out the existence

[7] Concurrent with the publication of this work, Joye and Passelégue [JP16] and Cash et al. [CLOZ16] also independently proposed practical ORE constructions (with leakage) based on bilinear groups where the ciphertexts naturally decompose into left and right components.

of such OPE schemes over a superpolynomial size message space, their lower bounds do not rule out the possibility of an OPE scheme over a polynomial-size domain that achieves best-possible security.

In this section, we show that even this is impossible. In particular, no OPE scheme whose plaintext space contains just three messages can satisfy the "best-possible" notion of security (IND-OCPA) unless the length of the ciphertexts is *superpolynomial* in the security parameter. In other words, the size of the ciphertext space for any such OPE scheme is at least $2^{2^{\omega(\log \lambda)}}$. We then show that our lower bound is tight by giving a construction of an IND-OCPA-secure OPE scheme with plaintext space $\{1, 2, 3\}$ and ciphertext space $[M]$ where $M = 2^{2^{\omega(\log \lambda)}}$. Our results thus show that there does not exist any efficient stateless, non-interactive OPE scheme that satisfies IND-OCPA security, even for small message spaces.

First, recall that an order-preserving encryption scheme [BCLO09, BCO11] is a special case of ORE where the ciphertext space is required to be a well-ordered range $\mathcal{R}$. Moreover, given two ciphertexts $\mathsf{ct}_1, \mathsf{ct}_2 \in \mathcal{R}$, the comparison algorithm outputs 1 if $\mathsf{ct}_1 < \mathsf{ct}_2$. In other words, an OPE scheme is an ORE scheme where the comparison function is the "natural" comparison operation on the ciphertext space. Formally we can specify an OPE scheme by a tuple of algorithms $\Pi_{\mathsf{OPE}} = (\mathsf{OPE.Setup}, \mathsf{OPE.Encrypt})$. We first review the formal definition of IND-OCPA security from [BCLO09].

Definition 6.1 (IND-OCPA Security [BCLO09])**.** Let $\Pi_{\mathsf{OPE}} = (\mathsf{OPE.Setup}, \mathsf{OPE.Encrypt})$ be an OPE scheme. Then, Π_{OPE} is IND-OCPA secure if for all efficient and admissible adversaries $\mathcal{A}$ and $\mathsf{sk} \leftarrow \mathsf{OPE.Setup}(1^\lambda)$,

$$\left| \Pr\left[b \xleftarrow{R} \{0,1\} : \mathcal{A}^{\mathbf{LoR}(\mathsf{sk}, b, \cdot, \cdot)}(1^\lambda) = b \right] - \frac{1}{2} \right| = \mathrm{negl}(\lambda),$$

where $\mathbf{LoR}(\mathsf{sk}, b, m_0, m_1)$ is the left-or-right encryption oracle which on input a key sk, a bit b, and two messages m_0, m_1, returns $\mathsf{OPE.Encrypt}(\mathsf{sk}, m_b)$. We say that an adversary $\mathcal{A}$ making q queries $(m_0^{(1)}, m_1^{(1)}), \ldots, (m_0^{(q)}, m_1^{(q)})$ to the $\mathbf{LoR}$ oracle is admissible if for all $i, j \in [q]$, $m_0^{(i)} < m_0^{(j)}$ if and only if $m_1^{(i)} < m_1^{(j)}$.

Lower bound for OPE schemes. We first show that any stateless OPE scheme with a plaintext space containing at least three messages cannot satisfy IND-OCPA security unless the ciphertext space has size $2^{2^{\omega(\log \lambda)}}$. In other words, the number of bits needed to represent a ciphertext is $2^{\omega(\log \lambda)}$, which is superpolynomial in the security parameter. This theorem effectively shows that there are *no* efficient OPE schemes when the message space contains even 3 elements.

Theorem 6.2. *Let Π_{OPE} be a stateless OPE scheme with plaintext space $[N]$ and ciphertext space $[M]$. If Π_{OPE} is IND-OCPA-secure and $N \geq 3$, then $M = 2^{2^{\omega(\log \lambda)}}$.*

Proof. By correctness of Π_{OPE}, the $\mathsf{OPE.Encrypt}(\mathsf{sk}, \cdot)$ function is deterministic with overwhelming probability over the randomness used to sample sk in $\mathsf{OPE.Setup}$. Thus, without loss of generality, we assume that $\mathsf{OPE.Encrypt}(\mathsf{sk}, \cdot)$ is *deterministic*. Since $N \geq 3$, define the random variable $y_i = \mathsf{OPE.Encrypt}(\mathsf{sk}, i)$ for $i \in [3]$, and let $\mathcal{D}_i$ be the distribution of y_i (taken over the randomness used to sample sk). For $1 \leq i < j \leq 3$, define random variables $d_{ij} = y_j - y_i$ to be random variables corresponding to the distance between ciphertexts. By definition, $d_{13} = d_{12} + d_{23}$. Let $\mathcal{D}_{ij}$ be the distribution of d_{ij}. By construction, each $\mathcal{D}_{ij}$ is a distribution over $[M]$. If Π_{OPE} is IND-OCPA secure, then it must be the case that $\mathcal{D}_{12} \overset{c}{\approx} \mathcal{D}_{23} \overset{c}{\approx} \mathcal{D}_{13}$. To complete the proof, we show the following two lemmas.

Lemma 6.3. *Suppose Π_{OPE} is IND-OCPA secure. Then, for any $M' \leq M$ where $\Pr[d_{12} \leq M'] = 1 - \mathrm{negl}(\lambda)$, it follows that $M' = 2^{\omega(\log \lambda)}$ (that is, M' is superpolynomial in λ).*

Proof. We proceed via contradiction. Suppose that Π_{OPE} is IND-OCPA-secure and that $M' = \mathrm{poly}(\lambda)$. Then, there must exist some $x \in [M']$ such that $\Pr[d_{12} = x] \geq 1/M' - \mathrm{negl}(\lambda)$, which is non-negligible. Let $x \in [M']$ be the smallest such x such that $\Pr[d_{12} = x]$ is non-negligible. Next, using the fact that $d_{13} = d_{12} + d_{23}$, we have

$$\begin{aligned}\Pr[d_{13} = x] &= \Pr[d_{12} = x] \cdot \Pr[d_{23} = 0 \mid d_{12} = x] + \sum_{z<x} \left(\Pr[d_{12} = z] \cdot \Pr[d_{23} = x - z \mid d_{12} = z]\right) \\ &= \mathrm{negl}(\lambda) + \mathrm{negl}(\lambda),\end{aligned}$$

where the first term is negligible since $\Pr[d_{23} = 0 \mid d_{12} = x] = \mathrm{negl}(\lambda)$ by correctness of the scheme, and the second term is negligible since x is the smallest value for which $\Pr[d_{12} = x]$ is non-negligible. By assumption, $\mathcal{D}_{13} \overset{c}{\approx} \mathcal{D}_{12}$, so it must be the case that $\Pr[d_{12} = x] \leq \Pr[d_{13} = x] + \mathrm{negl}(\lambda) = \mathrm{negl}(\lambda)$, which contradicts the assumption that $\Pr[d_{12} = x]$ is non-negligible. □

Lemma 6.4. *Let $M' \leq M$ be such that $\Pr[d_{12} \leq M'] = 1 - \mathrm{negl}(\lambda)$. Then, $\Pr[d_{12} > M'/2] = \mathrm{negl}(\lambda)$.*

Proof. Suppose by contradiction that $\Pr[d_{12} > M'/2] = \varepsilon_1$ for some non-negligible ε_1. By the law of total probability,

$$\begin{aligned}\Pr[d_{23} \leq M'/2] &= \Pr[d_{23} \leq M'/2 \mid d_{12} \leq M'/2] \cdot \Pr[d_{12} \leq M'/2] \\ &\quad + \Pr[d_{23} \leq M'/2 \mid d_{12} > M'/2] \cdot \Pr[d_{12} > M'/2] \\ &= \overbrace{(1 - \varepsilon_1) \Pr[d_{23} \leq M'/2 \mid d_{12} \leq M'/2]}^{\text{at least } 0} + \varepsilon_1 \Pr[d_{23} \leq M'/2 \mid d_{12} > M'/2] \qquad (6.1)\end{aligned}$$

By assumption $\Pr[d_{12} > M'] = \mathrm{negl}(\lambda)$, and since $\mathcal{D}_{13} \overset{c}{\approx} \mathcal{D}_{12}$, it follows that $\Pr[d_{13} > M'] = \mathrm{negl}(\lambda)$. Since $d_{13} > M'$ with negligible probability, and $d_{12} > M'/2$ with non-negligible probability ε_1, and $d_{13} = d_{12} + d_{23}$, it must be the case that $\Pr[d_{23} \leq M'/2 \mid d_{12} > M'/2] = 1 - \mathrm{negl}(\lambda)$. We conclude from Eq. (6.1) that

$$\begin{aligned}\Pr[d_{23} \leq M'/2] &\geq \varepsilon_1 \Pr[d_{23} \leq M'/2 \mid d_{12} > M'/2] \\ &= \varepsilon_1 - \mathrm{negl}(\lambda).\end{aligned}$$

Next, we use the fact that $d_{13} \leq M'/2$ only if $d_{12} \leq M'/2$ and $d_{23} \leq M'/2$. Let $\varepsilon_2 = \Pr[d_{13} \leq M'/2]$. Then,

$$\begin{aligned}\varepsilon_2 &\leq \Pr[d_{12} \leq M'/2] \cdot \Pr[d_{23} \leq M'/2 \mid d_{12} \leq M'/2] \\ &= (1 - \varepsilon_1) \cdot \Pr[d_{23} \leq M'/2 \mid d_{12} \leq M'/2]. \qquad (6.2)\end{aligned}$$

Substituting Eq. (6.2) into Eq. (6.1), we have that

$$\begin{aligned}\Pr[d_{23} \leq M'/2] &= \overbrace{(1 - \varepsilon_1) \Pr[d_{23} \leq M'/2 \mid d_{12} \leq M'/2]}^{\text{at least } \varepsilon_2} + \overbrace{\varepsilon_1 \Pr[d_{23} \leq M'/2 \mid d_{12} > M'/2]}^{\text{equal to } \varepsilon_1 - \mathrm{negl}(\lambda)} \\ &\geq \varepsilon_1 + \varepsilon_2 - \mathrm{negl}(\lambda).\end{aligned}$$

Again using the fact that $\mathcal{D}_{13} \overset{c}{\approx} \mathcal{D}_{23}$, we conclude that $\Pr[d_{13} \leq M'/2] \geq \varepsilon_1 + \varepsilon_2 - \mathrm{negl}(\lambda)$. By definition, $\varepsilon_2 = \Pr[d_{13} \leq M'/2]$, so we obtain the relation $\varepsilon_2 \geq \varepsilon_1 + \varepsilon_2 - \mathrm{negl}(\lambda)$. By assumption, ε_1 is non-negligible, so this is impossible. The claim follows. □

The theorem now follows by a straightforward invocation of Lemma 6.3 and 6.4. Let Π_{OPE} be an IND-OCPA secure OPE scheme with ciphertext space $[M]$ where $M = 2^{\lambda^c}$ for some $c \in \mathbb{N}$. In other words, $\log M = \lambda^c = \mathrm{poly}(\lambda)$. Define $M_0 = M$ and for $i \in [\lambda^c]$, set $M_i = M/2^i$. By assumption, $\Pr[d_{12} \leq M] = 1$, so invoking Lemma 6.4, $\Pr[d_{12} > M/2] = \mathrm{negl}(\lambda)$. This means that $\Pr[d_{12} \leq M_1] = 1 - \mathrm{negl}(\lambda)$. We can now inductively apply Lemma 6.4 (a polynomial number of times) to conclude that $\Pr[d_{12} \leq M_{\lambda^c}] = 1 - \mathrm{negl}(\lambda)$. However, if Π_{OPE} is IND-OCPA secure, then invoking Lemma 6.3, we require that $M_{\lambda^c} = 2^{\omega(\log \lambda)}$. But $M_{\lambda^c} = M/2^{\lambda^c} = O(1)$, so this is impossible. The claim follows. □

Upper bound for OPE schemes. We now give an explicit construction of a stateless OPE scheme for a 3-message plaintext space that achieves best-possible security and whose ciphertext space has size $2^{2^{\omega(\log \lambda)}}$. This matches the lower bound from Theorem 6.2.

Take any function $f(\cdot)$ where $f(\lambda) = \omega(\log \lambda)$, and set $M = 2^{2^{f(\lambda)+1}}$. The ciphertext space in our scheme will be $[M]$. Let $\Pi_{\mathsf{OPE}} = (\mathsf{OPE.Setup}, \mathsf{OPE.Encrypt})$ be an OPE scheme with plaintext space $\{1, 2, 3\}$, where the algorithms are given as follows:

- $\mathsf{OPE.Setup}(1^\lambda)$: On input the security parameter λ, the setup algorithm chooses a value $z \xleftarrow{R} [M]$, and $\delta \leftarrow [2^{f(\lambda)}]$. It outputs the secret key $\mathsf{sk} = (z, \delta)$.
- $\mathsf{OPE.Encrypt}(\mathsf{sk}, x)$: On input a secret key $\mathsf{sk} = (z, \Delta)$ and a message $x \in \{1, 2, 3\}$, the encryption algorithm writes x as $x = 2 + i$ for some $i \in \{-1, 0, 1\}$, and computes $y = z + i \cdot 2^\delta$. The algorithm outputs y if $y \in [M]$, and $\perp$ otherwise.

Correctness. Since the offset 2^δ is always positive, it suffices to argue that $\mathsf{OPE.Encrypt}(\mathsf{sk}, x)$ does not output $\perp$ with overwhelming probability. By construction, $2^\delta \leq 2^{2^{f(\lambda)}}$. Therefore, the quantity $z - 2^\delta$ is less than 1 only in cases where $z \leq 2^{2^{f(\lambda)}}$. But this happens with probability $2^{2^{f(\lambda)}}/2^{2^{f(\lambda)+1}} = 1/2^{2^{f(\lambda)}} = \mathrm{negl}(\lambda)$. Similarly, $z + 2^\delta$ is greater than $2^{2^{f(\lambda)+1}}$ only in cases where $z \geq 2^{2^{f(\lambda)+1}} - 2^{2^{f(\lambda)}}$, which again happens with probability $1/2^{2^{f(\lambda)}} = \mathrm{negl}(\lambda)$. Thus, correctness holds with overwhelming probability. □

Security. For $i \in \{1, 2, 3\}$, define the random variable $y_i = \mathsf{OPE.Encrypt}(\mathsf{sk}, i)$, and let $\mathcal{D}_i$ be the distribution of y_i taken over the randomness used to sample sk. For $1 \leq i < j \leq 3$, let $\mathcal{D}_{ij}$ be the distribution of $y_j - y_i$. First, we argue that for all $i \in \{1, 2, 3\}$, $\mathcal{D}_i \overset{s}{\approx} \mathsf{Unif}([M])$. By construction, $\mathcal{D}_2 \equiv \mathsf{Unif}([M])$. To show $\mathcal{D}_1 \overset{s}{\approx} \mathsf{Unif}([M])$, we examine the quantity $\Pr[y_1 = t]$ for $t \in [M]$. Since δ is uniform over $2^{f(\lambda)}$, we have that

$$\begin{aligned}\Pr[y_1 = t] &= \frac{1}{2^{f(\lambda)}} \sum_{\delta' \in [2^{f(\lambda)}]} \Pr[y_1 = t \mid \delta = \delta'] \\ &= \frac{1}{2^{f(\lambda)}} \sum_{\delta' \in [2^{f(\lambda)}]} \Pr[z = t + 2^{\delta'}].\end{aligned}$$

If $t + 2^{\delta'} \leq M$, then by the fact that z is uniform over $[M]$, $\Pr[z = t + 2^{\delta'}] = 1/M$. Thus, for all $t \leq M - 2^{2^{f(\lambda)}}$, $\Pr[y_1 = t] = 1/M$. More generally, we have for all $t \in [M]$, $\Pr[y_1 = t] \leq 1/M$. The statistical distance between $\mathcal{D}_1$ and $\mathsf{Unif}([M])$ can then be bounded as follows:

$$\begin{aligned}\sum_{t \in [M]} \left| \Pr[y_1 = t] - \frac{1}{M} \right| &= \sum_{M - 2^{2^{f(\lambda)}} < t \leq M} \left| \Pr[y_1 = t] - \frac{1}{M} \right| \\ &\leq \frac{2^{2^{f(\lambda)}}}{M} + \frac{2^{2^{f(\lambda)}}}{M} = \mathrm{negl}(\lambda),\end{aligned}$$

where we used the triangle inequality in the second line. A similar argument shows that $\mathcal{D}_3 \stackrel{s}{\approx} \mathsf{Unif}([M])$.

To conclude the proof, we argue that for $1 \leq i < j \leq 3$, $\mathcal{D}_{ij} \stackrel{s}{\approx} \mathsf{Unif}(S_\lambda)$, where $S_\lambda = \left\{2^1, 2^2, \ldots, 2^{2^{f(\lambda)}}\right\}$. By construction, $\mathcal{D}_{12} \equiv \mathsf{Unif}(S_\lambda) \equiv \mathcal{D}_{23}$, so it suffices to consider $\mathcal{D}_{13}$. By construction, $\mathcal{D}_{13} \equiv \mathsf{Unif}(S'_\lambda)$ where $S'_\lambda = \left\{2^2, 2^3, \ldots, 2^{2^{f(\lambda)}+1}\right\}$. Since $|S_\lambda| = 2^{f(\lambda)} = |S'_\lambda|$, the statistical distance between $\mathsf{Unif}(S_\lambda)$ and $\mathsf{Unif}(S'_\lambda)$ is $2/2^{f(\lambda)} = \mathrm{negl}(\lambda)$. □

7 Experimental Evaluation

To assess the practicality of our order-revealing encryption scheme from Section 4, we give a full implementation of our scheme and measure its performance on a wide range of parameter settings. We then compare the performance against the Boldyreva et al. [BCLO09] OPE scheme and the Chenette et al. [CLWW16] ORE scheme. In our implementation, we use the technique from Remark 3.1 to shrink the ciphertexts.

Instantiating primitives. Our implementation is entirely written in C. We operate at 128-bits of security ($\lambda = 128$). We instantiate the PRF with AES-128. To construct a PRP on $2d$-bit domains (for $d < 128$), we use a 3-round Feistel network using a PRF on d-bit inputs [LR88].[8] In our experiments, we only consider $d < 128$, and thus, can instantiate the PRF using AES (where the d-bit input is padded to 128-bits). For the random oracle, we consider two candidate constructions. In the first, we use SHA-256, a standard cryptographic hash function commonly modeled as a random oracle.

For our second instantiation of the random oracle, we use an AES-based construction. This allows us to leverage the AES-NI instruction set for hardware-accelerated evaluation of AES. Recall from Section 4 that our construction requires a random oracle mapping from a domain $\{0,1\}^{2\lambda} = \{0,1\}^{256}$ to $\mathbb{Z}_2$ (after applying the modification from Remark 3.1). On an input $(k, x) \in \{0,1\}^{128} \times \{0,1\}^{128}$, we take the output of the random oracle to be the least significant bit of $\mathrm{AES}(k, x)$. Certainly, if we model AES as an ideal cipher, then this construction implements a random oracle. We note that modeling AES as an idealized object such as a random permutation or an ideal cipher has been used in many other recent works such as constructing efficient garbling schemes [BHKR13] or the Simpira family of permutations [GM16].

[8] The security of a Feistel-based PRP is (roughly) proportional to the block size. When the block size d is small, the Feistel construction does not provide the desired level of security. We refer to [BKR18] for an updated instantiation of our ORE scheme with a secure implementation of the small-domain random permutation based on the Knuth shuffle [Knu98].

In our implementation, we use the OpenSSL [The03] implementations of AES and SHA-256 as well as the GMP [Gt12] library for big integer arithmetic. Our full implementation contains approximately 750 lines of code. For our implementation of Boldyreva et al.'s OPE scheme, we use the C++ implementation from CryptDB [PRZB11],[9] and for our implementation of Chenette et al.'s ORE scheme, we use the C implementation FastORE.[10] In our benchmarks, we substitute AES for HMAC as the underlying PRF used in the FastORE library. We believe this provides a more balanced comparison of the performance tradeoffs between the Chenette et al. scheme and our new ORE scheme.

Benchmarks and evaluation. We run all of our experiments on a laptop running Ubuntu 14.04 with a 2.3 GHz Intel Core i7 CPU (Haswell microarchitecture) and 16 GB of RAM. Although our encryption algorithm is easily parallelizable, we do not leverage parallelism in our benchmarks. The processor supports the AES-NI instruction set, hence our decision to base as many primitives as possible on AES. Our micro-benchmarks for encrypting and comparing 32-bit integers are summarized in Table 1. In Figure 1, we compare the cost of encryption for the different schemes across different-sized message spaces.

From Table 1, the time needed to compare two ORE ciphertexts is similar to the time needed to compare two integers (in the OPE setting). Thus, while it is the case that deploying ORE in encrypted database systems would require implementing a custom comparator in the database management system, in practice, this incurs a very small computational overhead.

Compared to OPE, our new ORE scheme is significantly faster. For instance, when processing byte-size blocks, encrypting a single 32-bit value requires just over 50 μs of computation and is over 65 times faster compared to vanilla OPE. Even our SHA-256-based implementation is about 10x faster compared to OPE. Moreover, as shown in [CLWW16, Remark 2.6 and §4], an ORE scheme which leaks the first bit that differs between two encrypted messages is provably more secure than any OPE scheme which behaves like a truly random order-preserving function. Since our new ORE scheme leaks strictly less information than the Chenette et al. scheme, we conclude that our new ORE scheme is both more secure and faster compared to OPE schemes. Of course, when compared to the bit-by-bit construction of [CLWW16], our new ORE scheme is much slower. However, in exchange, our new ORE scheme confers stronger security as well as lends itself nicely towards a range query system that provides robustness against inference attacks.

One of the main limitations of our new ORE scheme is the increase in the ciphertext size. Both OPE and the Chenette et al. ORE schemes are able to achieve ciphertexts where the overhead is an additive or (small) multiplicative factor in the length of the messages. In our setting, because our main construction relies critically on a small-domain ORE scheme that offers best-possible security, and the existing small-domain ORE scheme have ciphertexts that grow linearly in the size of the message space, the size of the ciphertexts in our composed scheme grows quickly in the block size. Nonetheless, when encrypting byte-by-byte, encrypting a 32-bit integer requires just 224 bytes, which is quite modest for many practical applications. An interesting direction for future work is to construct a more compact small-domain ORE with best-possible security. Such a construction can be extended to a large-domain ORE with shorter ciphertexts by applying our techniques from Section 4.

[9] `https://github.com/CryptDB/cryptdb`
[10] `https://github.com/kevinlewi/fastore`

Scheme	d	Encrypt	Compare	$\lvert\mathsf{ct}\rvert$	Leakage
Boldyreva et al. OPE [BCLO09]	–	3601.82 μs	0.36 μs	8 bytes	(Hard to quantify)
Chenette et al. ORE [CLWW16]	1	2.06 μs	0.48 μs	8 bytes	First bit that differs
Our ORE scheme (RO: SHA-256)	4	54.48 μs	0.38 μs	192 bytes	First block of d-bits that differs
	8	361.04 μs	0.98 μs	224 bytes	
	12	4370.64 μs	3.20 μs	1612 bytes	
Our ORE scheme (RO: AES)	4	16.50 μs	0.31 μs	192 bytes	First block of d-bits that differs
	8	54.87 μs	0.63 μs	224 bytes	
	12	721.37 μs	2.61 μs	1612 bytes	

Table 1: Performance comparison between our ORE scheme from Section 4 and existing OPE and ORE schemes. We consider two variants of our scheme: one where the random oracle is instantiated using an AES-based construction and one where the random oracle is instantiated with SHA-256. We describe these two instantiations in greater detail in Section 7. In these benchmarks, we use a 32-bit plaintext space, and measure the time needed to encrypt a (randomly chosen) message and the time needed to compare two ciphertexts. The parameter d is the block size (in bits) in our ORE scheme. Our micro-benchmarks are averaged over 50–10^7 iterations (the precise number is adjusted based on the approximate runtime of the algorithm).

8 Related Work

In this section, we survey some of the literature on order-revealing and order-preserving encryption, as well as the existing work on searching over encrypted data.

OPE and ORE. The concept of order-preserving encryption was first introduced by Agrawal, Kiernan, Srikant, and Xu [AKSX04], who explored the application of OPE for performing encrypted database queries. The first explicit OPE construction was formalized in the seminal work of Boldyreva et al. [BCLO09], and has subsequently been expanded on in a multitude of works [BCO11, PR12, PLZ13, TYM14, KS14, Ker15, MCO$^+$15, RACY15, BPP16]. Some of these works [BCO11, TYM14] have focused on exploring the security properties of order-preserving encryption. Others [PLZ13, KS14, Ker15, RACY15, BPP16] have considered stateful or interactive OPE solutions which avoid both the lower bounds in [BCLO09, BCO11, PLZ13] as well as our strengthened lower bound from Section 6. However, synchronizing state and coordinating multi-round interactions in distributed, large-scale execution environments is often difficult, and consequently, nearly all existing OPE deployments (e.g., SkyHigh Networks, CipherCloud) use stateless variants of OPE for sorting and filtering on encrypted data. Numerous ad hoc OPE schemes [BHF09, KAK10] have also been proposed in recent years, but they often lack a formal security analysis.

The notion of order-revealing encryption was first introduced by Boldyreva et al. [BCO11] under the name "efficiently-orderable encryption" (EOE). Subsequently, Boneh et al. [BLR$^+$15] gave the first construction of an order-revealing encryption scheme satisfying the best-possible notion of semantic security from multilinear maps. More generally, ORE is a special case of multi-input functional encryption (MIFE) [GGG$^+$14]. To date, the only constructions of general-purpose MIFE rely on heavy primitives such as indistinguishability obfuscation [BGI$^+$12, GGH$^+$13b] and are far too inefficient to deploy. Chenette et al. [CLWW16] recently proposed an efficient ORE scheme, which

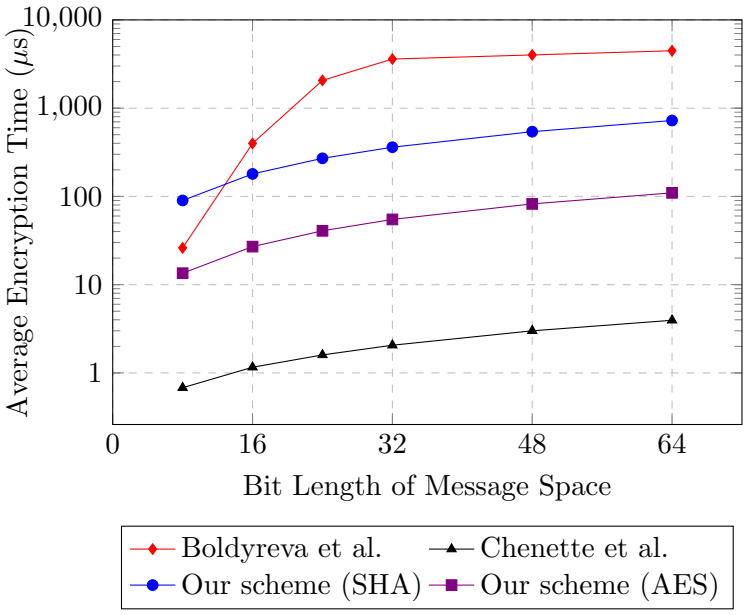

Figure 1: Performance comparison between our ORE scheme (Section 4) and existing OPE and ORE schemes. We use a fixed base representation $d = 8$ for our ORE scheme in these experiments. The two variants of our scheme, labeled SHA and AES, refer to how we instantiate the random oracle in our construction.

we improve upon and generalize in our work. In the small-domain setting, it is possible to construct ORE from either symmetric or public-key encryption [AJ15, BKS15] or bilinear maps [KLM$^+$16], but these constructions are far less efficient compared to our small-domain ORE from Section 3, which just relies on PRFs.

Concurrent work on ORE. Concurrent to this work, Joye and Passelégue [JP16] also construct a small-domain ORE scheme with best-possible security from one-way functions. Ciphertexts in their scheme are asymptotically longer that those in our scheme (by a multiplicative factor $\Omega(\log N)$ where N is the size of the plaintext space), but at the same time, their construction achieves simulation security without random oracles. In addition, they construct an ORE scheme (with leakage) from the decisional linear (DLIN) assumption over bilinear groups. The leakage of their scheme is incomparable to that of our new PRF-based ORE scheme from Section 4.

In another concurrent work, Cash, Liu, O'Neill, and Zhang [CLOZ16] give a pairing-based construction of an order-revealing encryption scheme that achieves strictly stronger security than the Chenette et al. construction. Their construction combines the bit-by-bit (or block-by-block) encryption approach of Chenette et al. with a novel property-preserving hash function to perform the comparisons. Their use of the property-preserving hash function enables hiding of the position of the first differing bit (or block). The resulting leakage of their scheme is whether two ciphertexts differ from a third ciphertext in the same bit (or block) position. Cash et al. instantiate the property-preserving hash function using the symmetric external Diffie-Hellman (SXDH) assumption over bilinear groups. While the leakage of their construction strictly improves upon both the work of Chenette et al. as well as this work (when looking at their blockwise generalization), ciphertexts

in their construction are significantly longer (by a multiplicative factor that grows linearly in the security parameter λ). The reliance on pairings also render their scheme less competitive in terms of practical efficiency compared to solutions that only require simple symmetric primitives. Nonetheless, the Joye-Passelégue and the Cash-Liu-O'Neill-Zhang constructions both represent important advancements in the study of efficient ORE constructions with limited leakage.

Searching on encrypted data. Numerous techniques, such as searchable symmetric encryption (SSE) [SWP00, CGKO06, CK10], property-preserving encryption (PPE) [BCLO09, PR12, CD15], fully homomorphic encryption (FHE) [Gen09], hidden vector encryption [BW07], oblivious RAMs (ORAM) [GO96], and others have been proposed for tackling the general problem of searching and querying on encrypted data. While tools such as FHE or ORAM can be used for searching on encrypted data [BGH$^+$13, YSK$^+$13], these methods are prohibitively expensive for nearly all real-world deployments. On the more practical side, numerous SSE schemes [SWP00, Goh03, CM05, CGKO06, CK10, JJK$^+$13, NPG14] have been proposed in the last 15 years, but these past works are limited to exact keyword searches, and generally do not handle the efficient computation of complex queries (such as *range queries*) over encrypted data. More recently, several works [CJJ$^+$13, CJJ$^+$14, PKV$^+$14, FJK$^+$15] describe constructions of SSE schemes that are able to handle more expressive queries. We survey these works below.

Cash et al. [CJJ$^+$13] give the first SSE scheme that supports Boolean queries (in time sublinear in the size of the database) with a small amount of leakage and security from the decisional Diffie-Hellman (DDH) assumption. Subsequently, Cash et al. [CJJ$^+$14] extend the construction to allow for updates to the encrypted database as well as support multiple, potentially dishonest clients. Handling updates requires the client to maintain a small amount of state (or requires additional rounds of communication and leads to increased leakage). Boolean queries alone, however, do not suffice for range queries, so in another follow-up work, Faber et al. [FJK$^+$15] show how the Cash et al. SSE scheme can be leveraged for range queries. Their resulting construction leaks some additional information about the database contents, namely the number of values that fall into certain subintervals within the requested range. Moreover, due to the use of universal covers, the size of the server's response set to a range query may be up to 66% larger than the size of the true response set. We do not know of any existing SSE scheme that can efficiently support range queries with optimal (minimal) leakage.

Concurrent to the work of Cash et al., Pappas et al. [PKV$^+$14] introduce BlindSeer, a private database management system that can support a wide-range of queries in sublinear time over an encrypted database. Their construction leverages generic two-party computation tools such as Yao's garbled circuits [Yao82], and their construction provides security in the semi-honest model.

Comparison to our techniques. To conclude, we highlight some of the key differences between existing SSE methods and our ORE-based construction for implementing range queries over an encrypted database:

- Like other PPE-based constructions, our ORE-based construction integrates well with *existing* database management systems—we just need to implement a custom comparator. With SSE, we would have to deploy a new, and oftentimes, complex database management system. This lacks legacy compatibility, which is a barrier to deployment in existing systems. Our approach provides a fast, simple, and direct solution for supporting range queries on encrypted data *without* requiring significant infrastructural changes.

- We explicitly model and analyze the leakage of our range query protocol assuming adaptive updates to the database.
- Our construction only requires symmetric primitives and does not require more expensive primitives such as public-key cryptography or oblivious transfer.

9 Conclusions

In this work, we gave two new constructions of order-revealing encryption schemes that rely only on symmetric primitives. Both of our constructions fit naturally into the left/right model for ORE (Section 1.1) and have the appealing property that in isolation, right ciphertexts are semantically secure. We leveraged this property to build an efficient range query protocol that is robust against inference attacks. Thus, our work shows that it is possible to leverage property-preserving encryption for searching on encrypted data while resisting offline inference attacks. As part of our analysis into the security of OPE and ORE, we also strengthen the lower bound on OPE schemes and show that there are no efficient OPE schemes on any message space containing at least three messages. To conclude, we present several interesting directions for future study:

- Can we construct a practical small-domain ORE with best-possible security and ciphertext length that is *sublinear* in the size of the message space? Such an ORE scheme can be combined with our domain-extension technique from Section 4 to obtain an ORE scheme with shorter ciphertexts or increased security (by allowing for larger block sizes).
- Can we construct a left/right ORE scheme (with similar or less leakage) from simple primitives where both the left ciphertexts and the right ciphertexts are semantically secure when taken in isolation? In our constructions from Sections 3 and 4, the left ciphertexts are deterministic and do not satisfy this property. The only ORE constructions that achieve semantic security for both left and right ciphertexts require multilinear maps [BLR$^+$15] or indistinguishability obfuscation [GGG$^+$14]. Concurrently with this work, both Joye and Passelégue [JP16] as well as Cash et al. [CLOZ16] proposed pairing-based constructions of ORE with limited leakage where the ciphertexts can be decomposed into left/right components such that the components are themselves semantically secure. It still remains an open question whether such a scheme can be constructed using weaker primitives.
- Can we strengthen our OPE lower bound to show that no OPE scheme can satisfy best-possible security even if they are stateful or interactive? Popa et al. [PLZ13] previously showed that even if we allow for state and interaction, the size of the ciphertext space must be exponential in the size of the plaintext space. However, their lower bound does not rule out the possibility of a stateful or interactive OPE scheme with best-possible security for *small domains*.

Acknowledgments

We thank Dan Boneh, Mark Zhandry, and Joe Zimmerman for insightful discussions about this work. We thank Dmytro Bogatov for pointing out the security issue of using a Feistel network with a small block size as a concrete instantiation of a small-domain PRP. We thank the members of the 2015 Stanford Theory Retreat for initiating our study of new OPE lower bounds. This work was

supported by the NSF, DARPA, the Simons foundation, a grant from ONR, and an NSF Graduate Research Fellowship. Opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of DARPA.

References

[AC15] Reed Abelson and Julie Creswell. Data breach at anthem may forecast a trend. *The New York Times*, 2015.

[AJ15] Prabhanjan Ananth and Abhishek Jain. Indistinguishability obfuscation from compact functional encryption. In *CRYPTO*, 2015.

[AKSX04] Rakesh Agrawal, Jerry Kiernan, Ramakrishnan Srikant, and Yirong Xu. Order-preserving encryption for numeric data. In *ACM SIGMOD*, 2004.

[BCLO09] Alexandra Boldyreva, Nathan Chenette, Younho Lee, and Adam O'Neill. Order-preserving symmetric encryption. In *EUROCRYPT*, 2009.

[BCO11] Alexandra Boldyreva, Nathan Chenette, and Adam O'Neill. Order-preserving encryption revisited: Improved security analysis and alternative solutions. In *CRYPTO*, 2011.

[BGH+13] Dan Boneh, Craig Gentry, Shai Halevi, Frank Wang, and David J. Wu. Private database queries using somewhat homomorphic encryption. In *ACNS*, 2013.

[BGI+12] Boaz Barak, Oded Goldreich, Russell Impagliazzo, Steven Rudich, Amit Sahai, Salil P. Vadhan, and Ke Yang. On the (im)possibility of obfuscating programs. *J. ACM*, 2012.

[BHF09] Carsten Binnig, Stefan Hildenbrand, and Franz Färber. Dictionary-based order-preserving string compression for main memory column stores. In *ACM SIGMOD*, 2009.

[BHKR13] Mihir Bellare, Viet Tung Hoang, Sriram Keelveedhi, and Phillip Rogaway. Efficient garbling from a fixed-key blockcipher. In *IEEE SP*, 2013.

[BKR18] Dmytro Bogatov, George Kollios, and Leo Reyzin. A comparative evaluation of order-preserving and order-revealing schemes and protocols. Cryptology ePrint Archive, Report 2018/953, 2018.

[BKS15] Zvika Brakerski, Ilan Komargodski, and Gil Segev. From single-input to multi-input functional encryption in the private-key setting. *IACR Cryptology ePrint Archive*, 2015.

[BLR+15] Dan Boneh, Kevin Lewi, Mariana Raykova, Amit Sahai, Mark Zhandry, and Joe Zimmerman. Semantically secure order-revealing encryption: Multi-input functional encryption without obfuscation. In *EUROCRYPT*, 2015.

[BPP16] Tobias Boelter, Rishabh Poddar, and Raluca Ada Popa. A secure one-roundtrip index for range queries. Cryptology ePrint Archive, Report 2016/568, 2016.

[BR93] Mihir Bellare and Phillip Rogaway. Random oracles are practical: A paradigm for designing efficient protocols. In *CCS*, 1993.

[BS03] Dan Boneh and Alice Silverberg. Applications of multilinear forms to cryptography. *Contemporary Mathematics*, 2003.

[BSW11] Dan Boneh, Amit Sahai, and Brent Waters. Functional encryption: Definitions and challenges. In *TCC*, 2011.

[BW07] Dan Boneh and Brent Waters. Conjunctive, subset, and range queries on encrypted data. In *TCC*, 2007.

[CD15] Sanjit Chatterjee and M. Prem Laxman Das. Property preserving symmetric encryption revisited. In *ASIACRYPT*, 2015.

[CGKO06] Reza Curtmola, Juan A. Garay, Seny Kamara, and Rafail Ostrovsky. Searchable symmetric encryption: improved definitions and efficient constructions. In *ACM CCS*, 2006.

[CJJ+13] David Cash, Stanislaw Jarecki, Charanjit S. Jutla, Hugo Krawczyk, Marcel-Catalin Rosu, and Michael Steiner. Highly-scalable searchable symmetric encryption with support for boolean queries. In *CRYPTO*, 2013.

[CJJ+14] David Cash, Joseph Jaeger, Stanislaw Jarecki, Charanjit S. Jutla, Hugo Krawczyk, Marcel-Catalin Rosu, and Michael Steiner. Dynamic searchable encryption in very-large databases: Data structures and implementation. In *NDSS*, 2014.

[CK10] Melissa Chase and Seny Kamara. Structured encryption and controlled disclosure. In *ASIACRYPT*, pages 577–594, 2010.

[CLOZ16] David Cash, Feng-Hao Liu, Adam O'Neill, and Cong Zhang. Reducing the leakage in practical order-revealing encryption. Cryptology ePrint Archive, Report 2016/661, 2016.

[CLT13] Jean-Sébastien Coron, Tancrède Lepoint, and Mehdi Tibouchi. Practical multilinear maps over the integers. In *CRYPTO*, 2013.

[CLWW16] Nathan Chenette, Kevin Lewi, Stephen A. Weis, and David J. Wu. Practical order-revealing encryption with limited leakage. In *FSE*, 2016.

[CM05] Yan-Cheng Chang and Michael Mitzenmacher. Privacy preserving keyword searches on remote encrypted data. In *ACNS*, 2005.

[FJK+15] Sky Faber, Stanislaw Jarecki, Hugo Krawczyk, Quan Nguyen, Marcel-Catalin Rosu, and Michael Steiner. Rich queries on encrypted data: Beyond exact matches. In *ESORICS*, 2015.

[FV15] Jim Finkle and Dustin Volz. Database of 191 million u.s. voters exposed on internet: researcher. *Reuters*, 2015.

[Gen09] Craig Gentry. Fully homomorphic encryption using ideal lattices. In *STOC*, 2009.

[GGG+14] Shafi Goldwasser, S. Dov Gordon, Vipul Goyal, Abhishek Jain, Jonathan Katz, Feng-Hao Liu, Amit Sahai, Elaine Shi, and Hong-Sheng Zhou. Multi-input functional encryption. In *EUROCRYPT*, 2014.

[GGH13a] Sanjam Garg, Craig Gentry, and Shai Halevi. Candidate multilinear maps from ideal lattices. In *EUROCRYPT*, 2013.

[GGH+13b] Sanjam Garg, Craig Gentry, Shai Halevi, Mariana Raykova, Amit Sahai, and Brent Waters. Candidate indistinguishability obfuscation and functional encryption for all circuits. In *FOCS*, 2013.

[GGM86] Oded Goldreich, Shafi Goldwasser, and Silvio Micali. How to construct random functions. *J. ACM*, 1986.

[GM84] Shafi Goldwasser and Silvio Micali. Probabilistic encryption. *J. Comput. Syst. Sci.*, 1984.

[GM16] Shay Gueron and Nicky Mouha. Simpira v2: A family of efficient permutations using the AES round function. *IACR Cryptology ePrint Archive*, 2016.

[GO96] Oded Goldreich and Rafail Ostrovsky. Software protection and simulation on oblivious rams. *J. ACM*, 1996.

[Goh03] Eu-Jin Goh. Secure indexes. *IACR Cryptology ePrint Archive*, 2003.

[Gt12] Torbjrn Granlund and the GMP development team. GNU MP: The GNU Multiple Precision Arithmetic Library. http://gmplib.org/, 2012.

[JJK+13] Stanislaw Jarecki, Charanjit S. Jutla, Hugo Krawczyk, Marcel-Catalin Rosu, and Michael Steiner. Outsourced symmetric private information retrieval. In *ACM CCS*, 2013.

[JP16] Marc Joye and Alain Passelégue. Practical trade-offs for multi-input functional encryption. Cryptology ePrint Archive, Report 2016/622, 2016.

[KAK10] Hasan Kadhem, Toshiyuki Amagasa, and Hiroyuki Kitagawa. A secure and efficient order preserving encryption scheme for relational databases. In *KMIS*, 2010.

[Kel14] Gordon Kelly. ebay suffers massive security breach, all users must change their passwords. *Forbes*, 2014.

[Ker15] Florian Kerschbaum. Frequency-hiding order-preserving encryption. In *ACM CCS*, 2015.

[KLM+16] Sam Kim, Kevin Lewi, Avradip Mandal, Hart William Montgomery, Arnab Roy, and David J. Wu. Function-hiding inner product encryption is practical. *IACR Cryptology ePrint Archive*, 2016.

[Knu98] Donald Ervin Knuth. *The art of computer programming, Volume II: Seminumerical Algorithms, 3rd Edition.* Addison-Wesley, 1998.

[KS14] Florian Kerschbaum and Axel Schröpfer. Optimal average-complexity ideal-security order-preserving encryption. In *ACM CCS*, 2014.

[LR88] Michael Luby and Charles Rackoff. How to construct pseudorandom permutations from pseudorandom functions. *SIAM J. Comput.*, 1988.

[MCO+15] Charalampos Mavroforakis, Nathan Chenette, Adam O'Neill, George Kollios, and Ran Canetti. Modular order-preserving encryption, revisited. In *ACM SIGMOD*, 2015.

[NKW15] Muhammad Naveed, Seny Kamara, and Charles V. Wright. Inference attacks on property-preserving encrypted databases. In *ACM CCS*, 2015.

[NPG14] Muhammad Naveed, Manoj Prabhakaran, and Carl A. Gunter. Dynamic searchable encryption via blind storage. In *IEEE SP*, 2014.

[PKV+14] Vasilis Pappas, Fernando Krell, Binh Vo, Vladimir Kolesnikov, Tal Malkin, Seung Geol Choi, Wesley George, Angelos D. Keromytis, and Steve Bellovin. Blind seer: A scalable private DBMS. In *IEEE SP*, 2014.

[PLZ13] Raluca A. Popa, Frank H. Li, and Nickolai Zeldovich. An ideal-security protocol for order-preserving encoding. In *IEEE SP*, 2013.

[PR12] Omkant Pandey and Yannis Rouselakis. Property preserving symmetric encryption. In *EUROCRYPT*, 2012.

[PRZB11] Raluca A. Popa, Catherine M. S. Redfield, Nickolai Zeldovich, and Hari Balakrishnan. Cryptdb: protecting confidentiality with encrypted query processing. In *ACM SOSP*, 2011.

[RACY15] Daniel S. Roche, Daniel Apon, Seung Geol Choi, and Arkady Yerukhimovich. POPE: partial order-preserving encoding. *IACR Cryptology ePrint Archive*, 2015.

[SWP00] Dawn Xiaodong Song, David Wagner, and Adrian Perrig. Practical techniques for searches on encrypted data. In *IEEE SP*, 2000.

[The03] The OpenSSL Project. OpenSSL: The open source toolkit for SSL/TLS. `www.openssl.org`, 2003.

[TYM14] Isamu Teranishi, Moti Yung, and Tal Malkin. Order-preserving encryption secure beyond one-wayness. In *ASIACRYPT*, 2014.

[Yao82] Andrew Chi-Chih Yao. Protocols for secure computations (extended abstract). In *FOCS*, 1982.

[YSK+13] Masaya Yasuda, Takeshi Shimoyama, Jun Kogure, Kazuhiro Yokoyama, and Takeshi Koshiba. Secure pattern matching using somewhat homomorphic encryption. In *CCSW*, 2013.

A Proof of Theorem 3.3

Let $\mathcal{A} = (\mathcal{A}_1, \ldots, \mathcal{A}_q)$ where $q = \text{poly}(\lambda)$ be an efficient adversary for the ORE security game (Definition 2.2). To show security, we construct an efficient simulator $\mathcal{S} = (\mathcal{S}_0, \ldots, \mathcal{S}_q)$ such that the two distributions $\mathsf{REAL}^{\text{ORE}}_{\mathcal{A}}(\lambda)$ and $\mathsf{SIM}^{\text{ORE}}_{\mathcal{A},\mathcal{S},\mathcal{L}_{\text{CMP}}}(\lambda)$ are computationally indistinguishable.

Description of the simulator. We begin by describing the simulator $\mathcal{S}$. In the security proof, we model H as a random oracle. Thus, in the ideal experiment, the simulator is also responsible for answering queries to the random oracle. First, on input the security parameter 1^λ, the simulator algorithm $\mathcal{S}_0$ initializes the following tables which will be used to ensure consistency throughout the simulation:

- The table $T_{\text{RO}} : \{0,1\}^\lambda \times \{0,1\}^\lambda \to \mathbb{Z}_3$ used to maintain the input-output mappings for the random oracle.
- The table $T_{\text{KEYS}} : \{0,1\}^\lambda \to [q] \times [N]$ used to maintain the mapping between keys to the corresponding message index and the permuted index of the message.

Both tables are initially empty at the beginning of the simulation. The simulator's initial state $\mathsf{st}_{\mathcal{S}}$ consists of the tuple $(T_{\text{KEYS}}, T_{\text{RO}})$. Then, for each $t \in [q]$, after the adversary outputs a query m_t, the simulation algorithm $\mathcal{S}_t$ is invoked on input $\mathsf{st}_{\mathcal{S}}$ and the leakage function $\mathcal{L}_{\text{CMP}}(m_1, \ldots, m_t)$. In the following, we write $\overline{\mathsf{ct}}^{(i)} = \left(\overline{\mathsf{ct}}^{(i)}_{\mathsf{L}}, \overline{\mathsf{ct}}^{(i)}_{\mathsf{R}}\right)$ to denote the simulator's response in the i^{th} query. We also assume that the simulator's state includes the ciphertexts $\overline{\mathsf{ct}}^{(i)}$ for all previous queries $i < t$. We now describe how $\mathcal{S}_t$ responds to the t^{th} query.

Simulating the left ciphertexts. We first show how $\mathcal{S}_t$ constructs the ciphertext components $\overline{\mathsf{ct}}^{(t)}_{\mathsf{L}}$. There are two cases:

- Suppose for some $i < t$, $\text{CMP}(m_i, m_t) = 0$. Then, the simulator sets $\overline{\mathsf{ct}}^{(t)}_{\mathsf{L}} = \overline{\mathsf{ct}}^{(i)}_{\mathsf{L}}$.
- Suppose for all $i < t$, $\text{CMP}(m_i, m_t) \neq 0$. Define $S \subset [N]$ to be the set of indices $\beta \in [N]$ for which there exists a mapping $k \mapsto (\alpha, \beta)$ in T_{KEYS} for some key $k \in \{0,1\}^\lambda$ and message index $\alpha \in [q]$. Then the simulator chooses a random key $k \xleftarrow{\text{R}} \{0,1\}^\lambda$ and a random index $j \xleftarrow{\text{R}} [N] \setminus S$. It then checks to see if there is a mapping $(k, r) \mapsto \rho$ for some $r \in \{0,1\}^\lambda$ and $\rho \in \mathbb{Z}_3$ in T_{RO}. If so, $\mathcal{S}_t$ aborts the simulation and outputs $\perp_1$. Otherwise, it sets $\overline{\mathsf{ct}}^{(t)}_{\mathsf{L}} = (k, j)$, and adds the mapping $k \mapsto (t, j)$ to T_{KEYS}.

Finally, if the simulator does not abort, it outputs the left ciphertext $\overline{\mathsf{ct}}^{(t)}_{\mathsf{L}}$ and an updated state.

Simulating the right ciphertexts. For the right ciphertexts, the simulator $\mathcal{S}_t$ first samples a nonce $\bar{r}_t \xleftarrow{\text{R}} \{0,1\}^\lambda$. It then checks to see if there is already a mapping of the form $(k, \bar{r}_t) \mapsto \rho$ in T_{RO} for some $k \in \{0,1\}^\lambda$ and $\rho \in \mathbb{Z}_3$. If so, $\mathcal{S}_t$ aborts the simulation and outputs $\perp_2$. Otherwise, for $i \in [N]$, it samples $\bar{v}^{(t)}_i \xleftarrow{\text{R}} \mathbb{Z}_3$, sets $\overline{\mathsf{ct}}^{(t)}_{\mathsf{R}} = (\bar{r}_t, \bar{v}^{(t)}_1, \ldots, \bar{v}^{(t)}_N)$, and outputs the ciphertext $\overline{\mathsf{ct}}^{(t)} = (\overline{\mathsf{ct}}^{(t)}_{\mathsf{L}}, \overline{\mathsf{ct}}^{(t)}_{\mathsf{R}})$ as well as an updated state.

Simulating the random oracle queries. To conclude the description of $\mathcal{S}$, we describe how it simulates the random oracle queries. Let $t \leq q$ be the number of encryption queries the adversary has made so far. Then, on an input (k, r), the simulator responds as follows:

- If there is a mapping $(k, r) \mapsto \bar{\rho}$ in T_{RO}, then the simulator replies with $\bar{\rho}$.
- If there is a mapping $k \mapsto (\alpha, \beta)$ in T_{KEYS} for some $\alpha \in [q]$ and $\beta \in [N]$ and $r = \bar{r}_j$ for some $j \in [t]$, then the simulator sets $\bar{\rho} = \bar{v}_\beta^{(j)} - \text{CMP}(m_\alpha, m_j)$, adds the mapping $(k, r) \mapsto \bar{\rho}$ to T_{RO}, and replies with $\bar{\rho}$.
- If there is no mapping $k \mapsto (\alpha, \beta)$ in T_{KEYS} for any $\alpha \in [q]$ and $\beta \in [N]$, or $r \neq \bar{r}_j$ for all $j \in [t]$, then the simulator chooses $\bar{\rho} \xleftarrow{R} \mathbb{Z}_3$, adds the mapping $(k, r) \mapsto \bar{\rho}$ to T_{RO}, and replies with $\bar{\rho}$.

Correctness of the simulation. To conclude the proof, we now show that the real and ideal experiments $\mathsf{REAL}^{\text{ORE}}_{\mathcal{A}}(\lambda)$ and $\mathsf{SIM}^{\text{ORE}}_{\mathcal{A},\mathcal{S},\mathcal{L}_{\text{CMP}}}(\lambda)$ are computationally indistinguishable. We begin by defining a series of hybrid experiments:

- Hybrid H_0: This is the real experiment $\mathsf{REAL}^{\text{ORE}}_{\mathcal{A}}(\lambda)$ (Definition 2.2).
- Hybrid H_1: Same as H_0, except the PRF $F(k, \cdot)$ is replaced by a truly random function f from $\{0,1\}^\lambda \to \{0,1\}^\lambda$.
- Hybrid H_2: Same as H_1, except the experiment aborts (with output $\perp_1$ or $\perp_2$) if one of the following events occur:
  - The adversary queries H on an input of the form $(f(\pi(m)), \cdot)$ before it issues an encryption query for the message m. In this case, the experiment outputs $\perp_1$.
  - The adversary queries H on an input of the form $(\cdot, r_j)$ before it makes its j^{th} encryption query. Here, r_j is the randomness the challenger samples when responding to the j^{th} encryption query. In this case, the experiment outputs $\perp_2$.
- Hybrid H_3: This is the ideal experiment $\mathsf{SIM}^{\text{ORE}}_{\mathcal{A},\mathcal{S},\mathcal{L}_{\text{CMP}}}(\lambda)$. (Definition 2.2).

We now argue that each consecutive pair of hybrid experiments are computationally indistinguishable.

Lemma A.1. *Hybrid H_0 and H_1 are computationally indistinguishable if F is a secure PRF.*

Proof. Follows immediately from PRF security. □

Lemma A.2. *Hybrids H_1 and H_2 are statistically indistinguishable if H is modeled as a random oracle.*

Proof. For each of the two abort events in H_2, we argue that the probability of the event occurring is negligible.

- **Case 1: The experiment outputs $\perp_1$.** Suppose the adversary has not issued an encryption query for a message $m \in [N]$. We argue that in this case, the adversary's view in the experiment is independent of $f(\pi(m))$. Consider the ciphertext $\mathsf{ct}' = (\mathsf{ct}'_{\mathsf{L}}, \mathsf{ct}'_{\mathsf{R}})$ the adversary obtains when it requests an encryption of some message $m' \neq m$. Then, $\mathsf{ct}'_{\mathsf{L}} = (f(\pi(m')), \pi(m'))$. Since π is a permutation, $\pi(m') \neq \pi(m)$. Next, because f is a truly random function, $f(\pi(m'))$ is independent of $f(\pi(m))$. We conclude that the components of $\mathsf{ct}'_{\mathsf{L}}$ are distributed independently of $f(\pi(m))$.

  Consider now $\mathsf{ct}'_R = (r', v'_1, \ldots, v'_N)$. Since r' is sampled uniformly at random from $\{0,1\}^\lambda$, it is distributed independently of $f(\pi(m))$. Next, for all $i \in [N]$, $v_i = \text{CMP}(\pi^{-1}(i), m') + H(f(i), r')$.

The value of $\textsc{cmp}(\pi^{-1}(i), m')$ is independent of the function f. Similarly, the output of the random oracle on $(f(i), r')$ is independent of its input, and thus, independent of $f(\pi(m))$. Thus, the components of ct'_R are distributed independently of $f(\pi(m))$.

Finally, the responses from the random oracle are distributed independently of f. We thus conclude that unless the adversary requests for an encryption of m, its view in hybrid H_1 is distributed independently of $f(\pi(m))$. Now, let $z_1, \ldots, z_\ell$ for $\ell = \text{poly}(\lambda)$ be the adversary's queries to the random oracle before it requests for an encryption of m. By our argument above, each z_i must be chosen independently of $f(\pi(m))$. Since f is a truly random function, the probability that there is some i such that $z_i = (f(\pi(m)), y)$ for any y is at most $\ell/2^\lambda = \text{negl}(\lambda)$. We conclude that H_2 outputs $\perp_1$ with negligible probability.

- **Case 2: The experiment outputs $\perp_2$.** Let $z_1, \ldots, z_\ell$ for $\ell = \text{poly}(\lambda)$ be the random oracle queries the adversary makes before making its j^{th} encryption query. When responding to the j^{th} encryption query, the challenger in H_2 samples $r_j \xleftarrow{\text{R}} \{0,1\}^\lambda$. In particular, r_j is independent of $z_1, \ldots, z_\ell$, and so the probability that there is some i such that $z_i = (x, r_j)$ for any x is at most $\ell/2^\lambda = \text{negl}(\lambda)$. We conclude that H_2 outputs $\perp_2$ with negligible probability, and the claim follows. □

Lemma A.3. *Hybrid H_2 and H_3 are statistically indistinguishable if H is modeled as a random oracle.*

Proof. Let $(\mathsf{ct}_1, \ldots, \mathsf{ct}_q)$ be the joint distribution of the ciphertexts output in H_2 and let $(\overline{\mathsf{ct}}_1, \ldots, \overline{\mathsf{ct}}_q)$ be the joint distribution of the ciphertexts output in H_3. We show that these two distributions are statistically indistinguishable, and moreover, that the outputs of the random oracle are properly simulated in H_3. Let $m_1, \ldots, m_q$ be the messages chosen by the adversary in H_2 and H_3. In the simulation, the table T_{KEYS} is used to maintain the mapping between keys to the message indices and the permuted positions of the messages. The proof proceeds via induction on the number of queries q. In the inductive step, we assume that the following conditions hold for some $t < q$:

- $(\mathsf{ct}_1, \ldots, \mathsf{ct}_t) \overset{s}{\approx} (\overline{\mathsf{ct}}_1, \ldots, \overline{\mathsf{ct}}_t)$.
- The outputs of the random oracle prior to the $(t+1)^{\text{th}}$ query are statistically indistinguishable in H_2 and H_3.

Consider the base case where $t = 0$. It suffices to argue that all of the random oracle queries are simulated properly in H_3. Suppose the adversary queries the random oracle H on an input (k, r). Without loss of generality, we can assume that each of the adversary's queries to the random oracle is unique (the random oracle responds consistently if an input is queried multiple times). Since the table T_{KEYS} is initially empty, the simulator in H_3 always replies with a uniform random element of $\mathbb{Z}_3$ in this case. In H_2, the outputs of the random oracle are distributed uniformly and independently in $\mathbb{Z}_3$, assuming that $k \neq f(i)$ for any $i \in [N]$ (otherwise, the experiment aborts with output $\perp_1$). However, as shown in the proof of Lemma A.2, the probability (taken over the randomness used to sample f) that $k = f(i)$ for $i \in [N]$ is negligible, and so we conclude that the outputs of the random oracle in H_2 and H_3 are statistically indistinguishable.

For the inductive step, suppose that both conditions outlined above hold for some $t < q$. We show that both conditions continue to hold for $t + 1$. We begin with some notation. For all $j \in [t]$, we write $\mathsf{ct}_j = \left(\mathsf{ct}_{\mathsf{L}}^{(j)}, \mathsf{ct}_{\mathsf{R}}^{(j)}\right)$ and similarly, $\overline{\mathsf{ct}}_j = \left(\overline{\mathsf{ct}}_{\mathsf{L}}^{(j)}, \overline{\mathsf{ct}}_{\mathsf{R}}^{(j)}\right)$, where $\mathsf{ct}_{\mathsf{L}}^{(j)} = (k_j, h_j)$, $\overline{\mathsf{ct}}_{\mathsf{L}}^{(j)} = (\bar{k}_j, \bar{h}_j)$,

$\mathsf{ct}_{\mathsf{R}}^{(j)} = \left(r_j, v_1^{(j)}, \ldots, v_N^{(j)}\right)$, and $\overline{\mathsf{ct}}_{\mathsf{R}}^{(j)} = \left(\bar{r}_j, \bar{v}_1^{(j)}, \ldots, \bar{v}_N^{(j)}\right)$. We now argue that the responses to the random oracle queries the adversary makes between its t^{th} and $(t+1)^{\text{th}}$ encryption query are statistically indistinguishable in H_2 and H_3. Let (k, r) be the adversary's query to the random oracle. We consider several possibilities:

- Suppose $k = k_i$ and $r = r_j$ for some $i, j \in [t]$. If there are multiple indices i where $k = k_i$, consider the smallest such i. In hybrid H_2, we have that $H(k, r)$ satisfies the relation

$$v_{h_i}^{(j)} = \textsc{cmp}(m_i, m_j) + H(k, r).$$

  By construction in H_3, if $k = \bar{k}_i$, the simulator must have added the mapping $\bar{k}_i \mapsto (i, \bar{h}_i)$ to $T_{\textsc{keys}}$ in response to the i^{th} encryption query (here, we rely on the fact that i is the smallest such i such that $k = \bar{k}_i$). In this case then, the simulator responds with $\bar{\rho}$ as follows:

$$\bar{\rho} = \bar{v}_{\bar{h}_i}^{(j)} - \textsc{cmp}(m_i, m_j).$$

  By the inductive hypothesis, $\mathsf{ct}_{\mathsf{L}}^{(i)} \equiv \overline{\mathsf{ct}}_{\mathsf{L}}^{(i)}$, so the simulator's response is identically distributed as the outputs of the random oracle in H_2.

- Suppose $k \neq k_i$ for all $i \in [t]$. In H_3, the simulator always responds with a uniformly random value in $\mathbb{Z}_3$. In H_2, there are two possibilities. If $k = f(j)$ for some $j \in [N]$, then the experiment aborts (with output $\perp_1$) because the adversary must not have issued an encryption query for $\pi^{-1}(j)$. However, as argued in the proof of Lemma A.2, the probability that $k \neq k_i$ for all $i \in [t]$, but $k = f(j)$ for some $j \in [N]$ is negligible. Thus, with overwhelming probability, in hybrid H_2, $k \neq f(j)$ for all $j \in [N]$. Then, the value $H(k, r)$ is distributed uniformly and independently of all other components of the adversary's view. Specifically, because $k \neq k_i$ for all $i \in [t]$, the value $H(k, r)$ is distributed independently of the ciphertexts $\mathsf{ct}_1, \ldots, \mathsf{ct}_t$. Moreover, the outputs of the random oracle are all distributed uniformly and independently, so we conclude that the distribution of $H(k, r)$ given the adversary's view in H_2 is uniform over $\mathbb{Z}_3$. This is precisely the distribution from which the simulator samples the value of $H(k, r)$ in H_3. We conclude that the random oracle outputs are statistically indistinguishable in H_2 and H_3.

- Finally, suppose that $k = k_i$ for some $i \in [t]$, but $r \neq r_j$ for all $j \in [t]$. Similar to the previous case, in hybrid H_2, the value $H(k, r)$ is distributed uniformly and independently of all other components of the adversary's view. Thus, conditioned on the adversary's view, the value $H(k, r)$ is uniform over $\mathbb{Z}_3$. This is precisely how $\mathcal{S}$ samples the random oracle output in H_3, so in this case, the responses in H_2 and H_3 are identically distributed.

Next, we show that the conditional distribution of ct_{t+1} given $\mathsf{ct}_1, \ldots, \mathsf{ct}_t$ is statistically indistinguishable from the conditional distribution of $\overline{\mathsf{ct}}_{t+1}$ given $\overline{\mathsf{ct}}_1, \ldots, \overline{\mathsf{ct}}_t$. Let m_{t+1} be the adversary's $(t+1)^{\text{th}}$ encryption query. First, we show that the conditional distribution of the left ciphertexts $\mathsf{ct}_{\mathsf{L}}^{(t+1)}$ and $\overline{\mathsf{ct}}_{\mathsf{L}}^{(t+1)}$ is statistically indistinguishable. In hybrid H_2, $\mathsf{ct}_{\mathsf{L}}^{(t+1)} = (k_{t+1}, h_{t+1}) = (f(\pi(m_{t+1})), \pi(m_{t+1}))$. We consider two possibilities:

- If $m_{t+1} = m_\ell$ for some $\ell \in [t]$, then $\mathsf{ct}_{\mathsf{L}}^{(t+1)} = \mathsf{ct}_{\mathsf{L}}^{(\ell)}$ in H_2. Since $m_{t+1} = m_\ell$, then $\textsc{cmp}(m_{t+1}, m_\ell) = 0$, so $\overline{\mathsf{ct}}_{\mathsf{L}}^{(t+1)} = \overline{\mathsf{ct}}_{\mathsf{L}}^{(\ell)}$. The claim then follows by the induction hypothesis.

- If $m_{t+1} \neq m_\ell$ for some $\ell \in [t]$, then conditioned on the adversary's view after its first t queries, we argue that $h_{t+1} = \pi(m_{t+1})$ is uniform over the set $[N] \setminus \{h_1, \ldots, h_t\}$, where by definition, $h_i = \pi(m_i)$ for all $i \in [t]$. To see this, we first note that $\mathsf{ct}_{\mathsf{L}}^{(1)}, \ldots, \mathsf{ct}_{\mathsf{L}}^{(t)}$ can be written entirely as a function that only depends on $\pi(m_1), \ldots, \pi(m_t)$. Certainly, the outputs of the random oracle are completely independent of π. Consider then the right ciphertext $\mathsf{ct}_{\mathsf{R}}^{(i)} = (r_i, v_1^{(i)}, \ldots, v_n^{(i)})$ for some $i \in [t]$. For all $j \notin \{\pi(m_1), \ldots, \pi(m_t)\}$, $v_j^{(i)}$ is blinded by $H(f(j), r_i)$. Moreover, in H_2, the adversary will never have queried H on $(f(j), r_i)$ prior to making its $(t+1)^{\text{th}}$ query (otherwise, the experiment aborts with output $\perp_1$). Thus, $v_j^{(i)}$ is uniformly distributed for all $j \notin \{\pi(m_1), \ldots, \pi(m_t)\}$. In particular, this means that conditioned on the view of the adversary, the values of the right ciphertexts $\mathsf{ct}_{\mathsf{R}}^{(i)}$ for $i \in [t]$ only depend on $\pi(m_1), \ldots, \pi(m_t)$. Since the permutation π is sampled uniformly, we conclude that given the output of the first t encryption queries, the value of $\pi(m_{t+1})$ is still distributed uniformly in $[N] \setminus \{\pi(m_1), \ldots, \pi(m_t)\}$. Similarly, by the same argument as that given in the proof of Lemma A.2, the value of $f(\pi(m_{t+1}))$ is independent of the adversary's view prior to its $(t+1)^{\text{th}}$ encryption query. Thus, in H_2, given the adversary's view up to the $(t+1)^{\text{th}}$ query, the left ciphertext $\mathsf{ct}_{\mathsf{L}}^{(t+1)} = (k_{t+1}, h_{t+1})$ is uniform over $\{0,1\}^\lambda \times ([N] \setminus \{h_1, \ldots, h_t\})$. In H_3, the simulator $\mathcal{S}_{t+1}$ samples $\bar{k}_{t+1}$ uniformly at random from $\{0,1\}^\lambda$ and $\bar{h}_{t+1}$ uniformly at random from the set $[N] \setminus \{\bar{h}_1, \ldots, \bar{h}_t\}$. Invoking the inductive hypothesis, we conclude that the conditional distributions of $\mathsf{ct}_{\mathsf{L}}^{(t+1)}$ and $\overline{\mathsf{ct}}_{\mathsf{L}}^{(t+1)}$ given the adversary's view in the respective experiments are statistically indistinguishable.

Finally, we show that the conditional distributions of the right ciphertext components $\mathsf{ct}_{\mathsf{R}}^{(t+1)}$ and $\overline{\mathsf{ct}}_{\mathsf{R}}^{(t+1)}$ are statistically indistinguishable. Certainly, r_{t+1} and $\bar{r}_{t+1}$ are identically distributed. Next, in H_2, if the adversary has already queried the random oracle on the input (x, r_{t+1}) for any $x \in \{0,1\}^\lambda$, then the experiment aborts with output $\perp_2$. Equivalently, if the adversary queries the random oracle on an input $(x, \bar{r}_{t+1})$ in H_3, the experiment also aborts (with the same output $\perp_2$). In H_2, each v_i is blinded by the value $H(f(i), r_{t+1})$, and since the adversary has not queried $H(f(i), r_{t+1})$ before seeing $\mathsf{ct}^{(t+1)}$, the components $v_1^{(i)}, \ldots, v_N^{(i)}$ are distributed uniformly and independently over $\mathbb{Z}_3$ to the adversary in H_2. This is precisely the distribution from which the simulator samples $\bar{v}_1^{(t+1)}, \ldots, \bar{v}_N^{(t+1)}$ in H_3. Thus, conditioned on the adversary's view in the experiment, the distributions of the right ciphertexts in Hybrids H_2 and H_3 are statistically indistinguishable. Lemma A.3 then follows by induction on t. □

Combining Lemmas A.1 through A.3, we conclude that $\Pi_{\mathsf{ore}}^{(\mathsf{s})}$ is secure with the best-possible leakage function $\mathcal{L}_{\text{CMP}}$.

B Proof of Theorem 4.1

Let $\mathcal{A} = (\mathcal{A}_1, \ldots, \mathcal{A}_q)$ where $q = \text{poly}(\lambda)$ be an efficient adversary for the ORE security game (Definition 2.2). We construct an efficient simulator $\mathcal{S} = (\mathcal{S}_0, \ldots, \mathcal{S}_q)$ such that the two distributions $\mathsf{REAL}_{\mathcal{A}}^{\text{ORE}}(\lambda)$ and $\mathsf{SIM}_{\mathcal{A},\mathcal{S},\mathcal{L}_{\text{BLK}}}^{\text{ORE}}(\lambda)$ are computationally indistinguishable.

B.1 Description of the Simulator

We begin by describing the simulator $\mathcal{S}$. As in the proof of Theorem 3.3 (Appendix A), we model H as a random oracle. Recall that the inputs to the encryption scheme are written in base d. First, on input the security parameter 1^λ, the simulator $\mathcal{S}_0$ maintains the following tables and sets which will be used to ensure consistency throughout the simulation:

- The table $T_{\text{RO}} : \{0,1\}^\lambda \times \{0,1\}^\lambda \to \mathbb{Z}_3$, used to maintain the input-output mappings to the random oracle.
- The collection of tables $T_{\text{KEYS}}[j,s] : \{0,1\}^\lambda \to [q] \times [d]$, for each $j \in [q]$ and $s \in [n]$ used to maintain mappings of keys $k \in \{0,1\}^\lambda$ to tuples containing a message index associated with k, along with the (permuted) position within the block associated with the key k.
- The collection of sets $S_{j,s} \subset [d]$, for each $j \in [q]$ and $s \in [n]$, used to lazily sample the random permutations for each block. Each of these sets $S_{j,s}$ will always be a subset of $[d]$.

The simulator's initial state $\mathsf{st}_{\mathcal{S}}$ consists of the (initially empty) tables T_{RO}, $T_{\text{KEYS}}[j,s]$, and the (initially empty) sets $S_{j,s}$, for all $j \in [q]$ and $s \in [n]$. Then, for each $t \in [q]$, after the adversary outputs a message m_t, the simulation algorithm $\mathcal{S}_t$ is invoked on the input $\mathsf{st}_{\mathcal{S}}$ and the leakage function $\mathcal{L}_{\text{BLK}}(m_1, \ldots, m_t)$. In particular, $\mathcal{L}_{\text{BLK}}(m_1, \ldots, m_t)$ includes both $\text{CMP}(m_i, m_t)$ and $\mathsf{ind}^{(d)}_{\mathsf{diff}}(m_i, m_t)$ for all $i < t$. In the following, we write $\overline{\mathsf{ct}}^{(i)} = (\mathsf{ct}_{\mathsf{L}}^{(i)}, \mathsf{ct}_{\mathsf{R}}^{(i)})$ to denote the simulator's response to the i^{th} query, and we will implicitly assume that the simulator's state $\mathsf{st}_{\mathcal{S}}$ also includes its responses $\overline{\mathsf{ct}}^{(1)}, \ldots, \overline{\mathsf{ct}}^{(t-1)}$ in the previous queries. We now describe how $\mathcal{S}_t$ responds to the t^{th} query.

Simulating the left ciphertexts. First, we describe how $\mathcal{S}_t$ simulates the left ciphertext $\mathsf{ct}_{\mathsf{L}}^{(t)}$. The left ciphertext $\mathsf{ct}_{\mathsf{L}}^{(t)}$ consists of a tuple of the form $(\bar{u}_1^{(t)}, \ldots, \bar{u}_n^{(t)})$, and for each $s \in [n]$, the simulator constructs $\bar{u}_s^{(t)}$ as follows:

- **Case 1:** There exists a $j < t$ such that $\mathsf{ind}^{(d)}_{\mathsf{diff}}(m_j, m_t) > s$. If there are multiple j for which $\mathsf{ind}^{(d)}_{\mathsf{diff}}(m_j, m_t) > s$, let j be the smallest one. In this case, the simulator sets $\bar{u}_s^{(t)} = \bar{u}_s^{(j)}$.
- **Case 2:** For each $\ell < t$, $\mathsf{ind}^{(d)}_{\mathsf{diff}}(m_\ell, m_t) \leq s$, and there exists some $j < t$ for which $\mathsf{ind}^{(d)}_{\mathsf{diff}}(m_j, m_t) = s$. If there are multiple j for which $\mathsf{ind}^{(d)}_{\mathsf{diff}}(m_j, m_t) = s$, let j be the smallest one. The simulator samples an index $\bar{i} \xleftarrow{\text{R}} [d] \setminus S_{j,s}$, and a key $\bar{k} \xleftarrow{\text{R}} \{0,1\}^\lambda$. If there exists a mapping of the form $(\bar{k}, y) \mapsto \rho$ in T_{RO} for some $y \in \{0,1\}^\lambda$ and $\rho \in \mathbb{Z}_3$, then the simulator aborts and outputs $\perp_1$. Otherwise, it adds the index $\bar{i}$ to $S_{j,s}$, and also adds the mapping $\bar{k} \mapsto (t, \bar{i})$ to $T_{\text{KEYS}}[j,s]$. Finally, it sets $\bar{u}_s^{(t)} = (\bar{k}, \bar{i})$.
- **Case 3:** For each $\ell < t$, $\mathsf{ind}^{(d)}_{\mathsf{diff}}(m_\ell, m_t) < s$. In this case, the simulator samples an index $\bar{i} \xleftarrow{\text{R}} [d]$ and a key $\bar{k} \xleftarrow{\text{R}} \{0,1\}^\lambda$. If there exists a mapping $(\bar{k}, y) \mapsto \rho$ in T_{RO} for any $y \in \{0,1\}^\lambda$ and $\rho \in \mathbb{Z}_3$, then the simulator aborts and outputs $\perp_1$. Otherwise, it adds the index $\bar{i}$ to $S_{t,s}$, and also adds the mapping $\bar{k} \mapsto (t, \bar{i})$ to $T_{\text{KEYS}}[j,s]$. Finally, it sets $\bar{u}_s^{(t)} = (\bar{k}, \bar{i})$.

Simulating the right ciphertexts. To simulate the right ciphertext $\overline{\mathsf{ct}}_{\mathsf{R}}^{(t)}$ for the t^{th} query, the simulator samples a random nonce $\bar{r}_t \xleftarrow{\text{R}} \{0,1\}^\lambda$. Next, it checks whether there is a mapping of the form $(k, \bar{r}_t) \mapsto \rho$ in T_{RO} for some $k \in \{0,1\}^\lambda$ and $\rho \in \mathbb{Z}_3$. If so, the simulator aborts and outputs $\perp_2$.

Otherwise, for $i \in [n]$ and $j \in [d]$, the simulator samples $\bar{z}_{i,j}^{(t)} \xleftarrow{\text{R}} \mathbb{Z}_3$, sets $\bar{v}_i^{(t)} = (\bar{z}_{i,1}^{(t)}, \ldots, \bar{z}_{i,d}^{(t)})$, and constructs $\overline{\mathsf{ct}}_{\mathsf{R}}^{(t)} = (\bar{r}_t, \bar{v}_1^{(t)}, \ldots, \bar{v}_n^{(t)})$.

Simulating the random oracle queries. To conclude the specification of the simulator $\mathcal{S}$, we describe how it responds to a random oracle query. Let $t \leq q$ be the number of encryption queries the adversary has made so far, and recall that $\bar{r}_1, \ldots, \bar{r}_t \in \{0,1\}^\lambda$ are the nonces chosen by the simulator when constructing the right ciphertexts for each encryption query. Then, on input $(k, r) \in \{0,1\}^\lambda \times \{0,1\}^\lambda$, the simulator responds as follows:

- If there is a mapping $(k, r) \mapsto \bar{\rho}$ in T_{RO} for some $\bar{\rho} \in \mathbb{Z}_3$, the simulator simply replies with $\bar{\rho}$.
- If there is a mapping $k \mapsto (\alpha, \beta)$ in $T_{\text{KEYS}}[j, s]$ for some $\alpha, j \in [q]$, $\beta \in [d]$, $s \in [n]$, and $r = \bar{r}_i$ for some $i \in [t]$, then the simulator responds as follows:
  - If $\mathsf{ind}_{\mathsf{diff}}^{(d)}(m_\alpha, m_i) < s$, then the simulator samples $\bar{\rho} \xleftarrow{\text{R}} \mathbb{Z}_3$.
  - If $\mathsf{ind}_{\mathsf{diff}}^{(d)}(m_\alpha, m_i) = s$, then the simulator sets $\bar{\rho} = \bar{z}_{s,\beta}^{(i)} - \textsc{cmp}(m_i, m_\alpha) \pmod 3$.
  - If $\mathsf{ind}_{\mathsf{diff}}^{(d)}(m_\alpha, m_i) > s$, then the simulator sets $\bar{\rho} = \bar{z}_{s,\beta}^{(i)}$.

  Finally, the simulator adds the mapping $(k, r) \mapsto \bar{\rho}$ to T_{RO} and replies with $\bar{\rho}$.
- For the final case, if there is either no mapping of the form $k \mapsto (\alpha, \beta)$ for some $\alpha \in [q]$ and $\beta \in [d]$ in $T_{\text{KEYS}}[j, s]$ for all $j \leq t$ and $s \in [n]$, or $r \neq \bar{r}_i$ for all $i \in [t]$, then the simulator chooses $\bar{\rho} \xleftarrow{\text{R}} \mathbb{Z}_3$, adds the mapping $(k, r) \mapsto \bar{\rho}$ to T_{RO}, and replies with $\bar{\rho}$.

B.2 Correctness of the Simulation

To complete the proof, we argue that the real and ideal experiments $\mathsf{REAL}_{\mathcal{A}}^{\text{ORE}}(\lambda)$ and $\mathsf{SIM}_{\mathcal{A},\mathcal{S},\mathcal{L}_{\text{BLK}}}^{\text{ORE}}(\lambda)$, respectively, are computationally indistinguishable. Similar to the proof of Theorem 3.3, we proceed via a hybrid argument:

- Hybrid H_0: The is the real experiment $\mathsf{REAL}_{\mathcal{A}}^{\text{ORE}}(\lambda)$ (Definition 2.2).
- Hybrid H_1: Same as H_0, except that the PRFs $F(k_1, \cdot)$ and $F(k_2, \cdot)$ are replaced by truly random functions $f_1, f_2 : \{0,1\}^\lambda \to \{0,1\}^\lambda$.
- Hybrid H_2: Same as H_1, except that for each $k \in \{0,1\}^\lambda$, we replace each of the PRPs $\pi(k, \cdot)$ with a truly random permutation τ_k over $[d]$. In other words, whenever there is an invocation to $\pi(k, \cdot)$, we replace it with an invocation to $\tau_k(\cdot)$. For distinct $k, k' \in \{0,1\}^\lambda$, the truly random permutations τ_k and $\tau_{k'}$ are independent.
- Hybrid H_3: Same as H_2, except that the experiment aborts and outputs either $\perp_1$ or $\perp_2$ if one of the following events occur:
  - If the adversary queries for an encryption of a message $m = m_1 m_2 \cdots m_n \in [N]$ and the adversary is able to query the random oracle H on a tuple $(f_1(m_{|i-1} \| \tau_k(m_i)), r')$, where $k = f_2(m_{|i-1})$ for some $i \in [n]$, and $r' \in \{0,1\}^\lambda$, *before* it has made an encryption query on some message $m' \in [N]$ for which $m'_{|i} = m_{|i}$. In this case, the experiment outputs $\perp_1$.

– If for some $j \in [q]$, the adversary queries H on an input of the form (k, r_j), for some $k \in \{0,1\}^\lambda$, *before* it makes the j^{th} encryption query. Recall that $r_j \in \{0,1\}^\lambda$ is the nonce sampled by the right encryption algorithm on the j^{th} encryption query. In this case, the experiment outputs $\perp_2$.

- Hybrid H_4: This is the ideal experiment $\mathsf{SIM}^{\text{ORE}}_{\mathcal{A},\mathcal{S},\mathcal{L}_{\text{BLK}}}(\lambda)$ (Definition 2.2).

Note that our sequence of hybrid experiments almost exactly mirrors the sequence used in the proof of Theorem 3.3. The main difference is the needing to switch from using the PRP to using a truly random permutation. This step was unnecessary in the small-domain setting because there, we required just a single permutation which could be sampled during setup. We now argue that each consecutive pair of hybrid arguments are computationally indistinguishable.

Lemma B.1. *Hybrids* H_0 *and* H_1 *are computationally indistinguishable if* F *is a secure PRF.*

Proof. Formally, we define an intermediate hybrid where we first replace $F(k_1, \cdot)$ with the truly random function f_1, but keep $F(k_2, \cdot)$ as normal. In the second hybrid, we replace $F(k_2, \cdot)$ with the truly random function f_2. For the first hybrid argument, we use the fact that k_1 is sampled uniformly at random from the keyspace $\mathcal{K}$ (during the setup procedure). Thus, we can invoke the PRF security of F to argue that $F(k_1, \cdot)$ is indistinguishable from a truly random function $f_1(\cdot) : \{0,1\}^\lambda \to \{0,1\}^\lambda$. The second hybrid argument proceeds similarly where we now use the fact that k_2 is sampled uniformly at random from the keyspace. The claim then follows by the PRF security of F. □

Lemma B.2. *Hybrids* H_1 *and* H_2 *are computationally indistinguishable if* π *is a secure PRP.*

Proof. In hybrid H_1, the keys used by the challenger to evaluate π are all derived from the outputs of the truly random function f_2. Using a sequence of hybrid arguments (one for each PRP key k), we invoke security of the PRP and replace $\pi(k, \cdot)$ with a truly random permutation $\tau_k(\cdot)$ on $[d]$.

Note that we only require a polynomial number of intermediate hybrids in this reduction, since we only need to invoke PRP security for each PRP key k that arises when responding to the adversary's queries. On each chosen message query, to construct the left ciphertexts, the challenger needs to evaluate the PRP π on up to $n = \text{poly}(\lambda)$ different keys (one for each digit in the message). Thus, if the adversary makes q queries, there are at most $qn = \text{poly}(\lambda)$ number of PRP keys that will be used to construct the ciphertexts in the real experiment. We conclude that the number of intermediate hybrids is polynomially-bounded, and so the claim follows from PRP security. □

Lemma B.3. *Hybrids* H_2 *and* H_3 *are statistically indistinguishable if* H *is modeled as a random oracle.*

Proof. The proof of this lemma proceeds very similarly to the proof of Lemma A.3. We argue that each of the abort events (represented by the simulator outputting either $\perp_1$ or $\perp_2$ in hybrid H_2 can only occur with negligible probability.

- **Case 1: The experiment outputs $\perp_1$.** Take any prefix $m_{|i}$ for some $m \in [N]$ and $i \in [n]$ and let $\mu = m_{|i-1} \| \tau_k(m_i)$ from the simulator for the left ciphertexts. Suppose the adversary has not yet queried for an encryption of any message m' where $m'_{|i} = m_{|i}$. Then, we claim that the adversary's view is completely independent of $f_1(\mu)$. Consider the ciphertext $\mathsf{ct}' = (\mathsf{ct}'_{\mathsf{L}}, \mathsf{ct}'_{\mathsf{R}})$ the adversary obtains when it requests an encryption of a message m'.

First, we write $\mathsf{ct}'_{\mathsf{L}}$ as $\mathsf{ct}'_{\mathsf{L}} = (u'_1, \ldots, u'_n)$. Since f_1 is a truly random function, each component u'_j for all $j \neq i$ is completely independent of $f_1(\mu)$. More precisely, the first component of u'_j is an output of f_1 on a different-lengthed prefix and the second component is the output of a random permutation independent of f_1. Finally, consider u'_i. Again, the second component of u'_i is the output of a random permutation independent of f_1 so it suffices to just consider the first component. The first component of u'_i is given by $f_1(m'_{|i-1} \| \tau_{k'}(m'_i))$ where $k' = f_2(m'_{|i-1})$. There are two possibilities. If $m_{|i-1} = m'_{|i-1}$ (so $\tau_k = \tau_{k'}$), then $m_i \neq m'_i$, and so $\tau_k(m'_i) \neq \tau_k(m_i)$. Independence of u'_i and $f_1(\mu)$ then follows from the fact that the outputs of f_1 are independently uniform in $\{0,1\}^\lambda$. If $m_{|i-1} \neq m'_{|i-1}$, then once again, we have that u'_i is independent of $f_1(\mu)$.

Next, we reason about the right ciphertext components $\mathsf{ct}'_{\mathsf{R}} = (r', v'_1, \ldots, r'_n)$. First r' is sampled uniformly at random, and thus, is independent of $f_1(\mu)$. Next, each of the components v'_j for $j \in [n]$ can be written as $\textsc{cmp}(j^*, m'_j) + H(\cdot)$ where j^* ranges over the values in $[d]$ in some order. Certainly, the comparison outputs are independent of f_1 and likewise for the outputs of the random oracle. We conclude that $\mathsf{ct}'_{\mathsf{R}}$ is independent of $f_1(\mu)$.

We have thus shown that as long as the adversary has not queried for an encryption of any message m' where $m_{|i} = m_{|i'}$, its view is independent of $f_1(\mu)$. Now, let $z_1, \ldots, z_\ell$ be the adversary's queries to the random oracle before it requests for an encryption of some m' where $m'_{|i} = m_{|i}$. By our argument above, each of the z_i's is necessarily chosen independently of $f(\mu)$. Since f is a truly random function, the probability that there is some i such that $z_i = (f_1(\mu), y)$ for any y, is at most $\ell/2^\lambda = \mathrm{negl}(\lambda)$, since $\ell = \mathrm{poly}(\lambda)$. Therefore, we conclude that experiment H_3 outputs $\perp_1$ with negligible probability.

- **Case 2: The experiment outputs $\perp_2$.** Let $z_1, \ldots, z_\ell$ for $\ell = \mathrm{poly}(\lambda)$ be the random oracle queries the adversary makes before making its j^{th} encryption query. When constructing the right ciphertext for the j^{th} encryption query, the real experiment samples $r_j \xleftarrow{\text{R}} \{0,1\}^\lambda$. In particular, r_j is independent of $z_1, \ldots, z_\ell$, and so the probability that there is some i such that $z_i = (x, r_j)$ for any x is at most $\ell/2^\lambda = \mathrm{negl}(\lambda)$, and so the experiment outputs $\perp_2$ with negligible probability. The claim follows. □

Lemma B.4. *Hybrid H_3 and H_4 are statistically indistinguishable if H is modeled as a random oracle.*

Proof. Let $(\mathsf{ct}_1, \ldots, \mathsf{ct}_q)$ be the joint distribution of the ciphertexts output in H_3 and let $(\overline{\mathsf{ct}}_1, \ldots, \overline{\mathsf{ct}}_q)$ be the joint distribution of the ciphertexts output in H_4. We show that these two distributions are statistically indistinguishable, and moreover, that the outputs of the random oracle are properly simulated in H_4. The structure of our proof proceeds very similarly to that of Lemma A.3.

Let $m_1, \ldots, m_q$ be the messages chosen by the adversary in H_3 and H_4. Recall that in the simulation, the tables $T_{\text{KEYS}}[\cdot, \cdot]$ are used to maintain the mapping of keys $k \in \{0,1\}^\lambda$ (the inputs to the random oracle) to tuples containing a message index associated with k, along with the (permuted) slot within the block associated with k. This is the analog of the table T_{KEYS} used in the proof of Theorem 3.3. The table T_{RO} is a mapping for the inputs and outputs of the random oracle.

We now proceed via induction on the number of queries q. In each step of the induction, we assume that the following invariants hold for each $t < q$:

- $(\mathsf{ct}_1, \ldots, \mathsf{ct}_t) \equiv (\overline{\mathsf{ct}}_1, \ldots, \overline{\mathsf{ct}}_t)$.

- The outputs of the random oracle queries prior to the $(t+1)^{\text{th}}$ query are statistically indistinguishable in H_3 and H_4.

Consider the base case where $t = 0$. If suffices to argue that all of the random oracle queries are simulated properly in H_4. Suppose the adversary queries the oracle on an input (k, r). Before the adversary makes a single encryption query, the outputs to the adversary's random oracle query are always a uniformly random draw from $\mathbb{Z}_3$ in both H_3 and H_4, which completes the base case.

For the inductive step, suppose that the two conditions hold for some $t < q$. We show that the same conditions hold for $t + 1$. First, we introduce some notation. For all $j \in [t]$, we write $\mathsf{ct}_j = \left(\mathsf{ct}_{\mathsf{L}}^{(j)}, \mathsf{ct}_{\mathsf{R}}^{(j)}\right)$ and similarly, $\overline{\mathsf{ct}}_j = \left(\overline{\mathsf{ct}}_{\mathsf{L}}^{(j)}, \overline{\mathsf{ct}}_{\mathsf{R}}^{(j)}\right)$. Next, we write $\mathsf{ct}_{\mathsf{L}}^{(j)} = \left(u_1^{(j)}, \ldots, u_n^{(j)}\right)$ and $\overline{\mathsf{ct}}_{\mathsf{L}}^{(j)} = \left(\bar{u}_1^{(j)}, \ldots, \bar{u}_n^{(j)}\right)$, where $u_s^{(j)} = \left(k_s^{(j)}, h_s^{(j)}\right)$ and $\bar{u}_s^{(j)} = \left(\bar{k}_s^{(j)}, \bar{h}_s^{(j)}\right)$ for each $s \in [n]$. Finally, we also write $\mathsf{ct}_{\mathsf{R}}^{(j)} = \left(r_j, v_1^{(j)}, \ldots, v_n^{(j)}\right)$ and $\overline{\mathsf{ct}}_{\mathsf{R}}^{(j)} = \left(\bar{r}_j, \bar{v}_1^{(j)}, \ldots, \bar{v}_n^{(j)}\right)$, where $v_s^{(j)} = \left(z_{s,1}^{(j)}, \ldots, z_{s,d}^{(j)}\right)$, and $\bar{v}_s^{(j)} = \left(\bar{z}_{s,1}^{(j)}, \ldots, \bar{z}_{s,d}^{(j)}\right)$.

We begin by showing that the responses to the random oracle queries the adversary makes between its t^{th} and $(t+1)^{\text{th}}$ encryption query are statistically indistinguishable in H_3 and H_4. Let (k, r) be a random oracle query made between the t^{th} and $(t+1)^{\text{th}}$ encryption query by the adversary. We consider several possibilities:

- Suppose in H_3 that $k = k_s^{(\alpha)}$ for some $s \in [n]$, $\alpha \in [t]$ and that $r = r_i \in \{0,1\}^\lambda$ for some $i \in [t]$. In H_4, this corresponds to the case where $k = \bar{k}_s^{(\alpha)}$ and $r = \bar{r}_i$. By construction of the simulation in H_4 this corresponds to the setting where there is some mapping of the form $k \mapsto (\alpha, \beta)$ in $T_{\text{KEYS}}[j, s]$ for some $\beta \in [d]$ and $j \in [q]$.

  In hybrid H_3, since the nonces $r_1, \ldots, r_t \in \{0,1\}^\lambda$ are sampled uniformly at random, they are distinct with overwhelming probability. Moreover, all of the ciphertexts ct_j for $j \neq i$ are constructed independently of r_i. Therefore, the outputs of the random oracle on an input $H(\cdot, r_i)$ are independent of ct_j for all $j \neq i$ with overwhelming probability. Additionally, all of the entries $v_j^{(i)}$ for $j \neq s$ are independent of $k = k_s^{(\alpha)}$ with overwhelming probability (since these keys are derived from the outputs of a truly random function on distinct inputs). We now consider several possibilities:

  - If $\mathsf{ind}_{\mathsf{diff}}^{(d)}(m_\alpha, m_i) < s$, then m_α and m_i do not share a prefix of length s. Note that $v_s^{(i)}$ is independent of k_s^α because m_α and m_i differ on the first $s-1$ bits, and so the keys used to blind the s^{th} block are distinct. We conclude that $k = k_s^{(\alpha)}$ is independent of $\mathsf{ct}_{\mathsf{R}}^{(i)}$, and correspondingly, ct_i. Since k is independent of all the ciphertexts, its value is thus uniformly distributed over $\mathbb{Z}_3$. In H_4, when $\mathsf{ind}_{\mathsf{diff}}^{(d)}(m_\alpha, m_i) < s$, the simulator replies with $\bar{\rho} \xleftarrow{\text{R}} \mathbb{Z}_3$, which is precisely the response in H_3.
  - If $\mathsf{ind}_{\mathsf{diff}}^{(d)}(m_\alpha, m_i) = s$, then m_α and m_i differ at position s. In H_3, this means that

    $$z_{s,\tau_{k'}(m_{i,s})}^{(i)} = \text{CMP}(m_\alpha, m_i) + H(k, r), \tag{B.1}$$

    where $m_{i,s}$ denotes the s^{th} digit of message m_i and k' is f_2 applied to the common prefix (of length $s-1$) of m_i. In H_4, the simulator uses the sets $S_{j,s}$ to maintain the permutation $\tau_{k'}$, and in particular, the value β corresponds to $\tau_{k'}$ on $m_{i,s}$. Thus, in H_4, the simulator's response $\bar{\rho}$ is precisely the value $\bar{\rho}$ such that Eq. (B.1) is satisfied.

– Finally, if $\mathsf{ind}^{(d)}_{\mathsf{diff}}(m_\alpha, m_i) > s$, then m_α and m_i agree on a prefix of length at least s. In H_3, this means that

$$z^{(i)}_{s,\tau_{k'}(m_{i,s})} = H(k, r),$$

where $m_{i,s}$ and k' are defined identically to the previous case. By the same argument as in the previous case, we have that the distributions of the output $H(k, r)$ in hybrids H_3 and H_4 are statistically indistinguishable.

- Otherwise, if $k \neq k^{(\alpha)}_s$ for all $s \in [n]$ and $\alpha \in [t]$ or $r \neq r_i$ for all $i \leq t$, then $H(k, r)$ is independent of all the ciphertexts $\mathsf{ct}_1, \ldots, \mathsf{ct}_t$ given out so far in H_3. In this case, the output of the random oracle is uniform and independent over $\mathbb{Z}_3$. By construction of the simulator, the same holds in H_4. In this case then, the outputs of the random oracle in H_3 and H_4 are identically distributed.

Next, we show that the conditional distributions of $\mathsf{ct}^{(t+1)}_{\mathsf{L}}$ and $\overline{\mathsf{ct}}^{(t+1)}_{\mathsf{L}}$ given the adversary's view in H_3 and H_4, respectively, are statistically indistinguishable. Let m_{t+1} be the adversary's $(t+1)^{\text{th}}$ encryption query. Since the components $u^{(t+1)}_1, \ldots, u^{(t+1)}_n$ and $\bar{u}^{(t+1)}_1, \ldots, \bar{u}^{(t+1)}_n$ in the left ciphertexts are constructed independently in H_3 and H_4, respectively, we reason about each component individually. For each $s \in [n]$, we consider the three possibilities highlighted in the simulation:

- **Case 1:** There exists a $j < t+1$ such that $\mathsf{ind}^{(d)}_{\mathsf{diff}}(m_j, m_{t+1}) > s$. If there are multiple j for which $\mathsf{ind}^{(d)}_{\mathsf{diff}}(m_j, m_{t+1}) > s$, let j be the smallest one.

  By construction of $\mathsf{ORE.Encrypt}_{\mathsf{L}}$, the component $u^{(t+1)}_s$ of the left ciphertext is a function of only the first $t+1$ blocks of the message m_{t+1}. Thus, if there is a message m_j for which the first $t+1$ blocks of m_j and m_{t+1} are identical, then correspondingly, $u^{(t+1)}_s = u^{(j)}_s$. In hybrid H_4, the simulator sets $\bar{u}^{(t+1)}_s = \bar{u}^{(j)}_s$, and so the claim follows from the inductive hypothesis.

- **Case 2:** For each $\ell < t$, $\mathsf{ind}^{(d)}_{\mathsf{diff}}(m_\ell, m_t) \leq s$, and there exists some $j < t$ for which $\mathsf{ind}^{(d)}_{\mathsf{diff}}(m_j, m_t) = s$. If there are multiple j where $\mathsf{ind}^{(d)}_{\mathsf{diff}}(m_j, m_t) = s$, let j be the smallest one.

  In hybrid H_3, $h^{(t+1)}_s = \tau_k(m_{t+1,s})$, where k is derived from f_2 applied to the first $s-1$ blocks of m_{t+1}. But since the first $s-1$ blocks of m_{t+1} match m_i, the index $h^{(t+1)}_s$ is derived from the same permutation used to construct $h^{(j)}_s$. In the simulation, the simulator samples a random index from the set $S_{j,s}$, which is used to lazily sample the permutation τ_k. This is precisely how the simulator lazily samples the permutation π in the proof of Theorem 3.3. In hybrid H_3, the key $k^{(t+1)}_s$ is computed as the output of f_1 on the prefix concatenated with the permuted index. By construction, this is the first time f_1 is evaluated on this input (otherwise, we would be in Case 1), and so the output of f_1 is uniformly and independently distributed. This is how $\bar{h}^{(t+1)}_s$ is sampled in H_4. Finally, both H_3 and H_4 abort (with output $\perp_1$) if the adversary has already queried the random oracle on $h^{(t+1)}_s$ and $\bar{h}^{(t+1)}_s$, respectively, as required.

- **Case 3:** For each $\ell < t$, $\mathsf{ind}^{(d)}_{\mathsf{diff}}(m_\ell, m_t) < s$.

  In H_3, $h^{(t+1)}_s = \tau_k(m_{t+1,s})$, where k is derived from f_2 applied to the first $s-1$ blocks of m_{t+1}. But since the first $s-1$ blocks differ from those of all other messages, this is the first time τ_k

is evaluated on *any* input, and so $h_s^{(t+1)}$ is distributed uniformly over $[d]$. Similarly, the key $k_s^{(t+1)}$ is computed as the output of f_1 on a unique input (not appearing in any of the previous queries), and so the output of f_1 is also uniformly distributed. In H_4, $\mathcal{S}_t$ samples $\bar{h}_s^{(t+1)}$ uniformly from $[d]$ and $\bar{k}_s^{(t+1)}$ uniformly from $\{0,1\}^\lambda$. Thus, the components $(k_s^{(t+1)}, h_s^{(t+1)})$ and $(\bar{k}_s^{(t+1)}, \bar{h}_s^{(t+1)})$ are identically distributed in this case. Finally, both H_3 and H_4 aborts with output $\perp_1$ if the adversary has already queried the random oracle on $h_s^{(t+1)}$ and $\bar{h}_s^{(t+1)}$.

We conclude from the above case analysis that the conditional distribution of the left ciphertexts in hybrids H_3 and H_4 is statistically indistinguishable.

To conclude the proof, we argue that the right ciphertext components are statistically indistinguishable in H_3 and H_4. Certainly r_{t+1} and $\bar{r}_{t+1}$ are identically distributed. By construction, in H_3 and H_4, the adversary must never have queried the random oracle on an input containing r_{t+1} and $\bar{r}_{t+1}$ (otherwise, the experiment aborts with output $\perp_2$). But now, each component in $\mathsf{ct}_{\mathsf{R}}^{(t+1)}$ and $\overline{\mathsf{ct}}_{\mathsf{R}}^{(t+1)}$ is blinded by the output the random oracle on an input containing r_{t+1} or $\bar{r}_{t+1}$. Thus, conditioned on the view of the adversary up to the point it issues its $t+1^{\text{th}}$ encryption query, in H_3, the components of $\mathsf{ct}_{\mathsf{R}}^{(t+1)}$ are perfectly hidden by the outputs of the random oracle, and thus, appear independently and uniformly random over $\mathbb{Z}_3$. This is precisely the distribution from which the simulator samples the elements of $\mathsf{ct}_{\mathsf{R}}^{(t+1)}$ in H_4. We conclude that the right ciphertexts are properly distributed in H_3 and H_4. The lemma now follows by induction on t. □

Combining Lemmas B.1 through B.4, we conclude that Π_{ore} is secure with leakage function $\mathcal{L}_{\text{BLK}}$. □