Deterministic Public-Key Encryption Revisited

Adam O'Neill[*]

University of Texas at Austin

Abstract

This paper revisits the notion of determinsitic public-key encryption (DE) introduced by Bellare, Boldyreva, and O'Neill (CRYPTO 2007) and further studied by Bellare et al. (CRYPTO 2008), Boldyreva et al. (CRYPTO 2008), and Hemenway et al. (ePrint 2010). First, we propose a general construction of DE from any trapdoor function admitting a special hardcore function we call *robust*, whose security proof is enabled by a more precise definitional equivalence for DE we give as compared to prior work. (We believe our notion of robustness for hardcore functions is a natural one that may also find additional applications.) Most known constructions of DE, both in the random oracle and standard models, can be explained as instantiations of our general construction or optimizations thereof; additionally, we obtain instantiations from assumptions not previously known to yield DE. Second, while achieving the strongest security notion for DE introduced by Bellare, Boldyreva, and O'Neill without random oracles remains open, we propose and achieve stronger notions of security in this context than prior work. Specifically, we propose a notion of "q-bounded" multi-message security for DE, and, following Boldyreva et al., extend it to a notion of unbounded multi-message security for what we call "q-block sources." We show how it can be achieved from any lossy trapdoor function that loses a $1 - o(1)$ fraction of its input.

Keywords: Deterministic encryption, trapdoor functions, hardcore functions, lossy trapdoor functions, q-bounded security.

[*] Email: `adamo@cs.utexas.edu`. Work done in part while the author was a Ph.D. student at Georgia Institute of Technology.

1 Introduction

We start with some background and motivation for our work and then overview our results in more detail.

1.1 Background and Motivation

One of the maxims of modern cryptography is that "randomness is needed for good encryption." In other words, the encryption algorithm of a scheme should be randomized. Such an encryption scheme is called *probabilistic*. Indeed, probabilistic encryption is necessary to meet the fundamental notion of semantic security introduced by Goldwasser and Micali [19]. An interesting question, however, is whether we can instead leverage *message entropy* to achieve security, rather than requiring the encryption to be probabilistic. To the best of our knowledge, the encryption of high-entropy messages was first considered by Russell and Wang [31] (and follow-up work by Dodis and Smith [14]) in the setting of one-time, information-theoretically secure symmetric-key encryption. (However, their motivation was somewhat different; namely, they showed that Shannon's famous lower-bound on the key-size for a one-time pad could be circumvented for high-entropy messages.)

The focus of our work is on the above question in the *public-key* (and thus computationally bounded) setting. In this setting, the encryption of high-entropy messages was first considered by Bellare, Boldyreva, and O'Neill [2]. Specifically, they proposed a notion of security for *deterministic* public-key encryption (DE) that essentially guarantees semantic security for high-entropy messages, and showed how to achieve it in the random oracle (RO) model of [5], in particular based on any semantically secure probabilistic scheme.[1] Subsequent works by Bellare et al. [4] and Boldyreva et al. [7] provided alternative security definitions and definitional equivalences for DE (building on [14, 10]), as well as constructions without random oracles (achieving somewhat weaker security in terms of allowed correlations for multiple messages). In particular, the former gave constructions from trapdoor permutations (that are one-way for high-entropy input distributions), and the latter gave constructions from the recent notion of lossy trapdoor functions by Peikert and Waters [28]. More recently, Hemenway et al. [23] showed that a "decisional" version of the notion of correlated product trapdoor function of Rosen and Segev [30] suffices.

Besides being interesting from a foundational standpoint, DE has a number of practical applications, such as efficient search on encrypted data (cf. [2]) and securing legacy protocols (made possible by the fact that DE allows length-preserving encryption). Furthermore, Bellare et al. [3] showed how it extends to a notion of "hedged" public-key encryption that reduces dependence on good randomness for probabilistic encryption more generally, and Dent et al. [9] adapted it to a notion of confidentiality for digital signatures, further testifying to the usefulness of the notion. As such, we would like to have a solid understanding of it, both in terms of how it can be constructed and what level of security it can achieve. Our work here makes contributions on both fronts. Namely, our first main contribution is to provide a general construction of DE from certain kinds of trapdoor functions that both explains existing constructions and leads to new ones, from assumptions not previously known to yield DE (namely exponentially-hard one-way trapdoor functions). Our second is to propose a new notion of "q-bounded" multi-message security for DE and show how it can be achieved without random oracles, based on lossy trapdoor functions that lose a $1 - o(1)$ fraction of their input. The notion is stronger than those previously achieved by standard-model DE. We next overview our results in more detail.

1.2 Our Results

A MORE PRECISE DEFINITIONAL EQUIVALENCE. The starting point of our work is to revisit the definitional equivalences for DE proven in [4] and [7]. At a high-level, they showed that the semantic-security style definition for DE (called PRIV) introduced in the intial work of [2] is equivalent to an indistinguishability-based "distribution hiding"

[1] Actually, one should note that Russell and Wang [31] did observe using their schemes in hybrid encryption would result in (probabilistic) public-key encryption schemes semantically secure for high-entropy messages. Thus, more accurately [2] were the first to consider the encryption of high-entropy messages in the public-key setting in its full generality and in particular to formulate a security notion which could be met by deterministic schemes.

notion for DE (called IND), which asks that the adversary cannot distinguish ciphertexts whose corresponding plaintexts are drawn from one of two possible distributions. Notice, while PRIV can be meaningfully said to hold for a given plaintext distribution, IND inherently talks of *pairs* of distributions on the plaintext space. The works of [4, 7] compensated for this by formulating their equivalences in terms of *entropy levels*. That is, they showed that PRIV for all plaintext distributions of (min-)entropy μ is equivalent to indistinguishability with respect to all pairs of plaintext distributions of entropy slightly less than μ. Ideally, however, we would like to be able to identify, for a *fixed* plaintext distribution X, a class of pairs of plaintext distributions on which IND is equivalent to PRIV on X. Via a careful re-examination of the equivalence proof of [4], we do exactly that. Namely, we show that PRIV on X is equivalent to IND on the class of pairs of *complementary induced distributions* of X, meaning each pair in the class is obtained by conditioning X on whether some (efficiently testable) event E, where both E and its complement occur with good (in fact constant) probability, happens or not. Besides being more technically precise, this equivalence enables simple and modular security proofs (as we will see).

A CONSTRUCTION FROM "ROBUST" HARDCORE FUNCTIONS. Now, consider the following natural construction of DE (which may be viewed as an extension of the "Encrypt-with-Hash" scheme of [2]). We start with a trapdoor function f and a semantically secure probabilistic encryption scheme $\Pi = (\mathcal{K}, \mathcal{E}, \mathcal{D})$. Suppose that furthermore f has a hardcore function h on input distribution X such that $h(x)$ has the same length as the number of coins consumed by $\mathcal{E}(pk, \cdot)$ on input x. Define the "Encrypt-with-Hardcore" DE scheme to encrypt plaintext x as the encryption of $f(x)$ under $\mathcal{E}$ using coins $h(x)$. (In other words, the encryption of x is $\mathcal{E}(pk, f(x); h(x))$; decryption is defined naturally.) Due to the pseudorandomness of $h(x)$, one might expect PRIV security of this scheme on X to follow. However, while $h(x)$ is guaranteed to look random given $f(x)$, that does not mean that the former does not leak any *partial information* on x (for example, h may be a "physical" hardcore function, i.e., output some of the input bits), and, since Π may not protect partial information about its coins, such information can propogate through. Thus, some extra condition on h is needed to ensure security of the resulting scheme. What is it? Our precise definitional equivalence discussed above reveals the answer. Namely, what does follow is *IND* security of the scheme on X. Thus, the condition needed is a new notion of "robustness" for hardcore functions that we introduce; intuitively, this means h remains hardcore on sub-distributions of X whose entropy is slightly reduced, by conditioning on an (efficiently testable) event that occurs with good probability. Indeed, in this case PRIV security of the Encrypt-with-Hardcore scheme on X follows via our precise definitional equivalence.

INSTANTIATIONS FROM ONE-WAYNESS. Our security analysis of the Encrypt-with-Hardcore scheme reduces the task of building a DE scheme for an input distribution to that of finding a trapdoor function with a large robust hardcore function for it.[2] Consider for now just a single high-entropy input. We explore how such hardcore functions may be obtained based on one-wayness of a trapdoor function on the latter.

In the RO model, any trapdoor function that is one-way on a given input distribution has a simple, large hardcore function, namely the RO hash itself. This follows because (1) any one-way function is also one-way on such induced input distributions and (2) a RO hash is a hardcore function for any one-way function, regardless of the input distribution on which it is one-way. This leads to an instantiation of the Encrypt-with-Hardcore scheme in the RO model that, as shown in [2], can be optimized to not use the "inner" trapdoor function at all.

To obtain robust hardcore functions in the standard model, we consider the well-known Goldreich-Levin (GL) bit [18], where the description of the trapdoor function is augmented to contain public coins r and the GL bit of an input x is the inner product of x with r.[3] Recall that [18] showed this bit is similarly hardcore for any one-way function, regardless of the input distribution. Robustness then follows by the same argument as above. This leads to two instantiations of the Encrypt-with-Hardcore scheme from one-wayness in the standard model, one based on trapdoor permutations and one based on trapdoor functions. In the former instantiation, in order to instantiate the

[2] In order to ensure the latter produces enough coins for the "outer" encryption scheme, we can use the output of the hardcore function as a seed for a pseudorandom generator to expand its length.

[3] Indeed, in the standard model no hardcore function h that does not use public coins can be robust. Intuitively, this is because otherwise we could find an induced distribution on which the function is no longer hardcore, for example restricting to the set of inputs x such that the first bit of $h(x)$ is 1.

Encrypt-with-Hardcore scheme on plaintext distribution X, we need to assume a trapdoor permutation that is one-way on all *permutation distributions* of X. Then we can use Blum-Micali-Yao iteration [6, 32], extracting a GL bit on each iteration. (Thus, the trapdoor function we use to instantiate the scheme is the iterated trapdoor permutation.) Here we obtain exactly the scheme of [4]. In the latter instantiation, in order to instantiate the Encrypt-with-Hardcore scheme on plaintext distribution X, we need to assume a trapdoor function that is very (i.e., exponentially) hard to invert on X (note that this is only possible if X has linear entropy). Then, Goldreich and Levin's result tells us the trapdoor function has a *large* hardcore function (which basically uses matrix instead of vector multiplication). Furthermore, it is robust because if a one-way function is very hard to invert on a given input distribution then it is very hard to invert on induced distributions with slightly less entropy. This latter instantiation is new to this work.

We make some remarks about the above two standard-model instantiations of the Encrypt-with-Hardcore scheme. First, their respective assumptions, namely a trapdoor permutation that is one way on all permutations of a given distribution, or a trapdoor function that is very hard to invert on the distribution, seem incomparable.[4] On the other hand, these assumptions are implied (respectively) by one-way permutations that are sufficiently hard to invert and one-way trapdoor functions that are sufficiently hard to invert, on the standard *uniform* distribution, using [16, Lemma 4]. (Of course, we require more "hardness" in the latter case since the trapdoor function must not only be one-way on input distributions of high entropy but also have a large GL hardcore function for them.) Finally, although the instantiation we obtain from trapdoor permutations is known, our approach enables a simpler and more modular proof of security as compared to [4]. The reason is that our more precise definitional equivalence allows making easy use of the IND definition in this case.

INSTANTIATIONS FROM LOSSINESS. Another example of a standard-model robust hardcore function comes from the theory of lossy trapdoor functions put forth by Peikert and Waters [28]. Recall that lossy trapdoor functions have a public description indistinguishable from that of a lossy function. Peikert and Waters showed that a lossy trapdoor function admits a simple, large hardcore function (whose output length is slightly less than the lossiness of the trapdoor function), namely a universal hash function. This follows from the fact that a universal hash function is a randomness extractor as per the leftover hash lemma [22] (and its average-case extension in [12]). Notice that (1) randomness extractors simply require a high-entropy input and (2) sub-distributions induced by an event that occurs with good probability also have high min-entropy if the starting distribution does. It follows that the hardcore function is again robust in this case (on any input distribution with sufficient entropy). We thus obtain an instantiation of the Encrypt-with-Hardcore scheme from any lossy trapdoor function losing a constant fraction of its input. Note that [7] obtained DE from such lossy trapdoor functions by making use of randomness extraction properties of (an augmented version of) the lossy trapdoor function directly. (More specifically, they first pre-process the input to the lossy trapdoor function with a pairwise-independent permutation.) Their construction can in some sense be viewed as an optimization of the scheme we obtain here that drops the "outer" encryption.

One can see the parallels between the argument for robustness of a universal hash in the case of a lossy trapdoor function and of the Goldreich-Levin function in the case of a one-way trapdoor function. We thus obtain a unified view of how (single-message) DE can be constructed under these assumptions.

RELATION TO DECISIONAL CORRELATED PRODUCT. We briefly discuss the relation of our general construction to the recent one of Hemenway et al. [23] from "decisional" 2-correlated product trapdoor functions. Essentially, these are trapdoor function families $\mathcal{F}$ for which $f_1(x_1), f_2(x_2)$ where x_1, x_2 are equal is indistinguishable from $f_1(x_1), f_2(x_2)$ where x_1, x_2 are sampled independently (for two independent public instances f_1, f_2 of $\mathcal{F}$). They show such a trapdoor function is a secure DE secure for uniform messages. It seems plausible to show that such $\mathcal{F}$ has a large robust hardcore function and thus can be used to instantiate the Encrypt-with-Hardcore scheme as well. Namely, we would augment $\mathcal{F}$ so that its public description contains f_1, f_2 sampled independently from $\mathcal{F}$ and a key K for a strong randomness extractor H. (Note that only f_1 is used for evaluation and invertion, in particular we do not need the trapdoor for f_2.) The hardcore function $\mathsf{hc}(x)$ would be defined as $H(K, f_2(x))$. It is clear that this

[4] This is the case even when we are interested in security on all distributions of a given entropy level, and hence we assume a trapdoor permutation that is (super-polynomially) one-way for all such distributions or a the trapdoor function is exponentially-hard to invert on all such distributions.

function is hardcore, but showing robustness is more subtle. For this, one might apply the techniques of [15, Lemma 3]. However, due to the non-standard nature of the assumption we have not worked out the details. (The authors of [23] only construct decisional correlated product functions without a trapdoor.)

SECURITY FOR MULTIPLE MESSAGES. An important point is that, as in [4, 7], we are only able to prove the above standard-model DE schemes secure for the encryption of a *single* high-entropy plaintext, or, what was shown equivalent in [7] when assuming single-message security for *all* distributions of a given entropy level (and under a conditional re-sampleability condition or by considering inefficient plaintext distributions[5]), an unbounded number of messages drawn from a *block source* (where each subsequent message meets the entropy level even conditioned on the previous messages). On the other hand, the strongest security model for DE introduced by [2] considers the encryption of an unbounded number of plaintexts that have *individual* high entropy but may not have any conditional entropy. They showed this notion is achievable in the RO model (in particular via the above-mentioned "Encrypt-with-Hash" scheme), leaving a large gap between what is achievable there and what is (known to be) achievable in the standard model. Note that the Encrypt-with-Hardcore scheme reduces building such stronger schemes in the standard model to finding robust hardcore functions that are "correlation secure" for arbitrary correlations among multiple inputs. We do not know of a construction of even such correlation-secure hardcore *bits* (even disregarding robustness).

BOUNDED MULTI-MESSAGE SECURITY. To help bridge this gap, we propose a natural notion of "q-bounded" multi-message security for DE, where up to q high entropy but arbitrarily correlated messages may be encrypted under the same public key. Following [7], we also consider a notion of unbounded multi-message security where the messages are drawn from what we call a "q-block source." Essentially, this is a block source where the "blocks" are of size q, and within each block the messages may be arbitrarily correlated (but have individual high entropy). Theorem 4.2 of [7] extends to this context to show (under similar restrictions to those mentioned above) that q-bounded multi-message security for a given entropy level and unbounded multi-message security for q-block sources are equivalent. While still weaker than the PRIV definition for unbounded messages achievable in the RO model, we feel this notion may be helpful in practice (indeed, it seems hard to guarantee in a given application that not even a small number of messages will have low conditional entropy). Additionally, it helps reveal where the technical "gap" between the RO and standard model constructions of DE really lies and should help guide future research.

IMPROVED CONSTRUCTIONS FROM LOSSY TRAPDOOR FUNCTIONS. Finally, we show constructions of DE meeting q-bounded multi-message security for any polynomial q, based on a lossy trapdoor function that loses a $1 - o(1)$ fraction of its input. (They can also be proven directly to meet the extension to q-block sources.) Such constructions are known from the decisional Diffie-Hellman [28] and the decisional composite residuosity [7, 17] assumptions.) In other words, it is when considering q-bounded multi-message security that parallels between constructions from lossy and one-way trapdoor functions break down and we exploit the stronger power of lossiness.

Our constructions here are inspired by an extension to the classical leftover hash lemma (LHL) given recently by Kiltz et al. [25, Lemma 3.2] to 4-wise independent hashing. Namely, their extension says that a 4-wise independent hash smoothes out two *correlated* but individually high entropy sources to two independent and uniform samples. Their lemma in fact generalizes to q such sources by using a $2q$-wise independent hash. This leads to an instantiation of the Encrypt-with-Hardcore scheme from lossy trapdoor functions that achieves q-bounded multi-message security, using a $2q$-wise independent hash function as the hardcore function. (That is, the latter is "correlation secure" for up to q inputs. The reason we need the lossiness to be $1 - o(1)$ is that the error bound in the generalization of the LHL is proportional to q *times* the output range of the function.) Moreover, we show how to optimize this instantiation by following [7] and considering an analogous extension to the "Crooked" LHL of Dodis and Smith [14]. The idea here is to get a direct construction from the lossy trapdoor function by first pre-processing an input plaintext with a $2q$-wise independent permutation (instead of pairwise as in [7]). We show q-bounded multi-message security of the resulting construction but with a much lower entropy requirement on the input. The catch is that permutations with independence greater than 3-wise are not known to exist. Fortunately, Kaplan et al. [24] give good constructions of

[5] For constructions based on lossy trapdoor function these restrictions are immaterial since they are actually secure for inefficient plaintext distributions. We also observe that this is true for some instantiations based on (standard) one-wayness; see Section 5.1.

almost $2q$-wise independent permutations that we show suffices for our purposes.

2 Preliminaries

An adversary is either an algorithm or a tuple of algorithms. Unless otherwise indicated, an adversary or algorithm may be randomized and must run in probabilistic polynomial-time (PPT) in its input size. In the case of a tuple of algorithms, each constituent must be PPT. By convention, the running-time of an adversary includes both its actual running-time and the time to run its overlying experiment. The security parameter is denoted by k, and 1^k denotes the string of k ones. We often surpress dependence of variables on k for readability. A function $f\colon \mathbb{N} \to [0,1]$ is called negligible if it approaches zero faster than any inverse polynomial.

ALGORITMIC NOTATION. If A is an algorithm then $x \overset{\$}{\leftarrow} A(\ldots)$ denotes that x is assigned the output of running A on the elided inputs and a fresh random tape, while if S is a finite set then $s \overset{\$}{\leftarrow} S$ denotes that s is assigned a uniformly random element of S. We often use the abbreviation $x_1, \ldots, x_n \overset{\$}{\leftarrow} A(\ldots)$ for $x_1 \overset{\$}{\leftarrow} A(\ldots)\ ;\ \ldots\ ;\ x_n \overset{\$}{\leftarrow} A(\ldots)$, for any $n \in \mathbb{N}$, and similarly for sets. If A is deterministic then we drop the dollar sign above the arrow. We let $A(\ldots) \Rightarrow y$ denote the event that A outputs y in the above experiment.

STRINGS. We denote by $\{0,1\}^*$ the set of all (binary) strings, and by $\{0,1\}^n$ the set of strings of length n, for any $n \in \mathbb{N}$. By $x_1 \| \cdots \| x_n$ we denote an encoding of strings $x_1, \ldots, x_n$ from which $x_1, \ldots, x_n$ are uniquely recoverable. We denote by $x \oplus y$ the bitwise exclusive-or (xor) of equal-length strings x, y. An n-bit string may also be interpreted as an n-dimensional vector over $GF(2)$. In particular, for two n-bit strings x, y we denote by $\langle x, y \rangle$ the inner-product of x and y over $GF(2)$.

VECTORS. Vectors are denoted in boldface, for example $\mathbf{x}$. If $\mathbf{x}$ is a vector then $|\mathbf{x}|$ denotes the number of components of $\mathbf{x}$ and $\mathbf{x}[i]$ denotes its ith component, for $1 \leq i \leq |\mathbf{x}|$. For convenience, we extend algorithmic and functional notation to operate on a vector of inputs component-wise. That is, if A is an algorithm and $\mathbf{x}, \mathbf{y}$ are vectors then $\mathbf{z} \overset{\$}{\leftarrow} A(\ldots, \mathbf{x}, \ldots, \mathbf{y}, \ldots)$ denotes that $\mathbf{z}[i] \overset{\$}{\leftarrow} A(\ldots, \mathbf{x}[i], \ldots, \mathbf{y}[i], \ldots)$ for all $1 \leq i \leq |\mathbf{x}|$, where the elided inputs are fixed across all invocations.

STATISTICAL NOTIONS. Let X be a random variable on a finite domain $\mathcal{X}$. We write P_X for the distribution of random variable X and $P_X(x)$ for the probability that X puts on value $x \in \mathcal{X}$, i.e., $P_X(x) = \mathrm{P}[X = x]$. We often identify X with P_X when there is no danger of confusion. By $x \overset{\$}{\leftarrow} X$ we denote that x is assigned a value drawn according to X. When this experiment is PPT we say that X is *efficiently sampleable*. We write $X \mid E$ for the conditional distribution of X on an event E with the same domain. The *min-entropy* of X is $\mathrm{H}_\infty(X) = -\log(\max_x P_X(x))$. The *(worst-case) conditional min-entropy* of X given Y is $\mathrm{H}_\infty(X|Y) = -\log(\max_{x,y} P_{X|Y=y}(x))$. The *average conditional* min-entropy of X given Y [12] is $\tilde{\mathrm{H}}_\infty(X|Y) = -\log(\sum_y P_Y(y) \max_x P_{X|Y=y}(x))$. The *collision probability* of X is $\sum_x P_X(x)^2$. The *statistical distance* between random variables X and Y with the same domain is $\Delta(X,Y) = \frac{1}{2}\sum_x |P_X(x) - P_Y(x)|$. If $\Delta(X,Y)$ is negligible then we say X and Y are *statistically close*. The *square of the 2-distance* is between X and Y is $D(X,Y) = \sum_x \big(P_X(x) - P_Y(x)\big)^2$.

t-WISE INDEPENDENT FUNCTIONS. Let $F\colon \mathcal{K} \times D \to R$ be a function. We say that F is *t-wise independent* if for all distinct $x_1, \ldots, x_t \in D$ and all $y_1, \ldots, y_t \in R$

$$\Pr\Big[\, F(K, x_1) = y_1 \ \wedge\ \ldots\ \wedge\ F(K, x_t) = y_t \ :\ K \overset{\$}{\leftarrow} \mathcal{K} \,\Big] \ = \ \frac{1}{|R|^t}\ .$$

In other words, $F(K, x_1), \ldots, F(K, x_t)$ are all uniformly and independently random over R.

PUBLIC-KEY ENCRYPTION. A *public-key encryption scheme* with plaintext-space PtSp is a triple of algorithms $\Pi = (\mathcal{K}, \mathcal{E}, \mathcal{D})$. The key-generation algorithm $\mathcal{K}$ takes input 1^k to return a public key pk and matching secret key sk. The encryption algorithm $\mathcal{E}$ takes pk and a plaintext m to return a ciphertext. The deterministic decryption algorithm

$\mathcal{D}$ takes sk and a ciphertext c to return a plaintext. We require that for all plaintexts $m \in \mathrm{PtSp}$

$$\Pr\Big[\, \mathcal{D}(sk, \mathcal{E}(pk, m)) = m \;:\; (pk, sk) \xleftarrow{\$} \mathcal{K}(1^k) \,\Big] \;=\; 1 \;.$$

Next we define semantic security aka. security against chosen-plaintext attack [19]. To an encryption scheme $\Pi = (\mathcal{K}, \mathcal{E}, \mathcal{D})$, an adversary $A = (A_1, A_2)$, and $k \in \mathbb{N}$ we associate

```
**Experiment** $\mathbf{Exp}^{\text{ind-cpa}}_{\Pi,A}(k)$:
$b \xleftarrow{\$} \{0,1\}$ ; $(pk, sk) \xleftarrow{\$} \mathcal{K}(1^k)$
$(m_0, m_1, state) \xleftarrow{\$} A_1(pk)$
$c \xleftarrow{\$} \mathcal{E}(pk, m_b)$
$d \xleftarrow{\$} A_2(pk, c, state)$
If $d = b$ return 1 else return 0
```

where we require A_1's output to satisfy $|m_0| = |m_1|$. Define the *IND-CPA advantage* of A against Π as

$$\mathbf{Adv}^{\text{ind-cpa}}_{\Pi,A}(k) = 2 \cdot \Pr\Big[\, \mathbf{Exp}^{\text{ind-cpa}}_{\Pi,A}(k) \Rightarrow 1 \,\Big] - 1 \;.$$

We say that Π is *IND-CPA secure* if $\mathbf{Adv}^{\text{ind-cpa}}_{\Pi,A}(\cdot)$ is negligible for any PPT adversary A.

LOSSY TRAPDOOR FUNCTIONS. A *lossy trapdoor function (LTDF) generator* [28] is a pair $\mathsf{LTDF} = (\mathcal{F}, \mathcal{F}')$ of algorithms. Algorithm $\mathcal{F}$ is a usual trapdoor function generator, namely on inputs 1^k outputs outputs (a description of a) function f on 1^k along with (a description of) its inverse f^{-1}, and algorithm $\mathcal{F}'$ outputs a (description of a) function f' on $\{0,1\}^k$. (More generally, a trapdoor function may have domain $\{0,1\}^n$ for a polynomial $n = n(k)$.) For a distinguisher D, define its *LTDF advantage* against LTDF as

$$\mathbf{Adv}^{\text{ltdf}}_{\mathsf{LTDF},D}(k) = \Pr\Big[\, D(f) \Rightarrow 1 \;:\; (f, f^{-1}) \xleftarrow{\$} \mathcal{F} \,\Big] - \Pr\Big[\, D(f') \Rightarrow 1 \;:\; f' \xleftarrow{\$} \mathcal{F}' \,\Big] \;.$$

We say that LTDF is *lossy* if $\mathbf{Adv}^{\text{ltdf}}_{\mathsf{LTDF},D}(\cdot)$ is negligible for any PPT D. We say LTDF has *residual leakage s* if for all f' output by $\mathcal{F}'$ we have $|R(f')| \leq 2^s$. The *lossiness* of LTDF is $\ell = k - s$.

3 Deterministic Encryption and its Precise Definitional Equivalence

We recall the notion of deterministic encryption and two security notions for it that have been introduced. We then give a more precise equivalence between these definitions than in prior work.

3.1 Deterministic Encryption and its Security

We say that an encryption scheme $\Pi = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ is *deterministic* if $\mathcal{E}$ is deterministic.

SEMANTIC SECURITY. We recall the semantic-security style PRIV notion for DE from [2].[6] To encryption scheme $\Pi = (\mathcal{K}, \mathcal{E}, \mathcal{D})$, an adversary $A = (A_0, A_1, A_2)$, and $k \in \mathbb{N}$ we associate

```
**Experiment** $\mathbf{Exp}^{\text{priv}}_{\Pi,A}(k)$:
$b \xleftarrow{\$} \{0,1\}$ ; $state \xleftarrow{\$} A_0(1^k)$
$(\mathbf{x}_0, t_0), (\mathbf{x}_1, t_1) \xleftarrow{\$} A_1(state)$
$\mathbf{c} \xleftarrow{\$} \mathcal{E}(pk, \mathbf{x}_b)$
$g \xleftarrow{\$} A_2(pk, \mathbf{c}, state)$
If $g = t$ return 1 else return 0
```

[6] More specifically, it is a "comparison-based" semantic-security style notion. This was shown equivalent to a "simulation-based" formulation in [4].

We require A_1's output to satisfy $|\mathbf{x}_0| = |\mathbf{x}_1|$ and $|\mathbf{x}_0[i]| = |\mathbf{x}_1[i]|$ for all i. Moreover, we require that $\mathbf{x}_0[i_1] = \mathbf{x}_0[i_2]$ if and only if $\mathbf{x}_1[i_1] = \mathbf{x}_1[i_2]$ for all i_1, i_2. (This reflects the fact that deterministic encryption leaks the equality pattern of the plaintexts; in fact, when the encryption scheme is deterministic we may assume without loss of generality that all the $\mathbf{x}_0[i]$, and similarly the $\mathbf{x}_1[i]$, are always distinct.) Define the *PRIV advantage* of A against Π as

$$\mathbf{Adv}^{\mathrm{priv}}_{\Pi,A}(k) = 2 \cdot \Pr\left[\, \mathbf{Exp}^{\mathrm{priv}}_{\Pi,A}(k) \Rightarrow 1 \,\right] - 1 \; .$$

Let $\mathbb{M}$ be a class of distributions on message vectors. Define $\mathbb{A}_{\mathbb{M}}$ to be the class of adversaries $\{A = (A_0, A_1, A_2)\}$ such that for each $A \in \mathbb{A}_{\mathbb{M}}$ there is a $\boldsymbol{M} \in \mathbb{M}$ for which $\mathbf{x}$ has distribution $\boldsymbol{M}$ over $(\mathbf{x}, t) \xleftarrow{\$} A_1(state)$ for any $state$ output by A_0. We say that Π is *PRIV secure for* $\mathbb{M}$ if $\mathbf{Adv}^{\mathrm{priv}}_{\Pi,A}(\cdot)$ is negligible for any PPT $A \in \mathbb{A}_{\mathbb{M}}$. Note that we can without loss of generality consider only those A with "empty" A_0, since A_1 can always be hardwired with the "best" state. However, following [4] we explicitly allow state because it greatly facilitates some proofs.

INDISTINGUISHABILITY. Next we recall the indistinguishability-based formulation of security for DE given (independently) by [4, 7] (and which is adapted from [14]). To an encryption scheme $\Pi = (\mathcal{K}, \mathcal{E}, \mathcal{D})$, an adversary $D = (D_1, D_2)$, and $k \in \mathbb{N}$ we associate

```
**Experiment** $\mathbf{Exp}^{\mathrm{ind}}_{\Pi,A}(k)$:
$b \xleftarrow{\$} \{0,1\}$ ; $(\mathbf{x}, t) \xleftarrow{\$} D_1(b)$
$\mathbf{c} \xleftarrow{\$} \mathcal{E}(pk, \mathbf{x})$
$d \xleftarrow{\$} D_2(pk, \mathbf{c})$
If $b = d$ return 1 else return 0
```

As before, we require D_1's output to satisfy $|\mathbf{x}_0| = |\mathbf{x}_1|$ and $|\mathbf{x}_0[i]| = \mathbf{x}_1[i]$ for all i. Moreover, we require that $\mathbf{x}_0[i_1] = \mathbf{x}_0[i_2]$ if and only if $\mathbf{x}_1[i_1] = \mathbf{x}_1[i_2]$ for all i_1, i_2. (Again, this reflects the fact that deterministic encryption leaks the equality pattern of the plaintexts.) Define the *IND advantage* of D against Π as

$$\mathbf{Adv}^{\mathrm{ind}}_{\Pi,D}(k) = 2 \cdot \Pr\left[\, \mathbf{Exp}^{\mathrm{ind}}_{\Pi,D}(k) \Rightarrow 1 \,\right] - 1 \; .$$

Let $\mathbb{M}^*$ be a class of *pairs* of distributions on message vectors. Define $\mathbb{D}_{\mathbb{M}^*}$ to be the class of adversaries $\{D = (D_1, D_2)\}$ such that for each $D \in \mathbb{D}_{\mathbb{M}^*}$, there is a pair of distributions $(\boldsymbol{M}_0, \boldsymbol{M}_1) \in \mathbb{M}^*$ such that for each $b \in \{0,1\}$ the distribution of $\mathbf{x} \xleftarrow{\$} D_1(b)$ is $\boldsymbol{M}_b$. We say that Π is *IND secure for* $\mathbb{M}^*$ if $\mathbf{Adv}^{\mathrm{ind}}_{\Pi,D}(\cdot)$ is negligible for any PPT $D \in \mathbb{D}_{\mathbb{M}^*}$.

3.2 A Precise Definitional Equivalence

Notice that, while the PRIV definition is meaningful with respect a single message distribution $\boldsymbol{M}$, the IND definition must inherently talk of *pairs* of different message distributions. Thus, in proving an equivalence between the two notions, the best we can hope to show is that PRIV security for a message distribution $\boldsymbol{M}$ is equivalent to IND security for some *class of pairs* of message distributions (depending on $\boldsymbol{M}$). However, prior works [4, 7] fell short of providing such a statement. Instead, they showed that PRIV security on *all* distributions of a given entropy μ is equivalent to IND security on all pairs of distributions of slightly less entropy.

INDUCED DISTRIBUTIONS. To state our result we first give some definitions relating to the notion of *induced distributions*. Let X, X' be distributions with a common domain. For $\alpha \in \mathbb{N}$, we say that X' is an *α-induced (sub-)distribution of* X if X' is a conditional distribution $X' = X \mid E$ for an event E such that $\Pr[\,E\,] \geq 2^{-\alpha}$. We call E the *corresponding event* for X'. We require that E is efficiently testable given an outcome of X, or more generally that the pair (X, E) is efficiently sampleable (where we view event E as a binary random variable). We denote by $X[\alpha]$ the class of all α-induced distributions of X. Furthermore, let X_0, X_1 be two α-induced distributions of X with corresponding events E_0, E_1 respectively. We say that X_0, X_1 are *complementary* if $E_1 = \overline{E_0}$. We denote by $X^*[\alpha]$ the class of all pairs of complementary α-induced distributions (X_0, X_1) of X. For technical reasons (cf. Proposition A.3), we also include in $X^*[\alpha]$ pairs (X'_0, X'_1) where X'_0 is statistically close to an X_0 and X'_1 is statistically close to an X_1 such that

(X_0, X_1) is a pair of complementary α-induced distributions of X. Since we will be interested in indistinguishability of functions of these distributions this will not make any difference, and hence we mostly ignore this issue in the remainder.[7]

THE EQUIVALENCE. We are now ready to state our equivalence result.

Theorem 3.1 Let $\Pi = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ be an encryption scheme. For any distribution $\boldsymbol{M}$ on message vectors, PRIV security of Π with respect $\boldsymbol{M}$ is equivalent to IND security of Π with respect to $\boldsymbol{M}^*[2]$. In particular, let $A \in \mathbb{A}_{\boldsymbol{M}}$ be a PRIV adversary against Π. Then there is a IND adversary $D \in \mathbb{D}_{\boldsymbol{M}^*[2]}$ such that for all $k \in \mathbb{N}$

$$\mathbf{Adv}^{\mathrm{priv}}_{\Pi,A}(k) \quad \leq \quad 36 \cdot \mathbf{Adv}^{\mathrm{ind}}_{\Pi,D}(k) + \left(\frac{3}{4}\right)^{-k} \ .$$

Furthermore, the running-time of D is the time for at most that for k executions of A (but 4 in expectation).

The theorem mostly follows from the techniques of [4]. Thus, our contribution here is not in providing any new technical tools used in proving this result but rather in extracting it from the techniques of [4].[8] For completeness, we give the entire proof (incorporating simplifications due to [9] that lead to better concrete security) in Appendix A.

One may wonder why our equivalence result is any better than those in prior work. After all, in practice it is natural to characterize data according to its min-entropy level (although in some applications one might be interested in other characterizations, which our equivalence allows), and we certainly do not intend to design a different DE scheme for each distribution. However, we argue that our equivalence is more desirable from a technical perspective. First, it is more technically precise, and it implies the results of [4, 7] since one can show that if $\boldsymbol{X}$ has min-entropy μ (here as in [4] the min-entropy of a distribution on message vectors refers to the *individual* min-entropy of each component) then any α-induced distribution of $\boldsymbol{X}$ has min-entropy at least $\mu - \alpha$ (cf. Lemma 5.8, which extends to this case). More importantly, our equivalence allows us to give simple and modular security proofs for many determinsitic encryption schemes, including the one based on trapdoor permutations from [4].

4 Deterministic Encryption from Robust Hardcore Functions

We show a general construction of secure deterministic encryption from trapdoor functions admitting what we call *robust* hardcore functions.

4.1 Robust Hardcore Functions

ONE-WAYNESS AND HARDCORE FUNCTIONS FOR NON-UNIFORM DISTRIBUTIONS. We extend the usual notions of one-wayness and hardcore functions to vectors of inputs drawn from non-unform and possibly correlated distributions, similar to the case of deterministic encryption. Let $\mathcal{F}$ be a trapdoor function generator and $\boldsymbol{X}$ be a distribution on input vectors. To $\mathcal{F}, \boldsymbol{X}$, an inverter I, and $k \in \mathbb{N}$ we associate

```
Experiment $\mathbf{Exp}^{\mathrm{owf}}_{\mathcal{F},\boldsymbol{X},I}(k)$:
$(f, f^{-1}) \xleftarrow{\$} \mathcal{F}$
$\mathbf{x} \xleftarrow{\$} \boldsymbol{X}$
$x' \xleftarrow{\$} I(f, f(\mathbf{x}))$
If $\exists i$ such that $\mathbf{x}[i] = x'$ return 1 else return 0
```

[7] This relaxation is reminiscent of the notion of *smooth* entropy [29] by Renner and Wolf in the context of randomness extraction.

[8] We note that the equivalence proof of [7], while it provides better concrete security and one bit fewer entropy loss, does not seem to yield such a statement.

Define the *OWF advantage* of I against $F, \boldsymbol{X}$ as

$$\mathbf{Adv}^{\mathrm{owf}}_{\mathcal{F},\boldsymbol{X},I}(k) = \Pr\left[\,\mathbf{Exp}^{\mathrm{owf}}_{\mathcal{F},\boldsymbol{X},I}(k) \Rightarrow 1\,\right] \ .$$

We say that $\mathcal{F}$ is *one-way* on a class of distributions on input vectors $\mathbb{X}$ if for every $\boldsymbol{X} \in \mathbb{X}$ and every PPT inverter I, $\mathbf{Adv}^{\mathrm{owf}}_{\mathcal{F},\boldsymbol{X},I}(\cdot)$ is negligible. We extend hardore functions in a similar way. Namely, to a trapdoor function generator $\mathcal{F}$, function $\mathsf{hc}\colon \{0,1\}^k \to \{0,1\}^n$, distribution on input vectors $\boldsymbol{X}$, a distinguisher D, and $k \in \mathbb{N}$ we associate

```
Experiment $\mathbf{Exp}^{\mathrm{hcf}}_{\mathcal{F},\mathsf{hc},\boldsymbol{X},D}(k)$:
$b \xleftarrow{\$} \{0,1\}$ ; $(f, f^{-1}) \xleftarrow{\$} \mathcal{F}$
$\mathbf{x} \xleftarrow{\$} \boldsymbol{X}$
$\mathbf{h}_0 \leftarrow \mathsf{hc}(f,\mathbf{x})$ ; $\mathbf{h}_1 \xleftarrow{\$} (\{0,1\}^n)^{\times|\mathbf{x}|}$
$d \xleftarrow{\$} D(f, f(\mathbf{x}), \mathbf{h}_b)$
If $d = b$ return 1 else return 0
```

Define the *HCF advantage* of D against $F, \mathsf{hc}, \boldsymbol{X}$ as

$$\mathbf{Adv}^{\mathrm{hcf}}_{\mathcal{F},\mathsf{hc},\boldsymbol{X},D}(k) = 2 \cdot \Pr\left[\,\mathbf{Exp}^{\mathrm{hcf}}_{\mathcal{F},\mathsf{hc},\boldsymbol{X},D}(k) \Rightarrow 1\,\right] - 1 \ .$$

We say that hc is *hardcore* for $\mathcal{F}$ on a class of distributions on input vectors $\mathbb{X}$ if for every $\boldsymbol{X} \in \mathbb{X}$ and every PPT distinguisher D, $\mathbf{Adv}^{\mathrm{hcf}}_{\mathcal{F},\mathsf{hc},\boldsymbol{X},D}(\cdot)$ is negligible.

Note that we depart somewhat from standard treatments in that we allow a hardcore function to also depend on the description of the trapdoor function (via the argument f). This allows us to simplify our exposition somewhat.

ROBUSTNESS. We are now ready to define our new notion of *robustness* for hardcore functions. Intuitively, robust hardcore functions are those that remain one-way when the min-entropy of the input is slightly reduced, by conditioning on an (efficiently testable) event that occurs with good probability.

Definition 4.1 Let $\mathcal{F}$ be a trapdoor function and let hc be a hardcore function such that hc is hardcore for $\mathcal{F}$ with respect to a distribution $\boldsymbol{X}$ on input vectors. For $\alpha \in \mathbb{N}$, we say hc is *α-robust for $\mathcal{F}$ on $\boldsymbol{X}$* if hc is also hardcore for $\mathcal{F}$ with respect to the class $\boldsymbol{X}[\alpha]$ of α-induced distributions of $\boldsymbol{X}$.

It is illustrative to consider the new notion just for single-input distinguishers on the uniform distribution (i.e., for $|\mathbf{x}| = 1$ in the HCF experiment where $\mathbf{x}$ is uniform). Note here for example that every bit of the input to RSA is well-known to be hardcore assuming RSA is one-way [1]. However, they are not even 1-robust, since when we decrease the min-entropy of the input the bit may become fixed. Indeed, no hardcore function that depends only on the input and not on the description of the function itself can be robust by a similar argument.

We comment that robustness seems like a natural notion for hardcore functions and it may be interesting to explore its consequences in other contexts. In particular, leakage resilience [27] and computational randomness extraction (or key derivation) [26] come to mind. In these applications, robustness for large α may also be interesting.

4.2 The Encrypt-with-Hardcore Scheme

THE SCHEME. Let $\Pi = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ be an encryption scheme, $\mathcal{F}$ be a trapdoor function generator, and hc be a hardcore function. Assume (e.g., by suitable padding) that hc has the property $\mathsf{hc}_f(x) \in \mathsf{Coins}_{pk}(|x|)$ for all pk output by $\mathcal{K}$ and all $x \in \{0,1\}^*$. Define the associated "*Encrypt-with-Hardcore*" deterministic encryption scheme $\mathsf{EwHCore}[\Pi, \mathcal{F}, \mathsf{hc}] = (\mathcal{K}, \mathcal{DE}, \mathcal{DD})$ with plaintext-space $\mathrm{PtSp} = \{0,1\}^k$ via

```
Algorithm $\mathcal{K}(1^k)$:
$(pk, sk) \xleftarrow{\$} \mathcal{K}(1^k)$
$(f, f^{-1}) \xleftarrow{\$} \mathcal{K}(1^k)$
Return $((pk, f), (sk, f^{-1})$
```

```
Algorithm $\mathcal{DE}((pk, f), x)$:
$r \leftarrow \mathsf{hc}(f, x)$
$c \leftarrow \mathcal{E}(pk, f(x); r)$
Return $c$
```

```
Algorithm $\mathcal{DD}((sk, f^{-1}), c)$:
$y \leftarrow \mathcal{D}(sk, c)$
$x \leftarrow f^{-1}(y)$
Return $x$
```

SECURITY ANALYSIS. To gain some intuition, suppose hc is hardcore for $\mathcal{F}$ on some distribution $\boldsymbol{X}$ on input vectors. One might think that PRIV security of $\mathsf{EwHCore} = \mathsf{EwHCore}[\Pi, \mathcal{F}, \mathsf{hc}]$ on $\boldsymbol{X}$ then follows by by IND-CPA security of Π. However, this is not true. To see this, suppose hc is a physical hardcore function (i.e., ouputs some bits of the input). Define $\Pi' = (\mathcal{K}, \mathcal{E}', \mathcal{D}')$ to be like $\Pi = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ except that the coins consumed by $\mathcal{E}'$ are extended by one bit, which $\mathcal{E}'$ outputs in the clear and $\mathcal{D}'$ ignores. That is, define $\mathcal{E}'(pk, x; r\|b) = \mathcal{E}(pk, x; r)\|b$ and $\mathcal{D}'(sk, y\|b) = \mathcal{D}(sk, y)$. Then IND-CPA security of Π' follows from that of Π, but a straightforward attack shows $\mathsf{EwHCore}$ is not PRIV on $\boldsymbol{X}$. This is where our notion of robustness comes into play.

Theorem 4.2 Suppose Π is IND-CPA secure and hc is *2-robust* for $\mathcal{F}$ on a distribution $\boldsymbol{M}$ on input vectors. Then $\mathsf{EwHCore}[\Pi, \mathcal{F}, \mathsf{hc}]$ is PRIV-secure on $\boldsymbol{M}$.

Note that Π can be built from any one-way trapdoor function, regardless of the input distribution on which the former is one-way [18]. Thus, the sole cryptographic assumption we need for security of $\mathsf{EwHCore}$ is security of hc. We remark that once we have a robust hardcore function that is sufficiently large, it can be expanded to any polynomial length by using its output as the seed for a pseudorandom generator, which in general can be built from any one-way function [22]. (However, more efficient constructions may be possible based on more specific assumptions; we comment on this for our instantiations later.) The theorem follows from the following lemma, which shows that what does follow if hc is hardcore (but not necessarily robust) is the *IND* security of $\mathsf{EwHCore}$.

Lemma 4.3 Suppose Π is IND-CPA, and that hc is hardcore for $\mathcal{F}$ on a distribution $\boldsymbol{M}$ on input vectors. Then $\mathsf{EwHCore} = \mathsf{EwHCore}[\Pi, \mathcal{F}, \mathsf{hc}]$ is IND secure on $\boldsymbol{M}$. In particular, let $D \in \mathbb{D}_{\boldsymbol{M}}$ be a IND adversary against $\mathsf{EwHCore}$. Then there is an IND-CPA adversary A against Π, and an adversary B against hc on $\boldsymbol{M}$, such that for all $k \in \mathbb{N}$

$$\mathbf{Adv}^{\mathrm{ind}}_{\mathsf{EwHCore},D}(k) \quad \leq \quad \mathbf{Adv}^{\mathrm{ind\text{-}cpa}}_{\Pi,A}(k) + 2 \cdot \mathbf{Adv}^{\mathrm{hcf}}_{\mathcal{F},\mathsf{hc},\boldsymbol{M},B}(k) \; . \tag{1}$$

Furthermore, the running-times of A, B are the time to run D.

The proof is in Appendix B. Theorem 4.2 now follows by combining Lemma 4.3 with Theorem 3.1.

A subtle point worth mentioning is where in the proof we use the fact that the Theorem 4.3 considers IND security of $\mathsf{EwHCore}$ rather than PRIV (which, as we have said, does not follow). It is in the step that uses security of the hardcore function. If we considered PRIV security, in this step the constructed HCF adversaries against $\mathcal{F}$ would need to test whether the output of the PRIV adversary against $\mathsf{EwHCore}$ is equal to a "target value" representing partial information on the input to $\mathcal{F}$, which these adversaries are not given. Indeed, this is exactly what caused complications in the original analysis of the scheme of [4].

5 Instantiations

Here we provide several instantiations of robust hardcore functions and hence of the Encrypt-with-Hardcore scheme.

5.1 Instantiations from One-Wayness

Our instantiations based on one-wayness rely on the following simple lemma, which says that "one-way hardness" (measured via an adversary's OWF advantage) on a given input distribution is preserved on sub-distributions induced by an event that occurs with good probability.

Lemma 5.1 Let $\mathcal{F}$ be a trapdoor function generator. Let $\boldsymbol{X}$ be a distribution on input vectors and let $\boldsymbol{X}'$ be a α-induced distribution of $\boldsymbol{X}$. Then for any inverter I against $\mathcal{F}$ on $\boldsymbol{X}$ there is an inverter I' against $\mathcal{F}$ on $\boldsymbol{X}'$ such that for all $k \in \mathbb{N}$

$$\mathbf{Adv}^{\mathrm{owf}}_{\mathcal{F},\boldsymbol{X},I}(k) \quad \leq \quad 2^{-\alpha} \cdot \mathbf{Adv}^{\mathrm{owf}}_{\mathcal{F},\boldsymbol{X}',I'}(k) \; . \tag{2}$$

Furthermore, the running-time of I' is the time to run I.

Proof: Let I' be the inverter that simply runs I on its input, and let E be the corresponding event to $\boldsymbol{X}'$. Let G be the event that $\mathbf{Exp}^{\mathrm{owf}}_{\mathcal{F},\boldsymbol{X}',I'}(k) \Rightarrow 1$. Then

$$\begin{aligned}\mathbf{Adv}^{\mathrm{owf}}_{\mathcal{F},\boldsymbol{X}',I'}(k) &= \Pr\left[\,G \mid E\,\right]\cdot\Pr\left[\,E\,\right] + \Pr\left[\,G \mid \overline{E}\,\right]\cdot\Pr\left[\,\overline{E}\,\right] \\ &\geq \Pr\left[\,G \mid E\,\right]\cdot\Pr\left[\,E\,\right] \\ &= \mathbf{Adv}^{\mathrm{owf}}_{\mathcal{F},\boldsymbol{X},I}(k)\cdot 1/2^{-\alpha}\ ,\end{aligned}$$

from which Equation 2 follows by re-arranging terms. Note that the proof did not need to use that E is efficiently testable. ▮

Note that when $\alpha = O(\log k)$ the reduction incurs a polynomial loss in advantage. Combining Theorem 4.2 and Theorem 3.1 with Lemma 5.1, we see that in order to instantiate the Encrypt-with-Hardcore scheme on input distribution $\boldsymbol{X}$ it suffices to find a trapdoor function $\mathcal{F}$ with a hardcore function hc for $\mathcal{F}$ on $\boldsymbol{X}$ which is furthermore hardcore on any input distribution $\boldsymbol{X}'$ such that the hardness of inverting $\mathcal{F}$ on $\boldsymbol{X}'$ is polynomially related to that on $\boldsymbol{X}$. Indeed, it follows that such a hc is $O(\log k)$-robust for $\mathcal{F}$ on $\boldsymbol{X}$ (though for our application we only need 2-robustness). We give examples of such hardcore functions below.

ROBUSTNESS OF A RANDOM ORACLE. In the random oracle (RO) model [5], the random oracle hash itself is a hardcore function satisfying the properties we need.

Proposition 5.2 Let $\mathcal{F}$ be a one-way trapdoor function on a distribution $\boldsymbol{X}$ on input vectors. Then a random oracle hash is $O(\log k)$-robust for $\mathcal{F}$ on $\boldsymbol{X}$.

The proof combines Lemma 5.1 with the techniques of [2, Theorem 5.1] and is omitted here. We thus obtain an instantiation of the Encrypt-with-Hardcore scheme in the RO model based on any one-way trapdoor function. In fact, the "Encrypt-with-Hash" scheme of [2] can be viewed as an optimization of this instantiation, where the trapdoor function itself is dropped (and only the outer encryption is used, with the hash of the message as the coins). Note the these RO model schemes are secure in the strongest sense of [2].

ROBUSTNESS OF GOLDREICH-LEVIN. Our standard-model robust hardcore functions are based on the Goldreich-Levin (GL) function [18]. To define this, let $\mathcal{F}$ be a trapdoor function generator and let $\mathcal{H}\colon \mathcal{K}\times\{0,1\}^k \to \{0,1\}^n$ be a function. Define its *H-padded version* $\mathcal{F}[H]$ that on input 1^k returns $(f,K),(f^{-1},K)$ where $(f,f^{-1}) \overset{\$}{\leftarrow} \mathcal{F}(1^k)$ and $K \overset{\$}{\leftarrow} \mathcal{K}$; evaluation is defined for $x \in \{0,1\}^k$ as $f(x)$ (i.e., evaluation just ignores K) and inversion is defined analogously. Define the *length-i Goldreich-Levin function* $\mathcal{GL}^i\colon \{0,1\}^{i\times k}\times\{0,1\}^k \to \{0,1\}^i$ as

$$\mathcal{GL}^i(M,x) = Mx$$

where Mx is the matrix-vector product of M and x over $GF(2)$. We recall the following.

Theorem 5.3 (**Goldreich-Levin Theorem** [18]) Let $\mathcal{F}[\mathcal{GL}^i]$ be as defined above and let X be a distribution on inputs to $\mathcal{F}$. Let D be a distinguisher against $\mathcal{GL}^i$. Then there is a inverter I such that for all $k \in \mathbb{N}$

$$\mathbf{Adv}^{\mathrm{hcf}}_{\mathcal{F}[\mathcal{GL}^i],\mathcal{GL}^i,X,D}(k) \quad\leq\quad 2^{i+3}\cdot\mathbf{Adv}^{\mathrm{owf}}_{\mathcal{F},X,I}(k)\ . \tag{3}$$

Furthermore, the running-time of I is the time for $O(\varepsilon^{-4}k^3)$ executions of D where $\varepsilon = \mathbf{Adv}^{\mathrm{hcf}}_{\mathcal{F}[\mathcal{GL}^i],\mathcal{GL}^i,X,D}(k)$.

We have the following result.

Proposition 5.4 Let $\mathcal{F}[\mathcal{GL}^i]$ be as defined above and suppose $\mathcal{GL}^i$ is hardcore for $\mathcal{F}[\mathcal{GL}^i]$ on single-input distribution X. Then $\mathcal{GL}^i$ is $O(\log k)$-robust for $\mathcal{F}[\mathcal{GL}^i]$ on X.

The proof follows by combining Lemma 5.1 and Theorem 5.3. We stress that while the above proposition gives us robust hardcore functions in the standard model, a major difference between Proposition 5.4 and Proposition 5.2 is that the former only considers distributions on *single inputs* to the given trapdoor function while the latter considers vectors of *correlated inputs*. This restriction is inherited by the standard-model instantiations of the Encrypt-with-Hardcore scheme we obtain based on one-wayness. We detail these instantiations next.

ENCRYPT-WITH-HARDCORE INSTANTIATIONS FROM TRAPDOOR FUNCTIONS. Suppose we want to instantiate the Encrypt-with-Hardcore scheme on plaintext distribution M, and we are given a trapdoor function $\mathcal{F}$ that is exponentially-hard to invert on M (which implies M has linear min-entropy) that is, we say that $\mathcal{F}$ is exponentially hard to invert on M if for every PPT inverter I, $\mathbf{Adv}^{\mathrm{owf}}_{\mathcal{F},M,I}(k) = 2^{-\Omega(k)}$. Theorem 5.3 combined with Proposition 5.3 tells us that $\mathcal{F}$ has a robust hardcore GL function with output size $\Omega(k)$. We would like to claim that this is sufficiently large to match the number of coins consumed by the "outer" encryption scheme. However, if we are interested in an instantiation under minimal assumptions, then we would also like to build latter from $\mathcal{F}$. A straightforward instantiation runs into a complication because the number of coins consumed by such a scheme is at least the length of an input to $\mathcal{F}$, which is greater than the output length of our hardcore function. As we mentioned, to solve this we can expand the length of the hardcore function by using its output as the seed for a pseudorandom generator. The latter can be built using an arbitrary one-way function [22] but such constructions have poor seed length, meaning the size of the inputs to $\mathcal{F}$ would need to be very long. However, since we are assuming exponential hardness of $\mathcal{F}$ we can use a construction of Haitner et al. [21] with seed length $O(k \log^2 k)$ (further improvements to the state-of-the-art here would yield corresponding improvements to our scheme). We make explicit the following corollary.

Corollary 5.5 Suppose $\mathcal{F}$ is a one-way trapdoor function generator that is exponentially-hard for a distribution M on $\{0,1\}^\ell$, where $\ell = \Omega(k \log^2 k)$ as determined by [21] and the security of $\mathcal{F}$. Then we obtain a PRIV secure DE scheme on M. In particular, if $\mathcal{F}$ is exponentially-hard on the class of all such distributions with min-entropy μ, then we obtain a PRIV secure DE scheme for this class.

Obtaining DE from such hardcore functions is new to this work.

ENCRYPT-WITH-HARDCORE INSTANTIATIONS FROM TRAPDOOR PERMUTATIONS. In the case of trapdoor *permutations*, instead of assuming the trapdoor permutation $\mathcal{F}$ is exponentially hard to invert on a given input distribution X we would like to use Blum-Micali-Yao iteration [6, 32] to extract many simultaneous bits assuming $\mathcal{F}$ is only super-polynomially hard to invert on X. The catch is that to use the latter we will actually need to assume that $\mathcal{F}$ is super-polynomially hard to invert on all (efficiently sampleable) *permutation distributions* of X (i.e., distributions obtained by a re-labeling of the points of X.) Namely, let $\mathcal{F}$ be a trapdoor permutation generator. For $i \in \mathbb{N}$ denote by $\mathcal{F}^i$ the trapdoor permutation generator that iterates $\mathcal{F}$ i-many times, and let $\mathcal{F}[\mathcal{GL}]$ be as defined above. For (f, r) output by $\mathcal{F}^i[\mathcal{GL}]$ define the Blum-Micali-Yao [6, 32], Goldreich-Levin [18] function $\mathcal{BMY}^i[f,r] \colon \{0,1\}^k \times \{0,1\}^k \to \{0,1\}^i$ via

$$\mathcal{BMY}^i[f,r](x) = \langle x, r\rangle \| \langle f(x), r\rangle \| \ldots \| \langle f^{i-1}(x), r\rangle \ .$$

Then we have the following result.

Proposition 5.6 Let $\mathcal{F}$ be a trapdoor permutation generator, and let X be a single-input distribution such that $\mathcal{F}$ is one-way on all efficiently sampleable permutation distributions of X. Then for any polynomial i, $\mathcal{BMY}^i[\mathcal{GL}]$ is $O(\log k)$-robust for $\mathcal{F}^i[\mathcal{GL}]$ on X.

The proof combines the hybrid argument of [6, 32] with Proposition 5.4 and the observation that a permutation distribution of an α-induced sub-distribution of X is an α-induced sub-distribution of a permutation distribution of X (in other words, the permuting and conditioning "commutes"). Again, we make explicit the following corollary.

Corollary 5.7 Suppose $\mathcal{F}$ is a trapdoor permutation generator that is one-way on all permutation distributions of an input distribution M. Then we obtain a PRIV secure DE scheme on M. In particular, if $\mathcal{F}$ is one-way on the class of all input distributions with min-entropy μ, then we obtain a PRIV secure DE scheme for this class.

Note that this instantiation of the Encrypt-with-Hardcore scheme is exactly the scheme of [4]. However, we believe our proof based on robustness properties of Goldreich-Levin is simpler and more modular, giving more insight into the scheme. It also reveals why the Goldreich-Levin bit "works" in the construction. (Indeed, any robust hardcore bit would do.)

INSTANTIATIONS FROM ONE-WAYNESS ON (JUST) UNIFORM INPUTS. We note that in the case of input distributions with linear entropy, the respective assumptions we need for the above schemes, namely trapdoor functions that are very hard to invert on a given distribution, or on trapdoor permutations that are hard to invert on all permutations of a given distribution (which preserves min-entropy), can be obtained simply based on (exponential) one-way hardness of trapdoor functions or permutations on the standard *uniform* distribution (though our assumptions above are more general). This follows from [16, Lemma 4], which says that *every* distribution with min-entropy α less than that of *uniform* can be viewed as an α-induced distribution of the latter, although the corresponding event may not be efficiently testable (which is not needed for Lemma 5.1, though). Indeed, it follows that the trapdoor function or permutation is one-way on all distributions of sufficient entropy. Of course, we require more "hardness" in the case of a trapdoor function, since it is required to have a *large* hardcore GL function.

An advantage of this approach is that it implies one-wayness even for input distributions that are not necessarily efficiently sampleable. Therefore, we are able to apply the equivalence between single-mesage PRIV security and unbounded multi-message PRIV security to for block sources proven in [7] to the resulting DE schemes without any "conditional re-sampleability" requirement on the block source.

5.2 Instantiations from Lossiness

Peikert and Waters [28] showed that lossy trapdoor functions admit a simple, large hardcore function in the standard model, namely a universal hash function. We also show robustness of the latter based on the following simple lemma, which says that min-entropy of a given input distribution is preserved on sub-distributions induced by an event that occurs with good probability.

Lemma 5.8 Let X be a distribution with $\mathrm{H}_\infty(X) \geq \mu$, and let X' be a α-induced distribution of X. Then $\mathrm{H}_\infty(X') \geq \mu - \alpha$.

Proof: Suppose not, and let E be the corresponding event to X'. Then there exists an x' such that $\Pr[\, X' = x' \,] > 2^{-\mu+\alpha}$. But then

$$\begin{aligned} \Pr[\, X = x' \,] &\geq \Pr[\, X = x' \mid E \,] \cdot \Pr[\, E \,] + \Pr[\, X = x' \mid \overline{E} \,] \cdot \Pr[\, \overline{E} \,] \\ &\geq \Pr[\, X = x' \mid E \,] \cdot \Pr[\, E \,] \\ &> 2^{-\mu+\alpha} \cdot 2^{-\alpha} \\ &= 2^{-\mu} \end{aligned}$$

a contradiction. ▌

Next we recall the Generalized Leftover Hash Lemma due to Dodis et al [12], which extends the original version of [22] to average conditional min-entropy.

Lemma 5.9 (Generalized Leftover Hash Lemma) [12] Let $\mathcal{H}\colon \mathcal{K} \times D \to R$ be a universal function. Let X be a random variable over D and Y be another random variable such that $\tilde{\mathrm{H}}_\infty(X \mid Y) \geq \log|R| + 2\log(1/\varepsilon)$. Then

$$\Delta\big((K, Y, \mathcal{H}(K, X)), (K, Y, U)\big) \leq \varepsilon \ ,$$

where $K \overset{\$}{\leftarrow} \mathcal{K}$ and U is uniform and independent on R.

By combining Lemma 5.9 with the "chain rule" for average conditional min-entropy [12, Lemma 2.2], it follows that if $\mathcal{F}$ is a lossy trapdoor function generator with residual leakage s, then a univeral hash function $\mathcal{H}$ is hardcore for $\mathcal{F}[\mathcal{H}]$ on any single-input distribution X with min-entropy at least $s + n + 2(\log 1/\varepsilon)$ for negligible ε (as compared to [28, Lemma 3.4] we simply observe that the argument does not require the input to be uniform). Then, using Lemma 5.8 we furthermore have the following.

Proposition 5.10 Let LTDF be a lossy trapdoor function generator with residual leakage s, and let $\mathcal{H}\colon K \times \{0,1\}^k \to \{0,1\}^n$ be a universal hash function. Then $\mathcal{H}$ is a $O(\log k)$-robust hardcore function for $\mathcal{F}[\mathcal{H}]$ on any single-input distribution X with min-entropy $s + n + 2(\log 1/\varepsilon)$ for negligible ε.

As in the case of exponentially-hard one-way trapdoor functions, if we want to also instantiate the"outer" encryption scheme in the Encrypt-with-Hardcore scheme using the given lossy trapdoor function, we need to expand the length of the hardcore function by using its output as the seed for a pseudorandom generator. However, Boldyreva et al. [7] give a direct construction of DE from lossy trapdoor functions anyway (which can be viewed as an optimization of the one we obtain here) that does not use the "outer" encryption scheme at all, so it is not as much of a concern.

Finally, one should note the parallel between the argument here and that to show robustness of the Goldreich-Levin function. In particular, the use of Lemma 5.8 is analogous to Lemma 5.1 and the use of Lemma 5.9 is analogous to Theorem 5.3. We thus obtain a unified view of how (single-message) DE can be derived from both one-wayness and lossiness.

6 Bounded Multi-Message Security for Deterministic Encryption

In this section, we propose a new q-bounded notion of multi-message security for DE and provide constructions meeting it from any lossy trapdoor function that loses a $1 - o(1)$ fraction of its input.

6.1 The New Notion and Variations

THE NEW NOTION. The notion of q-bounded multi-message security for DE is quite natural. Recall that in the PRIV or IND experiments defining security for deterministic encryption (see Section 3), a PRIV or IND adversary outputs a message vector $\mathbf{x}$. In q-bounded multi-message security, we require that the size of $\mathbf{x}$ is at most some polynomial $q = q(k)$ (note that this is fixed ahead of time prior to public key generation, and thus the latter may grow with q). Otherwise, the security definitions remain identical. In this case, we say that a distribution $\boldsymbol{X}$ on vectors of size q *has min-entropy* μ if $\mathrm{H}_\infty(\mathbf{x}[i]) \geq \mu$ for all $\mathbf{x}$ in the support of $\boldsymbol{X}$ and all $1 \leq i \leq |\mathbf{x}|$. (In other words, min-entropy for a message vector refers to the *individual* min-entropy of each component.) We may assume that $\mathbf{x}[i] \neq \mathbf{x}[j]$ for all $\mathbf{x}$ in the support of $\boldsymbol{X}$ and all $1 \leq i \neq j \neq |\mathbf{x}|$ as well.

SEMANTIC SECURITY VERSUS INDISTINGUISHABILITY. Note that Theorem 3.1 tells us that PRIV on a distribution $\boldsymbol{M}$ of message vectors of size q is equivalent to IND on the class of pairs of complementary 2-induced distributions of $\boldsymbol{M}$, the latter also being on message vectors of size q.

UNBOUNDED MULTI-MESSAGE SECURITY FOR q-BLOCK SOURCES. Generalizing the approach of Boldyreva et al. [7] for single-message security, we also consider unbounded multi-message security for what we call a *q-block source*, a generalization of a block-source [8] where every q messages introduces some "fresh entropy." That is, we call a sequence of random variables $(X_1, \ldots, X_{qn})$ on a set $\mathcal{X}$ a *q-block-source of entropy* μ for all $1 \leq i \leq n$, all $0 \leq j \leq q - 1$, and all $x_1, \ldots, x_{qi-1} \in \mathcal{X}$, $\mathrm{H}_\infty(X_{qi+j} \mid X_1 = x_1, \ldots, X_{qi-1} = x_{qi-1}) \geq \mu$. Using a similar argument to [7, Theorem 4.2], one can show equivalence of q-bounded multi-message security for *all* distributions of a given min-entropy μ (here min-entropy refers to the *individual* min-entropy of each component) and unbounded multi-message security for all q-block sources of min-entropy μ (essentially, this is done by viewing each "block" as consisting of q components in the hybrid arguments). However, as in [7], when considering efficiently sampleable message distributions one must make some extra restrictions for this to hold (which are immaterial for our actual constructions since they are secure even for inefficient distributions).

6.2 Extensions to the Crooked Leftover Hash Lemma

Note that we cannot trivially achieve q-bounded multi-message security by running say q copies of a scheme secure for one message in parallel (and encrypting the i-th message under the i-th public key), since this approach (if it works) would lead to a stateful scheme. The main technical tool we use to achieve the notion is a new extension to Crooked Leftover Hash Lemma due to Dodis and Smith [13]. (The latter was used by Boldyreva et al. [7] in the context of single-message security.) For our results we extend the lemma in two ways. First, we strengthen the lemma to the case of t-wise independent functions and t-many correlated sources, inspired by [25] who give a similar strengthening to the standard Leftover Hash Lemma to the case of 4-wise independence. Second, we consider a relaxation to *almost* t-wise independence, inspired by [11] who show a similar relaxation for the standard Leftover Hash Lemma. Namely, say that H is *δ-almost q-wise independent* if for all distinct $x_1, \ldots, x_q \in D$ and all $y_1, \ldots, y_q \in R$

$$\Delta((F(K,x_1),\ldots,F(K,x_q)),(U_1,\ldots,U_q)) \;\leq\; \delta\ .$$

Our extensions to the Crooked LHL are captured in the following.

Lemma 6.1 (Crooked Leftover Hash Lemma for Correlated Sources) Let $\mathcal{H}\colon \mathcal{K} \times D \to R$ be a $2t$-wise independent function for $t > 0$ with range R, and let $f : R \to S$ be a function. Let $\mathbf{X} = (X_1, \ldots, X_t)$ where the X_i are random variables over D such that $\mathrm{H}_\infty(X_i) \geq \mu$ for all $1 \leq i \leq n$ and moreover $\Pr[\, X_i = X_j \,] = 0$ for all $1 \leq i \neq j \leq t$. Then

$$\Delta((K, f(\mathcal{H}(K,\mathbf{X}))),(K,f(\mathbf{U}))) \;\leq\; \frac{1}{2}\sqrt{|S|^t(t^2 2^{-\mu} + 3\delta)}$$

where $K \xleftarrow{\$} \mathcal{K}$ and $\mathbf{U} = (U_1, \ldots, U_t)$ where the U_i are all uniform and independent over R (recall that functions operate on vectors component-wise).

The proof, which extends the proof of the Crooked LHL in [7], is in Appendix C.

6.3 Constructions

THE BASIC SCHEME. Lemma 6.1 (more precisely, its extension to average conditional min-entropy following [12]) implies that we can achieve q-bounded multi-message secure DE for any polynomial q by using a $2q$-wise independent hash function as the hardcore function in the instantiation of the Encrypt-with-Hardcore scheme from lossy trapdoor functions we obtained in Section 5.2. (Notice that the Crooked Leftover Hash Lemma is a generalization of the standard Leftover Hash Lemma, by taking the "outer" function f in the lemma to be the identity.) In other words, this hardcore function is "correlation secure" in the sense of Section 4.1 for a bounded number of arbitrarily correlated inputs. Note that we need the lossy trapdoor function to lose a $1 - o(1)$ fraction of its input, namely the residual leakage should be less than k/q. The DDH-based construction of Peikert and Waters [28] and the Paillier-based one from [7, 17] satisfy the latter requirement. Following [7] we give a more efficient DE scheme meeting q-bounded multi-message security that achieves much better parameters below, for which we give more details.

THE OPTIMIZED SCHEME. Intuitively, for the optimized scheme we achieve q-bounded multi-message security by modifying the scheme of [7] to first pre-process an input message using a $2q$-wise (instead of pairwise) independent permutation, and appealing to Lemma 6.1 in the security proof. The catch is that for $q > 1$ such a permutation is not known to exist (in an explicit and efficiently computable sense). However, there are good constructions of *almost* $2q$-wise independent permutations. Namely, for any $t, \delta > 0$, Kaplan et al. [24] construct a t-wise δ-almost independent permutation whose key length is $O(tk + \log(1/\delta))$.

More formally, let $\mathsf{LTDF} = (\mathcal{F}, \mathcal{F}')$ be a lossy trapdoor function and let $\mathcal{P}\colon \mathcal{K} \times \{0,1\}^k \to \{0,1\}^k$ be a family of permutations on k bits. Define the associated deterministic encryption scheme $\Pi[\mathsf{LTDF}, \mathcal{P}] = (\mathcal{K}, \mathcal{DE}, \mathcal{DD})$ with plaintext-space $\mathrm{PtSp} = \{0,1\}^k$ via

```
Algorithm $\mathcal{K}(1^k)$:
    $(f, f^{-1}) \xleftarrow{\$} \mathcal{F}(1^k)$ ; $K \xleftarrow{\$} \mathcal{K}$
    Return $((f, K), (f^{-1}, K))$
```

```
Algorithm $\mathcal{DE}((f, K), x)$:
    $c \leftarrow f(\mathcal{P}(K, x))$
    Return $c$
```

```
Algorithm $\mathcal{DD}((sk, f^{-1}), c)$:
    $x \leftarrow f^{-1}(\mathcal{P}^{-1}(K, c))$
    Return $x$
```

The following theorem follows by Lemma 6.1.

Theorem 6.2 Suppose LTDF is a lossy trapdoor funtion with residual leakage s, and let $q, \varepsilon > 0$. Set $\delta = 2\varepsilon^2/(3 \cdot 2^{qs})$ and suppose $\mathcal{P}$ is δ-almost $2q$-wise independent. Then for any q-message IND adversary D with min-entropy $\mu \geq qs + 2\log q + 2\log(1/\varepsilon) - 1$, there is a LTDF distinguisher D such that for all $k \in \mathbb{N}$

$$\mathbf{Adv}^{\mathrm{ind}}_{\Pi[\mathsf{LTDF},\mathcal{P}],A}(k) \quad \leq \quad \mathbf{Adv}^{\mathrm{ltdf}}_{\mathsf{LTDF},D}(k) + \varepsilon\ .$$

Furthermore, the running-time of D is the time to run A.

Now, combining Theorem 6.2 with Theorem 3.1 and Lemma 5.8 (extended to message vectors rather than single-input distributions) gives us bounded multi-message PRIV (rather than IND) security for any distribution on message vectors of size q with sufficient entropy. We make explicit the following corollary.

Corollary 6.3 Suppose LTDF is a lossy trapdoor function on $\{0,1\}^\ell$, where $\ell = \Omega(kq)$, with residual leakage s. Then we obtain a q-bounded multi-message PRIV secure DE scheme for the class of distributions on $\{0,1\}^{\ell \times q}$ with min-entropy $\mu \geq qs + 2\log q + 2\log(1/\varepsilon) - 1$ for negligible ε.

Note that, as for the basic scheme, to achieve q-bounded security we need a lossy trapdoor function that loses more than a $1 - 1/q$ fraction of its input (due to the factor q on s in the entropy bound in the above theorem). The resulting schemes have input lengths proportional to kq; it is an open problem to give a DE scheme achieving q-bounded message security that allows shorter messages. We also we can prove that the optimized scheme meets unbounded multi-message PRIV security on q-block sources of the same entropy directly by using our precise definitional equivalence, as follows. First, its IND security on q-block sources follows by extending Lemma 6.1 to q-block sources by a hybrid argument as in the case of the original Leftover Hash Lemma [33]. Then, its PRIV security on q-block sources (of 2 bits greater entropy) follows by Theorem 3.1 after extending Lemma 5.8 to show that a 2-induced distribution of a q-block source with min-entropy μ is a q-block source with min-entropy $\mu - 2$.

DISCUSSION. We believe that q-bounded multi-message security better illuminates the technical gap between achieving unbounded multi-message security for DE in the RO and standand models. In particular, note that for an unbounded number of arbitrarily correlated messages the information-theoretic approach we use to achieve q-bounded multi-message security breaks down (since there would not be enough randomness in the key plus the inputs to extract). One approach towards achieving unbounded security in the standard model would be to give a computational analogue of Lemma 6.1 (or rather, the version in the context of the "standard" LHL) for an unbounded number of arbitrarily correlated sources. We consider finding a hash function (in the standard model) that satisfies such an analogue to be a very interesting open problem. Some partial results in this direction were obtained recently in [20].

Acknowledgements

We are grateful to Mihir Bellare for encouraging us to write up the results and to Alexandra Boldyreva, Marc Fischlin, Serge Fehr, Payman Mohassel, Krzysztof Pietrzak, Adam Smith, and Ramarathnam Venkatesan for helpful comments and discussions. In particular we thank Adam Smith for pointing us to [24].

References

[1] Werner Alexi, Benny Chor, Oded Goldreich, and Claus-Peter Schnorr. RSA and Rabin functions: Certain parts are as hard as the whole. *SIAM J. Comput.*, 17(2), 1988.

[2] Mihir Bellare, Alexandra Boldyreva, and Adam O'Neill. Deterministic and efficiently searchable encryption. In *CRYPTO*, pages 535–552, 2007.

[3] Mihir Bellare, Zvika Brakerski, Moni Naor, Thomas Ristenpart, Gil Segev, Hovav Shacham, and Scott Yilek. Hedged public-key encryption: How to protect against bad randomness. In *ASIACRYPT*, pages 232–249, 2009.

[4] Mihir Bellare, Marc Fischlin, Adam O'Neill, and Thomas Ristenpart. Deterministic encryption: Definitional equivalences and constructions without random oracles. In *CRYPTO*, pages 360–378, 2008.

[5] Mihir Bellare and Phillip Rogaway. Random oracles are practical: A paradigm for designing efficient protocols. In *ACM Conference on Computer and Communications Security*, pages 62–73, 1993.

[6] Manuel Blum and Silvio Micali. How to generate cryptographically strong sequences of pseudo-random bits. *SIAM J. Comput.*, 13(4):850–864, 1984.

[7] Alexandra Boldyreva, Serge Fehr, and Adam O'Neill. On notions of security for deterministic encryption, and efficient constructions without random oracles. In *CRYPTO*, pages 335–359, 2008.

[8] Benny Chor and Oded Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM J. Comput.*, 17(2), 1988.

[9] Alexander W. Dent, Marc Fischlin, Mark Manulis, Martijn Stam, and Dominique Schröder. Confidential signatures and deterministic signcryption. In *Public Key Cryptography*, pages 462–479, 2010.

[10] Simon Pierre Desrosiers. Entropic security in quantum cryptography. *Quantum Information Processing*, 8(4):331–345, 2009.

[11] Yevgeniy Dodis, Rosario Gennaro, Johan Håstad, Hugo Krawczyk, and Tal Rabin. Randomness extraction and key derivation using the cbc, cascade and hmac modes. In *CRYPTO*, pages 494–510, 2004.

[12] Yevgeniy Dodis, Rafail Ostrovsky, Leonid Reyzin, and Adam Smith. Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. *SIAM J. Comput.*, 38(1):97–139, 2008.

[13] Yevgeniy Dodis and Adam Smith. Correcting errors without leaking partial information. In *STOC*, pages 654–663, 2005.

[14] Yevgeniy Dodis and Adam Smith. Entropic security and the encryption of high entropy messages. In *TCC*, pages 556–577, 2005.

[15] Stefan Dziembowski and Krzysztof Pietrzak. Leakage-resilient cryptography. In *FOCS*, pages 293–302, 2008.

[16] Sebastian Faust, Eike Kiltz, Krzysztof Pietrzak, and Guy N. Rothblum. Leakage-resilient signatures. In *TCC*, pages 343–360, 2010.

[17] David Mandell Freeman, Oded Goldreich, Eike Kiltz, Alon Rosen, and Gil Segev. More constructions of lossy and correlation-secure trapdoor functions. In *Public Key Cryptography*, pages 279–295, 2010.

[18] Oded Goldreich and Leonid A. Levin. A hard-core predicate for all one-way functions. In *STOC*, pages 25–32, 1989.

[19] Shafi Goldwasser and Silvio Micali. Probabilistic encryption. *J. Comput. Syst. Sci.*, 28(2):270–299, 1984.

[20] Vipul Goyal, Adam O'Neill, and Vanishree H. Rao. Correlated-input secure hash functions, 2010. Unpubished Manuscript.

[21] Iftach Haitner, Danny Harnik, and Omer Reingold. Efficient pseudorandom generators from exponentially hard one-way functions. In *ICALP (2)*, pages 228–239, 2006.

[22] Johan Håstad, Russell Impagliazzo, Leonid A. Levin, and Michael Luby. A pseudorandom generator from any one-way function. *SIAM J. Comput.*, 28(4):1364–1396, 1999.

[23] Brett Hemenway, Steve Lu, and Rafail Ostrovsky. Correlated product security from any one-way function and the new notion of decisional correlated product security. Cryptology ePrint Archive, Report 2010/100, 2010. `http://eprint.iacr.org/`.

[24] Eyal Kaplan, Moni Naor, and Omer Reingold. Derandomized constructions of *k*-wise (almost) independent permutations. *Algorithmica*, 55(1):113–133, 2009.

[25] Eike Kiltz, Krzysztof Pietrzak, Martijn Stam, and Moti Yung. A new randomness extraction paradigm for hybrid encryption. In *EUROCRYPT*, pages 590–609, 2009.

[26] Hugo Krawczyk. Cryptographic extraction and key derivation: The HKDF scheme. In *CRYPTO*, pages 631–648, 2010.

[27] Silvio Micali and Leonid Reyzin. Physically observable cryptography (extended abstract). In *TCC*, pages 278–296, 2004.

[28] Chris Peikert and Brent Waters. Lossy trapdoor functions and their applications. In *STOC*, pages 187–196, 2008.

[29] Renato Renner and Stefan Wolf. Smooth Renyi entropy and applications. In *IEEE International Symposium on Information Theory — ISIT 2004*, page 233. IEEE, June 2004.

[30] Alon Rosen and Gil Segev. Chosen-ciphertext security via correlated products. In *TCC*, pages 419–436, 2009.

[31] Alexander Russell and Hong Wang. How to fool an unbounded adversary with a short key. *IEEE Transactions on Information Theory*, 52(3):1130–1140, 2006.

[32] Andrew Chi-Chih Yao. Theory and applications of trapdoor functions (extended abstract). In *FOCS*, pages 80–91, 1982.

[33] David Zuckerman. Simulating BPP using a general weak random source. *Algorithmica*, 16(4/5):367–391, 1996.

A Proof of Theorem 3.1

We focus on the IND to PRIV implication; the other direction is straightforward.[9] Following [4], the high-level intuition for the proof is as follows. For the given distribution $\boldsymbol{M}$ on message vectors, we first show that it suffices to consider PRIV adversaries for which A_2 outputs $(\mathbf{x}, t)$ where t is *boolean*. Now, we would like to use the fact if t is easy to guess from the encryption of $\mathbf{x}$ then the encryption of $\mathbf{x}$ conditioned on (1) the output $(\mathbf{x}, t)$ of A_2 being such that $t = 1$, or (2) the output $(\mathbf{x}, t)$ of A_2 being such that $t = 0$ are easy to distinguish; indeed, these are induced distributions of $\boldsymbol{M}$ (viewing the binary t as the random variable indicating the event E). However, one of these distributions may be hard to sample from and have low entropy. Therefore, we show it additionally suffices to consider PRIV adversaries on $\boldsymbol{M}$ for which t is not just boolean but also *balanced*, meaning the probability it is 0 or 1 is about the same. Then, we can easily sample from the above-mentioned distributions by repeatedly running A.

REDUCTION TO THE BOOLEAN CASE. Call a PRIV adversary A *boolean* if it outputs test strings of length 1. We first show that is suffices to consider boolean PRIV adversaries (this was previously shown in both [4] and [7]).

[9] The idea for the latter is to have the constructed PRIV adversary according to $\boldsymbol{M}$ and let the partial information be whether the corresponding event for the induced complementary distributions of the given IND adversary occured or not.

Proposition A.1 Let $\Pi = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ be an encryption scheme and $A \in \mathbb{A}_{\boldsymbol{M}}$ be a PRIV adversary that outputs test strings of length ℓ. Then there is a boolean PRIV adversary $B \in \mathbb{A}_{\boldsymbol{M}}$ such that

$$\mathbf{Adv}^{\mathrm{priv}}_{\Pi,A}(k) \quad \leq \quad 2 \cdot \mathbf{Adv}^{\mathrm{priv}}_{\Pi,B}(k) \; .$$

Furthermore, the running-time of B is the time to run A plus $O(\ell)$.

Proof: The proof is identical to an argument in [10] for the information-theoretic setting. Adversary B works as follows:

```
Algorithm $B_1(1^k)$:
$r \xleftarrow{\$} \{0,1\}^n$
Return $r$
```

```
Algorithm $B_2(r)$:
$(\mathbf{x}, t) \xleftarrow{\$} A_1(1^k)$
Return $(\mathbf{x}, \langle t, r \rangle)$
```

```
Algorithm $B_3(pk, \mathbf{c}, r)$:
$g \xleftarrow{\$} A_3(pk, \mathbf{c})$
Return $\langle g, r \rangle$
```

For $d \in \{0,1\}$, let A_d denote the event $\mathbf{Exp}^{\text{priv-d}}_{\Pi,A}(k) \Rightarrow 1$ and similarly B_d denote $\mathbf{Exp}^{\text{priv-d}}_{\Pi,B}(k) \Rightarrow 1$. Then

$$\begin{aligned}
\mathbf{Adv}^{\mathrm{priv}}_{\Pi,B}(k) \quad &= \quad \Pr\left[\, B_1 \,\right] - \Pr\left[\, B_0 \,\right] \\
&= \quad \Big(\Pr\left[\, A_1 \,\right] + \frac{1}{2} \cdot (1 - \Pr\left[\, A_1 \,\right]) \Big) - \Big(\Pr\left[\, A_0 \,\right] + \frac{1}{2} \cdot (1 - \Pr\left[\, A_0 \,\right]) \Big) \\
&= \quad \frac{1}{2} \cdot (\Pr\left[\, A_1 \,\right] - \Pr\left[\, A_0 \,\right]) \\
&= \quad \frac{1}{2} \cdot \mathbf{Adv}^{\mathrm{priv}}_{\Pi,A}(k) \; .
\end{aligned}$$

The claimed running-time of B is easy to verify. ▌

REDUCTION TO THE BALANCED BOOLEAN CASE. As in [4] the next step is to show that it in fact suffices to consider boolean PRIV adersaries that are *balanced*, meaning the probability the partial information is 1 or 0 is approximately $1/2$. Namely, call a boolean PRIV adversary $A = (A_0, A_1, A_2)$ δ-*balanced* [4] if for all $b \in \{0,1\}$

$$\left| \Pr\left[\, t = b \;:\; (\mathbf{x}, t) \xleftarrow{\$} A_1(1^k, state) \,\right] - \frac{1}{2} \right| \leq \delta$$

for all $state$ output by A_0.

Proposition A.2 Let $\Pi = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ be an encryption scheme and $B \in \mathbb{A}_{\boldsymbol{M}}$ be a boolean PRIV adversary. Then for any $0 \leq \delta < 1/2$ there is a δ-balanced boolean PRIV adversary $B' \in \mathbb{A}_{\boldsymbol{M}}$ such that

$$\mathbf{Adv}^{\mathrm{priv}}_{\Pi,B}(k) \quad \leq \quad \left(\frac{2}{\delta} + 1 \right) \cdot \mathbf{Adv}^{\mathrm{priv}}_{\Pi,B'}(k) \; .$$

Furthermore, the running-time of B' is the time to run B plus $O(1/\delta)$.

Proof: As compared to [4] we give a simplified proof due to [9] (which also leads to better concrete security), where for simplicity we assume $1/\delta$ is an integer. Adversary B' works as follows:

```
Algorithm $B_1(1^k)$:
$(\mathbf{x}, t) \xleftarrow{\$} A_1(1^k)$
$i \xleftarrow{\$} [2(1/\delta) + 1]$
If $i \leq 1/\delta$ then return $(\mathbf{x}, 0)$
Else if $i \leq 2(1/\delta)$ then return $(\mathbf{x}, 1)$
Else return $(\mathbf{x}, t)$
```

```
Algorithm $B_2(pk, \mathbf{c})$:
$g \xleftarrow{\$} A_2(pk, \mathbf{c})$
$j \xleftarrow{\$} [2(1/\delta) + 1]$
If $j \leq \delta$ then return $0$
Else if $j \leq 2(1/\delta)$ then return $1$
Else return $g$
```

Note that B is δ-balanced, since for all $b \in \{0,1\}$

$$\left|\Pr\left[\, t=b \,:\, (\mathbf{x},t) \xleftarrow{\$} A_1(1^k) \,\right] - \frac{1}{2}\right| \;\le\; \frac{1}{2(1/\delta)+1}\;.$$

As before, for $d \in \{0,1\}$, let A_d denote the event $\mathbf{Exp}^{\text{priv-d}}_{\Pi,A}(k) \Rightarrow 1$ and similarly B_d denote $\mathbf{Exp}^{\text{priv-d}}_{\Pi,B}(k) \Rightarrow 1$. Then

$$\begin{aligned}
\mathbf{Adv}^{\text{priv}}_{\Pi,B}(k) &= \Pr[\,B_1\,] - \Pr[\,B_0\,] \\
&= \Pr[\,B_1 \mid E\,] - \Pr[\,B_0 \mid E\,] + \Pr\left[\,B_1 \mid \overline{E}\,\right] - \Pr\left[\,B_0 \mid \overline{E}\,\right] \\
&= \Pr[\,B_1 \mid E\,] - \Pr[\,B_0 \mid E\,] + \frac{1}{2} - \frac{1}{2} \\
&= \frac{1}{2} \cdot \mathbf{Adv}^{\text{priv}}_{\Pi,A}(k)\;.
\end{aligned}$$

As before, the claimed running-time of B' is easy to verify. ▌

REDUCTION TO DISTRIBUTION HIDING. Similar to [4] the final component for the proof is as follows.

Proposition A.3 Let $\Pi = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ be an encryption scheme and $B \in \mathbb{A}_{\boldsymbol{M}}$ be a δ-balanced boolean PRIV adversary. Then there is an IND adversary $D \in \mathbb{D}_{\boldsymbol{M}^*[\log(1/(1/2-\delta))]}$ such that

$$\mathbf{Adv}^{\text{priv}}_{\Pi,B}(k) \quad\le\quad 2 \cdot \mathbf{Adv}^{\text{ind}}_{\Pi,D}(k) + \left(\frac{1}{2} + \delta\right)^{-k}\;.$$

In particular, D samples from message distributions that are statistically $2^{\Omega(k)}$-close to complementary $\log(1/(1/2-\delta))$-induced message distributions of B. Furthermore,the running-time of D is the time for at most k executions of B.

Proof: Adversary D works as follows.

```
Algorithm $D_1(b)$:
For $i = 1$ to $k$ do:
    $(\mathbf{x}, t) \xleftarrow{\$} A_1(1^k)$
    If $t = b$ then return $\mathbf{x}$
Return $\mathbf{x}$
```

```
Algorithm $D_2(pk, \mathbf{c})$:
$g \xleftarrow{\$} A_2(pk, \mathbf{c})$
Return $g$
```

Let BAD denote the event that the final return statement is executed. Let $\mathsf{CORRECT}_D$ be the event that $b = d$ when D is executed in the PRIV experiment with Π and simiarly let $\mathsf{CORRECT}_B$ denote the event that $t = g$ when B is executed in the PRIV experiment with Π. Then

$$\begin{aligned}
\mathbf{Adv}^{\text{priv}}_{\Pi,D}(k) &\le \Pr[\,\mathsf{CORRECT}_D \mid b=1\,] + \Pr[\,\mathsf{CORRECT}_D \mid b=0\,] \\
&\le \Pr\left[\,\mathsf{CORRECT}_D \mid b=1 \wedge \overline{\mathsf{BAD}}\,\right] \\
&\quad + \Pr\left[\,\mathsf{CORRECT}_D \mid b=0 \wedge \overline{\mathsf{BAD}}\,\right] + \Pr\left[\,\overline{\mathsf{BAD}}\,\right] \\
&= \Pr\left[\,\mathsf{CORRECT}_D \mid b=1 \wedge \overline{\mathsf{BAD}}\,\right] \\
&\quad + \Pr\left[\,\mathsf{CORRECT}_D \mid b=0 \wedge \overline{\mathsf{BAD}}\,\right] + \left(\frac{1}{2} + \delta\right)^{-k}
\end{aligned}$$

where the last line uses that B is δ-balanced. Now we can lower-bound

$$\Pr\left[\,\mathsf{CORRECT}_D \mid b=1 \wedge \overline{\mathsf{BAD}}\,\right] + \Pr\left[\,\mathsf{CORRECT}_D \mid b=0 \wedge \overline{\mathsf{BAD}}\,\right]$$

by

$$\begin{aligned}
&= \Pr\left[\mathbf{Exp}^{\mathrm{ind}}_{\Pi,A}(k) \Rightarrow 1 \;\middle|\; t=1, b=1\right] - \Pr\left[\mathbf{Exp}^{\mathrm{ind}}_{\Pi,A}(k) \Rightarrow 1 \;\middle|\; t=0, b=1\right] \\
&\geq \Pr\left[\,\mathsf{CORRECT}_B \mid t=1\,\right] + \Pr\left[\,\mathsf{CORRECT}_B \mid t=0\,\right] \\
&\geq \Pr\left[\mathbf{Exp}^{\mathrm{priv}}_{\Pi,A}(k) \Rightarrow 1 \;\middle|\; b=1\right] \\
&\geq \Pr\left[\mathbf{Exp}^{\mathrm{priv}}_{\Pi,A}(k) \Rightarrow 1 \;\middle|\; b=1\right] - \Pr\left[\mathbf{Exp}^{\mathrm{priv}}_{\Pi,A}(k) \Rightarrow 1 \;\middle|\; b=0\right] \\
&= \mathbf{Adv}^{\mathrm{priv}}_{\Pi,B}(k) \; .
\end{aligned}$$

To complete the proof, let $\boldsymbol{M}_i$ be the message distribution sampled by D_1 on input i for $i \in \{0,1\}$. Observe that $\boldsymbol{M}_0 \mid \overline{\mathsf{BAD}}$ and $\boldsymbol{M}_1 \mid \overline{\mathsf{BAD}}$ are complementary $\log(1/(1/2-\delta))$-induced distributions of the message distribution of B, with corresponding events $t=0$ and $t=1$. Since $\Pr[\,\mathsf{BAD}\,] \leq (1/2+\delta)^{-k}$ it follows that $\boldsymbol{M}_i \mid \overline{\mathsf{BAD}}$ is statistically $2^{-\Omega(k)}$-close to $\boldsymbol{M}_i$ for $i \in \{0,1\}$ as required.[10] ▮

Theorem 3.1 now follows by combining Propositions A.1, A.2, and A.3 with $\delta = 1/4$. ▮

B Proof of Lemma 4.3

Let Game G_1 correspond to the IND experiment with D against $\mathsf{EwHCore}$, and let Game G_2 be like G_1 except that the coins used to encrypt the challenge plaintext vector are truly random. For $i \in \{0,1\}$ let $B^i = (B^i_1, B^i_2)$ be the HCF adversary against $\mathcal{F}$ hc defined via

```
Algorithm $B^i_1(1^k)$:
    $\mathbf{x} \xleftarrow{\$} D_1(i)$
    Return $\mathbf{x}$

Algorithm $B^i_2(pk, \mathbf{y}, \mathbf{h})$:
    $\mathbf{c} \leftarrow \mathcal{E}(pk, \mathbf{y}; \mathbf{h})$
    $d \xleftarrow{\$} D_2(pk, \mathbf{c})$
    Return $d$
```

Then

$$\begin{aligned}
\Pr\left[\,G_1^D \Rightarrow b\,\right] &= \Pr\left[\,G_1^D \Rightarrow b \mid b=1\,\right] + \Pr\left[\,G_1^D \Rightarrow b \mid b=0\,\right] \\
&= \Pr\left[\,G_2^D \Rightarrow b \mid b=1\,\right] + \mathbf{Adv}^{\mathrm{hcf}}_{\mathcal{F},\mathsf{hc},B^1}(k) \\
&\quad + \Pr\left[\,G_2^D \Rightarrow b \mid b=0\,\right] + \mathbf{Adv}^{\mathrm{hcf}}_{\mathcal{F},\mathsf{hc},B^0}(k) \\
&\leq \Pr\left[\,G_2^D \Rightarrow b\,\right] + 2 \cdot \mathbf{Adv}^{\mathrm{hcf}}_{\mathcal{F},\mathsf{hc},B}(k)
\end{aligned}$$

where we take B to be whichever of B^0, B^1 has the larger advantage. Now define IND-CPA adversary A against Π via

```
Algorithm $A_1(pk)$:
    $\mathbf{x}_0 \xleftarrow{\$} D_1(0)$
    $\mathbf{x}_1 \xleftarrow{\$} D_1(1)$
    Return $(\mathbf{x}_0, \mathbf{x}_1)$

Algorithm $A_2(pk, \mathbf{c})$:
    $d \xleftarrow{\$} D_2(pk, \mathbf{c})$
    Return $d$
```

Then Equation 1 follows from taking into account the definition of the advantages of D, A. ▮

[10] Note that as compared to [4] our approach avoids having to analyze the min-entropy of D, which is more involved.

C Proof of Lemma 6.1

Writing $\mathbf{E}_k$ for the expectation over the choice of k according to the distribution of K, it follows that

$$\begin{aligned}\Delta\big((K,f(\mathcal{H}(K,\mathbf{X}))),(K,f(\mathbf{U}))\big) &= \mathbf{E}_k\big[\Delta\big(f(\mathcal{H}(k,\mathbf{X})),f(\mathbf{U})\big)\big] \\ &\leq \frac{1}{2}\mathbf{E}_k\Big[\sqrt{|S|^t\cdot D\big(f(\mathcal{H}(k,\mathbf{X})),f(\mathbf{U})\big)}\Big] \\ &\leq \frac{1}{2}\sqrt{|S|^t\cdot \mathbf{E}_k\big[D\big(f(\mathcal{H}(k,\mathbf{X}))),f(\mathbf{U})\big)\big]}\end{aligned}$$

where the first inequality is by Cauchy-Schwarz and the second inequality is due to Jensen's inequality. We will show that

$$\mathbf{E}_k\big[D\big(f(\mathcal{H}(k,\mathbf{X})),f(\mathbf{U})\big)\big] \leq t^2 2^{-\mu}+3\delta\ ,$$

which completes the proof. Write $\mathbf{Y}=\mathcal{H}(k,\mathbf{X})$ for an arbitrary but fixed k. Then

$$\begin{aligned}D\big(f(\mathbf{Y}),f(\mathbf{U}))\big) &= \sum_{\mathbf{s}}\big(P_{f(\mathbf{Y})}(\mathbf{s})-P_{f(\mathbf{U})}(\mathbf{s})\big)^2 \\ &= \sum_{\mathbf{s}}P_{f(\mathbf{Y})}(\mathbf{s})^2-2\sum_{\mathbf{s}}P_{f(\mathbf{Y})}(\mathbf{s})P_{f(\mathbf{U})}(\mathbf{s})+\mathrm{Col}(f(\mathbf{U}))\ .\end{aligned}$$

For a set $Z\subseteq R^t$, define $\delta_{\mathbf{r},Z}$ to be 1 if $\mathbf{r}\in Z$ and else 0. For $\mathbf{s}\in S^t$ we can write $P_{f(\mathbf{Y})}(\mathbf{s})=\sum_{\mathbf{x}}P_{\mathbf{X}}(\mathbf{x})\delta_{\mathcal{H}(k,\mathbf{x}),f^{-1}(\mathbf{s})}$ and thus

$$\begin{aligned}\sum_{\mathbf{s}}P_{f(\mathbf{Y})}(\mathbf{s})^2 &= \sum_{\mathbf{s}}\Big(\sum_{\mathbf{x}}P_{\mathbf{X}}(\mathbf{x})\delta_{\mathcal{H}(k,\mathbf{x}),f^{-1}(\mathbf{s})}\Big)\Big(\sum_{\mathbf{x}'}P_{\mathbf{X}}(\mathbf{x}')\delta_{\mathcal{H}(k,\mathbf{x}'),f^{-1}(\mathbf{s})}\Big) \\ &= \sum_{\mathbf{s},\mathbf{x},\mathbf{x}'}P_{\mathbf{X}}(\mathbf{x})P_{\mathbf{X}}(\mathbf{x}')\delta_{\mathcal{H}(k,\mathbf{x}),f^{-1}(\mathbf{s})}\delta_{\mathcal{H}(k,\mathbf{x}'),f^{-1}(\mathbf{s})}\ ,\end{aligned}$$

so that

$$\begin{aligned}\mathbf{E}_k\Big[\sum_{\mathbf{s}}P_{f(\mathbf{Y})}(\mathbf{s})^2\Big] &= \sum_{\mathbf{s}}\sum_{\mathbf{x},\mathbf{x}'}P_{\mathbf{X}}(\mathbf{x})P_{\mathbf{X}}(\mathbf{x}')\mathbf{E}_k[\delta_{\mathcal{H}(k,\mathbf{x}),f^{-1}(\mathbf{s})}\delta_{\mathcal{H}(k,\mathbf{x}'),f^{-1}(\mathbf{s})}] \\ &= \sum_{\mathbf{s}}\sum_{\exists i,j,\ \mathbf{x}[i]=\mathbf{x}'[j]}P_{\mathbf{X}}(\mathbf{x})P_{\mathbf{X}}(\mathbf{x}') \\ &\quad+\sum_{\mathbf{s}}\sum_{\forall i,j,\ \mathbf{x}[i]\neq\mathbf{x}'[j]}P_{\mathbf{X}}(\mathbf{x})P_{\mathbf{X}}(\mathbf{x}')\mathbf{E}_k[\delta_{\mathcal{H}(k,\mathbf{x}),f^{-1}(\mathbf{s})}\delta_{\mathcal{H}(k,\mathbf{x}'),f^{-1}(\mathbf{s})}] \\ &\leq t^2 2^{-\mu}+\mathrm{Col}(f(U))+\delta\end{aligned}$$

where the first term is by a union bound over all $1\leq i,j\leq t$ and for the remaining terms we use the δ-almost $2t$-wise independence of $\mathcal{H}$ and note that

$$E_k[\delta_{\mathcal{H}(k,\mathbf{x}),f^{-1}(\mathbf{s})}\delta_{\mathcal{H}(k,\mathbf{x}'),f^{-1}(\mathbf{s})}]=\Pr\big[\,f(\mathcal{H}(K,\mathbf{x}))=f(\mathcal{H}(K,\mathbf{x}'))\,\big]\ .$$

Similarly,

$$\begin{aligned}\sum_{\mathbf{s}}P_{f(\mathbf{Y})}(\mathbf{s})P_{f(\mathbf{U})}(\mathbf{s}) &= \sum_{\mathbf{s}}\Big(\sum_{\mathbf{x}}P_{\mathbf{X}}(\mathbf{x})\delta_{\mathcal{H}(k,\mathbf{x}),f^{-1}(\mathbf{s})}\Big)\Big(\frac{1}{|R|}\sum_{\mathbf{u}}\delta_{\mathbf{u},f^{-1}(\mathbf{s})}\Big) \\ &= \frac{1}{|R|}\sum_{\mathbf{s}}\sum_{\mathbf{u},\mathbf{x}}P_{\mathbf{X}}(\mathbf{x})\delta_{\mathcal{H}(k,\mathbf{x}),f^{-1}(\mathbf{s})}\delta_{\mathbf{u},f^{-1}(\mathbf{s})}\end{aligned}$$

so that

$$\begin{aligned}\mathbf{E}_k\Big[\sum_{\mathbf{s}} P_{f(\mathbf{Y})}(\mathbf{s})P_{f(\mathbf{U})}(\mathbf{s})\Big] &= \frac{1}{|R|}\sum_{\mathbf{s}}\sum_{\mathbf{u},\mathbf{x}} P_{\mathbf{X}}(\mathbf{x})\,\mathbf{E}_k[\delta_{\mathcal{H}(k,\mathbf{x}),f^{-1}(\mathbf{s})}\delta_{\mathbf{u},f^{-1}(\mathbf{s})}] \\ &\geq \mathrm{Col}(f(\mathbf{U}))-\delta\end{aligned}$$

using δ-almost t-wise independence of $\mathcal{H}$. By combining the above, it follows that

$$\mathbf{E}_k\big[D\big(f(\mathbf{Y}),f(\mathbf{U})\big)\big] \;\leq\; t^2 2^{-\mu}+3\delta$$

which was to be shown. ▌