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) — which hides all possible partial information about high-entropy messages. Our results serve to unify as well as generalize/strengthen the prior work. First, we propose a general construction of DE that reduces the latter to trapdoor functions admitting certain kinds of hardcore functions. Roughly, the latter should be *robust*, meaning remain hardcore when the input is conditioned on an event that occurs with good probability; for the resulting DE scheme to be multi-message secure, the hardcore function should also be "correlated-input secure." We then show that most known DE schemes, both in the random oracle and standard models, can be viewed as instantiations of our general construction or optimizations thereof, thereby explaining how these different schemes come about. We also give novel instantiations in the standard model, in both the single and multi-message cases:

- In the single-message case, we give an instantiation one-way trapdoor functions, whereas previous constructions based on one-wayness due to Bellare et al. required a trapdoor permutation. In particular, we can use a trapdoor function that is super-polynomially hard to invert on high-entropy distributions but exponentially-hard to invert on uniform inputs.
- In the multi-message case, we give an instantiation meeting a new notion of "q-bounded" multi-message security we introduce, for any polynomial q, based on lossy trapdoor functions losing an $\Omega(1 - 1/q)$ fraction of their input. Prior work achieved only $q = 1$ in the standard model. We also give an optimized version of this instantiation by extending some ideas of Boldyreva et al. Both of these results are based on generalizations of the (Crooked) Leftover Hash Lemma (LHL) that build on that of Kiltz et al. (EUROCRYPT 2009). Interestingly, the optimized scheme relies on the fact that the Crooked LHL is more tolerant of "imperfection" of the hash function than the classical one.

An additional contribution is to give a more precise definitional equivalence (between semantic security and indistinguishability style security notions) for DE as compared to prior work. In particular, this makes our security proofs for instantiations based on one-wayness simpler than in prior work.

Keywords: Deterministic encryption, trapdoor functions, hardcore functions, lossy trapdoor functions, Leftover Hash Lemma, 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.

1.1 Background and Motivation

One of the maxims of modern cryptography is that "randomness is needed for good encryption," meaning 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 [22]. 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 explicitly considered by Russell and Wang [36] in the setting of one-time, information-theoretically secure symmetric-key encryption, and later in various follow-up works [15, 11]. (Other works have, however, targeted computationally-bounded security of a similar spirit [33, 34].)

The focus of our work is instead 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 arbitrarily correlated, 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 [15, 11]), as well as constructions without random oracles that achieve single-message rather than multi-message security (except in the case of block-sources [7]). In particular, the former gave constructions from one-way trapdoor permutations for high-entropy input distributions, and the latter gave constructions from the recent notion of lossy trapdoor functions [31]. More recently, Hemenway et al. [26] showed that a "decisional" version of the notion of correlated product trapdoor function of Rosen and Segev [35] suffices.

Besides being interesting from a foundational standpoint, DE has a number of practical applications, such as efficient search on encrypted data and securing legacy protocols (cf. [2]). Additionally, its study has proven useful in other contexts: Bellare et al. [3] showed how it extends to a notion of "hedged" public-key encryption that reduces dependence on external randomness for probabilistic encryption more generally, and Dent et al. [10] adapted its notion of privacy to a notion of confidentiality for digital signatures. However, our current understanding of DE is a somewhat lacking. The constructions of [2, 4, 7, 26], as well as their analysis techniques, are rather disparate. Furthermore, it is unclear to what extent the single-message security achieved by [4, 7, 26] represents an inherent limitation of standard model schemes. Accordingly, in this work our main goals are to provide a *unified framework* for the construction of DE and to achieve *as-strong-as-possible* notions of multi-message security without ROs. We view our main contributions as mostly conceptual; in particular, we show how (by abstracting out new constructions and notions) various DE schemes can be explained in a common way, their security proofs can be made simpler and more modular, and their associated techniques can be pushed further.

1.2 Our Results

A MORE PRECISE DEFINITIONAL EQUIVALENCE. We begin by revisiting 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], which asks that a scheme hides all public-key independent[2] functions of the data is in some sense equivalent to an indistinguishability-based notion for DE (called IND), which asks that a scheme hides the

[1] Actually, it should be noted that Russell and Wang [36] 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 can be met by deterministic schemes.

[2] As shown in [2], the restriction to public-key independent functions is inherent here.

"source" from which the data is drawn, meaning it is hard to distinguish ciphertexts whose corresponding plaintexts are drawn from one of two possible distributions. Notice that while PRIV can be meaningfully said to hold for a given plaintext distribution, IND inherently talks of *pairs* of distributions. The works of [4, 7] compensated for this by giving an equivalences in terms of *entropy levels*. That is, they showed that PRIV for all distributions on plaintext vectors of min-entropy μ[3] is equivalent to indistinguishability with respect to all pairs of plaintext distributions of min-entropy slightly less than μ. However, a more precise equivalence would identify, for a *fixed* distribution $\mathbf{X}$, a class of pairs of distributions on which IND is equivalent to PRIV on $\mathbf{X}$. Via a re-examination of the equivalence proof of [4], we do exactly that. Namely, we show that PRIV on $\mathbf{X}$ is equivalent to IND on the class of pairs of *complementary induced distributions* of $\mathbf{X}$, meaning each pair in the class is obtained by conditioning $\mathbf{X}$ on whether some (efficiently testable) event E occurs or not, where both possibilities happen with good probability. Besides being more technically precise, this equivalence enables simpler and more modular security proofs, as we will see.

A CONSTRUCTION FROM "ROBUST" HARDCORE FUNCTIONS. Consider the following natural construction of DE. We start with a one-way trapdoor function (OW-TDF) f on an input distribution X, a pseudorandom generator (PRG) g, and a semantically-secure *probabilistic* encryption scheme $\Pi = (\mathcal{K}, \mathcal{E}, \mathcal{D})$. Furthermore, suppose that f has a hardcore function (HCF) h whose output length is equal to the seed length of g, and that the output length of the latter is equal to the number of random coins consumed by $\mathcal{E}$. Then define the "Encrypt-with-Hardcore" DE scheme to encrypt plaintext x as $\mathcal{E}(pk, f(x); g(h(x)))$; decryption is defined naturally. For now, let us just consider security of the scheme for a single high-entropy message. Due to the pseudorandomness of h, 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 *partial information* on x (for example, h may output some of the input bits) and Π may not protect partial information about its coins. Thus, some extra condition on h is needed to ensure security of the resulting scheme. What is it? It is a new notion of "robustness" we introduce, roughly meaning h remains on distribution when we condition the input to f on any (efficiently testable) event that occurs with good probability, which we call "slightly induced" distributions. Security then follows according to our precise definitional equivalence.

Note that semantically secure encryption can be constructed from a OW-TDF on any input distribution. A PRG with seed length k can also be constructed from any injective OWF on on $\{0,1\}^k$ (the trapdoor is not needed here) that is one-way on the standard uniform inputs, which is a natural starting requirement for the OW-TDF that we will anyway need to get security of the resulting DE scheme on all distributions of a certain min-entropy. So our security analysis of the Encrypt-with-Hardcore scheme effectively reduces existence of a DE scheme for a given input distribution to that of a TDF with a large robust HCF for it.

INSTANTIATIONS FROM ONE-WAYNESS. In the RO model, any trapdoor function that is one-way on a given input distribution has a simple, large HCF on it, namely the RO hash itself. This follows because (1) any one-way function (OWF) is also one-way on slightly induced input distributions and (2) a RO hash is a HCF for any OWF, regardless of the input distribution on which the latter 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. (However, as observed in [2], due to [20] such an optimization is not possible in the standard model, which explains why we use a TDF in our general construction.)

To obtain robust HCF based on one-wayness without ROs, inspired by [4] we consider the well-known Goldreich-Levin (GL) bit [21], 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.[4] Recall that [21] showed this bit is similarly hardcore for any OWF, regardless of the input distribution on which the latter is one-way. Robustness then follows by the same argument as above. This leads to instantiations of the Encrypt-with-Hardcore scheme from one-wayness in the standard model based respectively on trapdoor permutations (TDPs) and trapdoor functions (TDFs).

In order to instantiate the Encrypt-with-Hardcore scheme on plaintext distribution X based on a TDP, we need a TDP that is one-way on all "permutation distributions" of X. Then we can use "BMY iteration" [6, 38], extracting

[3] This means that each individual component has min-entropy μ.

[4] Indeed, in the standard model no hardcore function h that does not use public coins can be robust; otherwise we could find an induced distribution on which the function is no longer hardcore.

a GL bit on each iteration; thus, the TDF we use to instantiate the scheme is the iterated TDP. Note that we do not need to use a separate PRG in this case. (One might view this as an optimization of the construction, but the PRG can always be "absorbed" into the HCF anyway.) This scheme is exactly that of [4]. In order to instantiate the Encrypt-with-Hardcore scheme on plaintext distribution X using a TDF, one possiblity is to assume a TDF that is exponentially-hard to invert on X.[5] Then [21] tells us the TDF has a linear-sized HCF on X, which we can expand using a PRG. The latter can be built assuming the TDF is also polynomially-hard on uniform inputs; more generally we can use any injective OWF (injectivity can be dropped for a much larger seed length). This HCF is robust because one-way hardness is "preserved" on slightly induced distributions. A more appealing (though technically incomparable) instantiation is to assume an injective OWF that is exponentially-hard on standard uniform inputs and construct an *exponentially-hard PRG* (here injectivity can be dropped for quadratic seed length [24]). Then the HCF of the TDF on X need only have super-logarithmic rather linear length, translating to a much more reasonable super-polynomial hardness assumption for the TDF on X. Achieving DE from one-way TDFs is new to our work.

We make some remarks about the above instantiations. First, while our security analyses allow us to "zero-in" on what makes these constructions secure on a fixed input distribution, as in [4, 7] we ultimately want to target security of these instantiations on all distributions of some min-entropy. What we then need to assume is (respectively) a TDP that is polynomially-hard on all such distributions or a TDF that is super-polynomially hard on all such distributions as well as exponentially-hard on the uniform distribution. Note that the latter is a potentially weaker assumption; for example Gertner et al. [19] show a black-box separation of OW-TDPs from OW-TDFs on uniform inputs. Finally, although the scheme we obtain from TDPs is known, our approach enables a simple and modular security proof as compared to [4] (which is facilitated by our precise definitional equivalence; see Appendix A for a discussion). Our analysis is also more general and shows that any robust hardcore bit (not just GL) would work.

INSTANTIATIONS FROM LOSSINESS. We next show that lossy trapdoor functions (LTDFs) [31], which have a public description indistinguishable from that of a lossy function, also admit a large robust hardcore function. Recall that [31] showed LTDFs admit a simple, large hardcore function, namely a pairwise-independent hash function as per the (generalized) Leftover Hash Lemma (LHL) [25, 13]. Notice that (1) randomness extractors simply require a high-entropy distribution, and (2) a slightly induced distribution of a high-entropy distribution is also high-entropy. It follows that the hardcore function here is also robust on any input distribution with sufficient entropy. We thus obtain another instantiation of the Encrypt-with-Hardcore scheme from any LTDF losing a constant fraction of its input. Boldyreva et al. [7] previously obtained DE from LTDFs by encrypting under (an augmented version of) the latter directly. Their construction can be viewed as an optimization of our instantiation that drops the "outer" encryption; on the other hand, their security proof requires the less-standard "Crooked" LHL [15].)

One should note the the argument to show robustness here is entirely analogous that to show robustness of the GL function. We believe this gives a unified view of how DE can be derived from both one-wayness and lossiness.

RELATION TO DECISIONAL CORRELATED PRODUCT. We briefly discuss the relation of our general construction to the recent one of Hemenway et al. [26] 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 pairwise independent hash function 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 function is hardcore, but showing robustness is more subtle. For this, one might apply the techniques of [16, Lemma 3]. However, due to the non-standard nature of the assumption we have not worked out the details. (Note that [26] only constructs decisional correlated product functions without a trapdoor.)

[5] Here s-hardness means that adversaries running in time s have advantage at most $1/s$. In particular exponential hardness means $s - hardness$ for exponential s. Standard polynomial hardness is p-hardness for all polynomials p.

SECURITY FOR MULTIPLE MESSAGES. An important point is that, as in [4, 7], we can only prove the above standard-model DE schemes secure for the encryption of a *single* high-entropy plaintext, or, what was shown equivalent in [7], an unbounded number of messages drawn from a *block source* [8], where each subsequent message brings "fresh" entropy. (This equivalence requires either efficient conditional re-sampleability of the blocks or single-message security even for plaintext distributions that are not efficiently sampleable[6]). On the other hand, the strongest and most practical 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 (i.e., "fresh") entropy. In order for our Encrypt-with-Hardcore scheme to achieve this, the hardcore function must also be secure on *correlated inputs*; see Section 4.1 for the definition. (A general study of correlated-input security for the more basic case of hash functions rather than hardcore functions was concurrently initiated in [23].) In particular, it follows from the techniques of [2] that a RO hash satisfies such a notion. This leads to a multi-message secure scheme in the RO model (as obtained in [2]). We thus have a large gap between what is (known to be) achievable there versus the standard model.

BOUNDED MULTI-MESSAGE SECURITY. To help bridge this gap, we propose a notion of "q-bounded" multi-message (or just q-bounded) security for DE, where up to q high entropy but arbitrarily correlated messages may be encrypted under the same public key. Following [7], we extend this to 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 an analogous requirement to that mentioned above) that q-bounded multi-message security and unbounded multi-message security for q-block sources are equivalent (for a given min-entropy). 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).

q-BOUNDED DE FROM LOSSY TRAPDOOR FUNCTIONS. We show q-bounded DE schemes (for long enough messages), for any polynomial q, based on a LTDFs losing an $\Omega(1 - 1/q)$ fraction of the input. For any polynomial q, such constructions are known from the decisional Diffie-Hellman [31], d-linear [18], and decisional composite residuosity [7, 18] assumptions. Even achieving any constant $q > 1$ was an open problem prior to our work.

We first give our "basic scheme," which is an instantiation of our general Encrypt-with-Hardcore scheme with such LTDFs, using a $2q$-wise independent hash function as the hardcore function. That is, we show that the a $2q$-wise independent hash function is a "correlated-input secure" hardcore function for such lossy TDFs for up to q inputs. To prove this, we introduce a generalization of the LHL to t-many correlated sources, following Kiltz et al. [28, Lemma 3.2] who considered $t = 2$. We then give an "optimized scheme" — which improves both ciphertext length and required input entropy — by extending some ideas of [7]. (On the other hand, it has a more complicated analysis; we view the fact that our basic scheme as a feasibility result has a simpler analysis as an additional benefit of our general Encrypt-with-Hardcore construction.) The idea is to drop the "outer" encryption and first pre-process a message with a $2q$-wise independent *permutation* (instead of a pairwise-indepndent permutation as in [7]). However, explicit and efficiently computable permutations with independence greater than 3-wise are not known to exist. So, we instead use a $2q$-wise "δ-dependent" permutations (see e.g. [27]) for appropriate δ. We prove security via a similar generalization of the "Crooked" LHL [14] (which was used in [7]) we introduce, but which also takes into account an "error" term for the $2q$-wise independent permutation. In fact, there are two error terms to deal with here: one from δ and one from the fact that in the proof we must "switch" from a random permutation to random function. While by using the constructions of [27] the former term can be made arbitrarily small, the latter term cannot. Interestingly, we are still able to tolerate it because the Crooked LHL turns out to be more tolerant than the classical one in this regard. (More explaination of this technical issue can be found in Section 6.4.)

Note that our constructions show that it is not non-block-sources but an *unbounded* number of messages with low conditional entropy that forms the "separating line" between what we can achieve in the RO and standard models using current techniques. (It also shows that it is when considering non-block-source multi-message security we need to

[6]For constructions based on lossy trapdoor function this requirement is immaterial since they are actually secure for plaintext distributions that are not efficiently sampleable. We also observe that this is true for some instantiations based on one-wayness; see Section 5.1.

exploit the extra power of lossiness over one-wayness.) Indeed, the above approach cannot extend to unbounded multi-message security since, information-theoretically, the messages plus the public key do not contain enough entropy. Achieving unbounded multi-message security without random oracles in the remains an interesting open problem. Some partial results in the case of hash functions (i.e., without supporting decryption) were recently obtained in [23].

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.

If A is an algorithm then $x \xleftarrow{\$} A(\ldots)$ 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 $s \xleftarrow{\$} S$ denotes that s is assigned a uniformly random element of S. We let $A(\ldots) \Rightarrow y$ denote the event that A outputs y in the above experiment. We use the abbreviation $x_1, \ldots, x_n \xleftarrow{\$} A(\ldots)$ for $x_1 \xleftarrow{\$} A(\ldots)\,; \ldots ; x_n \xleftarrow{\$} A(\ldots)$, and similarly for sets. If A is deterministic then we drop the dollar sign above the arrow. We denote by $\{0,1\}^*$ the set of all (binary) strings, and by $\{0,1\}^n$ the set of strings of length n. By $x_1 \| \cdots \| x_m$ we denote an encoding of strings $x_1, \ldots, x_m$ from which $x_1, \ldots, x_m$ 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 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 $1 \le i \le |\mathbf{x}|$. For convenience, we extend algorithmic and functional notation to operate on each vector of inputs component-wise. For example, if A is an algorithm and $\mathbf{x}, \mathbf{y}$ are vectors then $\mathbf{z} \xleftarrow{\$} A(\mathbf{x}, \mathbf{y})$ denotes that $\mathbf{z}[i] \xleftarrow{\$} A(\mathbf{x}[i], \mathbf{y}[i])$ for all $1 \le i \le |\mathbf{x}|$.

STATISTICAL NOTIONS. Let X be a random variable on a finite set $\mathcal{X}$. 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 $x \xleftarrow{\$} X$ we denote that x is assigned a value drawn according to P_X. When this experiment is PPT we say that X is *efficiently sampleable*. We write $X \mid \mathcal{E}$ for the random variable X conditioned on an event $\mathcal{E}$. When X is vector-valued we denote it in boldface, for example $\mathbf{X}$. 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 $\mathrm{H}_\infty(X|Y) = -\log(\max_{x,y} P_{X|Y=y}(x))$, and the *average conditional* min-entropy of X given Y [13] is $\tilde{\mathrm{H}}_\infty(X|Y) = -\log(\sum_y P_Y(y) \max_x P_{X|Y=y}(x))$. Following [2, 4], for vector-valued $\mathbf{X}$ the min-entropy is the minimum *individual* min-entropy of the components, i.e., $\mathrm{H}_\infty(\boldsymbol{X}) = -\log(\max_{\boldsymbol{x},i} P_{\boldsymbol{X}[i]}(\boldsymbol{x}[i]))$. The *collision probability* of X is $\mathrm{Col}(X) = \sum_x P_X(x)^2$. 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*, and 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 $y_1, \ldots, y_t \in R$

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

In other words, $F(K, x_1), \ldots, F(K, x_t)$ are all uniformly and independently random over R. 2-wise independence is also called *pairwise independence*.

LEFTOVER HASH LEMMA. Next we recall the Generalized Leftover Hash Lemma due to Dodis et al [13], which extends the original version of [25] to average conditional min-entropy.

Lemma 2.1 (Generalized Leftover Hash Lemma) [13] Let $\mathcal{H}\colon \mathcal{K} \times D \to R$ be a pairwise-independent function. Let X be a random variable over D and Y be another random variable. Then

$$\Delta\big((K, Y, \mathcal{H}(K, X)), (K, Y, U)\big) \leq \frac{1}{2}\sqrt{|R|\tilde{\mathrm{H}}_\infty(X|Y)}\ ,$$

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

PUBLIC-KEY ENCRYPTION. A *(probabilistic) 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 $m \in \mathrm{PtSp}$

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

Next we define security against chosen-plaintext attack [22]. 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

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

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* [31] is a pair $\mathsf{LTDF} = (\mathcal{F}, \mathcal{F}')$ of algorithms. Algorithm $\mathcal{F}$ is a usual trapdoor function (TDF) generator, namely on inputs 1^k outputs outputs (a description of a) function f on $\{0,1\}^n$ for $n = n(k)$ along with (a description of) its inverse f^{-1}, and algorithm $\mathcal{F}'$ outputs a (description of a) function f' on $\{0,1\}^n$. For a distinguisher D, define its *LTDF advantage* against LTDF as

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

We say that LTDF is *secure* 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 $\ell = n - s$.

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]. (More specifically, it is a "comparison-based" semantic-security style notion; this was shown equivalent to a "simulation-based" formulation in [4].) 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 $\mathbf{Exp}^{\text{priv-1}}_{\Pi,A}(k)$:	**Experiment** $\mathbf{Exp}^{\text{priv-0}}_{\Pi,A}(k)$:
$state \xleftarrow{\$} A_0(1^k)$	$state \xleftarrow{\$} A_0(1^k)$
$(\mathbf{x}_1, t_1) \xleftarrow{\$} A_1(state)$	$(\mathbf{x}_1, t_1), (\mathbf{x}_0, t_0) \xleftarrow{\$} A_1(state)$
$\mathbf{c} \xleftarrow{\$} \mathcal{E}(pk, \mathbf{x}_1)$	$\mathbf{c} \xleftarrow{\$} \mathcal{E}(pk, \mathbf{x}_0)$
$g \xleftarrow{\$} A_2(pk, \mathbf{c}, state)$	$g \xleftarrow{\$} A_2(pk, \mathbf{c}, state)$
If $g = t_1$ Return 1 Else Return 0	If $g = t_1$ Return 1 Else Return 0

We require that there are functions $v = v(k), \ell = \ell(k)$ such that (1) $|\mathbf{x}| = v$, (2) $|\mathbf{x}[i]| = \ell$ for all $1 \le i \le v$, and (3) the $\mathbf{x}[i]$ are all distinct with probability 1 over $(\mathbf{x}, t) \xleftarrow{\$} A_1(state)$ for any $state$ output by A_0. (Since in this work we only consider the definition relative to deterministic Π requirement (3) is without loss of generality.) In particular we say *A outputs vectors of size v* for v as above. Define the *PRIV advantage* of A against Π as

$$\mathbf{Adv}^{\text{priv}}_{\Pi,A}(k) = \Pr\left[\, \mathbf{Exp}^{\text{priv-1}}_{\Pi,A}(k) \Rightarrow 1 \,\right] - \Pr\left[\, \mathbf{Exp}^{\text{priv-0}}_{\Pi,A}(k) \Rightarrow 1 \,\right] .$$

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 $\mathbf{Adv}^{\text{priv}}_{\Pi,A}(\cdot)$ is negligible for any PPT $A \in \mathbb{A}_{\mathbb{M}}$. Note that (allowing non-uniform adversaries as usual) 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 [15]). 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 $\mathbf{Exp}^{\text{ind}}_{\Pi,A}(k)$:
$b \xleftarrow{\$} \{0,1\}$; $(\mathbf{x}, t) \xleftarrow{\$} D_1(b)$
$\mathbf{c} \xleftarrow{\$} \mathcal{E}(pk, \mathbf{x})$
$d \xleftarrow{\$} D_2(pk, \mathbf{c})$
If $b = d$ return 1 else return 0

We make the analgous requirements on D_1 as on A_1 in the PRIV definition. Define the *IND advantage* of D against Π as

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

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 $\mathbf{Adv}^{\text{ind}}_{\Pi,D}(\cdot)$ 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 a notion of "induced distributions." Let X, X' be distributions (or random variables) on the same domain. For $\alpha \in \mathbb{N}$, we say that X' is an

α-induced distribution of X if X' is a conditional distribution $X' = X \mid \mathcal{E}$ for an event $\mathcal{E}$ such that $\Pr[\mathcal{E}] \geq 2^{-\alpha}$. We call $\mathcal{E}$ the *corresponding event* to X'. We require that the pair (X, E) is efficiently sampleable (where we view event E as a binary random variable). Define $X[\alpha]$ to be 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 call X_0, X_1 *complementary* if $E_1 = \overline{E_0}$. Define $X^*[\alpha] = \{(X_0, X_1)\}$ to be the class of all pairs (X_0, X_1) for which there is a pair (X_0', X_1') of complementary α-induced distributions of X such that X_0 (resp. X_1) is statistically close to X_0' (resp. X_1').[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 to $\boldsymbol{M}$ 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 $k \in \mathbb{N}$

$$\mathbf{Adv}^{\mathrm{priv}}_{\Pi,A}(k) \quad \leq \quad 18 \cdot \mathbf{Adv}^{\mathrm{ind}}_{\Pi,D}(k) + \left(\frac{3}{4}\right)^{-k} .$$

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

The theorem essentially 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]. For completeness, we give the entire proof (incorporating simplifications due to [10] that lead to better concrete security) in Appendix B. We note the fact that $\boldsymbol{M}^*[2]$ contains pairs of *complementary* 2-induced distributions of $\boldsymbol{M}$ is needed for our equivalence proof but is not used for our constructions.

DISCUSSION. One may wonder why our equivalence result is any better than that in prior work. After all, in practice it is natural to characterize data according to its min-entropy, 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 (recall that this refers to the *individual* min-entropy of any 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 simpler and more modular and unified security proofs for various determinsitic encryption schemes, giving more insight into their security. It especially simplifies the security proofs for constructions based on one-wayness 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 TDF 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

[7] We need to allow a negligible statistical distance for technical reasons; cf. Proposition B.3. (This relaxation is reminiscent of the notion of *smooth* entropy [32] by Renner and Wolf.) Since we will be interested in indistinguishability of functions of these distributions this will not make any appreciable difference, and hence we mostly ignore this issue in the remainder of the paper.

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

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

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

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, $\mathbf{Adv}^{\text{owf}}_{\mathcal{F},\boldsymbol{X},I}(\cdot)$ is negligible. We extend hardore functions (HCFs) 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 $\mathbf{Exp}^{\text{hcf}}_{\mathcal{F},\mathsf{hc},\boldsymbol{X},D}(k)$:
$b \xleftarrow{\$} \{0,1\}$; $(f, f^{-1}) \xleftarrow{\$} \mathcal{F}$
$\mathbf{x} \xleftarrow{\$} \boldsymbol{X}$
$\mathbf{h}_0 \leftarrow \mathsf{hc}(f, \mathbf{x})$; $\mathbf{h}_1 \xleftarrow{\$} (\{0,1\}^n)^{\times|\mathbf{x}|}$
$d \xleftarrow{\$} D(f, f(\mathbf{x}), \mathbf{h}_b)$
If $d = b$ return 1 else return 0

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

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

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, $\mathbf{Adv}^{\text{hcf}}_{\mathcal{F},\mathsf{hc},\boldsymbol{X},D}(\cdot)$ is negligible.

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

ROBUSTNESS. We are now ready to define our new notion of *robustness* for HCFs. Intuitively, robust HCFs are those that remain one-way when the input is conditioned on an (efficiently testable) event that occurs with good probability. We expand on this below.

Definition 4.1 Let $\mathcal{F}$ be a TDF generator and let hc be a HCF such that hc is hardcore for $\mathcal{F}$ with respect to a distribution $\boldsymbol{X}$ on input vectors. For $\alpha = \alpha(k)$, 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 $\boldsymbol{X}$.

Robustness as we defined it is actually a combination of two different notions. To see this, first consider the classical definition of HCFs, where hc is boolean and a single uniform input x is generated in the secuirty experiment. Here robustness means that hc remains hardcore even when x is conditioned on an event that occurs with good probability. (Note for example that while every bit of the input to RSA is well-known to be hardcore assuming RSA is one-way [1], they are not even 1-robust since when we may condition on a particular bit of the input being a fixed value.) Now observe that the definition furthermore asks for a notion of *correlated-input security*, meaning that outputs of hc on correlated inputs look uniformly and independently random. (In fact, robustness asks that hc remain hardcore even when such an input vector is conditioned on an event that occurs with good probability.) We note that a treatment of correlated-input security for the more basic case of hash functions was concurrently initiated in [23].

It may also be interesting to explore robustness in other contexts. In particular, leakage resilience [30] and computational randomness extraction (or key derivation) [29] come to mind. In these applications, robustness for large α may be interesting.

4.2 The Encrypt-with-Hardcore Scheme

THE SCHEME. Let $\Pi = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ be a probabilistic encryption scheme, $\mathcal{F}$ be a TDF generator, g be a pseudorandom generator (PRG) — see e.g. [4] for the standard definition) — and hc be a HCF. Assume that hc, g have the property $g(\mathsf{hc}_f(x)) \in \mathsf{Coins}_{pk}(|x|)$ for all pk output by $\mathcal{K}$ and all $x \in \{0,1\}^*$. Define the associated "*Encrypt-with-Hardcore*" deterministic encryption scheme $\mathsf{EwHCore}[\Pi, \mathcal{F}, \mathsf{hc}, g] = (\mathcal{K}, \mathcal{DE}, \mathcal{DD})$ with plaintext-space $\mathrm{PtSp} = \{0,1\}^k$ via

Algorithm $\mathcal{K}(1^k)$:	**Algorithm** $\mathcal{DE}((pk, f), x)$:	**Algorithm** $\mathcal{DD}((sk, f^{-1}), c)$:
$(pk, sk) \xleftarrow{\$} \mathcal{K}(1^k)$	$r \leftarrow g(\mathsf{hc}(f, x))$	$y \leftarrow \mathcal{D}(sk, c)$
$(f, f^{-1}) \xleftarrow{\$} \mathcal{K}(1^k)$	$c \leftarrow \mathcal{E}(pk, f(x); r)$	$x \leftarrow f^{-1}(y)$
Return $((pk, f), (sk, f^{-1}))$	Return c	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 $\mathsf{EwHCore} = \mathsf{EwHCore}[\Pi, \mathcal{F}, \mathsf{hc}, g]$ on $\boldsymbol{X}$ then follows by pseudorandomness of g 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 how our notion of robustness comes into play.

Theorem 4.2 Suppose Π is IND-CPA secure, hc is *2-robust* for $\mathcal{F}$ on a distribution $\boldsymbol{M}$ on input vectors, and g is pseudorandom. Then $\mathsf{EwHCore}[\Pi, \mathcal{F}, \mathsf{hc}, g]$ 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. The PRG g (which we may allow to use public randomness put in the public key of the DE scheme) can similarly be be built from any injective one-way function (which we are already assuming), although we need the latter to be one-way on the standard uniform distribution, which is a natural basic requirement for it that we will need anyway to achieve security on all distributions of a given min-entropy. Thus, essentially the only cryptographic assumption we need for security of $\mathsf{EwHCore}$ is security of hc. The theorem follows from combining Theorem 3.1 with 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, hc is hardcore for $\mathcal{F}$ on a distribution $\boldsymbol{M}$ on input vectors, and that g is pseudorandom. Then $\mathsf{EwHCore} = \mathsf{EwHCore}[\Pi, \mathcal{F}, \mathsf{hc}, g]$ 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, an adversary B against hc on $\boldsymbol{M}$, and an adversary B' against g such that for all $k \in \mathbb{N}$

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

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

The proof is in Appendix C.

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], who used the PRIV notion directly. (The reason they did so is that their definitional equivalence is awkward to apply in arguing security of schemes based on one-wayness; see Section 5.1 and Appendix A for further discussion.)

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 describes how "one-way hardness" on an input distribution is preserved on induced distributions.

Lemma 5.1 Let $\mathcal{F}$ be a TDF 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 $k \in \mathbb{N}$

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

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

The proof is in Appendix D.

Note that when $\alpha = O(\log k)$ the reduction incurs a polynomial loss in advantage.

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 OW-TDF 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 an input distribution $\boldsymbol{X}$ based on any trapdoor function that is one-way on $\boldsymbol{X}$. 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 in fact secure in the strongest multi-message sense of [2], as a RO is correlated-input secure.

ROBUSTNESS OF GOLDREICH-LEVIN. In the standard model, for single-message security we can replace the RO with the Goldreich-Levin (GL) hardcore function [21]. To define the latter, 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 $(f,f^{-1}) \xleftarrow{\$} \mathcal{F}(1^k)$ and $K \xleftarrow{\$} \mathcal{K}$; 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 (GL) function* $\mathcal{GL}^i\colon \{0,1\}^{i\times k} \times \{0,1\}^k \to \{0,1\}^i$ as $GL^i(M,x) = Mx$, where Mx is the matrix-vector product of M and x over $GF(2)$. We recall the following.

Theorem 5.3 (**Goldreich-Levin Theorem** [21]) 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 $k \in \mathbb{N}$

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

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

Then 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. In the case of trapdoor *functions*, we cannot use BMY iteration, and in general a OW-TDF is only guarateed in general to have a GL HCF with logarithmic-sized output. Hence we consider trapdoor functions that are super-polynomially hard to invert. Namely, say that a TDF $\mathcal{F}$ is s-hard for $s = s(k)$ on input distribution X if for every inverter that runs in time s, $\mathbf{Adv}^{\mathrm{owf}}_{\mathcal{F},X,I}(k) \leq 1/s$. The standard notion of polynomial one-wayness asks for p-hardness for every polynomial s. Theorem 5.3 combined with Proposition 5.3 tells us that if $\mathcal{F}$ is s-hard on M then it has a robust HCF (namely the GL function) on M with output-size $c' \log s$ for some constant c'.

To instantiate the Encrypt-with-Hardcore scheme on plaintext distribution M, one possibility is to use a TDF $\mathcal{F}$ that is exponentially-hard to invert on M, meaning 2^{ck}-hard for some constant c. We then obtain a robust HCF with output size $c'k$ for another constant c', which we can expand using a PRG. To realize the latter from $\mathcal{F}$, we can also assume $\mathcal{F}$ is polynomially-hard to invert on the standard uniform distribution. More generally, we can assume any injective OWF that satisfies this requirement (injectivity could be dropped at the price of a much longer seed length and hence message length here using existing PRG constructions).

A more appealing (but strictly incomparable) way to instantiate the Encrypt-with-Hardcore scheme on plaintext distribution M is to use an *exponentially-hard PRG* ("exponentially-hard" for a PRG is defined analogously) in the scheme rather than one that is just polynomially-hard. To build such a PRG we can assume an injective OWF (which may also be the TDF for our construction) that is exponentially-hard to invert on standard uniform inputs (in fact, without injectivity Haitner et al. [24] construct an exponentially-hard PRG with seed length only $O(k^2)$). When using the resulting PRG in the Encrypt-with-Hardcore construction, the TDF then only needs to have a robust HCF on M with output-size $\omega(\log k)$, since the latter can then be expanded via the exponentially-hard PRG to any polynomial length with polynomial security. For this we only need the TDF to be super-polynomially hard, i.e., s-hard on M for some $s = \omega(\log k)$, which is much milder. We make explicit the following corollary.

Corollary 5.5 Suppose $\mathcal{F}$ is a TDF generator that is s-hard for some $s = \omega(\log k)$ on an input distribution M, as well as exponentially-hard on the uniform distribution. Then we obtain a PRIV secure DE scheme on M. As a consequence, if $\mathcal{F}$ is s-hard for some $s = \omega(\log k)$ on the class of all input distributions with min-entropy μ, as well as exponentially-hard on the uniform distribution, then we obtain a PRIV secure DE scheme for this class.

Obtaining DE from one-way TDFs rather than permutations is new to this work.

ENCRYPT-WITH-HARDCORE INSTANTIATIONS FROM TRAPDOOR PERMUTATIONS. In the case of trapdoor *permutations*, instead of assuming a TDP generator $\mathcal{F}$ that is exponentially hard to invert on input distribution X, we would like to use Blum-Micali-Yao iteration [6, 38] to extract many simultaneously hardcore bits assuming $\mathcal{F}$ is only polynomially hard to invert on X. The catch is that to use the latter we will actually need to assume that $\mathcal{F}$ is 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. (But still there is no inherent restriction on $\mathrm{H}_\infty(X)$ other than $\mathrm{H}_\infty(X) = \omega(\log k)$.) Namely, let $\mathcal{F}$ be a trapdoor permutation generator and hc be a hardcore bit for $\mathcal{F}$. For $i \in \mathbb{N}$ denote by $\mathcal{F}^i$ the trapdoor permutation generator that iterates $\mathcal{F}$ i-many times. Define the Blum-Micali-Yao (BMY) [6, 38] hardcore function for $\mathcal{F}^i$ via

$$\mathcal{BMY}^i[\mathcal{GL}](f, x) = \mathsf{hc}(x) \| \mathsf{hc}(f(x)) \| \ldots \| \mathsf{hc}(f^{i-1}) \ .$$

The following shows how Blum-Micali-Yao iteration expands one robust hardcore bit to many.

Proposition 5.6 Let $\mathcal{F}$ be a trapdoor permutation generator, and let X be an input distribution such thathc is α-robust for $\mathcal{F}$ on all efficiently sampleable permutation distributions of X. Then for any polynomial i, $\mathcal{BMY}^i[\mathsf{hc}]$ is α-robust for $\mathcal{F}^i[\mathsf{hc}]$ on X.

The proof combines the hybrid argument of [6, 38] 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").

In particular, by taking $\mathsf{hc} = \mathcal{GL}$ (the Goldreich-Levin hardcore bit) above, we have the following corollary.

Corollary 5.7 Suppose $\mathcal{F}$ is a TDP generator that is one-way on all permutation distributions of an input distribution M. Then we obtain a PRIV secure DE scheme on M. As a consequence (previously shown in [4]), 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 we do not need to use a separate PRG for this instantiation, since we already obtain a HCF with arbitrary polynomial length. This scheme is exactly the one given in [4]. However, we believe our proof based on robustness properties of Goldreich-Levin is simpler and more modular, giving more insight into the scheme (though we note that the robustness aspect is somewhat implicit in the security proof of [4], in particular in their notion of "PRGs with help" and their using the "balanced boolean" version of PRIV in showing how to achieve it via GL bits). It also reveals why the Goldreich-Levin bit "works" in the construction. Indeed, any robust hardcore bit for the TDP would do.

INSTANTIATIONS FROM ONE-WAYNESS ON (JUST) UNIFORM INPUTS. We note that to instantiate the above schemes on all input distributions of min-entropy μ, we can in fact obtain the required assumptions from (respectively) a TDF or TDP that is sufficiently hard to invert on the standard uniform distribution, where the required hardness depends on μ. This follows by Lemma 5.1 and a technical lemma [17, Lemma 4] saying that every distribution on $\{0,1\}^k$ with min-entropy $k - \alpha$ can be viewed as an α-induced distribution of the uniform distribution on $\{0,1\}^k$, although the corresponding event may not be efficiently testable (which is not needed for Lemma 5.1, though). Indeed, it follows that, on inputs of min-entropy $\mu = k - \alpha$, for we can use (respectively) TDF or TDP that is $2^{-\alpha-\omega(\log k)}$-hard on uniform inputs (where in the TDF case we assume this is exponential hardness).

An advantage of these instantiations is that they achieve security even for input distributions that are not necessarily efficiently sampleable. As a consequence, 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 [31] showed that LTDFs admit a simple, large hardcore function in the standard model, namely a pairwise-independent hash function. We 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 random variable with $\mathrm{H}_\infty(X) \geq \mu$, and let X' be a random variable where $P_{X'}$ is a an α-induced sub-distribution of P_X. Then $\mathrm{H}_\infty(X') \geq \mu - \alpha$.

The proof is in Appendix D.

By combining Lemma 2.1 with the "chain rule" for average conditional min-entropy [13, Lemma 2.2], it follows that if $\mathcal{F}$ is a lossy trapdoor function generator with residual leakage s, then a pairwise-independent hash function $\mathcal{H}\colon \mathcal{K}\times\{0,1\}^k \to \{0,1\}^r$ is hardcore for $\mathcal{F}[\mathcal{H}]$ on any single-input distribution X with min-entropy $s+r+2(\log 1/\varepsilon)$ for negligible ε (as compared to [31, 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 5.9 Let LTDF be a LTDF generator with residual leakage s, and let $\mathcal{H}\colon \mathcal{K} \times \{0,1\}^k \to \{0,1\}^r$ be a pairwise-independent 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 $s + r + 2(\log 1/\varepsilon)$ for negligible ε.

The corresponding instantiation of the Encrypt-with-Hardcore scheme requires an LTDF with residual leakage $s \leq \mathrm{H}_\infty(X) - 2\log(1/\varepsilon) - r$. In particular, we need $r = k$ so the hardcore function can be expanded using a PRG, thus the LTDF should lose a constant fraction of its input. Previously, Boldyreva et al. [7] gave a construction of DE from lossy trapdoor functions that can be viewed as an optimization of the one we obtain here; namely it encrypts under (an augmented version of) the LTDF directly and does not use the "outer" encryption scheme at all. Its analysis requires the "Crooked" LHL of Dodis and Smith [14] rather than the standard one but gets rid of r in the above bound leading to a better requirement on lossiness or input entropy.

6 Bounded Multi-Message Security for Deterministic Encryption

In this section, we propose a new notion of "q-bounded " multi-message security for DE and provide constructions meeting it (for long enough messages) from any sufficiently lossy trapdoor function.

6.1 The New Notion and Variations

THE NEW NOTION. The notion of q-bounded multi-message security (or just q-bounded security) for DE is quite natural, and parallels the treatment of "bounded" security in other contexts (e.g. [9]). In a nutshell, it asks for security on up to q arbitrarily correlated but high-entropy messages (where we allow the public-key size to depend on q). More formally, fix an encryption scheme $\Pi = (\mathcal{K}, \mathcal{E}, \mathcal{D})$. For $q = q(k)$ and $\mu = \mu(k)$, let $\mathbb{M}^{q,\mu}$ be the class of distributions on message vectors $M^{\mu,q} = (M_1^{\mu,q}, \ldots, M_q^{\mu,q})$ where $\mathrm{H}_\infty(M_i^{\mu,q}) \geq \mu$ and for all $1 \leq i \leq q$ and $M_{1,q}^{\mu}, \ldots, M_{q,q}^{\mu}$ are distinct with probability 1. We say that Π is *q-bounded multi-message PRIV (resp. IND) secure for μ-sources* if it is PRIV (resp. IND) secure for $\mathbb{M}^{q,\mu}$. Note that our definitional treatment here is less general than in Section 3 in that we just define the strongest possible notion of q-bounded multi-message security, i.e., one that puts the least possible restrictions on the message-vector distributions subject to the fact that they have a most q components. We do this for simplicity since all our results in this section concern such a notion.

SEMANTIC SECURITY VERSUS INDISTINGUISHABILITY. We note that Theorem 3.1 (combined with Lemma 5.8) tells us that PRIV on $\mathbb{M}^{q,\mu}$ is equivalent to IND on $\mathbb{M}^{q,\mu-2}$.

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-th message introduces some "fresh" entropy. More formally, fix an encryption scheme $\Pi = (\mathcal{K}, \mathcal{E}, \mathcal{D})$. For $q = q(k)$, $n = n(k)$, and $\mu = \mu(k)$, let $\mathbb{M}^{q,n,\mu}$ be the class of distributions on message vectors $M^{q,n,\mu} = (M_1^{q,n,\mu}, \ldots, M_{qn}^{q,n,\mu})$ such that $\mathrm{H}_\infty(X_{qi+j} \mid X_1 = x_1, \ldots, X_{qi-1} = x_{qi-1}) \geq \mu$ for all $1 \leq i \leq n$, all $0 \leq j \leq q-1$, and all outcomes $x_1, \ldots, x_{qi-1}$ of $X_1, \ldots, X_{qi-1}$. We say that Π is *q-bounded multi-message PRIV (resp. IND) secure for (μ, n)-block-sources* if Π is PRIV (resp. IND) secure on $\mathbb{M}^{q,n,\mu}$. Using a similar argument to [7, Theorem 4.2], one can show equivalence of PRIV on $\mathbb{M}^{q,n,\mu}$ to IND on $\mathbb{M}^{q,n,\mu}$ (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). In any case, we omit formalizing this since one can prove our constructions secure for q-block sources directly.

6.2 Extensions to the Crooked Leftover Hash Lemma

Note that we cannot trivially achieve q-bounded 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 generalization of the classical Leftover Hash Lemma (LHL) to t-many correlated sources, following [28] who considered the case $t = 2$.

It turns out that for our "optimized" scheme we need to use a similar generalization of the "Crooked" LHL due to Dodis and Smith [14] that is even further generalized to the case that the hash function is a t-wise 'δ-dependent"

function. Namely, say that $H\colon \mathcal{K} \times D \to R$ is *t-wise δ-dependent* if for all distinct $x_1, \ldots, x_t \in D$

$$\Delta((H(K,x_1),\ldots,H(K,x_t)),(U_1,\ldots,U_t)) \;\leq\; \delta\;,$$

where $K \xleftarrow{\$} \mathcal{K}$ and $U_1, \ldots, U_t$ are independent and uniform over R.

Since our generalization of the classical LHL is a special case of our generalization of the Crooked LHL, we just state the latter here.

Lemma 6.1 (CLHL for Correlated Sources) Let $\mathcal{H}\colon \mathcal{K} \times D \to R$ be a $2t$-wise δ-dependent 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

$$\Delta((K, f(\mathcal{H}(K,\mathbf{X}))),(K, f(\mathbf{U}))) \;\leq\; \frac{1}{2}\sqrt{|S|^t(t^2 2^{-\mu} + 3\delta)} \tag{4}$$

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).

Note that the lemma implies the corresponding generalization of the classical LHL by taking $\mathcal{H}$ to have range S and f to be the identity function.

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

Remark 6.2 We can further extend Lemma 6.1 to the case of *average conditional min-entropy* using the techniques of [13]. Such generalization will be needed for our basic scheme. Such a generalization (without considering correlated sources) is similarly useful in the context of randomized encryption from lossy TDFs [31].

6.3 The Basic Scheme

By Theorem 4.2, in order to instantiate the Encrypt-with-Hardcore scheme to achieve q-bounded security, it suffices for the trapdoor function $\mathcal{F}$ to have a hardcore function hc that is robust for $\mathcal{F}$ on $\mathbb{M}^{q,\mu}$. We call such a hardcore function *q-bounded correlated-input secure*. The following shows that in the case of sufficiently lossy TDFs, this notion is achievable for any polynomial q.

Proposition 6.3 For any q, let $\mathsf{LTDF} = (\mathcal{F}, \mathcal{F}')$ be an LTDF generator with input length n and residual leakage s, and let $\mathcal{H}\colon \mathcal{K} \times D \to R$ where $r = \log|R|$ be a $2q$-wise indpendent hash function. Then $\mathcal{H}$ is a q-bounded correlated-input secure hardcore function for $\mathcal{F}$ on any input distribution $X = (X_1, \ldots, X_q)$ such that $\mathrm{H}_\infty(X) \geq q(s+r) + 2\log q + 2\log(1/\varepsilon) - 2$ for negligible ε.

Proof: The first step in the proof is to switch the HCF experiment to execute not $(f, f^{-1}) \xleftarrow{\$} \mathcal{F}(1^k)$ but $f' \leftarrow \mathcal{F}'(1^k)$. Now, by the "chain rule" for average conditional min-entropy from [13], we have $\tilde{\mathrm{H}}_\infty(X \mid f'(X_1)\ldots, f'(X_q)) \geq \mathrm{H}_\infty(X) - qs$. We can conclude by applying Lemma 6.1 coupled with Remark 6.2 with $t = q$. ∎

Note that in the corresponding instantiation of the Encrypt-with-Hardcore scheme, we take $r = k$ in order to expand the latter using a PRG, and thus need $(\mathrm{H}_\infty(X) - 2\log q - \log(1/\varepsilon))/q - k \geq s$. In particular, $\mathcal{F}$ must lose a $\Omega(1 - 1/q)$ fraction of its input; concretely, the r term roughly puts a factor 2 on q here (or we need to increase n by a factor 2). The DDH-based construction of Peikert and Waters [31], the Paillier-based one of [7, 18], and the one from d-linear of [18] can all satisfy this requirement for any polynomial q.

6.4 The Optimized Scheme

We show that by extending some ideas of [7], we obtain a more efficient DE scheme meeting q-bounded security that achieves better parameters.

INTUITION AND PRELIMINARIES. Intuitively, for the optimized scheme we modifying the scheme of [7] to first pre-process an input message using a $2q$-wise independent permutation (instead of pairwise as in [7]). However, there are two issues to deal with here. First, for $q > 1$ such a permutation is not known to exist (in an explicit and efficiently computable sense). Second, Lemma 6.1 applies to t-wise independent *functions* rather than permutations. (In the case $t = 2$ as considered in [7] the difference turns out to be immaterial.)

To solve the first problem, we turn to $2q$-wise "δ-dependent" permutations (as constructed in e.g. [27]). Namely, say that a permutation $H\colon \mathcal{K} \times D \to D$ is *t-wise δ-dependent* if for all distinct $x_1, \ldots, x_t \in D$

$$\Delta((H(K, x_1), \ldots, H(K, x_t)), (P_1, \ldots, P_t)) \;\le\; \delta \;,$$

where $K \xleftarrow{\$} \mathcal{K}$ and $P_1, \ldots, P_t$ are defined iteratively by taking P_1 to be uniform on D and, for all $2 \le i \le t$, taking P_i to be uniform on $R \setminus \{p_1, \ldots, p_{i-1}\}$ where $p_1, \ldots, p_{i-1}$ are the outcomes of $P_1, \ldots, P_{i-1}$ respectively.

To solve the second problem, we use the following lemma, which says that a t-wise δ-dependent permutation is a t-wise δ'-dependent function where δ' is a bit bigger than δ.

Lemma 6.4 Suppose $H\colon \mathcal{K} \times D \to D$ is a t-wise δ-dependent permutation for some $t \ge 1$. Then $\mathcal{H}$ is a t-wise δ-dependent *function* for $\delta' = \delta + t^2/|D|$.

Thr proof uses the fact that the distribution of $(P_1, \ldots, P_t)$ equals the distribution of $(U_1, \ldots, U_t) \mid \mathsf{DIST}$ where DIST is the event that $U_1, \ldots, U_t$ are all distinct and then applies a union bound. It will be useful to now restate Lemma 6.1 in terms of δ-dependent permutations, which follows by combining Lemma 6.1 and Lemma 6.4.

Lemma 6.5 (CLHL for Correlated Sources with Permutations) Let $\mathcal{H}\colon \mathcal{K} \times D \to D$ be a δ-dependent t-wise permutation for $t > 0$ with range R, where $\delta = t^2/|D|$. 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 $\mathrm{H}_\infty(X_i) \ge \mu$ for all $1 \le i \le n$ and moreover $\Pr[\, X_i = X_j \,] = 0$ for all $1 \le i \ne j \le t$. Then

$$\Delta((K, f(\mathcal{H}(K, \mathbf{X}))), (K, f(\mathbf{U}))) \;\le\; \frac{1}{2}\sqrt{7|S|^t t^2 2^{-\mu}} \tag{5}$$

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).

It is interesting to note here that the the bound in Equation 5 is essentially as good as the one in Equation 4. At first one might not expect this to be the case. Indeed, when the classical LHL is extended to "imperfect" hash functions [37, 12], the error probability must be taken much smaller than $1/|R|$, where R is the range of the hash function. But in Lemma 6.1 we have $\delta = t^2/|D|$, which is large compared to $1/|D|$ (where D the range of the hash function in our case as it is a permutation). The reason we can tolerate this is that it is enough for $t^2/|D|$ to be much smaller than $1/|R|$ where R is the output range of f in the lemma, which is indeed the case in applications. In other words, the Crooked LHL turns out to be more tolerant than the classical one in this respect.

THE CONSTRUCTION. We now detail our construction. Let $\mathsf{LTDF} = (\mathcal{F}, \mathcal{F}')$ be an LTDF and let $\mathcal{P}\colon \mathcal{K} \times \{0,1\}^k \to \{0,1\}^k$ be an efficiently invertible 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 $\mathcal{K}(1^k)$:	**Algorithm** $\mathcal{DE}((f, K), x)$:	**Algorithm** $\mathcal{DD}((sk, f^{-1}), c)$:
$(f, f^{-1}) \xleftarrow{\$} \mathcal{F}(1^k)\,;\; K \xleftarrow{\$} \mathcal{K}$	$c \leftarrow f(\mathcal{P}(K, x))$	$x \leftarrow f^{-1}(\mathcal{P}^{-1}(K, c))$
Return $((f, K), (f^{-1}, K))$	Return c	Return x

We have the following result.

Theorem 6.6 Suppose LTDF is a lossy trapdoor funtion on $\{0,1\}^n$ with residual leakage s, and let $q, \varepsilon > 0$. Suppose $\mathcal{P}$ is a $2q$-wise δ-dependent permutation on $\{0,1\}^n$ for $\delta = t^2/2^n$. Then for any q-message IND adversary D with min-entropy $\mu \geq qs + 2\log q + \log(1/\varepsilon) + 5$, there is a LTDF distinguisher D such that for all $k \in \mathbb{N}$

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

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

Proof: The first step in the proof is to switch the HCF experiment to execute not $(f, f^{-1}) \xleftarrow{\$} \mathcal{F}(1^k)$ but $f' \leftarrow \mathcal{F}'(1^k)$. We can conclude by applying Lemma 6.5 with $t = 2q$ and $\mathcal{H} = \mathcal{P}$. ▌

An efficiently invertible $2q$-wise δ-dependent permutation on $\{0,1\}^n$ for $\delta = t^2/2^n$ can be obtained from [27] using key length $nt + \log(1/\delta) = n(t+1) - 2t$.

Now, combining Theorem 6.6 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 6.7 Suppose LTDF is a lossy trapdoor function on $\{0,1\}^n$ with residual leakage s. Then we obtain a q-bounded multi-message PRIV secure DE scheme for the class of distributions on $\{0,1\}^n$ with min-entropy $\mu \geq qs + 2\log q + 2\log(1/\varepsilon) + 7$ for negligible ε.

Comparing to Proposition 6.3, we see that we have dropped the r in the entropy bound (indeed, there is no hardcore function here). This translates to savings on the input entropy or lossiness requirement on the trapdoor function. Namely, while we still need to lose a $\Omega(1 - 1/q)$ fraction of the input, we get rid of the factor 2 on q. We also note that we can prove that the optimized scheme meets our notion of 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 LHL [39]. 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 security better illuminates the technical gap between achieving unbounded 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 [23].

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 (who pointed us to [27]), and Ramarathnam Venkatesan for helpful discussions. We are also thankful to an anonymous reviewer for pointing out that we were using the wrong definition of "δ-dependent" permutations in a prior version of the paper.

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] Ronald Cramer, Goichiro Hanaoka, Dennis Hofheinz, Hideki Imai, Eike Kiltz, Rafael Pass, Abhi Shelat, and Vinod Vaikuntanathan. Bounded cca2-secure encryption. In *ASIACRYPT*, pages 502–518, 2007.

[10] 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.

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

[12] 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.

[13] 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.

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

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

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

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

[18] 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.

[19] Yael Gertner, Sampath Kannan, Tal Malkin, Omer Reingold, and Mahesh Viswanathan. The relationship between public key encryption and oblivious transfer. In *FOCS*, pages 325–335, 2000.

[20] Yael Gertner, Tal Malkin, and Omer Reingold. On the impossibility of basing trapdoor functions on trapdoor predicates. In *FOCS*, pages 126–135, 2001.

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

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

[23] Vipul Goyal, Adam O'Neill, and Vanishree Rao. Correlated-input secure hash functions. In *TCC*, 2011.

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

[25] 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.

[26] 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/`.

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

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

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

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

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

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

[33] Phillip Rogaway. Nonce-based symmetric encryption. In *FSE*, pages 348–359, 2004.

[34] Phillip Rogaway and Thomas Shrimpton. A provable-security treatment of the key-wrap problem. In *EUROCRYPT*, pages 373–390, 2006.

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

[36] 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.

[37] Aravind Srinivasan and David Zuckerman. Computing with very weak random sources. In *FOCS*, pages 264–275, 1994.

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

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

A Comparison with the Approach of Prior Work

Here we discuss how to prove PRIV security of the Encrypt-with-Hardcore scheme when working with the original definitional equivalence in [4, 7] in terms of *entropy levels*. In particular, we argue it then becomes awkward to prove security of instantiations based on one-wayness. Testifying to this fact [4] did not use their definitional equivalence to IND at all in proving security of their scheme. For simplicity, we just deal here with single-message security.

First, let us rephrase some definitions from Section 4.1 in terms of entropy levels. Let $\mathcal{F}$ be a trapdoor function and hc a hardcore function. Denote by $\mathbb{X}_\ell$ the class of all input distributions X where $\mathrm{H}_\infty(X) \geq \ell$, and suppose hc is hardcore for $\mathcal{F}$ on $\mathbb{X}_\ell$. Then we say hc is *α-entropically robust for* $\mathbb{X}_\ell$ if hc is also hardcore for $\mathcal{F}$ on $\mathbb{X}_{\ell-\alpha}$. Using the definitional equivalences proven in [4, 7] and an analogous argument to the proof of Theorem 4.2, one can show that if hc is 2-entropically robust for $\mathbb{X}_\ell$ then $\mathsf{EwHCore}$ is PRIV-secure on $\mathbb{X}_\ell$.

Now let us consider proving PRIV security of the instantiations based on one-wayness in Section 5.1. For this, we need to show that if a trapdoor function is one-way on $\mathbb{X}_\ell$ then it remains one-way on $\mathbb{X}_{\ell-2}$. By Lemma 5.1, this reduces showing that given any input distribution $X' \in \mathbb{X}_{\ell-\alpha}$ there is an input distribution $X \in \mathbb{X}_\ell$ that is distributed like X' with probability about $2^{-\alpha}$. A way to do this is to "mix" (i.e., take a convex combination of) X' with the uniform distribution. (We stress that [4] did not take such an approach, but instead used the PRIV definition directly in the security proof of their scheme; the following represents a way they could have applied their definitional equivalence.) We then get the following:

Proposition A.1 Let $X' \in \mathbb{X}_{\ell-\alpha}$ be a random variable on $\{0,1\}^k$, for any $1 \leq \ell < k$ and $0 \leq \alpha \leq \ell - 1$. Then there is a random variable $X \in \mathbb{X}_\ell$ and an event $\mathcal{E}$ such that $\Pr[\,\mathcal{E}\,] = 2^{-\alpha-1}$ and $P_{X'} = P_{X \mid \mathcal{E}}$.

Proof: Define X to be such that $P_X = 2^{-\alpha-1} \cdot P_{X'} + (1 - 2^{-\alpha-1}) \cdot P_U$ where U is uniform on $\{0,1\}^k$. Then there is an event $\mathcal{E}$ as required: we can sample from X by first flipping a coin that is heads with probability $2^{-\alpha-1}$ and then sample from X' if so and otherwise U, so $\mathcal{E}$ is the event that the coin is heads. Furthermore, for any $x \in \{0,1\}^k$

$$\begin{aligned} P_X(x) &= 2^{-\alpha-1} \cdot P_{X'}(x) + (1 - 2^{-\alpha-1}) \cdot P_U(x) \\ &\leq 2^{-\alpha-1} \cdot 2^{-\ell+\alpha} + 2^{-k} \\ &= 2^{-\ell-1} + 2^{-k} \\ &\leq 2^{-\ell}\,, \end{aligned}$$

where on the last line we use that $\ell < k$. So $\mathrm{H}_\infty(X) \geq \ell$ as required. ▌

This gives us what we want for $\ell < k$. However, the proposition does *not* easily extend to the case that $\ell = k$. Intuitively, the problem is that when mixing X with the uniform distribution we always end up with something that has min-entropy appreciably less than k. The case $\ell = k$ is needed to show security of DE schemes for uniform plaintexts based on the standard assumptions of one-way trapdoor functions or permutations secure for uniform inputs; this case has been recognized as important in [4, 26]. The case $\ell = k$ can ultimately be handled by appealing to a completely different technical lemma [17, Lemma 4]. On the other hand, our precise definitional equivalence allows us to avoid using any of these auxilliary claims (which have nothing to do with one-wayness) in all cases entirely and apply Lemma 5.1 directly. Indeed, to show PRIV security on $\mathbb{X}_\ell$ we only need to show on-wayness on *induced* distributions of slightly less min-entropy (rather than *all* such distributions), which is very easy and is what Lemma 5.1 gives us by definition.

B Proof of Theorem 3.1

We focus on the IND to PRIV implication; the other direction is straightforward.[8] 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]).

Proposition B.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 [11] for the information-theoretic setting. Adversary B works as follows:

Algorithm $B_1(1^k)$:	**Algorithm** $B_2(r)$:	**Algorithm** $B_3(pk, \mathbf{c}, r)$:
$r \xleftarrow{\$} \{0,1\}^n$	$(\mathbf{x}, t) \xleftarrow{\$} A_1(1^k)$	$g \xleftarrow{\$} A_3(pk, \mathbf{c})$
Return r	Return $(\mathbf{x}, \langle t, r \rangle)$	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

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

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.

[8] The idea for the latter is to have the constructed PRIV adversary sample 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; note that here we use that these induced distributions are *complementary*.

Proposition B.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

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

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 [10] (which also leads to better concrete security), where for simplicity we assume $1/\delta$ is an integer. Adversary B' works as follows:

Algorithm $B_1(1^k)$:	**Algorithm** $B_2(pk, \mathbf{c})$:
$(\mathbf{x}, t) \xleftarrow{\$} A_1(1^k)$	$g \xleftarrow{\$} A_2(pk, \mathbf{c})$
$i \xleftarrow{\$} [2(1/\delta) + 1]$	$j \xleftarrow{\$} [2(1/\delta) + 1]$
If $i \leq 1/\delta$ then return $(\mathbf{x}, 0)$	If $j \leq \delta$ then return 0
Else if $i \leq 2(1/\delta)$ then return $(\mathbf{x}, 1)$	Else if $j \leq 2(1/\delta)$ then return 1
Else return $(\mathbf{x}, t)$	Else return g

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

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

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

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

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 B.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

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

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 $D_1(b)$:	**Algorithm** $D_2(pk, \mathbf{c})$:
For $i = 1$ to k do:	$g \xleftarrow{\$} B_2(pk, \mathbf{c})$
$\quad (\mathbf{x}, t) \xleftarrow{\$} B_1(1^k)$	Return g
$\quad$ If $t = b$ then return $\mathbf{x}$	
Return $\mathbf{x}$	

For the analysis, 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

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

where the second-to-last line uses that B is δ-balanced. The claimed running-time of D is easy to verify. It remains to argue that $D \in \mathbb{D}_{\boldsymbol{M}^{*}[\log(1/(1/2-\delta))]}$. Let $\boldsymbol{M}_{D,i}$ be the message distribution sampled by D_1 on input $b = i$ for $i \in \{0,1\}$ and similarly let $\boldsymbol{M}_{B,i}$ be the message distribution sampled by B_1 when $t = i$ in its output for $i \in \{0,1\}$. Observe that $\boldsymbol{M}_{B,0}$ and $\boldsymbol{M}_{B,1}$ are complementary $\log(1/(1/2-\delta))$-induced distributions of the message distribution of B, with corresponding events $t = 0$ and $t = 1$ respectively. Furthermore, we have $\boldsymbol{M}_{D,i} \mid \overline{\mathsf{BAD}} = \boldsymbol{M}_{B,i}$ for $i \in \{0,1\}$. Since $\Pr[\,\mathsf{BAD}\,] \leq (1/2 + \delta)^{-k}$, it follows that $\boldsymbol{M}_{D,i} \mid \overline{\mathsf{BAD}}$ is statistically $2^{-\Omega(k)}$-close to $\boldsymbol{M}_{B,i}$ for $i \in \{0,1\}$, which concludes the proof.[9] ▌

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

C Proof of Lemma 4.3

We first note that it is not hard to see that if a robust hardcore function is expanded via a pseudorandom generator it remains robust. Thus, in the following we ignore the pseudorandom generator and "absorb" it into the hardcore function.

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 $B^i_1(1^k)$:	**Algorithm** $B^i_2(pk, \mathbf{y}, \mathbf{h})$:
$\mathbf{x} \xleftarrow{\$} D_1(i)$	$\mathbf{c} \leftarrow \mathcal{E}(pk, \mathbf{y}; \mathbf{h})$
Return $\mathbf{x}$	$d \xleftarrow{\$} D_2(pk, \mathbf{c})$
	Return d

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

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 $A_1(pk)$:	**Algorithm** $A_2(pk, \mathbf{c})$:
$\mathbf{x}_0 \xleftarrow{\$} D_1(0)$	$d \xleftarrow{\$} D_2(pk, \mathbf{c})$
$\mathbf{x}_1 \xleftarrow{\$} D_1(1)$	Return d
Return $(\mathbf{x}_0, \mathbf{x}_1)$	

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

D Proofs of Simple Lemmas

Here we provide the proofs of Lemma 5.1 and Lemma 5.8. One should note the commonalities in the proofs, showing the connection between the analysis of our corresponding DE instantiations from one-wayness and lossiness.

Proof: (of Lemma 5.1) 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

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

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

Proof: (of Lemma 5.8) Suppose not, and let E be the corresponding event to X'. Then there exists an x' such that $P_{X'}(x') > 2^{-\mu+\alpha}$. But then

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

a contradiction. ▌

E 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

$$\mathbf{E}_k\big[D\big(f(\mathcal{H}(k,\mathbf{X})), f(\mathbf{U})\big)\big] \leq\ t^2 2^{-\mu} + 6t^2 2^{-r} + 3\delta\ ,$$

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

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

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

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

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. ▌