Honey Encryption
Beyond Message Recovery Security

Joseph Jaeger[1], Thomas Ristenpart[2], and Qiang Tang[3]

[1] University of California, San Diego
jsjaeger@eng.ucsd.edu
[2] Cornell Tech
ristenpart@cornell.edu
[3] Cornell University
qt44@cornell.edu

Abstract. Juels and Ristenpart introduced honey encryption (HE) and showed how to achieve message recovery security even in the face of attacks that can exhaustively try all likely keys. This is important in contexts like password-based encryption where keys are very low entropy, and HE schemes based on the JR construction were subsequently proposed for use in password management systems and even long-term protection of genetic data. But message recovery security is in this setting, like previous ones, a relatively weak property, and in particular does not prohibit an attacker from learning partial information about plaintexts or from usefully mauling ciphertexts.

We show that one can build HE schemes that can hide partial information about plaintexts and that prevent mauling even in the face of exhaustive brute force attacks. To do so, we introduce target-distribution semantic-security and target-distribution non-malleability security notions. We prove that a slight variant of the JR HE construction can meet them. The proofs require new balls-and-bins type analyses significantly different from those used in prior work. Finally, we provide a formal proof of the folklore result that an unbounded adversary which obtains a limited number of encryptions of known plaintexts can always succeed at message recovery.

1 Introduction

Password-based encryption (PBE) suffers from the threat of brute-force attacks. People pick poor, easy-to-predict passwords and so an attacker, given a ciphertext, can try decrypting it with the most likely password, the next most likely, and so on. It is easy to determine when the right password is found, and so as long as the password falls in this list the attacker wins, recovering the password and the full plaintext. Unfortunately, studies indicate that the most common password is typically selected by almost 1% of users [10], meaning that passwords have less than $\mu = 7$ bits of min-entropy. The straightforward attack succeeds with probability a bit more than $q/2^{\mu}$ where q is the number of decryption attempts. Bellare, Ristenpart, and Tessaro [7] proved a closely matching

upper bound, perhaps suggesting that the case was closed and that, for PBE, one cannot do better.

Honey encryption. Juels and Ristenpart (JR) [23], however, showed how one might provably achieve security for relatively low-entropy keys—even when attackers can try decrypting a ciphertext with all possible keys. Intuitively, their approach makes attacks unsuccessful by ensuring that all plaintexts generated during a brute-force attack look plausible. This approach was used previously for the special case of uniformly random plaintexts by Kausik and Hoover [26]. JR proposed a more general cryptographic primitive that they called *honey encryption* (HE). An HE scheme is tailored to an estimate of the (possibly non-uniform) distribution of messages for which it will be employed. We refer to this distribution as the target distribution. Decrypting an HE ciphertext with an incorrect key yields a decoy (or honey) message that appears, to the attacker, to be a fresh sample from the target distribution. An attacker that knows no further information about the true message will be unable to pick it out from the decoys.

JR gave a framework for building HE schemes that composes a distribution-transforming encoder (DTE) with an encryption scheme. A DTE is a kind of randomized encoding scheme tailored to the target distribution. They propose that HE schemes should achieve security in two distinct settings, what we will call the high-entropy key setting and the low-entropy key setting. The former is the conventional setting in which security rests on the adversary being unable to do work proportional to 2^{μ}. Here they show that DTE-then-Encrypt can use standard mechanisms to provably achieve the conventional goals of [7].

The novelty lies in the low-entropy setting, where we assume that keys have some entropy μ but that adversaries can nevertheless do work much greater than 2^{μ}. For simplicity here one most often just assumes unbounded attackers. In this context, JR formalized a message recovery security goal. They then proved that in some useful cases DTE-then-Encrypt constructions can achieve close to optimal message recovery security: for a (relatively high-entropy) message encrypted under a key whose maximum probability of taking on any particular value is at most $1/2^{\mu}$, then an unbounded adversary's ability to guess the correct message, even given the ciphertext, is at most $1/2^{\mu}$ plus a negligible amount. Given that an attacker can always output the decryption of the challenge ciphertext under the most likely key, the JR result is essentially tight.

The DTE-then-Encrypt construction provides a recipe for building HE for particular applications, as one need only build a custom DTE for the setting by way of some estimate of the message distribution. Chatterjee et al. [11] showed how to do so for messages that are themselves lists of human-chosen passwords and built a prototype password vault system based on HE. Huang et al. [21] showed how to construct DTEs for messages that describe a person's genetic information. The application was for building a secure, long-term genetic information store. In both contexts they rely on JR's goal of MR security.

But MR security has several deficiencies from the viewpoint of modern security goals for conventional symmetric encryption (SE), and even for the ap-

plications for which researchers have explored use of HE. For SE one strives for authenticated encryption security [6], or its robust variants [19, 29]. These notions allow chosen message and ciphertext attacks. Informally speaking, they demand that not even a single bit of information about plaintexts can be learned by an adversary and that ciphertexts cannot be forged. We are therefore left with a significant gap between the JR results and what we might like in terms of security. In the genetic store application, for example, it could be that using an only MR-secure HE scheme would leak most of your genome. All this begs the question of whether there exist stronger security goals for HE and constructions that meet them.

Our contributions. In this work, we provide a systematic study of stronger notions of security for HE schemes in the low-entropy key setting. The bad news first: we formally rule out the ability to strengthen the JR security notions to allow known-message attacks when attackers can exhaust the key space. While this result seems intuitively obvious, and was taken for granted in [23], showing it formally for arbitrary HE schemes required a surprising amount of care. Having done so, we return to unknown message attack settings, but here provide good news in the way of stronger security goals and proofs that simple constructions meet them. First, we give a semantic security-style notion suitable for unknown message attacks and, second, a notion of target-distribution non-malleability. We show how the JR construction meets the first, and a new construction that achieves both. In the remainder of the introduction we provide more overview of these results.

Impossibility of known-message attack security. The JR security message recovery (MR) definition works as follows. A challenge message is drawn from the target distribution, encrypted under a key, and the resulting ciphertext is given to the adversary. It wins if it can output the challenge message. While the adversary knows the target distribution and the distribution from which keys are drawn, it does not get access to any known message, ciphertext pairs under the key. We extend the notion to additionally give the adversary an oracle from which it can obtain message-ciphertext pairs more messages drawn from the target distribution, yielding a known-message attack variant. We denote this notion by MR-KMA.

Intuitively MR-KMAshould be unachievable when the adversary can exhaustively search the key space. The adversary simply queries the oracle on several different messages, runs a brute-force attack to find the key that is consistent with all the message-ciphertext pairs, and uses that to decrypt the challenge ciphertext. While this attack might seem to work against all schemes, in fact there exist many for which there will be a large set of consistent keys. In the most extreme case, all keys will be consistent after any number of queries when encryption is the identity function for each key. One approach to deal with this is to make assumptions about the underlying scheme that allow one to show that after sufficiently many queries the consistent set will shrink to one. For example,

if the encryption scheme has "sufficiently random" mappings for distinct keys. But we would like to make no assumptions about the HE scheme.

Our attacker instead simply embraces that there may be a large set of consistent keys, and just uses one of them at random to decrypt the challenge ciphertext. We then have to lower bound the probability that a random key from the consistent set decrypts the challenge ciphertext to the target plaintext. In fact we do not know how to (or whether one can) prove this for an adversary that makes a fixed number of queries. Rather we show that there exists some number of queries between zero and κ, where 2^κ is the size of the key space, for which an adversary will achieve advantage at least $1/2\kappa$.

In the end, our result rules out security against known-message attacks. We also note that the proof techniques here already apply to (non-stateful) symmetric encryption as they do not take advantage of any properties specific to HE. We are, in fact, unaware of any previous general lower bound on message recovery for exhaustive key search attacks against conventional symmetric encryption schemes. Finally, the proof technique can generalize as well to message authentication goals, such as unforgeability under chosen-message attack.

Protecting partial information. We now return to unknown message attacks, but seek to strengthen the security goals along two dimensions. First, we consider partial information leakage. MR security is potentially adequate in settings for which the encrypted message is, say, an authentication credential which must be supplied in full elsewhere (the original motivating settings in [23]). It is likely to prove insufficient more generally. Schemes meeting MR might trivially leak a significant amount of information about messages. The seminal work of Goldwasser and Micali [17] argued (in the context of public-key encryption) that one should instead prefer encryption to hide all partial information about plaintexts. This stronger goal, called semantic security, was subsequently adapted to (at least) the settings of symmetric encryption [3], deterministic symmetric encryption [15, 30], and deterministic and hedged public-key encryption [1, 2, 4].

Unfortunately the traditional symmetric encryption semantic security notion (denoted SS below) [3], along with its variants of indistinguishability under chosen plaintext attack [3], are unachievable when keys are low entropy. (This is a corollary of our negative results about MR-CPA.)

We therefore introduce a new semantic-security style notion suitable for the low-entropy key setting. We call it target-distribution semantic security (TDSS). In it, an adversary is given the encryption of a message drawn from a target distribution and must predict a boolean function applied to the plaintext. It needs to do this better than is possible when predicting the predicate without the ciphertext. The key difference from SS is that it is asked to hold only for a specific message distribution, the target, and not for all message distributions. Interestingly we could find no meaningful indistinguishability-style variant of TDSS (unlike in the conventional setting, where we have the notion of IND-CPA and, moreover, an equivalence between it and SS [3]).

We relate the MR and TDSS notions, in particular using a result from Dodis and Smith [15] (see also [4]) that straightforwardly adapts to our setting. We use

it as an intermediate step to show that predicting predicates implies predicting functions for TDSS. Since MR security is equivalent to predicting the identity function, we obtain that TDSS implies MR security. There exists a simple separation showing that MR does not imply TDSS.

We go on to analyze the DTE-then-Encrypt scheme due to JR, showing via a new balls-and-bins analysis an upper bound of about $2\omega_k^{\frac{7}{16}} + 2e^{-\frac{1}{3\omega_k{}^{1/8}}}$ on the advantage of unbounded TDSS attackers. Like the MR proof by JR, ours is in the random oracle model [8]. Because TDSS focuses on predicates, the new balls-and-bins analysis necessarily focuses on the trickier setting of having many more balls (representing keys here) than the two bins (the possible predicate outputs). Our proof crucially relies, as did JR's, on a majorization lemma due to Berenbrink et al. [9] to transition the balls-and-bins analysis from non-uniform keys to uniform ones. In comparison to MR security our new bound is quantitatively weaker: JR showed MR advantage upper bounded by ω_k (when message distribution entropy is sufficiently large). Here we instead lose about half the entropy of the key. Nevertheless our result may be close to optimal (see Remark 1).

Non-malleability. The JR message recovery security goal, as well as the TDSS goal above, do not rule out active attackers manipulating ciphertexts. Indeed, DTE-then-Encrypt instantiations used in [11, 21, 23] are trivially malleable as they encrypt the DTE output by XOR'ing it with a pad derived from a hash of the key. An attacker can flip particular bits of the ciphertext and know that the resulting ciphertext will be decrypted to a plaintext related in a predictable way to the original. This is true regardless of the unpredictability of either the key or message.

Complicating matters, achieving MR or TDSS security seems to rule out preventing manipulation by including in an HE scheme typical mechanisms such as authentication tags or redundancy. Intuitively, this is because they would seem to always help the attacker rule out incorrect keys. We therefore turn to weaker notions like non-malleability [16,25], which again are unachievable in the low-entropy key setting (by our negative results above) but may be adaptable to unknown message settings because their goals do not seem to inherently conflict with confidentiality goals like MR and TDSS.

We introduce a target-distribution non-malleability (TDNM) notion for HE schemes when used with low-entropy keys. Informally, an attacker should not be able maul a ciphertext C to produce a new ciphertext $\tilde{C}$ in a way that some fixed relation R over the associated plaintexts is met with probability higher than one can achieve without access to C. All this holds for C being the encryption of a message taken from the target distribution.

We propose a simple construction that we call DTE-then-Encipher. It composes a DTE with a block cipher with sufficiently large domain. Modeling the cipher as ideal allows us to prove both TDSS and TDNM security. The TDNM proof shares some similarity to the TDSS proof of DTE-then-Encrypt, but requires additional techniques. In particular, the balls-in-bins analysis here cannot

use the majorization lemma of [9], and so we perform a new majorization-style analysis that exploits Schur convexity [22].

Further related work. Entropic security was considered in [15,30] as a statistical analogue of semantic security, and like HE they can also resist unbounded attackers. They show security against can be achieved when $\mu_k + \mu_m \geq n$, where μ_k, μ_m are the min-entropy of the key and message distribution, respectively, and n is the message length in bits. They show one can do no better in their setting, which requires security to hold over all distributions with the indicated min-entropy. HE low-entropy key security instead relaxes this to focus on specific target distributions, thereby skirting their lower bounds on required entropy, and providing meaningful security even when $\mu_k + \mu_m < n$.

2 Notation and Definitions

Notation. If n is an integer we let $\mathbb{Z}_n$ be the set $\{0, \ldots, n-1\}$. We use $y \leftarrow\!\!\$\ A(x)$ to denote running randomized algorithm A on input x and setting y equal to its output. If instead A is deterministic we write $y \leftarrow A(x)$. If G is a game we let $\Pr[\mathrm{G} \Rightarrow \mathsf{true}]$ denote the probability that G outputs true.

Let $\mathcal{S}$ be a set. A distribution on $\mathcal{S}$ is a function $p : \mathcal{S} \to [0,1]$ such that $\sum_{s \in \mathcal{S}} p(s) = 1$. The maximum probability ω of a distribution p is defined to be $\omega = \max_{s \in \mathcal{S}} p(s)$. The min-entropy μ of p is defined to be $\mu = -\log \omega$. When referencing min-entropy and maximum probability the distribution will always be clear from context. By $s \leftarrow_p \mathcal{S}$ we denote sampling an element $s \in \mathcal{S}$ according to the distribution p. That is, each $s \in \mathcal{S}$ is chosen with probability $p(s)$. For $B \subseteq \mathcal{S}$ we overload notation and let $p(B) = \sum_{s \in B} p(s)$.

Hash functions. A hash function H is a function $\mathsf{H} : \{0,1\}^* \to \{0,1\}^n$ which maps strings of arbitrary length to strings of some fixed length n. The length n will always be clear from context. In this work, we model hash functions as random oracles.

Symmetric encryption. A symmetric encryption scheme $\mathsf{SE} = (\mathtt{Enc}, \mathtt{Dec})$ is a pair of algorithms defined relative to a key space $\mathcal{K}$ and message space $\mathcal{M}$. The randomized encryption algorithm $\mathtt{Enc}$ takes as input a key $K \in \mathcal{K}$ and a message $M \in \mathcal{M}$ and outputs a ciphertext $C \in \mathcal{C}$. The deterministic decryption algorithm $\mathtt{Dec}$ takes as input a key $K \in \mathcal{K}$ and a ciphertext $C \in \mathcal{C}$ and outputs a message $M \in \mathcal{M}$. We require that a symmetric encryption scheme must be correct, meaning that for all $K \in \mathcal{K}$ and all $M \in \mathcal{M}$, $\Pr[\mathtt{Dec}(K, \mathtt{Enc}(K, M)) = M] = 1$.

Majorization. We say $\bar{p}$ majorizes $\bar{q}$ (denoted as $\bar{p} \succ \bar{q}$), if the two vectors $\bar{p} = \langle p_1, p_2, \ldots, p_n \rangle$, and $\bar{q} = \langle q_1, q_2, \ldots, q_n \rangle$ (written in descending order such that for all $i \in [1, n-1]$ that $p_i \geq p_{i+1}$, and $q_i \geq q_{i+1}$) satisfy $\sum_{i=1}^{k} p_i \geq \sum_{i=1}^{k} q_i$ for all $k \in \{1, \ldots, n\}$. When $\bar{p}$ denotes the probabilities of a distribution with support size n, it is easy to see that $\bar{q} \succ \bar{p}$, if $\bar{q}$ is defined as $q_i = p_1$ for $1 \leq i \leq \lceil 1/p_1 \rceil$ and $q_i = 0$ for $\lceil 1/p_1 \rceil + 1 \leq i \leq n$.

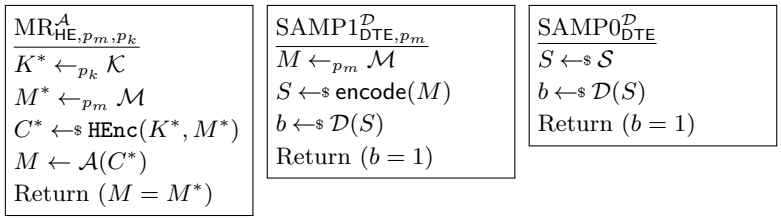

Fig. 1. Left: Game defining message recovery security. **Middle and Right:** Games defining security of a DTE.

3 Background on Honey Encryption

Honey encryption schemes. An HE scheme $\mathsf{HE} = (\mathtt{HEnc}, \mathtt{HDec})$ is a symmetric encryption scheme for some key space $\mathcal{K}$ and message space $\mathcal{M}$. Typically $\mathcal{K}$ will be strings representating human-chosen passwords, but HE can be applied in other settings as well. HE schemes should meet conventional security goals for password-based symmetric encryption [7, 24]. Differentiating HE schemes from conventional ones, however, is that they are designed relative to a specific (estimated) distribution over $\mathcal{M}$. This allows schemes that achieve a level of security even when the keys are relatively predictable, or have low min-entropy, from an attacker's perspective. Again, human-chosen passwords are the canonical example of such keys.

We let p_m represent the message distribution on the message space $\mathcal{M}$ and μ_m, ω_m denote its min-entropy and maximum probability respectively. Similarly we let p_k represent the key distribution on the key space $\mathcal{K}$ and let μ_k, ω_k denote its min-entropy and maximum probability respectively. In the low-entropy settings we focus on, we assume that ω_k is large enough that an attacker can easily perform work proportional to 2^{μ_k}. For simplicity in fact we will in our treatment simply assume adversaries can run in unbounded time. Our results extend to this setting, but also can be translated to computationally bounded settings in a straightforward manner.

MR security. Juels and Ristenpart [23] formalized and built schemes to achieve message recovery (MR) security. Their MR security game is defined in Figure 1 for a scheme HE and distributions p_m, p_k. An MR adversary $\mathcal{A}$ takes as input a ciphertext encrypting a challenge message chosen according to p_m and outputs a message $M \in \mathcal{M}$. The adversary wins if it outputs the challenge message. More precisely, we measure the advantage of a (computationally unbounded) MR adversary $\mathcal{A}$ against scheme HE and distributions p_m and p_k by

$$\mathbf{Adv}^{\mathrm{mr}}_{\mathsf{HE},p_m,p_k}(\mathcal{A}) = \Pr\left[\,\mathrm{MR}^{\mathcal{A}}_{\mathsf{HE},p_m,p_k} \Rightarrow \mathsf{true}\,\right] .$$

Distribution-transforming encoders. A distribution-tranforming encoder (let us use DTE for short) is a pair of algorithms $\mathsf{DTE} = (\mathsf{encode}, \mathsf{decode})$ defined relative to a message space $\mathcal{M}$ and a set $\mathcal{S}$ called the seed space. Via $S \leftarrow\!\!\$\ \mathsf{encode}(M)$ the randomized encoding algorithm encode taking a message

$\underline{\mathtt{HEnc}(K, M)}$	$\underline{\mathtt{HDec}(K, C)}$
$S \leftarrow\!\!\$\ \mathsf{encode}(M)$	$(R, C_2) \leftarrow C$
$R \leftarrow\!\!\$\ \{0,1\}^{rl}$	$S \leftarrow \mathsf{H}(R\|\|K) \oplus C_2$
$C_2 \leftarrow \mathsf{H}(R\|\|K) \oplus S$	$M \leftarrow \mathsf{decode}(S)$
Return (R, C_2)	Return M

Fig. 2. The DTE-then-Encrypt construction $\mathsf{HE}[\mathsf{DTE}, \mathsf{H}]$, using hash function H and DTE $\mathsf{DTE} = (\mathsf{encode}, \mathsf{decode})$.

$M \in \mathcal{M}$ as input and outputs a seed $S \in \mathcal{S}$. A DTE must satisfy correctness, that for any message $M \in \mathcal{M}$, $\Pr[\mathsf{decode}(\mathsf{encode}(M)) = M] = 1$. Like HE schemes, a DTE is designed for a specific message distribution p_m.

Following [23], the security property desired for a DTE is that it is hard for an adversary to distinguish between $S \in \mathcal{S}$ chosen uniformly at random and chosen by first picking a message according to p_m and then applying encode. This property is formalized by two of the games shown in Figure 1. We measure the advantage of an adversary $\mathcal{D}$ against DTE and distribution p_m by

$$\mathbf{Adv}^{\mathrm{dte}}_{\mathsf{DTE},p_m}(\mathcal{D}) = \Pr[\mathrm{SAMP1}^{\mathcal{D}}_{\mathsf{DTE},p_m}] - \Pr[\mathrm{SAMP0}^{\mathcal{D}}_{\mathsf{DTE}}] \ ,$$

and the DTE-goodness is defined by $\mathbf{Adv}^{\mathrm{dte}}_{\mathsf{DTE},p_m} = \max_{\mathcal{D}}\ \mathbf{Adv}^{\mathrm{dte}}_{\mathsf{DTE},p_m}(\mathcal{D})$ and where the the maximization is over all, even computationally unbounded, adversaries $\mathcal{D}$. When the DTE in question is clear we let p_d represent the distribution induced on $\mathcal{M}$ by sampling a random seed from $\mathcal{S}$ and applying decode. Formally,

$$p_d(M) = \Pr[M' = M : S \leftarrow\!\!\$\ \mathcal{S}; M' \leftarrow \mathsf{decode}(S)] \ .$$

DTE-then-Encrypt. JR introduced a framework of constructing HE schemes for a target distribution p_m from a symmetric encryption scheme SE and a distribution-transforming encoder DTE. More specifically, the DTE-then-Encrypt framework encrypts a message by applying the DTE encoding first and then encrypting the encoding using SE. Security requires some easy-to-meet properties of SE, such as that it does not pad out inputs.

In more detail, let H be a hash function and r an integer representing the number of random bits to be used by encryption. Then the scheme which we will denote by $\mathsf{HE}[\mathsf{DTE}, \mathsf{H}]$ is shown in Figure 2.

Note that as written, this scheme does not achieve the password-based encryption security goals of [7] for the high-entropy key setting. It is easy to modify the scheme to do so: simply replace the hash function with an appropriate password-based key derivation function (PBKDF). One can also deal with using fixed-output-length hash functions with large seed spaces by appropriate use of a mode of operation. See [23] for more detailed discussion.

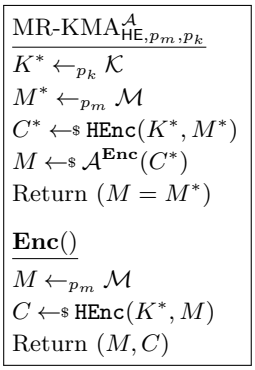

Fig. 3. Game defining message recovery security under a known message attack.

4 Impossibility of KMA Security with Low-entropy Keys

Recall that the MR security notion is a relatively weak goal in various ways. One such weakness is that it is only an unknown-message attack and provides adversaries with no plaintext-ciphertext examples. In this section we show that one cannot hope to achieve security in the low-entropy key setting when given a relatively small number of plaintext-ciphertext examples in a known-message attack. Making this claim formal required a surprising amount of care.

MR-KMA security definition. Let game MR-KMA be defined as in Figure 3 for scheme HE and distributions p_m, p_k. This game is exactly the same as the MR security game except the adversary additionally has access to an encryption oracle which samples a message M according to p_m and returns an encryption of M under the secret key. We measure the advantage of a (computationally unbounded) adversary $\mathcal{A}$ against HE with distributions p_m and p_k by

$$\mathbf{Adv}^{\text{mr-kma}}_{\mathsf{HE},p_m,p_k}(\mathcal{A}) = \Pr[\text{MR-KMA}^{\mathcal{A}}_{\mathsf{HE},p_m,p_k} \Rightarrow \mathsf{true}].$$

The "obvious" attack strategy. A straightforward strategy for an MR-KMA adversary is to use its encryption oracle to receive q distinct message-ciphertext pairs. Then, use test decryption under all keys to find those keys that correctly decrypt all the ciphertexts correctly. We refer to such a key as being consistent. The intuition is that even for small q the set of consistent keys will be a singleton and that, necessarily, it is the key chosen by the experiment. This intuition stems from the fact that for a "reasonable" scheme, the probability that the wrong key decrypts all the ciphertexts correctly is low.

Of course formally this logic is meaningless as it makes unspecified assumptions on the scheme. Indeed there are many examples of schemes for which the set of consistent keys will be large, no matter how large q gets. In the most egregious case, where $\mathtt{HEnc}$ and $\mathtt{HDec}$ implement the identity function for all keys, then the set of consistent keys will always be $\mathcal{K}$. Clearly this scheme is not MR secure, but the point is that when giving a proof that holds for all schemes we

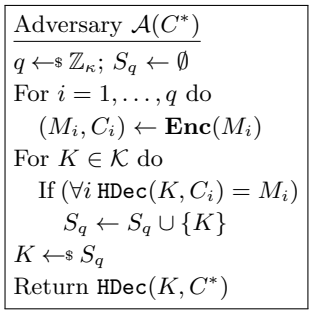

Fig. 4. Adversary for MR-KMA making at most $\kappa = \lceil \log |\mathcal{K}| \rceil$ encryption queries.

must handle such degenerate cases. This issue (and the particular degenerate example just given) is related to the well-known fact that key recovery security does not imply message recovery security for all schemes. Nevertheless, we are unaware of any proofs showing that no SE scheme can resist message recovery attacks that exhaustively search the key space.

A lower bound on MR-KMA security. Given the example that ruling out keys may not work very well, we give a slightly different adversary. Our adversary, shown in Figure 4, runs the attack as described above, but simply finishes by decrypting the challenge using a uniformly chosen key from the set of consistent keys. It is clear, for example, that in the trivial identity-function scheme mentioned above all keys will be consistent with the challenge and this attack achieves advantage one.

We must lower-bound the success probability for any scheme. Doing so requires showing that with high probability the uniformly selected consistent key must be consistent also with the challenge ciphertext. Due to technical difficulties relating to our proof, we cannot give an exact number of oracle queries for which this attack has a high advantage. Instead we show that for some number of queries which is at most $\kappa = \lceil \log |\mathcal{K}| \rceil$ this attack has a high success probability of at least $1/(2\kappa)$. For concreteness we then say that the advantage of an adversary who picks the number of queries at random from $0, \ldots, \kappa$ will have advantage at least $1/(2\kappa^2)$. These results give us the following theorem.

Theorem 1. *Let* HE *be an encryption scheme and* $\kappa = \lceil \log |\mathcal{K}| \rceil$*. Then for any* p_m, p_k *the adversary* $\mathcal{A}$ *shown in Figure 4 makes at most* $\kappa - 1$ *oracle queries and has advantage*

$$\mathbf{Adv}^{\text{mr-kma}}_{\mathsf{HE},p_m,p_k}(\mathcal{A}) \geq \frac{1}{2\kappa^2}. \tag{1}$$

The idea of the proof is to note that the advantage of the adversary $\mathcal{A}$ for a particular value of q is equal to the probability that a randomly chosen key that is consistent with q message-ciphertext pairs is also consistent with a $(q+1)$-th pair (the challenge message and ciphertext). Then letting S_q denote the set of keys consistent after q pairs. We have that the advantage of $\mathcal{A}$ for a particular

value of q is $\mathbb{E}[|S_{q+1}|/|S_q|]$, where the expectation is taken over the appropriate experiment (defined below). Intuitively, this ratio can only be really small for a small number of q's because each S_q must contain between 1 and 2^κ keys.

Before presenting the full proof we formalize the above intuition with the following lemma about random variables.

Lemma 1. *If $s_0, \ldots, s_\kappa$ are positive integer-valued random variables such that $s_0 \leq 2^\kappa$ and $s_{q+1} \leq s_q$ for $q \in \mathbb{Z}_\kappa$, then $\max_{q \in \mathbb{Z}_\kappa} \mathbb{E}[s_{q+1}/s_q] \geq \frac{1}{2\kappa}$.*

Proof. Let $\epsilon = \max_{q \in \mathbb{Z}_\kappa} \mathbb{E}[s_{q+1}/s_q]$. We will use an inductive argument to prove that $\Pr[s_q \geq 2^{\kappa-q}] \leq 2q\epsilon$ for $1 \leq q \leq \kappa$. Then considering when q is κ and noting that $s_\kappa \geq 1$ always we have $1 = \Pr[s_\kappa \geq 1] \leq 2\kappa\epsilon$. Solving for ϵ gives the desired bound.

We now give the inductive argument. First, Markov's inequality can be used to bound the probability that s_{q+1} is at least half s_q by $\Pr[s_{q+1}/s_q \geq 1/2] \leq 2\mathbb{E}[s_{q+1}/s_q]$. Rewriting and bounding $\mathbb{E}[s_{q+1}/s_q]$ by ϵ we get

$$\Pr[s_{q+1} \geq (1/2)s_q] \leq 2\epsilon \tag{2}$$

for all $q \in \mathbb{Z}_\kappa$.

Recalling that $s_0 \leq 2^\kappa$, the base case is easily derived by $\Pr[s_1 \geq 2^{\kappa-1}] \leq \Pr[s_1 \geq (1/2)s_0] \leq 2\epsilon$.

Now suppose $1 < q \leq \kappa$ and $\Pr[s_{q-1} \geq 2^{\kappa-(q-1)}] \leq 2(q-1)\epsilon$. By definition we have,

$$\begin{aligned}\Pr[s_q \geq 2^{\kappa-q}] = \Pr[s_q \geq 2^{\kappa-q} | s_{q-1} < 2^{\kappa-(q-1)}] \Pr[s_{q-1} < 2^{\kappa-(q-1)}] \\ + \Pr[s_q \geq 2^{\kappa-q} | s_{q-1} \geq 2^{\kappa-(q-1)}] \Pr[s_{q-1} \geq 2^{\kappa-(q-1)}];\end{aligned}$$

The first part of the equation can be bounded using our inductive assumption:

$$\begin{aligned}\Pr[s_q \geq 2^{\kappa-q} | s_{q-1} \geq 2^{\kappa-(q-1)}] \Pr[s_{q-1} \geq 2^{\kappa-(q-1)}] &\leq \Pr[s_{q-1} \geq 2^{\kappa-(q-1)}] \\ &\leq 2(q-1)\epsilon;\end{aligned}$$

To bound the second part note that conditioned on the fact that s_{q-1} is less than $2^{\kappa-(q-1)}$, it can only hold that s_q is greater than $2^{\kappa-q}$ if s_q is greater than $(1/2)s_{q-1}$. This gives us

$$\Pr[s_q \geq 2^{\kappa-q} | s_{q-1} < 2^{\kappa-(q-1)}] \leq \Pr[s_q \geq (1/2)s_{q-1} | s_{q-1} < 2^{\kappa-(q-1)}] \,.$$

Then from the definition of conditional probability and using (2) we get that

$$\begin{aligned}\Pr[s_q \geq (1/2)s_{q-1} | s_{q-1} < 2^{\kappa-(q-1)}] \Pr[s_{q-1} < 2^{\kappa-(q-1)}] &\leq \Pr[s_q \geq (1/2)s_{q-1}] \\ &\leq 2\epsilon;\end{aligned}$$

Putting the above equations together we get $\Pr[s_q \geq 2^{\kappa-q}] \leq 2q\epsilon$, completing the proof. □

We now use the above result to prove Theorem 1. The proof proceeds by showing that the advantage of adversary $\mathcal{A}$ for a particular q is $\mathbb{E}[|S_{q+1}|/|S_q|]$ where S_q is the set of consistent keys after q message-ciphertext pairs and then noting that the size of these sets fulfill the conditions of the lemma above.

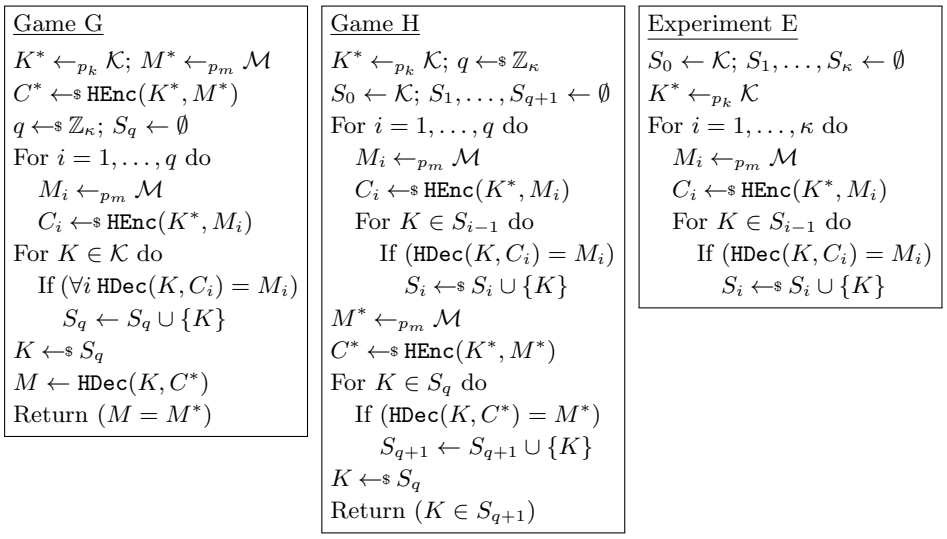

Fig. 5. Left and Middle: Games used in MR-KMA proof. **Right:** Experiment used in MR-KMA proof.

Proof (of Theorem 1). First note that $\mathbf{Adv}^{\text{mr-kma}}_{\mathsf{HE},p_m,p_k}(\mathcal{A}) = \Pr[\text{G} \Rightarrow \mathsf{true}]$, where game G is defined on the left side of Figure 5. This is clear because G is simply the game $\text{MR-KMA}^{\mathcal{A}}_{\mathsf{HE},p_m,p_k}$ with the code of $\mathcal{A}$ inserted.

Now consider game H shown in the middle of Figure 5. Game H is obtained from G via a few simple transforms. In it S_q is computed iteratively one (M, C) pair at a time, the choice of M^* and C^* is deferred until they are used, and instead of checking whether $M = M^*$ the game equivalently checks whether the randomly chosen K falls in the subset of S_q that decrypts C^* to M^* which is called S_{q+1}. It is thus clear that $\Pr[\text{H} \Rightarrow \mathsf{true}] = \Pr[\text{G} \Rightarrow \mathsf{true}]$

Noting that $S_{q+1} \subseteq S_q$ holds for every q, it is clear from the last two lines of H that once q, S_q, and S_{q+1} are chosen, the probability that H will output true is $\mathbb{E}[|S_{q+1}|/|S_q|]$. Thus we have that $\Pr[\text{H} \Rightarrow \mathsf{true}] = \sum_{q=0}^{\kappa}(1/\kappa)\mathbb{E}[|S_{q+1}|/|S_q|]$.

Next we transition our analysis to considering the experiment E shown in Figure 5. Note that the distribution of S_{q+1} and S_q for any $q \in \mathbb{Z}_\kappa$ in E is identical to the distribution in H. For $0 \leq q \leq \kappa$, let s_q be the random variable representing $|S_q|$ in E and ϵ be $\max_{q \in \mathbb{Z}_\kappa} \mathbb{E}[s_{q+1}/s_q]$ where the expectation is taken in experiment E.

Since all S_q always contains at least K^*, each s_q must be positive. Thus $s_0, \ldots, s_\kappa$ are positive integer-valued random variables which fulfill the conditions of Lemma 1 so we have $\epsilon \geq \frac{1}{2\kappa}$. Then the following sequence of inequalities exhibits (1):

$$\mathbf{Adv}^{\text{mr-kma}}_{\mathsf{HE},p_m,p_k}(\mathcal{A}) = \Pr[\text{H} \Rightarrow \mathsf{true}] = \sum_{q=0}^{\kappa} \frac{1}{\kappa}\mathbb{E}[s_{q+1}/s_q] \geq \frac{1}{\kappa} \cdot \epsilon \geq \frac{1}{2\kappa^2}. \qquad \square$$

Extensions. While we focused above on known-message attacks, our proof techniques carry over to the more typical setting of chosen-plaintext attacks. Here the adversary has access instead to an encryption oracle that takes as input an adversarially chosen message, encrypts it using the secret key, and returns the ciphertext.

Furthermore, the ideas behind our proof can be extended to cover unforgeability under chosen-message attacks for, e.g., message authentication codes [5, 18]. Here an adversary with access to a tagging oracle tries to come up with a valid message-tag pair for a message it has not queried yet. The adversary used to prove its impossibility would use a fixed sequence of messages $M_1 \ldots, M_\kappa$ and use a random key consistent with the first q messages to sign the next message (a fixed sequence of messages is used here to avoid the problem of the adversary trying to tag a message it has already been given the correct tag for). Then essentially the exact same analysis shows that this adversary will succeed with high probability. We omit the details for brevity.

5 Stronger Message Privacy for HE Schemes

Given the impossibility result of the last section, we turn to explore achievable but still meaningful security notions that capture the goal of hiding partial informations about the messages encrypted by an HE scheme. In this section, we propose a semantic security-style definition tailored to the low-entropy key setting. We call it targeted-distribution semantic security (TDSS). We will also investigate its relationship with MR security. We then go on to show that the DTE-then-Encrypt construction meets this stronger notion of security, though with concrete security bounds slightly worse than what could be proved in the MR case.

5.1 TDSS security and its relation to MR security

Recall that semantic security style notions ask, roughly, that an attacker given the encryption of a message cannot predict a predicate on it with probability better than is possible without the encryption. In the symmetric encryption setting, semantic security was first formalized by Bellare et al. [3] where they give the adversary a chosen-message encryption oracle. By our impossibility results in the last section, we cannot do so, and instead return to an unknown-message only attack setting for the target message distribution. We refer to this as target-distribution semantic security (TDSS).

Let $\mathcal{M}$ be a message space and p_m be an associated target distribution. Let HE be an HE scheme for $\mathcal{M}$. We let $f : \mathcal{M} \to \{0,1\}$ be a predicate on messages. Let $p_f(b) = \Pr[f(M) = b \mid M \leftarrow_{p_m} \mathcal{M}]$ and let $\omega_f = \max(p_f(0), p_f(1))$.

The TDSS security games are shown in Figure 6. In game $\mathrm{TDSS1}^{\mathcal{A},f}_{\mathsf{HE},p_m,p_k}$ an adversary $\mathcal{A}$ is charged with predicting $f(M)$ given an encryption of it. In game $\mathrm{TDSS0}^{\mathcal{A}_s,f}_{p_m}$ an adversary $\mathcal{A}_s$, called the simulator, which attempts to guess

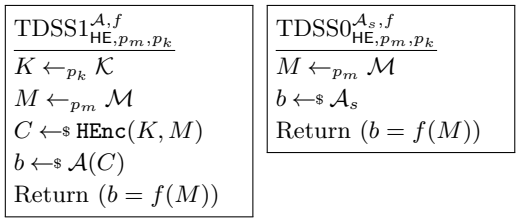

Fig. 6. Games defining TDSS security.

$f(M)$ without access to a ciphertext. The optimal simulator $\mathcal{A}_s$ for any p_m, f pair simply outputs most likely value of $f(M)$ given the message distribution and predicate f. This forces $\Pr[\text{TDSS0}^{\mathcal{A}_s,f}_{p_m} \Rightarrow \mathsf{true}] = \omega_f$. We therefore define the advantage of adversary $\mathcal{A}$ against the TDSS security of an HE scheme HE with respect to distributions p_m, p_k and predicate f by

$$\mathbf{Adv}^{\text{tdss}}_{\mathsf{HE},p_m,p_k}(\mathcal{A}, f) = \Pr\left[\text{TDSS1}^{\mathcal{A},f}_{\mathsf{HE},p_m,p_k} \Rightarrow \mathsf{true}\right] - \omega_f \ .$$

When working with a random oracle H, the game allows the adversary and the encryption algorithm to query H but f must be independent of H.
And the TDSS security of HE is measured by

$$\mathbf{Adv}^{\text{tdss}}_{\mathsf{HE},p_m,p_k} = \max_{\mathcal{A},f} \mathbf{Adv}^{\text{tdss}}_{\mathsf{HE},p_m,p_k}(\mathcal{A}, f) \ .$$

The maximization is over all, even unbounded adversaries $\mathcal{A}$ and arbitrary predicates f. It is easy to derive ways of measuring restricted versions of TDSS security, such as by placing computational limits on $\mathcal{A}$ or restricting the class of predicates. We will not consider such weaker notions further.

TDSS and MR. We will now consider the relation between MR security and TDSS security. It is not hard to see that MR security does not imply TDSS security. We can easily construction an HE scheme such that one bit of the message is revealed completely but the rest is secure using a good HE scheme, thus making the resulting scheme secure in the MR sense but not in the TDSS sense.

Intuitively, TDSS should imply MR, but proving this is not as easy as the other direction. Consider the trivial reduction in which a TDSS adversary $\mathcal{B}$ runs an MR adversary $\mathcal{A}$ and then computes the predicate on the message returned by $\mathcal{A}$. It's clear that $\Pr[\text{TDSS0}^{\mathcal{B},f}_{\mathsf{HE},p_m,p_k} \Rightarrow \mathsf{true}] \geq \mathbf{Adv}^{\text{mr}}_{\mathsf{HE},p_m,p_k}(\mathcal{A})$, but this might be *smaller* than ω_f even if $\mathcal{A}$ is a very good MR adversary.

Fortunately, Dodis and Smith [15] showed that in the information theoretic setting, a good predictor for a function can be turned into a good predictor for a boolean predicate. Viewing a MR adversary as a predictor for the identity function, we can use this to convert a good MR adversary into a good TDSS adversary. We defer the proof of the following theorem to the full version.

Theorem 2. *Let HE be a honey encryption scheme for message distribution p_m.*
(i) If $\mathbf{Adv}^{\text{mr}}_{\mathsf{HE},p_m,p_k} \geq \omega_m + \omega_m^{2/3}$, then $\mathbf{Adv}^{\text{mr}}_{\mathsf{HE},p_m,p_k} \leq \omega_m + 4 \cdot \mathbf{Adv}^{\text{tdss}}_{\mathsf{HE},p_m,p_k}$.

(ii) There exists message distribution p'_m, honey encryption scheme HE', predicate f, and TDSS adversary $\mathcal{A}$ such that for any p_k, HE' satisfies $\mathbf{Adv}^{\mathrm{mr}}_{\mathsf{HE}',p'_m,p_k} = \mathbf{Adv}^{\mathrm{mr}}_{\mathsf{HE},p_m,p_k}$ *and* $\mathbf{Adv}^{\mathrm{tdss}}_{\mathsf{HE}',p'_m,p_k}(\mathcal{A},f) = \frac{1}{2}$.

5.2 TDSS Security of DTE-then-Encrypt

We turn to showing that the DTE-then-Encrypt construction (refer back to Figure 2 in Section 3) achieves TDSS security in the random oracle model.

Our analysis proceeds in a modular fashion similar to the JR proof of MR security for this construction, but with important differences. First we use DTE security to transition to a game in which the ciphertext is chosen uniformly at random and the challenge key is not sampled until after the adversary has run (one might look ahead to Figure 8 for the games). In this game, we can show that the advantage of any adversary is no better than the advantage of an adversary $\mathcal{A}^*$ that decrypts the ciphertext using all possible keys, computes the predicate value on the resulting plaintext, and outputs the bit which has the higher cumulative mass of keys that resulted in this bit.

One can then view the game measuring this optimal adversary's success equivalently as a balls-and-bins experiment. The detailed experiment is shown in Figure 7. Here the balls represent keys and each ball has weight indicated by p_k. There are two bins B_0 and B_1, and throwing a ball into a bin corresponds to seeing the predicate value arrived at by decrypting the fixed ciphertext under the key associated to that ball. Ball throws are independent because H is modeled as a RO.

To our knowledge, in the case that the number of balls is much larger than the number of bins, existing analyses of balls-and-bins experiments only provide an asymptotic bound [28] and in the case that bins are chosen uniformly. We instead analyze the maximum load in the case of non-uniform bin selection and uniformly weighted balls (with the same weights). We can finally then apply a majorization lemma [9] to get a concrete upper bound in the general case of non-uniform balls. We break down the analysis into a series of lemmas, and give the final theorem at the end of this section.

The following lemma captures the first part of our analysis, reducing the security of $\mathsf{HE}[\mathsf{DTE},\mathsf{H}]$ to the security of DTE, the expected maximum load $\mathbb{E}[\mathrm{L}^{\mathsf{H},\mathsf{DTE},f}_{p_k}]$ in experiment $\mathrm{E}^{\mathsf{H},\mathsf{DTE},f}_{p_k}$, and the bias ω_f of the predicate f on p_m.

Lemma 2. *Let* HE *be* $\mathsf{HE}[\mathsf{DTE},\mathsf{H}]$ *as defined in Section 3 for distributions* p_m, p_k. *Let* f *be a predicate on* $\mathcal{M}$, $\mathcal{A}$ *be any adversary, then we have:*

$$\mathbf{Adv}^{\mathrm{tdss}}_{\mathsf{HE},p_m,p_k}(\mathcal{A},f) \leq \mathbf{Adv}^{\mathrm{dte}}_{\mathsf{DTE},p_m} + \mathbb{E}[\mathrm{L}^{\mathsf{H},\mathsf{DTE},f}_{p_k}] - \omega_f.$$

Proof. We will use the sequence of games shown in Figure 8 to transition to a game in which the optimal strategy is clearly $\mathcal{A}^*$ (shown in Figure 9) which outputs the bit most likely to be the output of f applied to the decryption of the challenge ciphertext under a randomly chosen key.

First note that $\Pr[\mathrm{G}_0 \Rightarrow \mathsf{true}] = \Pr[\mathrm{TDSS0}^{\mathcal{A},f,0}_{\mathsf{HE},p_m,p_k} \Rightarrow \mathsf{true}]$, which is clear because game G_0 is simply the TDSS0 game with the code of HE inserted. Thus,

$$\mathbf{Adv}^{\mathrm{tdss}}_{\mathsf{HE},p_m,p_k}(\mathcal{A},f) = \Pr\left[\,\mathrm{G}_0 \Rightarrow \mathsf{true}\,\right] - \omega_f\ .$$

Experiment $\mathrm{E}^{\mathsf{H},\mathsf{DTE},f}_{p_k}$

$R \xleftarrow{\$} \{0,1\}^r$; $C_2 \xleftarrow{\$} \mathcal{S}$

For $K \in \mathcal{K}$ do

  $S \leftarrow \mathsf{H}(R||K) \oplus C_2$

  $M \leftarrow \mathsf{decode}(S)$

  $b \leftarrow f(M)$

  $B_b \leftarrow B_b \cup \{K\}$

$\mathrm{L}^{\mathsf{H},\mathsf{DTE},f}_{p_k} \leftarrow \max_{b\in\{0,1\}} p_k(B_b)$

Fig. 7. Balls-into-bins experiment used to analyze the security of $\mathsf{HE}[\mathsf{DTE},\mathsf{H}]$.

We can then use the security of DTE to transition to game G_1 because G_1 is identical to G_0 except instead of a random message being sampled and then encoded, a random seed is sampled and then decoded. Consider the adversary $\mathcal{D}$ against the security of DTE shown on the left side of Figure 9. Adversary $\mathcal{D}$ uses its input S to simulate the view of $\mathcal{A}$, returning 1 if $\mathcal{A}$ selects the correct bit and 0 otherwise. It's easy to verify that $\Pr[\mathrm{G}_0 \Rightarrow \mathsf{true}] = \Pr[\mathrm{SAMP1}^{\mathcal{D}}_{\mathsf{DTE},p_m} \Rightarrow \mathsf{true}]$ and $\Pr[\mathrm{G}_1 \Rightarrow \mathsf{true}] = \Pr[\mathrm{SAMP0}^{\mathcal{D}}_{\mathsf{DTE}} \Rightarrow \mathsf{true}]$. This gives us:

$$\Pr[\,\mathrm{G}_0 \Rightarrow \mathsf{true}\,] \le \mathbf{Adv}^{\mathrm{dte}}_{\mathsf{DTE},p_m} + \Pr[\,\mathrm{G}_1 \Rightarrow \mathsf{true}\,] .$$

Next we will see that G_2 is equivalent to G_1. In game G_2 the ciphertext C is sampled uniformly at random while the sampling of K and computation of M are delayed until the adversary has already executed. Note that in G_1 because S was a uniformly chosen element of $\mathcal{S}$, C_2 was also a uniform element of $\mathcal{S}$ independent of the choice of K. Thus we can instead select C_2 at random and defer the choice of K (and thus S and M) until after $\mathcal{A}$ is executed. Consequently,

$$\Pr[\,\mathrm{G}_1 \Rightarrow \mathsf{true}] = \Pr[\mathrm{G}_2 \Rightarrow \mathsf{true}\,] .$$

Now we argue that $\mathcal{A}^*$ is the best possible adversary in game G_2. To see this note that in G_2 the choice of the challenge key K is independent of the input to $\mathcal{A}$, so the values L_0 and L_1 calculated by $\mathcal{A}^*$ are exactly the probabilities that 0 and 1 will be the correct output, respectively. Thus it's clear that the b^* output by $\mathcal{A}^*$ is the optimal output. Letting $\Pr[\mathrm{G}^*_2 \Rightarrow \mathsf{true}]$ denote the probability that $\mathcal{A}^*$ succeeds in G_2 we have $\Pr[\mathrm{G}_2 \Rightarrow \mathsf{true}] \le \Pr[\mathrm{G}^*_2 \Rightarrow \mathsf{true}]$.

Finally we note that the weight L^* of the maximally loaded bin in the balls-in-bins experiment $\mathrm{L}^{\mathsf{H},\mathsf{DTE},f}_{p_k}$ is identical to the probability that the output of $\mathcal{A}^*$ is correct for the chosen (R, C_2). So we have $\Pr[\mathrm{G}^*_2 \Rightarrow \mathsf{true}] = \mathbb{E}[\mathrm{L}^{\mathsf{H},\mathsf{DTE},f}_{p_k}]$.

Putting everything together gives the desired theorem. □

Before we move onto the next step, we first simplify notation. Recall that p_d is the distribution on $\mathcal{M}$ given by applying decode to uniform samples from $\mathcal{S}$. When H is a random oracle each of its outputs is a uniform and independent sample S from $\mathcal{S}$. Thus we can view each message M as being independently sampled according to p_d. Now we can see that experiment $\mathrm{E}^{\mathsf{H},\mathsf{DTE},f}_{p_k}$ is equiv-

Game G_0	Game G_1	Game G_2
$K \leftarrow_{p_k} \mathcal{K}$	$K \leftarrow_{p_k} \mathcal{K}$	$R \leftarrow\$ \{0,1\}^r$
$M \leftarrow_{p_m} \mathcal{M}$	$S \leftarrow\$ \mathcal{S}$	$C_2 \leftarrow\$ \mathcal{S}$
$S \leftarrow\$ \mathsf{encode}(M)$	$M \leftarrow \mathsf{decode}(S)$	$C \leftarrow (R, C_2)$
$R \leftarrow\$ \{0,1\}^r$	$R \leftarrow\$ \{0,1\}^r$	$b \leftarrow\$ \mathcal{A}(C)$
$C_2 \leftarrow \mathsf{H}(R\|\|K) \oplus S$	$C_2 \leftarrow \mathsf{H}(R\|\|K) \oplus S$	$K \leftarrow_{p_k} \mathcal{K}$
$C \leftarrow (R, C_2)$	$C \leftarrow (R, C_2)$	$S \leftarrow \mathsf{H}(R\|\|K) \oplus C_2$
$b \leftarrow\$ \mathcal{A}(C)$	$b \leftarrow\$ \mathcal{A}(C)$	$M \leftarrow \mathsf{decode}(S)$
Return $(b = f(M))$	Return $(b = f(M))$	Return $(b = f(M))$

Fig. 8. Games used in proof of Theorem 2.

Adversary $\mathcal{D}(S)$	Adversary $\mathcal{A}^*(C)$
$K \leftarrow_{p_k} \mathcal{K}$	$(R, C_2) \leftarrow C$
$M \leftarrow \mathsf{decode}(S)$	For $K \in \mathcal{K}$ do
$R \leftarrow\$ \{0,1\}^r$	$\quad S \leftarrow \mathsf{H}(R\|\|K) \oplus C_2$
$C_2 \leftarrow \mathsf{H}(R\|\|K) \oplus S$	$\quad M \leftarrow \mathsf{decode}(S)$
$C \leftarrow (R, C_2)$	$\quad L_{f(M)} \leftarrow L_{f(M)} + p_k(K)$
$b \leftarrow\$ \mathcal{A}(C)$	$b^* \leftarrow \mathrm{argmax}_{b \in \{0,1\}} L_b$
If $(b = f(M))$	Return b^*
$\quad$ Return 1	
Return 0	

Fig. 9. Adversaries used in proof of Theorem 2.

alent to a new experiment $\mathrm{E}_{p_k}^{p_d,f}$ in Figure 10, which is more intuitive. Thus $\mathbb{E}[\mathrm{L}_{p_k}^{\mathsf{H},\mathsf{DTE},f}] = \mathbb{E}[\mathrm{L}_{p_k}^{p_d,f}]$.

Experiment $\mathrm{E}_{p_k}^{p_d,f}$
For $K \in \mathcal{K}$ do
$\quad M \leftarrow_{p_d} \mathcal{M}$
$\quad b_K \leftarrow f(M)$
$\quad B_{b_K} \leftarrow B_{b_K} \cup \{K\}$
$\mathrm{L}_{p_k}^{p_d,f} \leftarrow \max_{b \in \{0,1\}} p_k(B_b)$

Fig. 10. Simplified balls-into-bin experiment.

Next, we will recall a majorization lemma so that we can transition to a balls and bins experiment with uniform ball weights. Let $K_1, \ldots, K_{|\mathcal{K}|}$ denote an ordering of $\mathcal{K}$ according to weight, that is, for all $i \in \{1, \ldots, |\mathcal{K}| - 1\}$ we have $p_k(K_i) \geq p_k(K_{i+1})$. Then we let p'_k be defined such that for $i \leq \lceil 1/\omega_k \rceil$ we have $p'_k(K_i) = \omega_k$ and $p'_k(K_i) = 0$ otherwise. (Note that p'_k may no longer define a distribution because it's elements may sum to more than one, but this is not

important for our analysis below.) Recalling the notion of majorization defined in Section 2, we see that p'_k majorizes p_k. The following is a special case of a lemma from [9].

Lemma 3 (BFHM08). *For all p_d, f, and weight vectors p'_k, p_k for which p'_k majorizes p_k it holds that $\mathbb{E}[\mathrm{L}_{p_k}^{p_d,f}] \leq \mathbb{E}[\mathrm{L}_{p'_k}^{p_d,f}]$.*

We can now concentrate on establishing an upper-bound on $\mathbb{E}[\mathrm{L}_{p'_k}^{p_d,f}]$, where p'_k consists of $a = \lceil 1/\omega_k \rceil$ weights all equal to ω_k. Note that we have here ignored the keys of weight zero, but this is clearly without loss of generality since they have no influence on bin loads. The following lemma is gives a bound on the expected maximum load.

Lemma 4. *Let f be a predicate, p_d be a distribution, and p_t be the distribution over $\{0,1\}$ defined by sampling from p_d and applying f. Let p'_k be a weight vector with $a = \lceil 1/\omega_k \rceil$ values each equal to $\omega_k \leq 1$. Then for all s satisfying $a^{s-1} \leq \omega_t$*

$$\mathbb{E}[\mathrm{L}_{p'_k}^{p_d,f}] \leq (1+\omega_k)(\omega_t + a^{s-1} + 2e^{\frac{-a^{2s-1}}{3}}).$$

Proof. As per the lemma statement, we have that p_t is defined by $p_t(b) = \Pr[f(M) = b : M \leftarrow_{p_d} \mathcal{M}]$ and that ω_t is the associated probability of the most probable value in p_t. That is, let $b^* = \mathrm{argmax}_{b\in\{0,1\}} p_t(b)$ and then $\omega_t = p_t(b^*)$. For simplicity we will assume without loss of generality that $b^* = 1$.

Referring to experiment $\mathrm{E}_{p'_k}^{p_d,f}$ (Figure 7 with p_k replaced by p'_k), b_K is a random variable which equals 1 if K_i is thrown into B_1 and 0 otherwise. Noting then that $|B_1| = \sum_{K\in\mathcal{K}} b_K$ it is easy to see that $\mathbb{E}[|B_1|] = a\omega_t$. Let $B = (1/\omega_k) \cdot \mathrm{L}_{p'_k}^{p_d,f}$ be the random variable corresponding to the number of balls that fall into the maximally loaded bin at the end of the experiment. Then we can see that for all n, $\Pr[B \geq n] \leq 2 \cdot \Pr[|B_1| \geq n]$ because if $B \geq n$ then either B_0 or B_1 must have size at least n and from our assumption that $b^* = 1$, $\Pr[|B_0| \geq n]$ is clearly less than $\Pr[|B_1| \geq n]$.

To complete the proofs we carefully chose a value of n, so that we can bound the probability that B is greater than n and obtain the desired result by pessimistically assuming that B is a whenever it is greater than n and n otherwise.

Recall that Chernoff's bound tells us that for any $0 \leq \delta \leq 1$, $\Pr[|B_1| \geq (1+\delta)\mathbb{E}[|B_{b^*}|]] \leq e^{\frac{-\delta^2\mathbb{E}[|B_1|]}{3}}$. Then setting $\delta = a^{s-1}/\omega_t$ (which is less than 1 from our choice s) and let we get $\Pr[|B_1| \geq a\omega_t + a^s] \leq e^{\frac{-a^{2s-1}}{3}}$. Then we get the following sequence of inequalities.

$$\begin{aligned}
\mathbb{E}[B] &= \sum_{i=1}^{a} i \cdot \Pr[B = i] \\
&\leq (a\omega_t + a^s)\Pr[B < a\omega_t + a^s] + a \cdot \Pr[B \geq a\omega_t + a^s] \\
&\leq (a\omega_t + a^s) + 2a \cdot \Pr[|B_1| \geq a\omega_t + a^s] \\
&\leq a(\omega_t + a^{s-1} + 2e^{\frac{-a^{2s-1}}{3}})
\end{aligned}$$

Multiplying both sides of the inequality by ω and noting that $\lceil 1/\omega_k \rceil \cdot \omega_k \leq 1+\omega_k$ gives the bound on $\mathbb{E}[\mathrm{L}^{p_d,f}_{p_k}]$. □

At this point we can combine Lemmas 2, 3, 4 in turn to derive the following sequence of inequalities:

$$\begin{aligned}
\mathbf{Adv}^{\mathrm{tdss}}_{\mathsf{HE},p_m,p_k}(\mathcal{A},p_m) &\leq \mathbf{Adv}^{\mathrm{dte}}_{\mathsf{DTE},p_m} + \mathbb{E}[\mathrm{L}^{\mathsf{H},\mathsf{DTE},f}_{p_k}] - \omega_f \\
&\leq \mathbf{Adv}^{\mathrm{dte}}_{\mathsf{DTE},p_m} + \mathbb{E}[\mathrm{L}^{p_d,f}_{p_k}] - \omega_f \\
&\leq \mathbf{Adv}^{\mathrm{dte}}_{\mathsf{DTE},p_m} + \mathbb{E}[\mathrm{L}^{p_d,f}_{p'_k}] - \omega_f \\
&\leq \mathbf{Adv}^{\mathrm{dte}}_{\mathsf{DTE},p_m} + (1+\omega_k)(\omega_t + a^{s-1} + 2e^{\frac{-a^{2s-1}}{3}}) - \omega_f
\end{aligned}$$

along with the restriction that $a^{s-1} \leq \omega_t$ needed for the last transition. To proceed we note that we can apply again the security of our DTE to transition ω_t to ω_f. Consider the adversary $\mathcal{D}_f$ who just decodes its input S and outputs f applied to the resulting message. It is easy to verify that $\Pr[\mathrm{SAMP1}^{\mathcal{D}_f}_{\mathsf{DTE},p_m}] = p_f(1)$ and $\Pr[\mathrm{SAMP0}^{\mathcal{D}_f}_{\mathsf{DTE}}] = p_t(1)$ which gives us:

$$|\omega_f - \omega_t| \leq \mathbf{Adv}^{\mathrm{dte}}_{\mathsf{DTE},p_m}(\mathcal{D}_f) \leq \mathbf{Adv}^{\mathrm{dte}}_{\mathsf{DTE},p_m}.$$

One can apply this to the last inequality above and rearrange, as well as to the restriction on a^{s-1}. Together all this, combined with maximizing over $\mathcal{A}, f$ yields a proof of the following theorem.

Theorem 3. *Let* HE *be* $\mathsf{HE}[\mathsf{DTE},\mathsf{H}]$ *as defined in Section 3 for distributions* p_m, p_k *and with* H *modeled as a random oracle.Then for any* $\mathcal{A}$*, and any* s *satisfying* $a^{s-1} \leq 1/2 - \mathbf{Adv}^{\mathrm{dte}}_{\mathsf{DTE},p_m}$*,*

$$\mathbf{Adv}^{\mathrm{tdss}}_{\mathsf{HE},p_m,p_k} \leq \omega_k + \lceil 1/\omega_k \rceil^{s-1} + 2\mathbf{Adv}^{\mathrm{dte}}_{\mathsf{DTE},p_m} + 2\exp\left(\frac{-\lceil 1/\omega_k \rceil^{2s-1}}{3}\right).$$

Remark 1. (1) This bound does not seem too far from optimal. From existing results about the estimation of the number of balls in the uniform bin case (see full version for details.) we can see that when $\omega_f = \frac{1}{2}$, choosing $\alpha = \frac{3}{4}$, $\mathbb{E}[L^{p_d,f}_{p'_k}] \geq \Pr[X > k_\alpha] \cdot k_\alpha \omega_k \geq (1-o(1))(\frac{1}{2} + \sqrt{\frac{\omega_k}{3}})$, thus the advantage of the TDSS adversary is at least at the order of $\omega_k^{\frac{1}{2}}$. (2). As long as p_k has more than 3 bits entropy, we can easily find s such that $\lceil 1/\omega_k \rceil^{s-1} \leq \frac{1}{2}$. (3). If we choose $s = \frac{9}{16}$, $\mathbf{Adv}^{\mathrm{tdss}}_{\mathsf{HE},p_m,p_k} \leq 2\omega_k^{\frac{7}{16}} + 2e^{-\frac{1}{3\omega_k{}^{1/8}}} + 2\epsilon$. When $\omega_k = 2^{-30}$, the bound is close to 2^{-13}, for which we lose about half of the entropy in the key. If we don't mind losing more entropy (choosing larger s), we can tolerate even smaller ω_k. (3). The condition that $a^{s-1} \leq 1/2 - \mathbf{Adv}^{\mathrm{dte}}_{\mathsf{DTE},p_m}$, simply comes from the condition that $a^{s-1} \leq \omega_f$ for all f.

6 Non-malleability for HE Schemes

The TDSS security goal provides stronger confidentiality properties than MR. But it still does not speak to the threat of attackers mauling ciphertexts. In particular, the DTE-then-Encrypt construction as instantiated with a hash function is trivially malleable: an attacker can, without any knowledge of the key or plaintext, flip bits in a ciphertext so that when it is later decrypted the resulting plaintext is different from the original in a predictable way. Unfortunately the negative results in Section 4 suggest that we cannot meet traditional non-malleability security goals [16,25], let alone ciphertext integrity notions [6], when attackers can exhaustively search the key space.

Analogously to the last section, we therefore provide a notion of target-distribution non-malleability (TDNM) for HE schemes for use in the low-entropy key setting. TDNM, like TDSS, is an unknown-message attack setting, and intuitively demands that even if the key space is searchable, the ability of an attacker to successfully maul a ciphertext is not improved by having access to the ciphertext. We then give a construction called DTE-then-Encipher and show it enjoys both TDSS and TDNM security in the ideal cipher model.

6.1 TDNM Security

We adjust the standard non-malleability notion for symmemtric encryption [25] to consider only messages from a target distribution. Informally, TDNM security requires that given a ciphertext, it is difficult to come up with a new ciphertext so that the underling messages satisfy some relation. This is formalized by the two games shown in Figure 11.

Both games are defined with respect to a binary relation $R\colon \mathcal{M} \times \mathcal{M} \to \{0,1\}$. To simplify notation we sometimes say $R(M, M')$ is true if $R(M, M') = 1$ and false otherwise. We let $p_R = \Pr[R(M, M') = 1 \mid M, M' \leftarrow_{p_m} \mathcal{M}]$. The first game has an adversary $\mathcal{A}$ attempt to maul a ciphertext C so as to satisfy R. The second game has another adversary $\mathcal{A}_s$, called the simulator, attempt to do so without access to C.

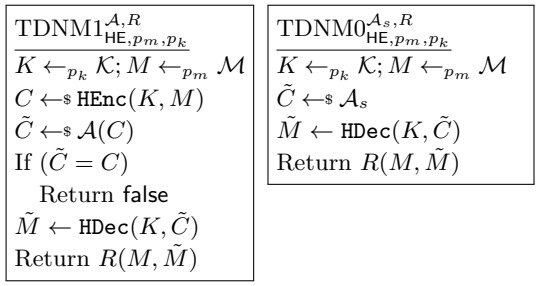

Fig. 11. Games defining TDNM security.

The TDNM advantage of an adversary $\mathcal{A}$ with respect to a binary relation R, HE scheme HE, and distributions p_m, p_k is defined by

$$\mathbf{Adv}^{\text{tdnm}}_{\mathsf{HE},p_m,p_k}(\mathcal{A},R) = \Pr\left[\text{TDNM1}^{\mathcal{A},R}_{\mathsf{HE},p_m,p_k} \Rightarrow \mathsf{true}\right] - \max_{\mathcal{A}_s} \Pr\left[\text{TDNM0}^{\mathcal{A}_s,R}_{\mathsf{HE},p_m,p_k} \Rightarrow \mathsf{true}\right]$$

We can then define the TDNM advantage of an HE with distributions p_m, p_k by

$$\mathbf{Adv}^{\text{tdnm}}_{\mathsf{HE},p_m,p_k} = \max_{\mathcal{A},R} \mathbf{Adv}^{\text{tdnm}}_{\mathsf{HE},p_m,p_k}(\mathcal{A},R).$$

Remark 2. There are two points we would like to note.

1. TDNM with n-ary relations: For simplicity we choose the simpler binary form of the TDNM definition. Of course, we may generalize it to an n-ary relations, but one must be careful about concrete security with respect to n. Imagine that n equals the size of the key space. Then TDNM can be broken for the relation that returns true if at least one of $M_1, \ldots, M_n$ is the challenge message. An adversary that generates each C_i by decrypting the challenge ciphertext using a different key and re-encrypting the message with the same key, sill succeed with probability 1.
2. Relationship between TDNM and TDSS (or MR): It is easy obvious that MR and TDSS do not imply TDNM, the encode-then-encrypt construction serves as an example. However, the other directions are less clear. Intuitively, an MR adversary can be used as an imperfect decryption oracle, this property may be explored for the TDNM adversary to compute a new ciphertext encrypting the same message. We conjecture that TDNM implies MR (and TDSS), at least under certain conditions. On the other hand, proving unconditional implication results would require new observations. The straightforward transformation of an MR adversary to find the secret key would likely incur a large reduction loss which we can not afford. In all those notions, the bound is already quite small.
   Note that this implication is in contrast with the classical notions when adversary has no access to the encryption oracle. A trivial scheme than outputs the plaintext directly together with a MAC is unforgeable, but has no message security.
   It would be an interesting open problem to give a complete characterization of the security notions of the honey encryption schemes.

6.2 The DTE-then-Encipher Construction

Intuitively, to achieve non-malleability we would like a scheme for which modifying any portion of a ciphertext would yield a ciphertext that will be decrypted to an independent message. Revisiting the DTE-then-Encrypt construction, a natural route to achieving this property is to replace the (malleable) encryption with one that is non-malleable. A good block cipher has the property that changing any bit of a ciphertext will randomize the decrypted plaintext. In our low-key

setting standard security properties like being a pseudorandom permutation are insufficient, and we will instead turn to the ideal cipher model. Here we model a deterministic, length-preserving encryption scheme $(\mathtt{Enc}, \mathtt{Dec})$ as a family of $|\mathcal{K}|$ random permutations, one for each key. The resulting DTE-then-Encipher construction is shown in Figure 12. We denote it by HE-NM.

$\mathtt{HEnc}(K, M)$:	$\mathtt{HDec}(K, C)$
$S \leftarrow \mathtt{encode}(M)$	$S \leftarrow \mathtt{Dec}(K, C)$
$C \leftarrow \mathtt{Enc}(K, S)$	$M \leftarrow \mathtt{decode}(S)$
return C	return M

Fig. 12. The TDNM construction HE-NM

To instantiate $\mathtt{Enc}, \mathtt{Dec}$ one could use a standard block cipher such as AES, but this will only work when the seed space of the DTE used is exactly the set of bit strings of length equal to the block size of the cipher, e.g., 128. One can turn to constructions that are proven indifferntiable [27] from an ideal cipher of the appropriate domain size. Coron et al. [12] show how to build form an ideal cipher $E\colon \{0,1\}^{2k} \times \{0,1\}^n$ an ideal cipher with domain $2n$ and key length k. One could repeatedly use this construction to extend the domain sufficiently. In theory one could also build large-domain ciphers using a hash function (modeled as a RO) within the Feistel-like constructions analyzed in [13, 14, 20], but the bounds are too loose to be of practical use. We leave as an open question finding more efficient constructions of TDNM constructions, and focus in the remainder on analysis assuming a suitable ideal cipher.

The TDSS security for $\mathsf{HE}[\mathsf{DTE}, \mathsf{H}]$ can be adjusted to apply to HE-NM in a straightforward manner. We focus below on establishing TDNM security.

Proof intuition. Intuitively, in an DTE-then-Encipher construction, any two different ciphertext would be decrypted to a pair of (nearly) uniform encoded strings, and they will thus decoded to two randomly sampled messages. However, to formally demonstrate the analysis for TDNM, it still requires us to show that the maximum probability that an (unbounded) adversary can generate a ciphertext that is correlated with a given ciphertext is not much better than without having it. In particular, the adversary can even enumerate all possible ciphertexts that are not equal to the given ciphertext C, and try decrypting each of them using all possible keys. Based on the decrypted message pairs, she may choose one to try to maximize the chance of success.

The nontrivial part of the analysis concentrates on bounding the maximal possible success probability in $\mathrm{TDNM1}^{\mathcal{A},R}_{\mathsf{HE\text{-}NM},p_m,p_k}$. Again we first do game changes so that the adversary would output the modified ciphertext before the key is selected. In this case, for each pair of ciphertext, we can define clearly a set of "preferable" keys for which the decrypted messages resulting from decrypting

using these keys satisfy the relation. After exhausting searching all possible ciphertexts, the maximum probability that an adversaries can win with is achieved by outputting the ciphertext $\tilde{C}$ which defines the set of "preferable" keys which has the maximum accumulated probability among all those sets, i.e., the largest possible probability that a randomly selected key will fall into a preferable set. Bounding the accumulated probability can again be transformed into bounding the maximum weight of balls in a bin. The difference compared to the TDSS analysis is that now in every experiment, we will throw $|\mathcal{K}|$ balls into two bins, but once for every single ciphertext. Letting N be the number of possible ciphertexts in the range of the scheme, we therefore analyze N experiments and find the maximum load among all N experiments.

It is not hard to see the expected load in one experiment would be p_R, however, directly bounding the load using, e.g., a Chernoff bound would not be very effective since the expected value is small but the "bad" event (load with significant deviation compared to the expected value) happens with a significant probability. To proceed forward in a similar way as the TDSS analysis, we would reduce the bound to the flat distribution. Unfortunately, in this case, the expected maximum load defined in the TDNM analysis is over several independent balls-into-bins experiments instead of a single such experiment we can not directly apply the majorization lemma from [9].

We turn to a more general majorization technique that uses the property of Schur convexity [22] (to be defined below). A Schur convex function preserves the order under majorization, i.e., if $\bar{p} \succ \bar{q}$, for a Schur convex function f, it holds that $f(\bar{p}) \geq f(\bar{q})$. To bound the expected maximum load in N experiments, we then proceed in two steps. First we argue the expected value of maximum weight across all bins in the N experiments as a function over the key distribution p_k is indeed Schur convex. Since the flat weight vector ($\lceil 1/\omega_k \rceil$ keys each with probability at most ω_k) majorizes the key distribution, we then bound the expected maximum load in N experiments for the flat weight vector which can be done by counting the maximum number of balls falling into a bin in N experiments.

First we introduce some notion we will use for the generalized majorization technique.

Schur convex functions. A function $g : \mathbb{R}^n \to \mathbb{R}$ is called Schur-convex if for any $p, q \in \mathbb{R}^n$, if $p \succ q$ then $g(p) \geq g(q)$. A useful result from Schur [22], tells us that any function satisfying two properties known as convex and symmetric must be Schur-convex.

We say g is convex if for any $t \in [0, 1]$ and $p, q \in \mathbb{R}^n$ we have $g(tp+(1-t)q)] \leq tg(p) + (1-t)g(q)$. Finally, g is symmetric if the value of the function does not change if the input vector is permuted, that is if $\phi : n \to n$ is a permutation and $p^\phi \in \mathbb{R}^n$ is define by $p^\phi(i) = p(\phi(i))$ for all $1 \leq i \leq n$ then $g(p) = g(p^\phi)$.

Lemma 5 (Schur23). *If a function $g : \mathbb{R}^n \to \mathbb{R}$ is symmetric and convex, then g is Schur-convex.*

6.3 Security Proofs for DTE-then-Encipher

We are now in position to formalize the proof intuition above. The main theorem of this section, given below, establishes an upper bound on the TDNM security of the DTE-then-Encipher construction.

Theorem 4. *Let HE-NM be defied as in Figure 12 for distributions p_m, p_k, where $(\mathtt{Enc}, \mathtt{Dec})$ is an ideal cipher with $\mathtt{Enc} : \mathcal{K} \times \{0,1\}^\ell \to \{0,1\}^\ell$. Let $\epsilon = \mathbf{Adv}^{\mathrm{dte}}_{DTE,p_m}$. Then for any s satisfying $\lceil 1/\omega_k \rceil^{s-1} \leq p_R - \epsilon - 2^\ell$ we have*

$$\mathbf{Adv}^{\mathrm{tdnm}}_{\mathsf{HE\text{-}NM},p_m,p_k} \leq \omega_k(1 + \lceil 1/\omega_k \rceil^s) + 2^{1-\ell} + 2^{\ell - \frac{\lceil 1/\omega_k \rceil^{2s-1}}{3}} + 2\epsilon$$

Remark 3. The given bound will typically be quantitatively similar to that of the TDSS advantage, since the message length ℓ is quite small comparing to $\lceil 1/\omega_k \rceil^{2s-1}$. When we take $\omega_k = 2^{-40}, \ell = 128, s = \frac{5}{8}$, we can get $\mathbf{Adv}^{\mathrm{tdnm}}_{\mathsf{HE\text{-}NM},p_m,p_k}$ is around 2^{-15}.

Proof. First we give a lower bound for $\max_{\mathcal{B}} \Pr[\mathrm{TDNM0}^{\mathcal{B},R}_{\mathsf{HE},p_m,p_k} \Rightarrow \mathsf{true}]$. Consider the simulator $\mathcal{A}_s$ that simply outputs a ciphertext $\tilde{C}$ randomly sampled from $\mathcal{C}$. It's easy to verify that the probability $\mathcal{A}_s$ succeeds is equal to the probability that a random sample according to p_m and a random sample according to p_d satisfy the relation, i.e., $\Pr[R(M, \tilde{M}) = 1 : M \leftarrow_{p_m} \mathcal{M}, \tilde{M} \leftarrow_{p_d} \mathcal{M}]$.

Denoting this quantity by p^d_R, let $\mathcal{D}_R$ be the adversary against the security of DTE which simply samples a random M according to p_m and decodes its input S to obtain $\tilde{M}$, then outputs 1 if M and $\tilde{M}$ satisfy R. It is easy to verify that $\Pr[\mathrm{SAMP1}^{\mathcal{D}_R}_{\mathsf{DTE},p_m}] = p_R$ and $\Pr[\mathrm{SAMP0}^{\mathcal{D}_R}_{\mathsf{DTE}}] = p^d_R$ which gives us:

$$\max_{\mathcal{A}_s} \Pr[\mathrm{TDNM0}^{\mathcal{B},R}_{\mathsf{HE},p_m,p_k} \Rightarrow \mathsf{true}] \geq p_R - \mathbf{Adv}^{\mathrm{dte}}_{\mathsf{DTE},p_m}.$$

Transitioning $\mathrm{TDNM1}^{\mathcal{A},R}_{\mathsf{HE},p_m,p_k}$. Now we analyze the maximum winning probability of an adversary $\mathcal{A}$ in $\mathrm{TDNM1}^{\mathcal{A},R}_{\mathsf{HE\text{-}NM},p_m,p_k}$. Consider the sequence of game shown in Figure 13. Game G_0 is the simply $\mathrm{TDNM1}^{\mathcal{A},R}_{\mathsf{HE\text{-}NM},p_m,p_k}$ with the encryption code of HE-NM inserted. Thus,

$$\Pr[\mathrm{G}_0 \Rightarrow \mathsf{true}] = \Pr[\mathrm{TDNM1}^{\mathcal{A},R}_{\mathsf{HE},p_m,p_k} \Rightarrow \mathsf{true}].$$

We can then use the security of DTE to transition to game G_1 because G_1 is identical to G_0 except instead of a random message being sampled and then encoded, a random seed is sampled and then decoded. Consider the adversary $\mathcal{D}$ against the security of DTE shown in the center of Figure 14. Adversary $\mathcal{D}$ uses its input to simulate the view of $\mathcal{A}$ returning 1 is $\mathcal{A}$ wins and 0 otherwise. It is easy to check that when in SAMP1, $\mathcal{D}$ perfectly simulated game G_0 for $\mathcal{A}$ and when in SAMP0 it perfect simulates game G_1. It is then clear that

$$\Pr[\mathrm{G}_1 \Rightarrow \mathsf{true}] \leq \Pr[\mathrm{G}_0 \Rightarrow \mathsf{true}] + \mathbf{Adv}^{\mathrm{dte}}_{\mathsf{DTE},p_m}.$$

Finally game G_2 is simply a rewriting of G_1 so that the sampling of K is delayed until after $\mathcal{A}$ has already executed. It is clear that,

$$\Pr[\mathrm{G}_2 \Rightarrow \mathsf{true}] = \Pr[\mathrm{G}_1 \Rightarrow \mathsf{true}].$$

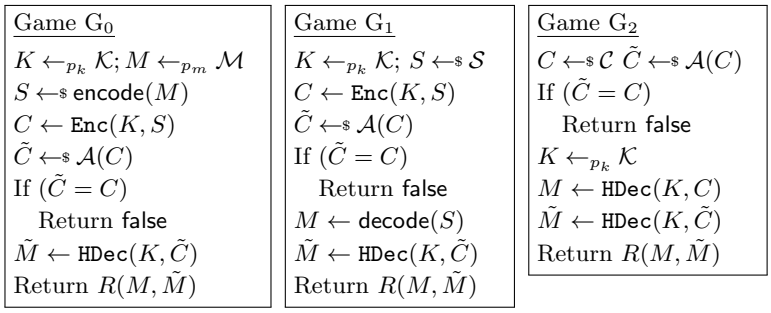

Fig. 13. Game transition for TDNM analysis

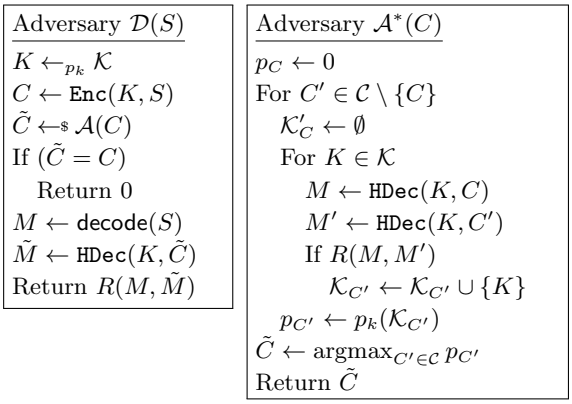

Fig. 14. Adversaries used in proof of Theorem 4.

Next, we will focus on bounding the winning probability of $\mathcal{A}$ in game G_2. Consider the attacking strategy described on the right side of Figure 14. The adversary $\mathcal{A}^*$ takes a ciphertext C as input. For each other $C' \in \mathcal{C}$, adversary $\mathcal{A}^*$ tries decrypting both C and C' using all possible keys and defines a set $\mathcal{K}_{C'}$ consisting of all the keys for which the corresponding decrypted messages M and M' satisfy the relation R. Then $\mathcal{A}^*$ defines a quantity $p_{C'}$ as the probability that a key sampled according to p_k will fall into $\mathcal{K}_{C'}$.

Recall that in G_2, the key is selected after the adversary outputs the ciphertext $\tilde{C}$, thus we can see that the winning probability of an adversary $\mathcal{A}$ in G_2 will be exactly the value $p_{C'}$ calculated by $\mathcal{A}^*$ corresponding to the output C'.

Thus because $\mathcal{A}^*$ outputs $\tilde{C}$ maximizing this value, it is clear that $\mathcal{A}^*$ is an optimal adversary for G_2. Thus letting G_2^* denote the game G_2 when run with $\mathcal{A}^*$ it is clear that:

$$\Pr[G_2^* \Rightarrow \mathsf{true}] \geq \Pr[G_2 \Rightarrow \mathsf{true}].$$

Furthermore we can clear see that $\Pr[G_2^* \Rightarrow \mathsf{true}]$ will be exactly the expected value of $\max_{C \in \mathcal{C}} p_C$, denoted by $\mathbb{E}[\max_{C \in \mathcal{C}} p_C]$.

Schur convexity of $\mathbb{E}[\max_{C\in\mathcal{C}} p_C]$. To apply the majorization technique, we will argue the Schur convexity of the quantity we want to bound. We will argue $\mathbb{E}[\max_{C\in\mathcal{C}} p_C]$ is symmetric and convex. then following lemma 5, it is Schur convex.

For a given key distribution p_k let P_{p_k} be a random variable denoting the value $\max_{C'\in\mathcal{C}} p'_C$ when $\mathcal{A}^*$ uses distribution p_k as the key distribution.

It is clear that $\mathbb{E}[P_p]$ is symmetric because the key are only used for the ideal cipher $(\texttt{Enc}, \texttt{Dec})$ whose a priori behavior of the key used as input. Thus permuting the corresponding probabilities of the keys will does not change the expected value of P.

To see that $\mathbb{E}[P_p]$ is convex let $p, q \in \mathbb{R}^{|\mathcal{K}|}$, $t \in \mathbb{R}$, and set $r = tp + (1-t)q$. We would like to show that $\mathbb{E}[P_r] \leq t \cdot \mathbb{E}[P_p] + (1-t) \cdot \mathbb{E}[P_q]$. Note that the corresponding executions of $\mathcal{A}^*$ differ only in the weights assigned to the keys, so the distributions of which keys are included in the various sets $\mathcal{K}_C$ are the same between them.

For a fixed choice of random coins, let C_r, C_p, C_q denote the respective output of $\mathcal{A}^*$ in the different experiments. Then from the definition of r and the fact that the ciphertexts are chosen to maximize the weights of the corresponding $\mathcal{K}_C$ we get:

$$\begin{aligned} P_r &= r(\mathcal{K}_{C_r}) \\ &= t \cdot p(\mathcal{K}_{C_r}) + (1-t)q(C_r) \\ &\leq t \cdot p(\mathcal{K}_{C_p}) + (1-t)q(C_q) \\ &= tP_p + (1-t)P_q. \end{aligned}$$

Because the above holds for every choice of random coins in the corresponding experiments it is clear that $\mathbb{E}[P_r] \leq t \cdot \mathbb{E}[P_p] + (1-t) \cdot \mathbb{E}[P_q]$, so $\mathbb{E}[P_p]$ is convex.

Having now shown that $\mathbb{E}[P_p]$ is symmetric and convex, Lemma 5 tells us it is Schur-convex.

Bounding $\mathbb{E}[P_{p'_k}]$ for flat distribution p'_k. Now as in our TDSS analysis let p'_k be defined such that for $i \leq \lceil 1/\omega_k \rceil$ we have $p'_k(K_i) = \omega_k$ and $p'_k(K_i) = 0$ otherwise, and note that p'_k majorizes p_k. Since $\mathbb{E}[P_p]$ is Schur convex, and p'_k majorizes p_k, we have:

$$\mathbb{E}[P_{p_k}] \leq \mathbb{E}[P_{p'_k}]$$

Next, we will focus on bounding $\mathbb{E}[P_{p'_k}]$.

Let us rephrase the quantity $p_{C'}$ using the terminology of a balls-into-bins game. Letting the challenge ciphertext C be fixed we can think of each $K \in \mathcal{K}$ as a ball into bins B_0 and B_1 according to the value of $R(M, M')$ where M and M' are obtained by decrypting C and C' respectively with K. Because each decryption uses the ideal cipher it is clear that each key is thrown independently.

Because decrypting applying $\texttt{Dec}$ to ciphertext results in a uniformly random S we would like to say that we can view M and M' as both being drawn independently according to p_d. However, this is not quite true because there is a small dependence between the samples because $\texttt{Dec}$ applied to C and C' results

Experiment $\mathrm{E}^{p'_d,R}_{p'_k}$

For $K \in \mathcal{K}$ do
  $(M, M') \leftarrow_{p'_d} \mathcal{M} \times \mathcal{M}$
  $b \leftarrow R(M, M')$
  $B_b \leftarrow B_b \cup \{K\}$
$\mathrm{L}^{p'_d,R}_{p_k} \leftarrow p_k(B_1)$

Fig. 15. Ball-into-bins experiments used to analyze the security of HE-NM.

in uniformly chosen S and S' with the restriction that $S \neq S'$. Let p'_d denote the distribution on $\mathcal{M} \times \mathcal{M}$ obtained by applying decode to two uniformly chosen seeds with the restriction that the seeds are not equal. Then we can view the values of M and M' for each K in the balls-into-bins experiment as being independent samples from p'_d.

Putting this together we can thing of the quantity $p_{C'}$ as the load $\mathrm{L}^{p'_d,R}_{p'_k}$ in the balls-into-bins experiment $\mathrm{E}^{p'_d,R}_{p_k}$ shown in Figure 15.

Let $p'_R = \Pr[R(M, M') | (M, M') \leftarrow_{p'_d} \mathcal{M} \times \mathcal{M}]$ denote the probability that (M, M') sampled according to p'_d satisfies R and $a = \lceil 1/\omega_k \rceil$ denote the number of balls thrown in experiment $\mathrm{E}^{p'_d,R}_{p'_k}$. Then it is clear that the expected number of balls that fall into bin B_1 is a ap'_R.

Then letting X denote the number of balls thrown into B_1 and $\delta = a^{s-1}/p'_R$ (which is less than 1 from our choice of s) we can apply Chernoff's bound to get:

$$\Pr[X \geq ap'_R + a^s] \leq e^{\frac{-a^{2s-1}}{3}}.$$

Now we can complete the proof by using this to bound the expected value of $\max_{C \in \mathcal{C}} p'_C$ for $\mathcal{A}^*$. For this to be greater than $\omega_k(ap'_R + a^s)$ it must be the case that for some $C \in \mathcal{C}$, p'_C is greater than $\omega_k(ap'_R + a^s)$. Then from the union bound we get

$$\begin{aligned}
\Pr\left[\max_{C \in \mathcal{C}} p'_C \geq \omega_k(ap'_R + a^s)\right] &\leq \sum_{C \in \mathcal{C}} \Pr[p'_C \geq \omega_k(ap'_R + a^s)] \\
&\leq \sum_{C \in \mathcal{C}} \Pr[\mathrm{L}^{p'_d,R}_{p'_k} \geq \omega_k(ap'_R + a^s)] \\
&= (|\mathcal{C}| - 1) \cdot \Pr[X \geq ap'_R + a^s] \\
&\leq (|\mathcal{C}| - 1) e^{\frac{-a^{2s-1}}{3}}.
\end{aligned}$$

Note that applying the union bound in this manner allows us to ignore the dependence that exists between difference p'_C for different C.

Finally we can bound the expected value of $\max_{C \in \mathcal{C}} p'_C$ by pessimistically assuming it is always 1 whenever it is greater than $\omega_k(ap'_R + a^s)$ and it is $\omega_k(ap'_R +$

a^s) otherwise. Recalling that $|\mathcal{C}| = 2^\ell$ and letting $P = \max_{C \in \mathcal{C}} p'_C$, this gives us the following sequence of inequalities:

$$\begin{aligned}\mathbb{E}[P] &\leq \omega_k(ap'_R + a^s)\Pr[P \leq \omega_k(ap'_R + a^s)] + 1 \cdot \Pr[P \geq \omega_k(ap'_R + a^s)] \\ &\leq \omega_k(ap'_R + a^s) + (2^\ell - 1)e^{\frac{-a^{2s-1}}{3}} \\ &\leq \omega_k(ap'_R + a^s) + 2^{\ell - \frac{a^{2s-1}}{3}}\end{aligned}$$

From the definition of p'_R it is clear that $p'_R \leq p_R + 1/|\mathcal{S}| = p_R + 2^{-}\ell$. Putting everything together we get the final bound of

$$\begin{aligned}\mathbf{Adv}^{\text{tdnm}}_{\mathsf{HE\text{-}NM},p_m,p_k} &\leq \omega_k(\lceil 1/\omega_k \rceil (p_R + 2^{-\ell}) + \lceil 1/\omega_k \rceil^s) + 2^{\ell - \frac{\lceil 1/\omega_k \rceil^{2s-1}}{3}} - p_R + 2\epsilon \\ &\leq (1 + \omega_k)(p_R + 2^{-\ell}) + \omega_k \lceil 1/\omega_k \rceil^s + 2^{\ell - \frac{\lceil 1/\omega_k \rceil^{2s-1}}{3}} - p_R + 2\epsilon \\ &\leq \omega_k(1 + \lceil 1/\omega_k \rceil^s) + 2^{1-\ell} + 2^{\ell - \frac{\lceil 1/\omega_k \rceil^{2s-1}}{3}} + 2\epsilon.\end{aligned}$$

This completes the proof. □

7 Conclusions and Open Problems

In this work, we initiated the study of security notions for honey encryption schemes stronger than the previously proposed goal of resistance to message recovery attacks. We, first, proved that message recovery is always possible with a known-message attack. Formally proving this folklore result was more nuanced than expected. We then defined semantic security and non-malleability for honey encryption schemes with respect to targeted message distributions, and we showed that the simple constructions of encode-then-encrypt and encode-then-encipher achieve targeted distribution semantic security and targeted distribution non-malleability, respectively. The general technique for balls-into-bins type of analysis using Schur convexity may be of independent interest.

Security notions for symmetric key encryption with low-entropy keys are still not yet fully understood. For honey encryption schemes, completely characterizing the relations among various security notions remains an open problem whose solution would expand on our results. Also, although replacing a random oracle with a k-wise independent hash function to get a standard model construction for TDSS seems intuitive, formally analyzing its security requires more delicate balls-into-bins analysis than we have provided here. Last, our TDNM construction relies on an ideal cipher with large block size. Obtaining a construction provably TDNM secure in the random oracle model therefore represents an important open question.

Acknowledgements : We thank the anonymous reviewers for valuable comments. Joseph Jaeger was supported in part by NSF grants CNS-1526801 and CNS-1228890, ERC Project ERCC FP7/615074 and a gift from Microsoft. Thomas Ristenpart was supported by NSF grants CNS-1514163, CNS-1546033, CNS-1065134, CNS-1330308 and a gift from Microsoft. Qiang Tang was supported by NSF grants CNS-1518765 and CNS-1514261.

References

1. M. Bellare, A. Boldyreva, and A. O'Neill. Deterministic and efficiently searchable encryption. In *Advances in Cryptology - CRYPTO 2007, 27th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 19-23, 2007, Proceedings*, pages 535–552, 2007.
2. M. Bellare, Z. Brakerski, M. Naor, T. Ristenpart, G. Segev, H. Shacham, and S. Yilek. Hedged public-key encryption: How to protect against bad randomness. In *Advances in Cryptology - ASIACRYPT 2009, 15th International Conference on the Theory and Application of Cryptology and Information Security, Tokyo, Japan, December 6-10, 2009. Proceedings*, pages 232–249, 2009.
3. M. Bellare, A. Desai, E. Jokipii, and P. Rogaway. A concrete security treatment of symmetric encryption. In *38th Annual Symposium on Foundations of Computer Science, FOCS '97, Miami Beach, Florida, USA, October 19-22, 1997*, pages 394–403, 1997.
4. M. Bellare, M. Fischlin, A. O'Neill, and T. Ristenpart. Deterministic encryption: Definitional equivalences and constructions without random oracles. In *Advances in Cryptology - CRYPTO 2008, 28th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 17-21, 2008. Proceedings*, pages 360–378, 2008.
5. M. Bellare, J. Kilian, and P. Rogaway. The security of the cipher block chaining message authentication code. *Journal of Computer and System Sciences*, 61(3):362–399, 2000.
6. M. Bellare and C. Namprempre. Authenticated encryption: Relations among notions and analysis of the generic composition paradigm. In *Advances in CryptologyASIACRYPT 2000*, pages 531–545. Springer, 2000.
7. M. Bellare, T. Ristenpart, and S. Tessaro. Multi-instance security and its application to password-based cryptography. In *Advances in Cryptology - CRYPTO 2012 - 32nd Annual Cryptology Conference, Santa Barbara, CA, USA, August 19-23, 2012. Proceedings*, pages 312–329, 2012.
8. M. Bellare and P. Rogaway. Random oracles are practical: A paradigm for designing efficient protocols. In *ACM Conference on Computer and Communications Security*, pages 62–73, 1993.
9. P. Berenbrink, T. Friedetzky, Z. Hu, and R. Martin. On weighted balls-into-bins games. *Theor. Comput. Sci.*, 409(3):511–520, Dec. 2008.
10. J. Bonneau. The science of guessing: Analyzing an anonymized corpus of 70 million passwords. In *Proceedings of the 2012 IEEE Symposium on Security and Privacy*, SP '12, pages 538–552, 2012.
11. R. Chatterjee, J. Bonneau, A. Juels, and T. Ristenpart. Cracking-resistant password vaults using natural language encoders. In *2015 IEEE Symposium on Security and Privacy, SP 2015, San Jose, CA, USA, May 17-21, 2015*, pages 481–498, 2015.
12. J.-S. Coron, Y. Dodis, A. Mandal, and Y. Seurin. A domain extender for the ideal cipher. In *Theory of Cryptography*, pages 273–289. Springer Berlin Heidelberg, 2010.
13. D. Dachman-Soled, J. Katz, and A. Thiruvengadam. 10-round feistel is indifferentiable from an ideal cipher. Cryptology ePrint Archive, Report 2015/876, 2015. `http://eprint.iacr.org/`.
14. Y. Dai and J. Steinberger. Feistel networks: Indifferentiability at 10 rounds. Cryptology ePrint Archive, Report 2015/874, 2015. `http://eprint.iacr.org/`.
15. Y. Dodis and A. Smith. Entropic security and the encryption of high entropy messages. In *TCC*, pages 556–577, 2005.

16. D. Dolev, C. Dwork, and M. Naor. Nonmalleable cryptography. *SIAM J. Comput.*, 30(2):391–437, 2000.
17. S. Goldwasser and S. Micali. Probabilistic encryption and how to play mental poker keeping secret all partial information. In *Proceedings of the 14th Annual ACM Symposium on Theory of Computing, May 5-7, 1982, San Francisco, California, USA*, pages 365–377, 1982.
18. S. Goldwasser, S. Micali, and R. Rivest. A digital signature scheme secure against adaptive chosen-message attacks. *SIAM J. Comput.*, 17(2):281–308, 1988.
19. V. T. Hoang, T. Krovetz, and P. Rogaway. Robust authenticated-encryption aez and the problem that it solves. In *Advances in Cryptology–EUROCRYPT 2015*, pages 15–44. Springer, 2015.
20. T. Holenstein, R. Künzler, and S. Tessaro. The equivalence of the random oracle model and the ideal cipher model, revisited. In *Proceedings of the forty-third annual ACM symposium on Theory of computing*, pages 89–98. ACM, 2011.
21. Z. Huang, E. Ayday, J. Fellay, J. Hubaux, and A. Juels. Genoguard: Protecting genomic data against brute-force attacks. In *2015 IEEE Symposium on Security and Privacy, SP 2015, San Jose, CA, USA, May 17-21, 2015*, pages 447–462, 2015.
22. I.Schur. U?ber eine klassevon mittelbildungen mit anwendungen die determinanten. *Theorie Sitzungsber. Berlin. Math. Gesellschaft*, 22:9–20, 1923.
23. A. Juels and T. Ristenpart. Honey encryption: Security beyond the brute-force bound. In *Advances in Cryptology - EUROCRYPT 2014 - 33rd Annual International Conference on the Theory and Applications of Cryptographic Techniques, Copenhagen, Denmark, May 11-15, 2014. Proceedings*, pages 293–310, 2014.
24. B. Kaliski. PKCS #5: Password-Based Cryptography Specification Version 2.0. RFC 2898 (Informational), September 2000.
25. J. Katz and M. Yung. Complete characterization of security notions for probabilistic private-key encryption. In *Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing, May 21-23, 2000, Portland, OR, USA*, pages 245–254, 2000.
26. B. Kausik. Method and apparatus for cryptographically camouflaged cryptographic key storage, certification and use, Jan. 2 2001. US Patent 6,170,058.
27. U. Maurer, R. Renner, and C. Holenstein. Indifferentiability, impossibility results on reductions, and applications to the random oracle methodology. In *Theory of cryptography*, pages 21–39. Springer, 2004.
28. M. Raab and A. Steger. "balls into bins" - a simple and tight analysis. In *Proceedings of the Second International Workshop on Randomization and Approximation Techniques in Computer Science*, RANDOM '98, pages 159–170, 1998.
29. P. Rogaway and T. Shrimpton. A provable-security treatment of the key-wrap problem. In *Advances in Cryptology-EUROCRYPT 2006*, pages 373–390. Springer, 2006.
30. A. Russell and H. Wang. How to fool an unbounded adversary with a short key. In *Advances in Cryptology - EUROCRYPT 2002, International Conference on the Theory and Applications of Cryptographic Techniques, Amsterdam, The Netherlands, April 28 - May 2, 2002, Proceedings*, pages 133–148, 2002.