Honey Encryption: Security Beyond the Brute-Force Bound

Ari Juels
ajuels@gmail.com

Thomas Ristenpart
University of Wisconsin
rist@cs.wisc.edu

January 29, 2014
Version 1.1

Abstract

We introduce *honey encryption* (HE), a simple, general approach to encrypting messages using low min-entropy keys such as passwords. HE is designed to produce a ciphertext which, when decrypted with any of a number of *incorrect* keys, yields plausible-looking but bogus plaintexts called *honey messages*. A key benefit of HE is that it provides security in cases where too little entropy is available to withstand brute-force attacks that try every key; in this sense, HE provides security beyond conventional brute-force bounds. HE can also provide a hedge against partial disclosure of high min-entropy keys.

HE significantly improves security in a number of practical settings. To showcase this improvement, we build concrete HE schemes for password-based encryption of RSA secret keys and credit card numbers. The key challenges are development of appropriate instances of a new type of randomized message encoding scheme called a *distribution-transforming encoder* (DTE), and analyses of the expected maximum loading of bins in various kinds of balls-and-bins games.

1 Introduction

Many real-world systems rely for encryption on low-entropy or weak secrets, most commonly user-chosen passwords. Password-based encryption (PBE), however, has a fundamental limitation: users routinely pick poor passwords. Existing PBE mechanisms attempt to strengthen bad passwords via salting, which slows attacks against multiple users, and iterated application of one-way functions, which slows decryption and thus attacks by a constant factor c (e.g., $c = 10{,}000$). Recent results [4] prove that for conventional PBE schemes (e.g., [30]), work q suffices to crack a single ciphertext with probability $q/c2^{\mu}$ for passwords selected from a distribution with min-entropy μ. This *brute-force bound* is the best possible for in-use schemes.

Unfortunately empirical studies show this level of security to frequently be insufficient. A recent study [9] reports $\mu < 7$ for passwords observed in a real-world population of 69+ million users. (1.08% of users chose the same password.) For any slowdown c small enough to support timely decryption in normal use, the security offered by conventional PBE is clearly too small to prevent message-recovery (MR) attacks.

We explore a new approach to PBE that provides security beyond the brute-force bound. The idea is to build schemes for which attackers are *unable to succeed in message recovery even after trying every possible password / key*. We formalize this approach by way of a new cryptographic primitive called *honey encryption* (HE). We provide a framework for realizing HE schemes and show scenarios useful in practice in which even computationally unbounded attackers can provably recover an HE-encrypted plaintext with probability at most $2^{-\mu} + \epsilon$ for negligible ϵ. Since there exists a trivial, fast attack that succeeds with probability $2^{-\mu}$ (guess the most probable password), we thus demonstrate that HE can yield optimal security.

While HE is particularly useful for password-based encryption (PBE), we emphasize that "password" here is meant very loosely. HE is applicable to *any* distribution of low min-entropy keys, including passwords, PINs, biometrically extracted keys, etc. It can also serve usefully as a hedge against partial compromise of high min-entropy keys.

Background. Stepping back, let us review briefly how brute-force message-recovery attacks work. Given an encryption $C = \mathsf{enc}(K, M)$ of message M, where K and M are drawn from known distributions, an attacker's goal

is to recover M. The attacker decrypts C under as many candidate keys as she can, resulting in a set of messages $M_1, \ldots, M_q$. Should one of the candidate keys be correct (i.e., K is from a low-entropy distribution), M is guaranteed to appear in this list, and at this stage the attacker wins with probability equal to her ability to pick out M from the q candidates. Conventional PBE schemes make this easy in almost all settings. As just one example, if M is a 16-digit credit card number encoded via ASCII and the PBE scheme behaves like an ideal cipher, the probability that any $M_i \neq M$ is a valid ASCII encoding of a 16-digit string is negligible, at $(10/256)^{16} < 2^{-74}$. An attacker can thus reject incorrect messages and recover M with overwhelming probability. In fact, cryptographers generally ignore the problem of identifying valid plaintexts and assume conservatively that if M appears in the list, the attacker wins.

Prior theoretical frameworks for analyzing PBE schemes have focused on showing strong security bounds for sufficiently unpredictable keys. A result for the PKCS#5 PBE schemes due to Bellare, Ristenpart, and Tessaro [4] proves that no attacker can break semantic security (learn any partial information about plaintexts) with probability greater than $q/(c2^{\mu})$; here, c is the time to perform a single decryption, μ is the min-entropy of the distribution of the keys, and negligible terms are ignored. As mentioned above, though, when $\mu = 7$, such a result provides unsatisfying security guarantees, and the formalisms and proof techniques of [4] cannot offer better results. It may seem that this is the best one can do and that providing security beyond this "brute-force barrier" remains out of reach.

Perhaps unintuitively (at least to the authors of the present paper), the bounds above are actually *not* tight for all settings, as they do not take into account the distribution of the challenge message M. Should M be a uniformly chosen bit-string of length longer than μ, for instance, then the best possible message recovery attack would appear to work with probability at most $1/2^{\mu}$. This is because for typical PBE schemes an attacker will have a hard time, in practice, distinguishing the result of $\mathsf{dec}(K, C)$ for any K from a uniform bit string. Said another way, the candidate messages $M_1, \ldots, M_q$ would all appear to be equally valid as plaintexts. Thus an adversary would seem to maximize her probability of message recovery simply by decrypting C using the key with the highest probability, which is at most $1/2^{\mu}$.

Previously proposed security tools have exploited exactly this intuition for special cases. Hoover and Kausik [24] consider the problem of encrypting a (uniformly-chosen) RSA or DSA secret exponent for authenticating a user to a remote system. Only the remote system holds the associated public key. To hedge against compromise of the user's machine, they suggest encrypting the secret exponent under a PIN (a short decimal-string password). They informally argue that brute-force attacks will only result in a set of valid-looking exponents, and so the best an attacker can do is use each candidate exponent to attempt login to the remote system, effectively requiring an online brute-force attack. Their work led to a commercially deployed system [26]. Other systems similarly seek to foil offline brute-force attacks, but mainly by means of hiding valid authentication credentials in an *explicitly stored list* of plausible-looking fake ones (often called "decoys" or "honeywords") [8, 25]. Similarly, detection of system breaches using "honeytokens," such as fake credit-card numbers, is a common industry practice [37].

Honey encryption (HE). Inspired by such decoy systems, we set out to build HE schemes that provide security beyond the brute-force barrier, in particular yielding candidate messages during brute-force attacks that are indistinguishable from the attacker's perspective. We refer to the incorrect plaintext candidates in HE as *honey messages*, following the long tradition of this sweet substance's role in computer security terminology.

We provide a formal treatment of HE. Functionally, an HE scheme is exactly like a PBE scheme: it takes arbitrary strings as passwords and uses them to perform randomized encryption of a message. We ask that HE schemes simultaneously target two security goals: message recovery (MR) security, as parameterized by a distribution over messages, and the more (multi-instance) semantic-security style goals of [4]. As we noted, the latter can only be achieved up to the brute-force barrier, and is thus meaningful only for high min-entropy keys; our HR schemes achieve the goals of [4] using standard techniques. The bulk of our efforts in this paper will be on MR security, where we target security better than $q/c2^{\mu}$. Our schemes will, in fact, achieve security bounds close to $1/2^{\mu}$ for unbounded attackers when messages are sufficiently unpredictable.

HE schemes also produce compact ciphertexts (unlike explicitly stored decoys). While lengths vary by construction, an HE ciphertext on M is typically a small constant multiple (e.g., 2) of the length of a conventional PBE ciphertext on M.

Framework for HE schemes. We provide a general methodology for building HE schemes. Its cornerstone is a new kind of (randomized) message encoding that we call a *distribution-transforming encoder (DTE)*. A DTE is designed

with an estimate of the message distribution p_m in mind, making it conceptually similar to arithmetic/Huffman coding [16]. The message space handled by a DTE is exactly the support of p_m (messages with non-zero probability). Encoding a message sampled from p_m yields a "seed" value distributed (approximately) uniformly. It is convenient in many cases for seeds to be binary strings. A DTE must have an efficient decoder that, given a seed, obtains the corresponding message. Applying the decoder to a uniformly sampled seed produces a message distributed (approximately) under p_m. A good (secure) DTE is such that no attacker can distinguish with significant probability between these two distributions: (1) a pair (M, S) generated by selecting M from p_m and encoding it to obtain seed S, and (2) a pair (M, S) generated by selecting a seed S uniformly at random and decoding it to obtain message M. Building DTEs is non-trivial in many cases, for example when p_m is non-uniform.

Encrypting a message M under HE involves a two-step procedure that we call *DTE-then-encrypt*. First, the DTE is applied to M to obtain a seed S. Second, the seed S is encrypted under a conventional encryption scheme enc using the key K, yielding an HE ciphertext C. This conventional encryption scheme enc must have message space equal to the seed space and all ciphertexts must decrypt under any key to a valid seed. Typical PBE schemes operating on bitstrings provide all of this (but authenticated encryption schemes do not). Appropriate care must be taken, however, to craft a DTE whose outputs require no padding (e.g., for CBC-mode encryption).

We prove a general theorem (Theorem 2) that upper bounds the MR security of any DTE-then-encrypt scheme by the DTE's security and a scheme-specific value that we call the expected maximum load. Informally, the expected maximum load measures the worst-case ability of an unbounded attacker to output the right message; we relate it to the expected maximum load of a bin in a kind of balls-and-bins game. Analyzing an HE scheme built with our approach (and a good DTE) therefore reduces to analyzing the balls-and-bins game that arises for the particular key and message distribution. Assuming the random oracle model or ideal cipher model for the underlying conventional encryption scheme enables us to assume balls are thrown independently in these games. (We conjecture that k-wise independent hashing, and thus k-wise independent ball placement, may achieve strong security in many cases as well.)

A DTE is designed using an estimate of the target message distribution p_m. If the DTE is only approximately right, we can nevertheless prove message-recovery security far beyond the brute-force-barrier. If the DTE is bad, i.e., based on a poor estimate of p_m, we fall back to normal security (up to the brute-force barrier), at least provably achieving the semantic security goals in [4]. This means we never do worse than prior PBE schemes, and, in particular, attackers must always first perform the work of offline brute-force attacks before HE security becomes relevant.

HE instantiations. We offer as examples several concrete instantiations of our general DTE-then-encrypt construction. We build HE schemes that work for RSA secret keys by crafting a DTE for uniformly chosen pairs of prime numbers. This enables us to apply HE to RSA secret keys as generated and stored by common tools such as OpenSSL. This improves on the scheme of Hoover and Kausik, which requires that RSA secret exponents be generated in a non-standard manner [24]. Interestingly, simple encoding strategies here fail. For example, encoding the secret keys directly as binary integers (in the appropriate range) would enable an attacker to rule out candidate messages resulting from decryption by running primality tests. Indeed, the DTE we design has decode (essentially) implement a prime number generation algorithm. (This approach slows down decryption significantly, but as noted above, in PBE settings we want decryption to not be very fast anyway.)

We also build HE schemes for password-based encryption of credit card numbers, their associated Card Verification Values (CVVs), and (user-selected) PINs. Encryption of PINs requires a DTE that handles a non-uniform distribution over messages, as empirical studies show a heavy user bias in PIN selection [6]. The resulting analysis consequently involves a balls-and-bins game with non-uniform bin capacities, a somewhat unusual setup in the literature.

In each of the cases above we are able to prove close to optimal MR security.

Limitations of HE. The security guarantees offered by HE come with some strings attached. First, HE security does not hold when the adversary has some side information about the target message. As a concrete example, the RSA secret key HE scheme will only provide the strong MR guarantees when the attacker does not already have the public key associated with the encrypted secret key. Thus the HE guarantees are not typically going to help protect normal HTTPS certificate keys. (The intended application for this HE scheme is client authorization, where the public key is stored only at the remote server, a typical setting for SSH users. Refer to [24] for more detailed deployment discussion.) Second, because decryption of an HE ciphertext under a wrong key produces fake but valid-looking

messages, typos in passwords might confuse legitimate users in some settings. We address this issue of "typo-safety" in Section 7. Third and finally, we assume in our HE analyses that the key and message distributions are independent. If they are correlated, an attacker may be able to identify a correct message by comparing it with the decryption key that produced it. Similarly, encrypting two correlated messages under the same key may enable an adversary to identify correct messages. (Encrypting independent messages under the same key is fine.) We however emphasize that, should any of these assumptions fail, HE security falls back to normal PBE security: there is never any harm in using HE.

2 Related Work

Our HE schemes (though not all HE schemes) provide a form of information-theoretic encryption, as their MR security does not rely on any computational hardness assumption. Information-theoretic encryption schemes, starting with the one-time pad [34], have seen extensive study. Most closely related is entropic security [18, 33], where the idea is to exploit high-entropy messages to perform encryption that leaks no predicate on the plaintext even against unbounded attackers (and hence beyond the brute-force bound). Their goal was to enable use of uniform, smaller (than one-time pads) keys yet achieve information-theoretic security. HE similarly exploits the entropy of messages, but also provides useful bounds (by targeting MR security) even when the combined entropy of messages and keys goes below the threshold needed to achieve entropic security. See also the discussion in Appendix A.

In natural applications of HE, the message space $\mathcal{M}$ must encompass messages of special format, rather than just bitstrings. In this sense, HE is somewhat related to format-preserving encryption (FPE) [2], although HE is randomized and has no preservation requirement (our ciphertexts are unstructured bit strings). An implication of our approach, however, is that some FPE constructions (e.g., for credit-card encryption) can be shown to achieve HE-like security guarantees when message distributions are uniform. HE is also conceptually related to collisionful hashing [7], the idea of creating password hashes for which it is relatively easy to find inverses and thus hard to identify the original, correct password (as opposed to identifying a correct message).

(Non-interactive) non-committing encryption [14, 29] are schemes for which a ciphertext can be "opened" to an arbitrary message by finding an appropriate key to do so. (For example, a one-time-pad is non-committing.) HE has no such requirement, rather HE schemes seek to ensure that decrypting a fixed ciphertext under different keys gives rise to independent-looking samples of the message space. Note that unlike non-committing encryption [29], HE is achievable in the non-programmable random oracle model. Deniable encryption [15] also allows opening a ciphertext to a chosen message; HE schemes do not in general provide deniability.

Canetti, Halevi, and Steiner [13] propose a protocol in which a password specifies a subset of CAPTCHAs that must be solved to decrypt a credential store. Their scheme creates ambiguity around where human effort can be most effectively invested, rather than around the correctness of the contents of the credential store, as HE would.

Perhaps most closely related to HE is a wealth of literature on deception and decoys in computer security. Honeypots, fake computer systems intended to attract and study attacks, are a stock-in-trade of computer security research [36]. Researchers have proposed honeytokens [17,37], which are data objects whose use signals a compromise, and honeywords [25], a system encompassing the use of passwords as honeytokens. Additional proposals include false documents [11], false network traffic [10], and many variants.

The Kamouflage system [8] is particularly relevant. It conceals a true password vault encrypted under a true master password among N bogus vaults encrypted under bogus master passwords. Kamouflage requires $O(N)$ storage. With a suitable DTE, HE offers the possibility of realizing similar functionality and security with $O(1)$ storage. Decoy deployment in Kamouflage and related systems requires the construction of plausible decoys. This problem has seen study specifically for protection of passwords in, e.g., [8, 25], but to the best of our knowledge, we are the first to formalize it with the concept of DTEs.

3 HE Overview

HE schemes. An HE scheme has syntax and semantics equivalent to that of a symmetric encryption scheme. Encryption maps a key and message to a ciphertext, and in our schemes encryption will be randomized. Decryption

recovers messages from ciphertexts. The departure from conventional symmetric encryption schemes will be in how HE decryption behaves when one uses the wrong key in attempting to decrypt a ciphertext. Instead of giving rise to some error, decryption will emit a plaintext that "looks" plausible.

Formally, let $\mathcal{K}$ and $\mathcal{M}$ be sets, the key space and message space. For generality, we assume that $\mathcal{K}$ consists of variable-length bit strings. (This supports, in particular, varying length passwords.) An HE scheme $\mathsf{HE} = (\mathsf{HEnc}, \mathsf{HDec})$ is a a pair of algorithms. Encryption HEnc takes input a key $K \in \mathcal{K}$, message $M \in \mathcal{M}$, some uniform random bits, and outputs a ciphertext C. We write this as $C \leftarrow\!\!\$\ \mathsf{HEnc}_K(M)$, where $\leftarrow\!\!\$$ denotes that HEnc may use some number of uniform random bits. Decryption HDec takes as input a key $K \in \mathcal{K}$, ciphertext C, and outputs a message $M \in \mathcal{M}$. Decryption, always deterministic, is written as $M \leftarrow \mathsf{HDec}_K(C)$.

We require that decryption works, formally that $\Pr[\mathsf{HDec}_K(\mathsf{HEnc}_K(M)) = M] = 1$ for all $K \in \mathcal{K}$, all $M \in \mathcal{M}$, and where the event is defined over the randomness used by HEnc.

We will write $\mathsf{SE} = (\mathsf{enc}, \mathsf{dec})$ to denote a conventional symmetric encryption scheme, but note that the syntax and semantics match those of an HE scheme.

Message and key distributions. We denote a distribution on set S by a map $p\colon S \to [0,1]$ and require that $\sum_{s\in S} p(s) = 1$. The min-entropy of a distribution is $-\log \max_{s\in S} p(s)$. Sampling according to such a distribution is written $s \leftarrow_p S$, and we assume all sampling is efficient. We use p_m to denote a message distribution over $\mathcal{M}$ and p_k for a key distribution over $\mathcal{K}$. Thus sampling according to these distributions is denoted $M \leftarrow_{p_m} \mathcal{M}$ and $K \leftarrow_{p_k} \mathcal{K}$. Note that we assume that draws from p_m and p_k are independent, which is not always the case but will be in our example applications; see Section 7. Whether HE schemes can provide security for any kind of dependent distributions is an interesting question for future work.

Message recovery security. To formalize our security goals, we use the notion of security against message recovery attacks. Normally, one aims that, given the encryption of a message, the probability of any adversary recovering the correct message is negligible. But this is only possible when both messages and keys have high entropy, and here we may have neither. Nevertheless, we can measure the message recovery advantage of any adversary concretely, and will do so to show (say) that attackers cannot achieve advantage better than $1/2^{\mu}$ where μ is the min-entropy of the key distribution p_k.

Formally, we define the MR security game as shown in Figure 1 and define advantage for an adversary $\mathcal{A}$ against a scheme HE by $\mathbf{Adv}^{\mathrm{mr}}_{\mathsf{HE},p_m,p_k}(\mathcal{A}) = \Pr[\mathrm{MR}^{\mathcal{A}}_{\mathsf{HE},p_m,p_k} \Rightarrow \mathsf{true}]$. When working in the random oracle (RO) model, the MR game additionally has a procedure implementing a random function that $\mathcal{A}$ may query. For our schemes, we allow $\mathcal{A}$ to run for an unbounded amount of time and make an unbounded number of queries to the RO. For simplicity we assume p_m and p_k are independent of the RO.

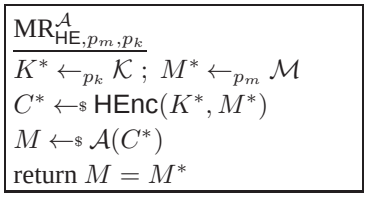

Figure 1: Game defining MR security.

Semantic security. In the case that keys are sufficiently unpredictable and adversaries are computationally bounded, our HE schemes will achieve semantic security [21]. Our schemes will therefore never provide worse confidentiality than conventional encryption, and in particular the MR advantage in this case equals the min-entropy of the message distribution p_m plus the (assumed) negligible semantic security term.

When combined with a suitable password-based key-derivation function [30], our schemes will also achieve the multi-instance security guarantees often desired for password-based encryption [4]. Note that the results in [4] still hold only for attackers that cannot exhaust the min-entropy of the key space.

In Appendix A we discuss why existing or naive approaches, such as conventional encryption or hiding the true plaintext in a list of fake ones, fail to be satisfactory HE schemes.

4 Distribution-Transforming Encoders

We introduce a new type of message encoding scheme that we refer to as a *distribution-transforming encoder* (DTE). Formally, it is a pair $\mathsf{DTE} = (\mathsf{encode}, \mathsf{decode})$ of algorithms. The usually randomized algorithm encode takes as

input a message $M \in \mathcal{M}$ and outputs a value in a set $\mathcal{S}$. We call the range $\mathcal{S}$ the *seed space* for reasons that will become clear in a moment. The deterministic algorithm decode takes as input a value $S \in \mathcal{S}$ and outputs a message $M \in \mathcal{M}$. We call a DTE scheme *correct* if for any $M \in \mathcal{M}$, $\Pr[\mathsf{decode}(\mathsf{encode}(M)) = M] = 1$.

A DTE encodes a priori knowledge of the message distribution p_m. One goal in constructing a DTE is that decode applied to uniform points provides sampling close to that of a target distribution p_m. For a given DTE (that will later always be clear from context), we define p_d to be the distribution over $\mathcal{M}$ defined by $p_d(M) = \Pr[\, M' = M \;:\; U \leftarrow\!\!\$\, \mathcal{S}\;;\; M' \leftarrow \mathsf{decode}(S)\,]$. We will often refer to p_d as the DTE distribution. Intuitively, in a good or secure DTE, the distributions p_m and p_d are "close."

Formally, we define this notion of DTE security or goodness, as follows. Let $\mathcal{A}$ be an adversary attempting to distinguish between the two games shown in Figure 2. We define advantage of an adversary $\mathcal{A}$ for a message distribution p_m and encoding scheme $\mathsf{DTE} = (\mathsf{encode}, \mathsf{decode})$ by

$$\mathbf{Adv}^{\mathrm{dte}}_{\mathsf{DTE},p_m}(\mathcal{A}) = \left|\Pr\left[\,\mathrm{SAMP1}^{\mathcal{A}}_{\mathsf{DTE},p_m} \Rightarrow 1\,\right] - \Pr\left[\,\mathrm{SAMP0}^{\mathcal{A}}_{\mathsf{DTE}} \Rightarrow 1\,\right]\right| .$$

While we focus mostly on adversaries with unbounded running times, we note that these measures can capture computationally-good DTEs as well. A perfectly secure DTE is a scheme for which the indistinguishability advantage is zero for even unbounded adversaries. In Appendix C we explore another way of measuring DTE goodness that, while more complex, sometimes provides slightly better bounds.

The inverse sampling DTE. We first build a broadly useful DTE based on inverse sampling, the well-known technique for converting uniform random variables into ones from some other distribution. Let F_m be the cumulative distribution function (CDF) associated with a known message distribution p_m according to some ordering of $\mathcal{M} = \{M_1, \ldots, M_{|\mathcal{M}|}\}$. Define $F_m(M_0) = 0$. Let the seed space be $\mathcal{S} = [0, 1)$. Inverse sampling picks a value according to p_m by selecting $S \leftarrow\!\!\$\, [0, 1)$ and outputs M_i such that $F_m(M_{i-1}) \le S < F_m(M_i)$; this amounts to computing the inverse CDF function $M = F_m^{-1}(S)$. The associated DTE scheme $\mathsf{IS\text{-}DTE} = (\mathsf{is\text{-}encode}, \mathsf{is\text{-}decode})$ encodes by picking uniformly from the range $[F_m(M_{i-1}), F_m(M_i))$ for input message M_i, and decodes by computing $F_m^{-1}(S)$.

$\mathrm{SAMP1}^{\mathcal{B}}_{\mathsf{DTE},p_m}$	$\mathrm{SAMP0}^{\mathcal{B}}_{\mathsf{DTE}}$
$M^* \leftarrow_{p_m} \mathcal{M}$	$S^* \leftarrow\!\!\$\, \mathcal{S}$
$S^* \leftarrow\!\!\$\, \mathsf{encode}(M^*)$	$M^* \leftarrow \mathsf{decode}(S^*)$
$b \leftarrow\!\!\$\, \mathcal{B}(S^*, M^*)$	$b \leftarrow\!\!\$\, \mathcal{B}(S^*, M^*)$
return b	return b

Figure 2: Games defining DTE goodness.

All that remains is to fix a suitably granular representation of the reals between $[0, 1)$. The representation error gives an upper bound on the DTE security of the scheme. We defer the details and analysis to Appendix B. The scheme works in time $\mathcal{O}(\log |\mathcal{M}|)$ using a tables of size $\mathcal{O}(|\mathcal{M}|)$, though its performance can easily be improved for many special cases (e.g., uniform distributions).

DTEs for RSA secret keys. We turn to building a DTE for RSA secret keys. Recall that a popular key generation algorithm (used in, e.g., OpenSSL) generates an RSA key of bit-length 2ℓ via rejection sampling of random values $p, q \in [2^{\ell-1}, 2^{\ell})$. The rejection criterion for either p or q is failure of a Miller-Rabin primality test [28, 31]; the resulting distribution of primes is (essentially) uniform over the range. The private exponent is computed as $d = e^{-1} \bmod (p-1)(q-1)$ for some fixed e (typically 65537), yielding secret key (N, d) and public key (N, e). Most often, however, the key as stored includes p, q, and some ancillary values (not efficiently recoverable from d) to speed up exponentiation via the Chinese Remainder Theorem. Since for fixed e, the secret key is fully defined by p, q, we now focus on building DTEs that take as input primes $p, q \in [2^{\ell-1}, 2^{\ell})$ for some ℓ and aim to match the message distribution p_m that is uniformly distributed over the primes in $[2^{\ell-1}, 2^{\ell})$.

One strawman approach is just to encode the input p, q as a pair of $(\ell - 2)$-bit strings (the leading '1' bit left implicit), but this gives a poor DTE. The prime number theorem indicates that an ℓ-bit integer will be prime with probability about $1/\ell$; thus an adversary $\mathcal{A}$ that applies primality tests to a candidate plaintext has a (very high) DTE advantage of about $1 - 1/\ell^2$.

We can instead adapt the rejection-sampling approach to prime generation itself as a DTE, $\mathsf{RSA\text{-}REJ\text{-}DTE} = (\mathsf{rsa\text{-}rej\text{-}encode}, \mathsf{rsa\text{-}rej\text{-}decode})$, which works as follows. Encoding $(\mathsf{rsa\text{-}rej\text{-}encode})$ takes a pair of primes (p, q), constructs a vector of t bitstrings selected uniformly at random from the range $[2^{\ell-1}, 2^{\ell})$, replaces the first (resp. sec-

ond) prime integer in the list by p (resp. q), and outputs the modified vector of t integers (each encoded using $\ell - 2$ bits). (If there's one prime and it's not the last integer in the vector, then that prime is replaced by p and the last integer is replaced by q. Should there be no primes in the vector, or one prime in the last position, then the last two integers in the vector are replaced by (p, q).) Decoding (rsa-rej-decode) takes as input a vector of the t integers, and outputs its first two primes. If there do not exist two primes, then it outputs some (hard-coded) fixed primes.[1] For simplicity, we assume a perfect primality testing algorithm; it is not hard to generalize to probabilistic ones.[2] We obtain the following security bound.

Theorem 1 *Let p_m be uniform over primes in $[2^{\ell-1}, 2^\ell)$ for some $\ell \geq 2$ and RSA-REJ-DTE be the scheme described above. Then $\mathbf{Adv}^{\mathrm{dte}}_{\mathsf{RSA\text{-}REJ\text{-}DTE},p_m}(\mathcal{A}) \leq (1 - 1/(3\ell))^{t-1}$ for any adversary $\mathcal{A}$.*

Proof: Let $\pi(x)$ be the number of primes less than or equal to x. Then Bertrand's postulate (c.f., [35]) states that $\pi(2^\ell) - \pi(2^{\ell-1}) > \frac{2^{\ell-1}}{3\ell}$ for $\ell > 2$. Thus the probability of each sample from $\{0,1\}^\ell$ being a prime is at least $1/3\ell$. One can verify that the $\mathrm{SAMP1}_{\mathsf{RSA\text{-}REJ\text{-}DTE},p_m}$ and $\mathrm{SAMP0}_{\mathsf{RSA\text{-}REJ\text{-}DTE},p_m}$ have identical distributions assuming at least two primes are chosen amongst the t. A standard argument gives that the advantage is bounded by $(1 - 1/(3\ell))^{t-1}$. ∎

This scheme is simple, but a small adversarial advantage does translate into a large encoding. For example with $\ell = 1024$ (2048-bit RSA), to achieve $\mathbf{Adv}^{\mathrm{dte}}_{\mathsf{RSA\text{-}REJ\text{-}DTE},p_m}(\mathcal{A}) < 10^{-5}$ requires $t \geq 35{,}361$, resulting in an encoding of about 4.5 megabytes. (Assuming keys of low entropy, 10^{-5} is small enough to contribute insignificantly to security bounds on the order of those in Section 7.) It may be tempting to try to save on space by treating S as a seed for a pseudorandom generator (PRG) that is then used to generate the t values during decoding. Encoding, though, would then need to identify seed values that map to particular messages (prime pairs), effectively inverting the PRG, which is infeasible.

Some RSA key generators do not use rejection-sampling, but instead use the classic algorithm that picks a random integer in $[2^{\ell-1}, 2^\ell)$ and increments it by two until a prime is found (c.f., [12, 22]). In this case, a DTE can be constructed (see Appendix D for details) that requires only $2(\ell - 2)$-bit seeds, and so is space-optimal. Other, more randomness-efficient rejection-sampling techniques [20] may also be used to obtain smaller encodings.

In some special settings it may be possible to hook existing key-generation software, extract the PRG key / seed κ used for the initial generation of an RSA key pair, and apply HE directly to κ. A good DTE (and thus HE scheme) can then be constructed trivially, as κ is just a short (e.g., 256-bit) uniformly random bitstring.

5 DTE-then-Encrypt Constructions

We now present a general construction for HE schemes for a target distribution p_m. Intuitively, the goal of any HE scheme is to ensure that the plaintext resulting from decrypting a ciphertext string under a key is indistinguishable from freshly sampling a plaintext according to p_m. Let $\mathsf{DTE} = (\mathsf{encode}, \mathsf{decode})$ be a DTE scheme whose outputs are in the space $\mathcal{S} = \{0,1\}^s$. Let $\mathsf{SE} = (\mathsf{enc}, \mathsf{dec})$ be a conventional symmetric encryption scheme with message space $\mathcal{S}$ and some ciphertext space $\mathcal{C}$.

Then DTE-then-Encrypt $\mathsf{HE}[\mathsf{DTE}, \mathsf{SE}] = (\mathsf{HEnc}, \mathsf{HDec})$ applies the DTE encoding first, and then performs encryption under the key. Decryption works in the natural way. It is easy to see that the resulting scheme is secure in the sense of semantic security (when keys are drawn from a large enough space) should SE enjoy the same property.

We fix a simple instantiation using a hash function $H : \{0,1\}^n \times \mathcal{K} \to \mathcal{S}$ to perform

$\mathsf{HEnc}^H(K, M)$
$S \leftarrow\!\!\$\ \mathsf{encode}(M)$
$R \leftarrow\!\!\$\ \{0,1\}^n$
$C_2 \leftarrow\!\!\$\ H(R, K) \oplus S$
return (R, C_2)

$\mathsf{HDec}^H(K, (R, C_2))$
$S \leftarrow C_2 \oplus H(R, K)$
$M \leftarrow \mathsf{decode}(S)$
return M

Figure 3: A particularly simple instantiation of DTE-then-Encrypt using a hash-function H to implement the symmetric encryption.

[1] We could also output bottom, but this would require allowing errors in decoding and HE decryption.

[2] Doing so would also require our definition of DTE correctness to allow errors.

symmetric encryption, see Figure 3. It is denoted as $\mathsf{HE}[\mathsf{DTE}, H]$. Of course, one should apply a password-based key-derivation function to K first, as per [30]; we omit this for simplicity.

To analyze security, we use the following approach. First we establish a general theorem (Theorem 2) that uses the goodness of the DTE scheme to move to a setting where, intuitively, the attacker's best bet is to always output the message M which maximizes the probability (over choice of key) of M being the result of decrypting a random challenge ciphertext. The attacker wins, then, with exactly the sum of the probabilities of the keys that map the ciphertext to that message. Second, we define a weighted balls-and-bins game with non-uniform bin sizes in a way that makes the expected load of the maximally loaded bin at the end of the game exactly the winning probability of the attacker. We can then analyze these balls-and-bins games for various message and key distributions combinations (in the random oracle model). We put all of this together to derive bounds for some concrete applications in Section 7, but emphasize that the results here provide a general framework for analyzing HE constructions that applies more broadly.

Applying DTE goodness. Let $\mathcal{K}_{M,C} = \{K \ : \ K \in \mathcal{K} \wedge M = \mathsf{HDec}(K, C)\}$ be the set of keys that decrypt a specific ciphertext to a specific message and (overloading notation slightly) let $p_k(\mathcal{K}_{M,C}) = \sum_{K \in \mathcal{K}_{M,C}} p_k(K)$ be the aggregate probability of selecting a key that falls in any such set. Then for any $C \in \mathcal{C}$ we define $L_{\mathsf{HE},p_k}(C) = \max_M p_k(\mathcal{K}_{M,C})$. Let L_{HE,p_k} represent the random variable $L_{\mathsf{HE},p_k}(C)$ defined over C uniformly chosen from $\mathcal{C}$ and any coins used to define HDec. (For example in the hash-based scheme, we take this over the coins used to define H when modeled as a random oracle.) We will later show, for specific message/key distributions and using balls-and-bins-style arguments, bounds on $\mathrm{E}\left[\, L_{\mathsf{HE},p_k} \,\right]$. We call this value the expected maximum load, following the terminology from the balls-and-bins literature.

For the following theorem we require from SE only that encrypting uniform messages gives uniform ciphertexts. More precisely, that $S \leftarrow\!\!\$\ \mathcal{S}\ ;\ C \leftarrow\!\!\$\ \mathsf{enc}(K, S)$ and $C \leftarrow\!\!\$\ \mathcal{C}\ ;\ S \leftarrow \mathsf{dec}(K, C)$ define identical distributions for any key $K \in \mathcal{K}$. This is true for many conventional schemes, including the hash-based scheme used in Figure 3, CTR mode over a block-cipher, and CBC-mode over a block cipher (assuming the DTE is designed so that $\mathcal{S}$ includes only bit strings of length a multiple of the block size). The proof of the following theorem is given in Appendix G.

Theorem 2 *Fix distributions p_m, p_k, an encoding scheme DTE for p_m, and a symmetric encryption scheme $\mathsf{SE} = (\mathsf{enc}, \mathsf{dec})$. Let $\mathcal{A}$ be an MR adversary against $\mathsf{HE}[\mathsf{DTE}, \mathsf{SE}]$. Then we give a specific adversary $\mathcal{B}$ in the proof such that $\mathbf{Adv}^{\mathrm{mr}}_{\mathsf{HE},p_m,p_k}(\mathcal{A}) \ \leq \ \mathbf{Adv}^{\mathrm{dte}}_{\mathsf{DTE},p_m}(\mathcal{B}) + \mathrm{E}\left[\, L_{\mathsf{HE},p_k} \,\right]$. Adversary $\mathcal{B}$ runs in time that of $\mathcal{A}$ plus the time of one enc operation.*

The balls-and-bins interpretation. What remains is to bound $\mathrm{E}\left[\, L_{\mathsf{HE},p_k} \,\right]$. To do so, we use the following equivalent description of the probability space as a type of balls-and-bins game. Uniformly pick a ciphertext $C \leftarrow\!\!\$\ \mathcal{C}$. Each ball represents one key K and has weight equal to $p_k(K)$. We let $a = |\mathcal{K}|$ be the number of balls. Each bin represents a message M and $b = |\mathcal{M}|$ is the number of bins.[3] A ball is placed in a particular bin should C decrypt under K to the message labeling that bin. Then L_{HE,p_k} as defined above is exactly the random variable defined as the maximum, over bins, sum of weights of all balls thrown into that bin. In this balls-and-bins game the balls are weighted, the bins have varying capacities, and the (in)dependence of ball throws depends on the details of the symmetric encryption scheme used.

To derive bounds, then, we must analyze the expected maximum load for various balls-and-bins games. For brevity in the following sections we focus on the hash-based HE scheme shown in Figure 3. By modeling H as a random oracle,[4] we get that all the ball throws are independent. At this stage we can also abstract away the details of the DTE, instead focusing on the distribution p_d defined over $\mathcal{M}$. The balls-and-bins game is now completely characterized by p_k and p_d, and we define the random variable L_{p_k,p_d} as the load of the maximally loaded bin at the end of the balls-and-bins game that throws $|\mathcal{K}|$ balls with weights described by p_k independently into $|\mathcal{M}|$ bins, choosing a bin according to p_d. The following lemma formalizes this transition.

[3] Convention is to have m balls and n bins, but we use a balls and b bins to avoid confusion since m connotes messages.

[4] Technically speaking we only require the non-programmable random oracle [19, 29].

Lemma 1 *Consider $\mathsf{HE}[\mathsf{DTE}, H]$ for H modeled as a RO and DTE having distribution p_d. For any key distribution p_k,* $\mathrm{E}[L_{\mathsf{HE},p_k}] \leq \mathrm{E}[L_{p_k,p_d}]$.

We give similar lemmas for block-cipher based modes (in the ideal cipher model) in Appendix E. Thus we can interchange the hash-based symmetric encryption scheme for other ones in the final results of Section 7 with essentially the same security bounds.

6 Balls-and-Bins Analyses

In this section we will derive bounds for various special cases of the balls-and-bins games , these cases motivated and used by the example applications of HE given in the next section. These cases are by no means exhaustive, and rather we strive only to illustrate the power of our general HE analysis framework. Treating p_k and p_d as vectors, we can write their dimension as $|p_k| = a$ and $|p_d| = b$.

In the special case of $a = b$ and both p_k and p_d uniform, the balls-and-bins game becomes the standard one. One can use the classic proof to show that $\mathrm{E}[\, L_{p_k,p_d}\,] \leq \frac{1}{b} + \frac{3 \ln b}{b \ln \ln b}$. HE schemes for real applications, however, are unlikely to coincide with this special case, and so we seek other bounds.

Majorization. To analyze more general settings, we exploit a result due to Berenrink, Friedetzky, Hu, and Martin [5] that builds on a technique called "majorization" earlier used for the balls-and-bins setting by Azar, Broder, Karlin, and Upfal [1].

Distributions such as p_k and p_d can be viewed as vectors of appropriate dimension over $\mathbb{R}$. In the below we assume that vectors have their components in decreasing order, e.g. that $p_k(i) \geq p_k(j)$ for $i < j$. Let m be a number and $p_k, p'_k \in \mathbb{R}^a$. Then p'_k *majorizes* p_k, denoted $p'_k \succ p_k$, if $\sum_{i=1}^{a} p'_k[i] = \sum_{i=1}^{a} p_k[i]$ and $\sum_{i=1}^{j} p'_k[i] \geq \sum_{i=1}^{j} p_k[i]$ for all $1 \leq j \leq a$.

Majorization intuitively states that p'_k is more "concentrated" than p_k: a prefix of any length of p'_k has cumulative weight at least as large as the cumulative weight of the same-length prefix of p_k. We have the following theorem from [5, Cor. 3.5], slightly recast to use our terminology. We also extend our definition of load to include the i highest loaded bins: let $L^i_{p_k,p_d}$ be the random variable which is the total weight in the i highest-loaded bins at the end of the balls-and-bins game.

Theorem 3 (BFHM08) *Let p_k, p'_k, p_d be distributions. If $p'_k \succ p_k$, then $\mathrm{E}[L^i_{p'_k,p_d}] \geq \mathrm{E}[L^i_{p_k,p_d}]$ for all $i \in [1, b]$.*

Consider the case that $i = 1$, which corresponds to the expected maximum bin loads for the two key distributions. As a concrete example, let $p_k = (1/2, 1/4, 1/4)$, $p'_k = (1/2, 1/2, 0)$. Then $p'_k \succ p_k$ and thus $\mathrm{E}[\, L(p'_k, p_d)\,] \geq \mathrm{E}[\, L(p_k, p_d)\,]$ because the "fusion" of the two 1/4-weight balls into one ball biases the expected maximum load upwards.

Our results will use majorization to shift from a setting with non-uniform key distribution p_k having max-weight w to a with uniform key distribution over $\lceil 1/w \rceil$.

Non-uniform key distributions. We turn now to giving a bound for the case that p_k has maximum weight w (meaning $p_k(M) \leq w$ for all M) and p_d is uniform. In our examples in the next section we have that $a \ll b$, and so we focus on results for this case. We start with the following lemma (whose proof is given in Appendix G).

Lemma 2 *Suppose p_k has maximum weight w and p_d is such that $b = ca$ for some positive integer c. Then for any positive integer $s > 2e/c$, where e is Euler's constant, it holds that*

$$\mathrm{E}[\, L_{p_k,p_d}\,] \leq w\left((s-1) + 2\left(\frac{a^2}{c^{s-1}}\right)\left(\frac{e}{s}\right)^s\right) .$$

For cases in which $b = \mathcal{O}(a^2)$, a convenient, somewhat tighter bound on $\mathrm{E}[\, L_{p_k,p_d}\,]$ is possible. We observe that in many cases of interest, the term $r(c, b)$ in the bound below will be negligible. Proof of this next lemma is given in Appendix G.

Lemma 3 *Suppose p_k has maximum weight w and p_d is such that $b = ca^2$ for some positive integer c. Then $\mathrm{E}[\, L_{p_k,p_d}\,] \leq w\left[1 + \frac{1}{2c} + r(c,b)\right]$, where e is Euler's constant and $r(c,b) = \left(\frac{e}{27c^2}\right)\left(1 - \frac{e}{cb}\right)^{-1}$.*

Non-uniform balls-and-bins. As a final analysis supportive of our examples in the next section, we must analyze settings in which p_d is non-uniform. The proof of this lemma is given in Appendix G.

Lemma 4 *Let $L_{\mathcal{B}}$ denote the maximum load yielded by throwing a balls (of weight 1) into a set $\mathcal{B}$ of b bins of non-uniform capacity at most $0 \leq \gamma \leq 3 - \sqrt{5}$. Let $L_{\mathcal{B}^*}$ denote the maximum load yielded by throwing $a^* = 3a$ balls (of weight 1) into a set $\mathcal{B}^*$ of $b^* = \lfloor 2/\gamma \rfloor$ bins of uniform capacity. Then $\mathrm{E}[L_{\mathcal{B}}] \leq \mathrm{E}[L_{\mathcal{B}^*}]$.*

7 Example Applications, Bounds, and Deployment Considerations

We now draw all of the results of the previous sections together in several concrete examples. We investigate in particular honey encryption of RSA secret keys and of credit card data. Throughout this section, for concreteness, we use password-based encryption of these secrets, though the proven results are much more general. Appealing again to Bonneau's Yahoo! study [9] in which the most common password was selected by $1.08\% \approx 1/100$ of users, we assume for simplicity that the maximum-weight password / key is selected with probability $w = 1/100$. Note that at this level of entropy, prior security results for PBE schemes are not very useful.

7.1 HE for Credit Card Numbers, PINs, and CVVs

We first consider application of HE to credit card numbers. For convenience, we evaluate HE as applied to a single value, e.g., an individual credit-card number. Recall, though, that HE security is unaffected by simultaneous encryption of multiple, independent messages drawn from the same distribution. So our security bounds would in principle apply equally well to encryption of a vault or repository of multiple credit-card numbers.

A (Mastercard or Visa) credit card number, known technically as a Primary Account Number (PAN), consists of sixteen decimal digits. Although structures vary somewhat, commonly nine digits constitute the cardholder's account number, and may be regarded as selected uniformly at random upon issuance. One digit is a (mod 10) checksum (known as the Luhn formula). A useful result then is the following theorem, whose proof is given in Appendix H.

Theorem 4 *Consider $\mathsf{HE}[\mathsf{IS\text{-}DTE}, H]$ with H modeled as a RO and $\mathsf{IS\text{-}DTE}$ using an ℓ-bit representation. Let p_m be a uniform distribution over b messages and p_k be a key-distribution with maximum weight w. Let $\alpha = \lceil 1/w \rceil$. Then for any adversary $\mathcal{A}$, $\mathbf{Adv}^{\mathrm{mr}}_{\mathsf{HE},p_m,p_k}(\mathcal{A}) \leq w(1+\delta) + \dfrac{1+\alpha}{2^\ell}$ where $\delta = \frac{\alpha^2}{2b} + \frac{e\alpha^4}{27b^2}\left(1 - \frac{e\alpha^2}{b^2}\right)^{-1}$.*

For many cases of interest, $b \gg \alpha^2$, and thus δ will be small. We can also set ℓ appropriately to make $(1+\alpha)/2^\ell$ negligible. Theorem 4 then yields a simple and useful bound, as for our next two examples.

As cardholder account numbers are uniformly selected nine-digit values, they induce a uniform distribution over a space of $b = 10^9$ messages. Given $w = 1/100$, then, $\alpha^2/b = 10^{-5}$ and so $\delta \approx 0$. The upper bound on MR advantage is $w = 1/100$. Note that this bound is essentially tight, as there exists an adversary $\mathcal{A}$ achieving advantage $w = \frac{1}{100}$. Namely, the adversary that decrypts the challenge ciphertext with the most probable key and then outputs the resulting message. This adversary has advantage at least w.

Often the last four digits of a credit-card number are treated as semi-public information. It is common, for example, for receipts and web sites to display them. Another interesting bound to consider, therefore, is the security of the previous HE scheme here assuming adversarial knowledge of these digits. Three digits form part of the customer account number and one is a checkdigit. Thus, the effective message space is reduced in this scenario to five digits, i.e., $b = 10^5$. Thus $\alpha^2/b = 1/10$ and Theorem 4 yields a message recovery bound of about 1.05%.

Finally, consider encrypting both 5-digits of the credit-card / debit-card account number (the last 4 digits still considered public) along with the user's PIN number. (Credit card PINs are used for cash withdrawals and to authorize debit-card transactions.) A detailed examination of a corpus of 3.4 million user-selected PINs is given in [6], and gives in particular a CDF that can be used to define an inverse sampling DTE. The most common user-selected PIN

is '1234'; it has an observed frequency of 10.713%. Thus, PINs have very little minimum entropy (roughly 3 bits). Combining a PIN with a five-digit effective account number induces a *non-uniform* message space, with maximum message probability $\gamma = 1.0713 \times 10^{-6}$. Consequently, Theorem 4 is not applicable to this example.

A variant of the proof of Theorem 4, however, that makes use of Lemma 4 for non-uniform bin sizes, establishes the following corollary.

Corollary 1 *Consider $\mathsf{HE}[\mathsf{IS\text{-}DTE}, H]$ with H modeled as a RO and IS-DTE using an ℓ-bit representation. Let p_m be a non-uniform distribution with maximum message probability $\gamma \leq 3 - \sqrt{5}$, and p_k be a key-distribution with maximum weight w. Let $\alpha = \lceil 1/w \rceil$. Then for any adversary $\mathcal{A}$, $\mathbf{Adv}^{\mathrm{mr}}_{\mathsf{HE},p_m,p_k}(\mathcal{A}) \leq w(1+\delta) + \dfrac{(1+\alpha)}{2^\ell}$ where $\delta = \frac{\overline{\alpha}^2}{2\overline{b}} + \frac{e\overline{\alpha}^4}{27\overline{b}^2}\left(1 - \frac{e\overline{\alpha}^2}{\overline{b}^2}\right)^{-1}$ and $\overline{\alpha} = \lceil 3/w \rceil$ and $\overline{b} = \lfloor 2/\gamma \rfloor$.*

Corollary 1 yields a bound defined by the expected maximum load of a balls-and-bins experiment with 300 balls (of weight $w = 1/100$) and $\lfloor 2/\gamma \rfloor$ = 1,866,890 uniform-capacity bins, with $c = \overline{\alpha}^2/\overline{b} = 1/20.74$. The final MR bound is therefore about 1.02%. This is slightly better than the bound of the previous example (at 1.05%). It shows, significantly, that Corollary 1 is tight enough to give improved bounds despite the scant minimum entropy in a PIN.

Credit cards often have an associated three- or four-digit *card verification value*, a secret used to conduct transactions. As a final case we investigate encrypting a three-digit, uniformly random CVV under a password. Here $\alpha = 100$ and $b = 1000$, which means that $\alpha^2/b = 10$. Applying Theorem 4 yields a loose bound of about 16.35%. For a tighter bound, we offer the following corollary, a variant of Theorem 4 whose proof makes use of Lemma 2:

Corollary 2 *Consider $\mathsf{HE}[\mathsf{IS\text{-}DTE}, H]$ with H modeled as a RO and IS-DTE using an ℓ-bit representation. Let p_m be a uniform distribution over b messages and let p_k be a key-distribution with maximum weight w. Let $\alpha = \lceil 1/w \rceil$ and $c = b/\alpha$. Then for any positive integer $s > 2e\alpha/b$, where e is Euler's constant, and for any adversary $\mathcal{A}$, it holds that $\mathbf{Adv}^{\mathrm{mr}}_{\mathsf{HE},p_m,p_k}(\mathcal{A}) \leq w\left((s-1) + 2\left(\dfrac{\alpha^2}{c^{s-1}}\right)\left(\dfrac{e}{s}\right)^s\right) + (1+\alpha)/2^\ell$.*

Application of Corollary 2 to our CVV example here with $c = 10$ and $s = 5$ yields the considerably improved bound of approximately 4.094%.

In cases with relatively small a and b, simulation yields a considerably better estimate of expected maximum loads than some of our upper bounds suggest. For the example of CVV encryption, a simulation over 100,000 runs yields a mean expected maximum load of 2.14% (mean number of balls = 2.14, min = 1, max = 5, std. dev. = 0.372), which makes our analytical upper bound of 4.094% appear to be loose. Future work might therefore seek improved bounds.

7.2 HE for RSA Secret Keys

We now show how to apply HE to RSA secret keys using the DTE introduced for this purpose in Section 4.

In some settings, RSA is used without making a user's public key readily available to attackers. A common example is RSA-based client authentication to authorize access to a remote service using HTTPS or SSH. The client stores an RSA secret / private key and registers the corresponding public key with the remote service.

Practitioners recommend encrypting the client's secret key under a password, as this provides defense-in-depth should the client's system be passively compromised.[5] Using traditional password-based encryption, though, means that an attacker can mount an offline brute-force attack against the encrypted secret key. Use of straightforward unauthenticated encryption wouldn't help here: as the secret key is usually stored as a pair of constituent primes p and q (to facilitate use of the Chinese Remainder Theorem), an attacker can quickly test the correctness of a candidate secret key by applying a primality test to its factors. Similarly, given the kinds of passwords used in practice (e.g., for $w = 1/100$), key-hardening mechanisms (e.g., iterative hashing) do not provide an effective slowdown against brute-force attack. Cracking a password-encrypted RSA secret key remains fairly easy.

HE is an attractive option in this setting. To build an HE scheme for 2ℓ-bit RSA secret keys we can use the DTE from Section 4. We have the following theorem.

[5] Obviously an active attacker can sniff the keyboard or otherwise capture the secret key. We also are ignoring the role of network attackers that may also gain access to transcripts dependent on the true secret key. See [24] for discussion.

Theorem 5 *Consider* $\mathsf{HE}[\mathsf{RSA\text{-}REJ\text{-}DTE}, H]$ *with* H *modeled as a RO and* RSA-REJ-DTE *the* 2ℓ*-bit RSA DTE using seed space vectors of size* t*. Let* p_m *be uniform over primes in* $[2^{-\ell-1}, 2^{\ell})$ *and let* p_k *be a key-distribution with maximum weight* w*. Let* $\alpha = \lceil 1/w \rceil$*. Then for any adversary* $\mathcal{A}$ *it holds that*

$$\mathbf{Adv}^{\mathrm{mr}}_{\mathsf{HE},p_m,p_k}(\mathcal{A}) \leq w(1+\delta) + (1+\alpha)\left(1 - \frac{1}{3\ell}\right)^{t-1}$$

where $\delta = \frac{\alpha^2}{2\lceil 2^{\ell-1}/\ell \rceil} + \left(\frac{e\alpha^4}{27\lceil 2^{\ell-1}/\ell \rceil^2}\right) \cdot \left(1 - \frac{e\alpha^2}{\lceil 2^{\ell-1}/\ell \rceil^2}\right)^{-1}$.

The proof is much like that of Theorem 4 (Appendix H): apply Theorem 2; plug in the advantage upper bound for the RSA rejection sampling DTE (Theorem 1); apply Lemma 1 to get independent ball tosses; majorize to get uniform-weighted balls (Theorem 3); apply a union bound to move from p_d back to uniform bin selection; and then finally apply the balls-and-bins analysis for uniform bins (Lemma 3).

The term δ is small when $-\log w \ll \ell$. For example, with $\ell = 1024$ and $w = 1/100$ and setting $t = 35{,}393$, we have that $\delta \approx 0$ and the overall MR advantage is upper bounded by 1.1%. The ciphertext size will still be somewhat large, at about 4.5 megabytes; one might use instead the DTEs discussed in Appendix D for which similar MR bounds can be derived yet ciphertext size ends up short.

7.3 Deployment considerations

A number of considerations and design options arise in the implementation and use of HE. To give some flavor of such issues, we briefly mention a couple involving the use of checksums.

Typo-safety. Decryption of an HE ciphertext C^* under an incorrect password / key K yields a fake but valid-looking message M. This is good for security, but can be bad for usability if a fake plaintext appears valid to a legitimate user.

One possible remedy, proposed in [25], is the use of error-detecting codes or checksums, such as those for ISBN book codes. For example, a checksum on the password / key K^* might be stored with the ciphertext C^*. Such checksums would reduce the size of the key space $\mathcal{K}$ and cause some security degradation, and thus require careful construction and application. Another option in some cases is online verification of plaintexts. For example, if a credit-card number is rejected by an online service after decryption, the user might be prompted to re-enter her password.

Honeytokens without explicit sharing. In [8], it is suggested that fake passwords / honeytokens be shared explicitly between password vault applications and service providers. Application of error-correcting codes to plaintexts in HE can create *honeytokens without explicit sharing*. As a naïve example (and crude error-correcting code), an HE scheme for credit-card numbers might explicitly store the first two digits of the credit-card account number. If a service provider then receives an invalid credit-card number in which these digits are correct, it gains evidence of a decryption attempt on the HE ciphertext by an adversary. This approach degrades security slightly by reducing the message space, and must be applied with care. But it offers an interesting way of coupling HE security with online security checks.

8 Conclusion

Low-entropy secrets such as passwords are likely to persist in computer systems for many years. Their use in encryption leaves resources vulnerable to offline attack. Honey encryption can offer valuable additional protection in such scenarios. HE yields plausible looking plaintexts under decryption with invalid keys (passwords), so that offline decryption attempts alone are insufficient to discover the correct plaintext. HE also offers a gracefully degrading hedge against partial disclosure of high min-entropy keys, and, by simultaneously meeting standard PBE security notions should keys be high entropy, HE never provides worse security than existing PBE schemes.

We showed applications in which HE security upper bounds are equal to an adversary's conditional knowledge of the key distribution, i.e., they min-entropy of keys. These settings have message space entropy greater than the entropy of keys, but our framework can also be used to analyze other settings.

A key challenge for HE—as with all schemes involving decoys—is the generation of plausible honey messages through good DTE construction. We have described good DTEs for several natural problems. For the case where

plaintexts consist of passwords, e.g., password vaults, the relationship between password-cracking and DTE construction mentioned above deserves further exploration. DTEs offer an intriguing way of potentially repurposing improvements in cracking technology to achieve improvements in encryption security by way of HE.

More generally, for human-generated messages (password vaults, e-mail, etc.), estimation of message distributions via DTEs is interesting as a natural language processing problem. Similarly, the reduction of security bounds in HE to the expected maximum load for balls-and-bins problems offers an interesting connection with combinatorics. The concrete bounds we present can undoubtedly be tightened for a variety of cases. Finally, a natural question to pursue is what kinds of HE bounds can be realized in the standard model via, e.g., k-wise independent hashing.

References

[1] Yossi Azar, Andrei Z Broder, Anna R Karlin, and Eli Upfal. Balanced allocations. *SIAM journal on computing*, 29(1):180–200, 1999.

[2] M. Bellare, T. Ristenpart, P. Rogaway, and T. Stegers. Format-preserving encryption. In *Selected Areas in Cryptography*, pages 295–312, 2009.

[3] Mihir Bellare and Chanathip Namprempre. Authenticated encryption: Relations among notions and analysis of the generic composition paradigm. In *Advances in CryptologyASIACRYPT 2000*, pages 531–545. Springer, 2000.

[4] Mihir Bellare, Thomas Ristenpart, and Stefano Tessaro. Multi-instance security and its application to password-based cryptography. In *Advances in Cryptology–CRYPTO 2012*, pages 312–329. Springer Berlin Heidelberg, 2012.

[5] Petra Berenbrink, Tom Friedetzky, Zengjian Hu, and Russell Martin. On weighted balls-into-bins games. *Theoretical Computer Science*, 409(3):511 – 520, 2008.

[6] N. Berry. PIN analysis. DataGenetics blog, 2012.

[7] T. A. Berson, L. Gong, and T.M.A. Lomas. Secure, keyed, and collisionful hash functions. Technical Report SRI-CSL-94-08, SRI International Laboratory, 1993 (revised 2 Sept. 1994).

[8] H. Bojinov, E. Bursztein, X. Boyen, and D. Boneh. Kamouflage: loss-resistant password management. In *ESORICS*, pages 286–302, 2010.

[9] Joseph Bonneau. The science of guessing: analyzing an anonymized corpus of 70 million passwords. In *Security and Privacy (SP), 2012 IEEE Symposium on*, pages 538–552. IEEE, 2012.

[10] B. M. Bowen, V. P. Kemerlis, P. Prabhu, A. D. Keromytis, and S. J. Stolfo. Automating the injection of believable decoys to detect snooping. In *WiSec*, pages 81–86. ACM, 2010.

[11] B.M. Bowen, S. Hershkop, A. D. Keromytis, and S. J. Stolfo. *Baiting Inside Attackers Using Decoy Documents*, pages 51–70. 2009.

[12] Jørgen Brandt and Ivan Damgård. On generation of probable primes by incremental search. In *Advances in Cryptology—Crypto'92*, pages 358–370. Springer, 1993.

[13] R. Canetti, S. Halevi, and M. Steiner. Hardness amplification of weakly verifiable puzzles. In *TCC*, pages 17–33, 2005.

[14] Ran Canetti, U Friege, Oded Goldreich, and Moni Naor. Adaptively secure multi-party computation. 1996.

[15] Rein Canetti, Cynthia Dwork, Moni Naor, and Rafail Ostrovsky. Deniable encryption. In *Advances in CryptologyCRYPTO'97*, pages 90–104. Springer, 1997.

[16] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to Algorithms*, pages 428–436. MIT Press, third edition, 2009.

[17] A. Paes de Barros. Ids mailing list, "RES: Protocol anomaly detection IDS - honeypots". `http://seclists.org/focus-ids/2003/Feb/95`, 21 Feb. 2003.

[18] Y. Dodis and A. Smith. Entropic security and the encryption of high entropy messages. In *Theory of Cryptography Conference (TCC)*, pages 556–577, 2005.

[19] Marc Fischlin, Anja Lehmann, Thomas Ristenpart, Thomas Shrimpton, Martijn Stam, and Stefano Tessaro. Random oracles with (out) programmability. In *Advances in Cryptology-ASIACRYPT 2010*, pages 303–320. Springer Berlin Heidelberg, 2010.

[20] Pierre-Alain Fouque and Mehdi Tibouchi. Close to uniform prime number generation with fewer random bits. Cryptology ePrint Archive, Report 2011/481, 2011. `http://eprint.iacr.org/`.

[21] Shafi Goldwasser and Silvio Micali. Probabilistic encryption. *Journal of computer and system sciences*, 28(2):270–299, 1984.

[22] John Gordon. Strong primes are easy to find. In *Advances in Cryptology*, pages 216–223. Springer, 1985.

[23] G.H. Hardy and J.E. Littlewood. Some problems of 'partitio numerorum'; iii: On the expression of a number as a sum of primes. *Acta Mathematica*, 44(1):1–70, 1923.

[24] Douglas N Hoover and BN Kausik. Software smart cards via cryptographic camouflage. In *Security and Privacy, 1999. Proceedings of the 1999 IEEE Symposium on*, pages 208–215. IEEE, 1999.

[25] A. Juels and R. Rivest. Honeywords: Making password-cracking detectable. Manuscript (v2.0), `people.csail.mit.edu/rivest/honeywords`, 2 May 2013.

[26] B. Kausik. Method and apparatus for cryptographically camouflaged cryptographic key. U.S. Patent 6,170,058, 2001.

[27] David A McGrew and John Viega. The security and performance of the galois/counter mode (gcm) of operation. In *Progress in Cryptology-INDOCRYPT 2004*, pages 343–355. Springer, 2005.

[28] Gary L Miller. Riemann's hypothesis and tests for primality. *Journal of computer and system sciences*, 13(3):300–317, 1976.

[29] Jesper Buus Nielsen. Separating random oracle proofs from complexity theoretic proofs: The non-committing encryption case. In *Advances in Cryptology—Crypto 2002*, pages 111–126. Springer, 2002.

[30] PKCS #5: Password-based cryptography standard (rfc 2898). RSA Data Security, Inc., September 2000. Version 2.0.

[31] Michael O Rabin. Probabilistic algorithms. *Algorithms and Complexity*, 21, 1976.

[32] Phillip Rogaway, Mihir Bellare, and John Black. Ocb: A block-cipher mode of operation for efficient authenticated encryption. *ACM Transactions on Information and System Security (TISSEC)*, 6(3):365–403, 2003.

[33] A. Russell and H. Wang. How to fool an unbounded adversary with a short key. In *EUROCRYPT*, pages 133–148, 2002.

[34] C.E. Shannon. Communication theory of secrecy systems. *Bell System Technical Journal*, 28(4):656–715, 1948.

[35] Victor Shoup. *A computational introduction to number theory and algebra*. Cambridge University Press, 2009.

[36] L. Spitzner. *Honeypots: Tracking Hackers*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2002.

[37] L. Spitzner. Honeytokens: The other honeypot. Symantec SecurityFocus, July 2003.

A Unsatisfying Approaches to HE

Here we discuss in more detail why existing or simple mechanisms fail to provide good HE schemes. Recall we want both (1) semantic security from our schemes when key entropy is high as well as (2) unpredictability of a target message even when keys have exhaustible entropy. (We assume here that p_m has some amount of uncertainty.)

Existing AE or PBE schemes. The first possible HE solution would be existing password-based encryption schemes [4, 30], which certainly satisfy criteria (1) but fail for goal (2). To see why, consider mounting a brute-force attack against a ciphertext C^* resulting from encrypting a message M^* under a target key K^*. Should a typical authenticated-encryption scheme SE have been used to generate C^* (e.g., Encrypt-then-MAC [3], OCB [32], GCM [27], etc.), then brute-force attacks can proceed as follows. Enumerate a dictionary of all potential keys D, meaning $K^* \in D$, and then, for each $K \in D$, execute $\mathsf{dec}(K, C^*)$ and see if the result is $\perp$. If not, meaning a message was produced, then with all but negligible probability[6] the message is the target M^*. This highlights how the strong authenticity guarantees of AE schemes *benefit* an attacker when D is small enough to enumerate because the attacker can quickly discard incorrect keys.

If, instead, encryption was performed using a scheme SE such as CTR-mode or CBC-mode (that are not AE-secure), then the above brute-force strategy does not work as-is because with these schemes decrypting C^* with any key returns a possible plaintext. This means attackers must somehow distinguish the true plaintext M^* from the set of $d = |D|$ messages $M_1, \ldots, M_d$ that result from the trial decryptions. Cryptographers often suggest that M^* can be picked out easily and programmatically, so that this is not a problem for the attacker. In the example of CTR-mode or CBC-mode, trial decryptions for the wrong key result in messages distributed uniformly (assuming the underlying block cipher is ideal). Thus if an attacker has partial knowledge of the structure of M^*, for example that the first few bytes are a fixed value, then the attacker can with reasonable probability pick out M^*.

Schemes with entropic security. Russell and Wang [33] and Dodis and Smith [18] offer symmetric encryption schemes with security against unbounded attackers for messages with some entropy, but they target the (stronger) goal that no partial information about plaintexts is leaked. In very low-entropy settings, their schemes suffer from the same brute-force attacks as other symmetric encryption schemes. For example, the scheme by Dodis and Smith encrypts by choosing a key R for an xor-universal hash, and then outputs $H_R(K) \oplus M$. In the spirit of our RSA HE example, assume K is sampled from a distribution with max-weight $w = 1/100$ (min-entropy $\mu = -\log w$) and M is a uniformly selected ℓ-bit prime number trivially encoded as an ℓ-bit integer. Then a brute-force message recovery attack will succeed with probability close to one (by checking primality). This is just a concrete example showing how, as Dodis and Smith discuss, security for this scheme holds only when $\gamma + \mu \geq |M| + 2\log(1/\epsilon) + 2$ where γ is the min-entropy of p_m. The problem is that in this example $\gamma \approx \ell - \log \ell$ while $|M| = \ell$, and so $\gamma + \mu$ comes up short, and security up to a bound of $2^{-\mu}$ (as HE is able to achieve for MR) cannot be achieved using these techniques. That said, they target a stronger notion than MR, and applying their techniques to HE could provide a middle ground security between full semantic security and MR security.

Explicitly stored decoy lists. Another possible approach would be to base HE schemes off the idea of generating decoys, for example by storing multiple fake plaintexts along with the legitimate one. The use of decoys is not new in security, and there exist several examples of schemes that use honey messages to attempt to limit the effect of offline brute-force attacks [8, 25, 26]. The simplest idea, similar in spirit to prior approaches, would be to build an HE scheme for some target message distribution p_m using the following "Hide-in-a-List" scheme. Let $\mathsf{HiaL} = (\mathsf{HEnc}, \mathsf{HDec})$ be parameterized by p_m and a security parameter t. It uses a hash function $H : \{0,1\}^* \to [0, t-1]$ as shown below:

$\mathsf{HEnc}(K, M)$
$i \leftarrow H(K)$
$(M_1, \ldots, M_{i-1}, M_{i+1}, \ldots M_t) \leftarrow_{p_m} \mathcal{M}^{t-1}$
$M_i \leftarrow M$
return $M_1, \ldots, M_t$

$\mathsf{HDec}(K, M_1, \ldots, M_t)$
$i \leftarrow H(K)$
return M_i

[6] This holds for all typical AE schemes, though does not necessarily hold for all schemes since the distribution of keys in D is adversarially specified.

The MR security of this scheme is upper-bound by $1/t$, regardless of how high the entropy of the key distribution p_k. Beyond giving poorer bounds than the schemes we will eventually build, it also means that (quite obviously) the scheme never achieves semantic security regardless of the quality of keys. Second, the space requirements are going to be high. The scheme therefore does not satisfy goal (1) and only very poorly helps in achieving goal (2).

One might attempt to fix the first issue by combining with an outer symmetric scheme. The merged construction would split the password K into two parts somehow, use the first part to choose i, and use the second part to do password-based encryption of the list of messages. (Reusing K across the two steps enables an attacker to completely win via offline-brute force attacks when keys have low entropy.) This approach, however, degrades the entropy available to both the outer encryption (reducing brute-force attack effort) and the inner hide-in-a-list (possibly reducing the message uncertainty to be lower than $1/t$). It also does not rectify the space issue.

Our schemes will do better, achieving semantic security when key distributions have high entropy. When not, attackers will have to mount an offline brute-force attack equal to the entropy of the key, and even then have uncertainty about the actual message that is approximately equal to the number of messages returned during the brute-force attack.

B Details of the Inverse Sampling DTE

The following DTE scheme $\mathsf{IS\text{-}DTE} = (\mathsf{is\text{-}encode}, \mathsf{is\text{-}decode})$ realizes inverse sampling using fixed-point arithmetic. Let g be the greatest common divisor (GCD) of the fractions in the image of the CDF, and assume use of a ℓ-bit fixed-point representation with $g \geq u$ where $u = 2^{-\ell}$. The seed space is $\mathcal{S} = \{0,1\}^\ell$ and a fraction $a \in [0,1]$ is represented by the value b such that $\mathsf{rep}_u(a) = \operatorname{argmin}_b |a - b \cdot u|$, i.e. we round to the nearest multiple of u and store the multiple. (Rounding ties are broken arbitrarily, e.g., by always rounding up.) The requirement that the GCD is larger than u ensures[7] that rep_u is unambiguous. Then $\mathsf{is\text{-}encode}(M_i)$ selects $S \leftarrow\!\!\$\ [\mathsf{rep}_u(F_m(M_{i-1})), \mathsf{rep}_u(F_m(M_i)) - 1]$ and outputs S. Finally $\mathsf{is\text{-}decode}(S)$ determines the value M_i such that $\mathsf{rep}_u(F_m(M_{i-1})) \leq S < \mathsf{rep}_u(F_m(M_i))$. Computation of $\mathsf{IS\text{-}DTE}$ is possible in time $\log |\mathcal{M}|$ and space $\mathcal{O}(\mathcal{M})$ (binary search over a table of precomputed CDF values), and often faster. For example if p_m is the uniform distribution over a set of integers $\mathcal{M}$, then encoding and decoding can be made fast. For decoding, simply compute $S \cdot u$ and find the nearest integer factor of $1/|\mathcal{M}|$.

The representation error of this encoder is the maximum, over $a \in \mathsf{Img}(F_m)$, of the value $|a - b \cdot u|$. Denote this maximal error by ϵ_{is}. We have that

$$\epsilon_{\text{is}} = \max_{a \in \mathsf{Img}(F_m)} \left| a - \mathsf{round}\left(\frac{a}{u}\right) \cdot u \right| \leq \max_{a \in \mathsf{Img}(F_m)} \left| a - \left(\frac{a}{u} - \frac{1}{2}\right) \cdot u \right| = \frac{u}{2}$$

where round is the rounding function. We can therefore make ϵ_{is} arbitrarily small, at the cost of encoding output size, by choosing u small (making ℓ large). The representation error gives that $|p_m(M) - p_d(M)| \leq 2\epsilon_{\text{is}} = u$ for all M. More formally we have the following theorem.

Theorem 6 *Let p_m be a message distribution and $\mathsf{IS\text{-}DTE} = (\mathsf{is\text{-}encode}, \mathsf{is\text{-}decode})$ be the inverse sampling DTE described above using ℓ bits. Let $\mathcal{A}$ be any sampling adversary, then $\mathbf{Adv}^{\text{dte}}_{\mathsf{IS\text{-}DTE},p_m}(\mathcal{A}) \leq 1/2^\ell$.*

Proof: We below write SAMP1 for $\text{SAMP1}_{\mathsf{DTE},p_m}$ and SAMP0 for $\text{SAMP0}_{\mathsf{DTE},p_m}$. We first observe that

$$\Pr\left[\ \text{SAMP1}^{\mathcal{A}} \Rightarrow 1 \ \mid\ M^* = M\ \right] = \Pr\left[\ \text{SAMP0}^{\mathcal{A}} \Rightarrow 1 \ \mid\ M^* = M\ \right]$$

where the event "$M^* = M$" is defined appropriately for each game. To see why the equality holds, note that for any particular message M we have that $\mathsf{is\text{-}encode}(M)$ by construction picks uniformly from the set of seed values S for which $\mathsf{is\text{-}decode}(S) = M$. In SAMP0 conditioning on a particular message M fixes the choice of S to be uniform

[7] Consider otherwise, that two points $a \neq a'$ are such that $\mathsf{rep}_u(a) = \mathsf{rep}_u(a')$. This implies that $a - a' = |mg - ng| = |m - n|g \leq u$ for appropriate m, n. But since $g \geq u$ it holds that $u/g \leq 1$ and we arrive at a contradiction.

over the same set. Then we have that

$$\begin{aligned}\Pr\left[\,\text{SAMP1}^{\mathcal{A}} \Rightarrow \mathsf{true}\,\right] &= \sum_{M\in\mathcal{M}} \Pr\left[\,\text{SAMP1}^{\mathcal{A}} \Rightarrow \mathsf{true} \mid M^* = M\,\right]\cdot p_m(M)\\ &\le \sum_{M\in\mathcal{M}} \Pr\left[\,\text{SAMP0}^{\mathcal{A}} \Rightarrow \mathsf{true} \mid M^* = M\,\right]\cdot (u + p_d(M))\\ &= \Pr\left[\,\text{SAMP0}^{\mathcal{A}} \Rightarrow \mathsf{true}\,\right] + u\ .\end{aligned}$$

∎

C A Ratio-based Advantage Measure for DTE Goodness

In Section 4 we defined DTE goodness using a standard indistinguishability advantage measure. Another approach is a ratio-based measure, defined for a message distribution p_m, encoding scheme $\mathsf{DTE} = (\mathsf{encode}, \mathsf{decode})$, and any adversary $\mathcal{A}$ by the equation

$$\mathbf{Adv}^{\text{dte-ratio}}_{\mathsf{DTE},p_m}(\mathcal{A}) = \Pr\left[\,\text{SAMP1}^{\mathcal{A}}_{\mathsf{DTE},p_m} \Rightarrow 1\,\right] \,/\, \Pr\left[\,\text{SAMP0}^{\mathcal{A}}_{\mathsf{DTE}} \Rightarrow 1\,\right]$$

when $\Pr\left[\,\text{SAMP0}^{\mathcal{A}}_{\mathsf{DTE}} \Rightarrow 1\,\right] \neq 0$ and defined to be $\mathbf{Adv}^{\text{dte-ratio}}_{\mathsf{DTE},p_m}(\mathcal{A}) = 1$ otherwise. The closer the advantage is to one, the better the DTE, and the further from one, the worse.

We can prove an analogue of Theorem 2 using the above advantage measure for DTE goodness. The statement is below.

Theorem 7 *Let p_m be a message distribution, p_k be a key distribution, and $\mathsf{HE}[\mathsf{DTE},\mathsf{SE}]$ be the DTE-then-Encrypt scheme using a suitable SE. Let $\mathcal{A}$ be an MR adversary against HE. Then we can give an explicit adversary $\mathcal{B}$ such that $\mathbf{Adv}^{\text{mr}}_{\mathsf{HE},p_m,p_k}(\mathcal{A}) \le \mathbf{Adv}^{\text{dte-ratio}}_{\mathsf{encode},p_m}(\mathcal{B})\cdot \mathrm{E}\left[\,L(p_k,p_d)\,\right]$. Adversary $\mathcal{B}$ runs in time that of $\mathcal{A}$ plus the time of one enc operation.*

The proof proceeds as in the proof of Theorem 2, except that when moving from game G_0 to G_1 using the adversary $\mathcal{B}$ we use instead that $\Pr[G_0^{\mathcal{A}} \Rightarrow 1] \le \Pr[G_1^{\mathcal{A}} \Rightarrow 1]\cdot \mathbf{Adv}^{\text{dte-ratio}}_{\mathsf{encode},p_m}(\mathcal{B})$. Comparing with Theorem 2, this leads to slightly stronger bound for some DTE schemes, such as the inverse sampling one of Section 4 (detailed in Appendix B). For example, we have the following for the inverse sampling DTE scheme.

Theorem 8 *Let p_m be a message distribution and $\mathsf{IS\text{-}DTE} = (\mathsf{is\text{-}encode}, \mathsf{is\text{-}decode})$ be the inverse sampling DTE described above using an ℓ-bit representation. Let $\mathcal{A}$ be any sampling adversary, then $\mathbf{Adv}^{\text{dte-ratio}}_{\mathsf{IS\text{-}DTE},p_m}(\mathcal{A}) \le 1 + 1/2^{\ell}$.*

Proof: We below write SAMP1 for $\text{SAMP1}_{\mathsf{DTE},p_m}$ and SAMP0 for $\text{SAMP0}_{\mathsf{DTE},p_m}$. We first observe that

$$\Pr\left[\,\text{SAMP1}^{\mathcal{A}} \Rightarrow 1 \mid M^* = M\,\right] = \Pr\left[\,\text{SAMP0}^{\mathcal{A}} \Rightarrow 1 \mid M^* = M\,\right]$$

where the event "$M^* = M$" is defined appropriately for each game. To see why the equality holds, note that for any particular message M we have that $\mathsf{is\text{-}encode}(M)$ by construction picks uniformly from the set of seed values S for which $\mathsf{is\text{-}decode}(S) = M$. In SAMP0 conditioning on a particular message M fixes the choice of S to be uniform over the same set. Let $r = \min_M p_d(M)/p_m(M)$. Then we have that

$$\begin{aligned} r\cdot\Pr\left[\,\text{SAMP1}^{\mathcal{A}} \Rightarrow \mathsf{true}\,\right] &= r\sum_{M\in\mathcal{M}} \Pr\left[\,\text{SAMP1}^{\mathcal{A}} \Rightarrow \mathsf{true} \mid M^* = M\,\right]\cdot p_m(M)\\ &\le \sum_{M\in\mathcal{M}} \Pr\left[\,\text{SAMP1}^{\mathcal{A}} \Rightarrow \mathsf{true} \mid M^* = M\,\right]\cdot p_m(M)\cdot\frac{p_d(M)}{p_m(M)}\\ &= \sum_{M\in\mathcal{M}} \Pr\left[\,\text{SAMP0}^{\mathcal{A}} \Rightarrow \mathsf{true} \mid M^* = M\,\right]\cdot p_d(M)\\ &= \Pr\left[\,\text{SAMP0}^{\mathcal{A}} \Rightarrow \mathsf{true}\,\right]\end{aligned}$$

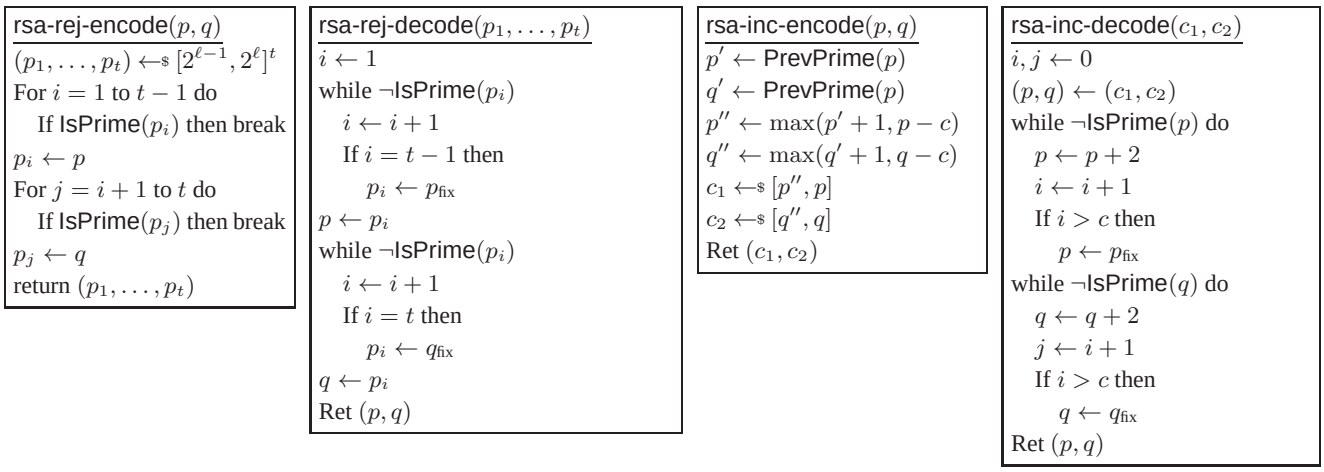

Figure 4: DTE schemes RSA-REJ-DTE (left) and RSA-INC-DTE (right) for pairs of uniform primes in $[2^{\ell-1}, 2^{\ell})$. Both decoding algorithms output some a priori fixed primes in case normal decoding fails. RSA-REJ-DTE is suitable for uniformly selected RSA primes, while RSA-INC-DTE is suitable for keys generated using the PRIMEINC algorithm. PrevPrime(x) returns the maximial prime $p' < x$.

where recall that p_d is the DTE distribution. Let $u = 2^{-\ell}$ and $\epsilon_{\text{is}} = u/2$. Let $\mathcal{M} = \{M_1, \ldots, M_{|\mathcal{M}|}\}$ and let $a_i = F_m(M_i)$ and $b_i = \text{argmin}_b\, |a_i - b \cdot u|$. Let $a_0 = b_0 = 0$. Rearranging the final inequality in the sequence above yields that

$$\mathbf{Adv}^{\text{dte-ratio}}_{\mathsf{DTE},p_m}(\mathcal{A}) \;\le\; \max_i \frac{p_m(M_i)}{p_d(M_i)} \;=\; \max_i \frac{a_i - a_{i-1}}{b_i u - b_{i-1} u} \;\le\; \max_i \frac{b_i u - b_{i-1} u + 2\epsilon_{\text{is}}}{b_i u - b_{i-1} u} \;\le\; 1 + 2\epsilon_{\text{is}}$$

which uses that $p_d(M_i) = (b_i - b_{i-1})u$ and $p_m(M_i) = a_i - a_{i-1}$. ▮

Combining Theorem 7 with Theorem 8 leads to a final MR bound of $\mathrm{E}\,[\,L(p_k, p_d)\,] + 2^{-\ell} \cdot \mathrm{E}\,[\,L(p_k, p_d)\,]$ as compared to the final bound of $2^{-\ell} + \mathrm{E}\,[\,L(p_k, p_d)\,]$ using the indistinguishability-based approach (Theorem 2 combined with Theorem 6). The former will be tighter, though the improvement admittedly may not matter much in many situations. The difference for the credit-card number application from Section 7, for example, is tiny.

D More Compact RSA DTEs

The rejection-sampling DTE RSA-REJ-DTE, whose pseudocode is shown in Figure 4, is not particularly space efficient. An alternative, with optimal compactness, arises when the pair p, q is generated by the classic PRIMENC algorithm [12, 22]. A DTE scheme RSA-INC-DTE is given for primes generated in this manner in Figure 4. The subroutine PrevPrime can be implemented by linearly scanning backwards at most t steps, checking primes, and outputing the last value checked if no prime is found. Decoding outputs some a priori fixed primes should the list of integers input as seed not have enough primes. (We could also abort in other ways.) Unlike RSA-REJ-DTE, RSA-INC-DTE requires just $2(\ell - 2)$ bits for encoding and hence is optimal. Observe that it is security-critical to use the most compact $(\ell - 2)$-bit representation of integers (for both schemes), and not use $\ell - 1$ bits and include a leading bit that is always 1. The error probability of this scheme can be analyzed using the results of Brandt and Damgård [12] which assume the Hardy-Littlewood prime r-tuples conjecture [23]; it is exponentially vanishing in t.

Unfortunately, it is not clear whether one can use the compact scheme RSA-INC-DTE for primes generated by rejection sampling, as PRIMEINC does not output primes that are statistically close to uniform [20]. To see why, note that the larger of two twin primes (ones that are separated by two) is very unlikely to be selected by PRIMEINC, while it is as likely as any other prime to be selected by rejection sampling. Fouque and Tibouchi show that, in fact, one can give a lower bound of 0.86 on the statistical distance between uniform primes and ones generated by PRIMEINC, suggesting this approach is unlikely to work.

Another approach is to use a construction due to Fouque and Tibouchi [20], whose rejection-sampling algorithm uses fewer bits of randomness than the standard rejection sampling approach, yet enjoys upper bounds on the statistical

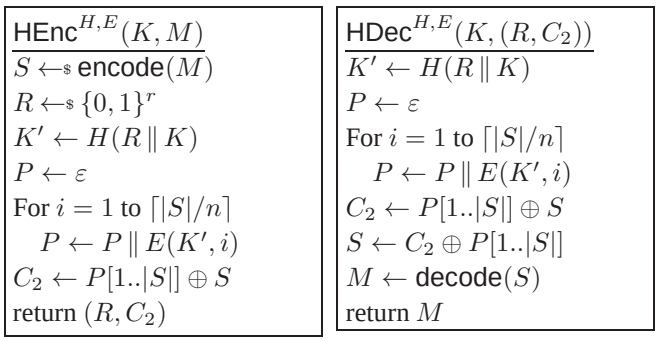

Figure 5: DTE-then-Encrypt using a CTR mode encryption. The notation $P[1..|S|]$ signifies taking the first $|S|$ bits of P.

distance of generated primes from uniform.

E HE Using Block Cipher Modes

We focus on showing on a variant of CTR mode encryption; similar analyses for other modes (e.g., CBC) are possible. The scheme $\mathsf{HE}[\mathsf{DTE}, \mathsf{CTR}]$ is shown in Figure 5. It uses a hash function $H: \{0,1\}^* \to \{0,1\}^k$ to derive a one-time key for CTR mode encryption using a block cipher $E: \{0,1\}^k \times \{0,1\}^n \to \{0,1\}^n$.

The following lemma shows that the balls-and-bins analysis for this CTR-mode based mechanism (in the random oracle and ideal cipher model) can be reduced to that of the hash-based scheme $\mathsf{HE}[\mathsf{DTE}, H]$ which was described in Section 5.

Lemma 5 *Let $\mathsf{HE}1 = \mathsf{HE}[\mathsf{DTE}, \mathsf{CTR}]$ and $\mathsf{HE}2 = \mathsf{HE}[\mathsf{DTE}, H]$ and model H as a random oracle and E as an ideal cipher. Let p_d be the DTE distribution for DTE and fix a key distribution p_k over key space $\mathcal{K}$. Then*

$$\mathrm{E}\left[\, L_{\mathsf{HE}1,p_k} \,\right] \leq \mathrm{E}\left[\, L_{\mathsf{HE}2,p_k} \,\right] + \frac{|\mathcal{K}|^2}{2^k}$$

Proof: (Sketch) Note that in $\mathsf{HE}2$ the pad values xor'd into the fixed ciphertext C_2 are uniform and independent. For $\mathsf{HE}1$ there is the chance that a collision in the output of H occurs, which would give rise to repeated P values. For the fixed R value of interest (in the challenge ciphertext), a standard birthday-bound argument gives that the probability of $H(R, K') = H(R, K'')$ for any two keys $K', K'' \in \mathcal{K}$ is at most $|\mathcal{K}|^2/2^k$ (the probability being over coins of H). Conditioned on there being no collisions, the pad values P are selectedly independently and uniformly (over the coins of the ideal cipher). ∎

Interestingly, the result above could get by without modeling H as a random oracle, and instead rely only on it being collision resistant (though E would still need to be ideal). This approach would lead to a proof of MR security for computationally bounded attackers.

F Proof of Theorem 2

We use a sequence of games to move from the message recovery setting to one in which the adversary can, at best, simply guess the message the challenge ciphertext decrypts to with highest probability. The games G_0, G_1, and G_2 are shown in Figure 6. Game G_0 is equivalent to the MR game, and so

$$\mathbf{Adv}^{\mathrm{mr}}_{\mathsf{HE},p_m,p_k}(\mathcal{A}) = \Pr\left[\, G_0^{\mathcal{A}} \Rightarrow \mathsf{true} \,\right] .$$

Game G_1 picks a uniform point S and then sets $M^* = \mathsf{decode}(Y)$. We bound this transition using the goodness of the DTE. Namely, we build an adversary $\mathcal{B}$ against the DTE scheme. This adversary takes as input (S^*, M^*) and uses these values to simulate the MR game for $\mathcal{A}$. Should $\mathcal{A}$ win the MR game, then $\mathcal{B}$ outputs 1 and otherwise it outputs 0. Then we have that $\Pr[G_0^{\mathcal{A}} \Rightarrow \mathsf{true}] = \Pr[\mathrm{SAMP1}^{\mathcal{B}}_{\mathsf{encode}} \Rightarrow 1]$ and that $\Pr[G_1^{\mathcal{A}} \Rightarrow \mathsf{true}] = \Pr[\mathrm{SAMP0}^{\mathcal{B}}_{\mathsf{encode}} \Rightarrow 1]$. Thus, $\Pr\left[\, G_0^{\mathcal{A}} \Rightarrow \mathsf{true} \,\right] \leq \mathbf{Adv}^{\mathrm{dte}}_{\mathsf{DTE},p_m}(\mathcal{B}) + \Pr\left[\, G_1^{\mathcal{A}} \Rightarrow \mathsf{true} \,\right]$.

Game G_0	Game G_1	Game G_2
$K^* \leftarrow_{p_k} \mathcal{K}$	$K^* \leftarrow_{p_k} \mathcal{K}$	$C^* \leftarrow_{\$} \mathcal{C}$
$M^* \leftarrow_{p_m} \mathcal{M}$	$S^* \leftarrow_{\$} \mathcal{S}$	$M \leftarrow_{\$} \mathcal{A}(C^*)$
$S^* \leftarrow_{\$} \mathsf{encode}(M^*)$	$M^* \leftarrow \mathsf{decode}(S^*)$	$K^* \leftarrow_{p_k} \mathcal{K}$
$C^* \leftarrow_{\$} \mathsf{enc}(K^*, S^*)$	$C^* \leftarrow_{\$} \mathsf{enc}(K^*, S^*)$	$S^* \leftarrow \mathsf{dec}(K^*, C^*)$
$M \leftarrow_{\$} \mathcal{A}(C^*)$	$M \leftarrow_{\$} \mathcal{A}(C^*)$	$M^* \leftarrow \mathsf{decode}(S^*)$
ret $M = M^*$	ret $M = M^*$	ret $M = M^*$

Figure 6: Games used in the proof of Theorem 2.

In game G_2, the ciphertext C^* is chosen uniformly, and S^* is then computed as $\mathsf{dec}(K^*, C^*)$. By our assumption on SE that decrypting a uniformly chosen ciphertext gives a uniform plaintext, we have that this modification does not change the distribution of any of the variables in the game as compared to G_1. We have also delayed computation of K^*, S^*, and M^* until after $\mathcal{A}$ executes; the execution of $\mathcal{A}$ being independent of those values. Note, however, that the choice of M^* is not independent of M, since the coins underlying the choice of M^* are, in part, known to $\mathcal{A}$.

In game G_2, we see that $\mathcal{A}$ wins exactly when it wins the game in which a ciphertext string is sampled uniformly, given to $\mathcal{A}$, and the message output by $\mathcal{A}$ matches the decryption of that ciphertext under a fresh key. In this game, $\mathcal{A}$ maximizes its probability of success by choosing the message with highest probability of being decrypted by C^*. Recall that $L_{\mathsf{HE},p_k}(C) = \max_M \sum_{K \in \mathcal{K}_{M,C}} p_k(K)$. We now argue that $\Pr\left[\, G_2^{\mathcal{A}} \Rightarrow \mathsf{true} \,\right] \leq \mathrm{E}\left[\, L_{\mathsf{HE},p_k} \,\right]$. We have that

$$\begin{aligned}
\Pr\left[\, G_2^{\mathcal{A}} \Rightarrow \mathsf{true} \,\right] &= \sum_{C \in \mathcal{C}} \Pr\left[\, M = M^* \mid C^* = C \,\right] \cdot \Pr\left[\, C^* = C \,\right] \\
&= \sum_{C \in \mathcal{C}} \Pr\left[\, M = \mathsf{decode}(K^*, C) \mid C^* = C \,\right] \cdot \frac{1}{|\mathcal{C}|} \\
&\leq \sum_{C \in \mathcal{C}} L_{\mathsf{HE},p_k}(C) \cdot \frac{1}{|\mathcal{C}|} \;=\; \mathrm{E}\left[\, L_{\mathsf{HE},p_k} \,\right]
\end{aligned}$$

where the events are defined in the straightforward way over the coins used in the execution of $G_2^{\mathcal{A}}$.

G Balls-and-Bins Proofs

In Section Section 6, we present a series of results bounding the expected maximum load for various balls-into-bins experiments. The first Lemmas 2 and 3 give bounds for cases involving uniform-capacity bins. Lemma 4 treats the case of bins with non-uniform capacity.

Lemma 2 *Suppose p_k has maximum weight w and p_d is such that $b = ca$ for some positive integer c. Then for any positive integer $s > 2e/c$, where e is Euler's constant, it holds that*

$$\mathrm{E}\left[\, L_{p_k,p_d} \,\right] \leq w\left((s-1) + 2\left(\frac{a^2}{c^{s-1}}\right)\left(\frac{e}{s}\right)^s\right) .$$

Proof: Let $q_{s,j}$ denote the probability that bin j contains exactly s balls. Then

$$q_{s,j} = \binom{a}{s}\left(\frac{1}{b}\right)^s\left(1-\frac{1}{b}\right)^{a-s} \leq \left(\frac{be}{s}\right)^s\left(\frac{1}{b}\right)^s\left(1-\frac{1}{b}\right)^{a-s} = \left(\frac{ae}{bs}\right)^s\left(1-\frac{1}{b}\right)^{a-s} < \left(\frac{ae}{bs}\right)^s .$$

Thus, a bound on the probability q_s that at least one bin contains at least s balls is

$$q_s \leq b\sum_{i=s}^{a} p_{i,j} < b\sum_{i=s}^{a}\left(\frac{ae}{bs}\right)^i < b\left(\frac{ae}{bs}\right)^s\left(1+\frac{ae}{bs}+\left(\frac{ae}{bs}\right)^2+\ldots\right) = b\left(\frac{ae}{bs}\right)^s\left(1-\frac{ae}{bs}\right)^{-1} .$$

This last step is achieved by letting $A = \frac{ae}{bs}$ and using the well-known equality $S = 1 + A + A^2 + \ldots = 1/(1-A)$ for $A \in [0,1)$. By assumption in the lemma, $s > 2e/c$, which implies $A = \frac{ae}{bs} < 1/2$, and thus $A \in [0,1)$. Additionally,

$s > 2e/c$ implies that $(1 - \frac{ae}{bs})^{-1} < 2$. Thus, $q_s < 2b\left(\frac{ae}{bs}\right)^s$. For any s, we can obtain a bound on $\mathrm{E}[\,L_{p_k,p_d}\,]$ by assuming pessimistically that: (1) At least one bin contains $s-1$ balls; (2) If there is a bin that at least s balls, it contains all a balls; and (3) All balls have weight w. The resulting bound is:

$$\mathrm{E}[\,L_{p_k,p_d}\,] \leq w((s-1) + aq_s) = w\left((s-1) + 2ab\left(\frac{ae}{bs}\right)^s\right).$$

Plugging in $b = ca$ yields the lemma. ∎

Lemma 3 *Suppose p_k has maximum weight w and p_d is such that $b = ca^2$ for some positive integer c. Then $\mathrm{E}[\,L_{p_k,p_d}\,] \leq w\left[1 + \frac{1}{2c} + r(c,b)\right]$, where e is Euler's constant and $r(c,b) = \left(\frac{e}{27c^2}\right)\left(1 - \frac{e}{cb}\right)^{-1}$.*

Proof: As in the proof of Lemma 2, let q_b denote the probability that at least one bin contains at least b balls. The well-known Birthday Bound states that $q_2 \leq \frac{a(a-1)}{2b} < \frac{a^2}{2b} = \frac{1}{2c}$. Now q_3 denotes the probability of at least one triple collision, i.e., three balls landing in the same bin. As shown in the proof of Lemma 2,

$$q_3 < b\left(\frac{ea}{3b}\right)^3\left(1 - \frac{ae}{3b}\right)^{-1} = \left(\frac{ea^3}{27b^2}\right)\left(1 - \frac{ea}{3b}\right)^{-1} = \left(\frac{e}{27c^2}\right)\left(1 - \frac{e}{cb}\right)^{-1}.$$

We have that $\mathrm{E}[\,L_{p_k,p_d}\,] < q_1 + q_2 + aq_3$, where the last term captures the pessimistic assumption that a triple collision results in a maximum load of a balls (and thus weight wa). This yields the lemma. ∎

Lemma 4 *Let $L_{\mathcal{B}}$ denote the maximum load yielded by throwing a balls (of weight 1) into a set $\mathcal{B}$ of b bins of non-uniform capacity at most $0 \leq \gamma \leq 3 - \sqrt{5}$. Let $L_{\mathcal{B}^*}$ denote the maximum load yielded by throwing $a^* = 3a$ balls (of weight 1) into a set $\mathcal{B}^*$ of $b^* = \lfloor 2/\gamma \rfloor$ bins of uniform capacity. Then $\mathrm{E}[L_{\mathcal{B}}] \leq \mathrm{E}[L_{\mathcal{B}^*}]$.*

Proof: Consider an arbitrary set of $b^{(0)}$ bins $\mathcal{B}^{(0)} = \{B_1^{(0)}, \ldots, B_{b^{(0)}}^{(0)}\}$. Suppose that two distinct bins $(B_{b^{(0)}-1}^{(0)}, B_{b^{(0)}}^{(0)})$ are "fused." This means that there results a set of $b^{(1)} = b^{(0)} - 1$ bins $\mathcal{B}^{(1)} = \{B_1^{(1)}, \ldots, B_{b^{(1)}}^{(1)}\}$ such that $c(B_i^{(0)}) = c(B_i^{(1)})$ for $1 \leq i < b^{(1)}$ and $c(B_{b^{(1)}}^{(1)}) = c(B_{b^{(0)}-1}^{(0)}) + c(B_{b^{(0)}}^{(0)})$.

Let X_B be a random variable on bin B denoting the number of balls it contains after a ball-throwing experiment. Consider the obvious coupling of ball-throwing events on $\mathcal{B}^{(0)}$ and $\mathcal{B}^{(1)}$ in which $X_{B_i^{(0)}} = X_{B_i^{(1)}}$ for $1 \leq i < b^{(1)}$ and $X_{B_{b^{(1)}}^{(1)}} = X_{B_{b^{(0)}-1}^{(0)}} + X_{B_{b^{(0)}}^{(0)}}$. As $\max(X_{B_{b^{(0)}-1}^{(0)}}, X_{B_{b^{(0)}}^{(0)}}) \leq X_{B_{b^{(1)}}^{(1)}}$, we have $\mathrm{E}[L_{\mathcal{B}^{(0)}}] \leq \mathrm{E}[L_{\mathcal{B}^{(1)}}]$.

Let $\mathcal{B} = \mathcal{B}^{(0)}$ and, w.l.o.g., let bins be ordered by monotonically decreasing capacity. Now, starting with $j = 0$, repeat the following procedure: While $c(B_{b^{(j)}-1}^{(j)}) + c(B_{b^{(j)}}^{(j)}) \leq \gamma$, do the following: (1) Fuse $B_{b^{(j)}-1}^{(j)}$ and $B_{b^{(j)}}^{(j)}$, yielding bin set $\mathcal{B}^{(j+1)}$; (2) Increment j; and (3) Reorder the bins in $\mathcal{B}^{(j)}$ by monotonically decreasing capacity.

Upon termination after t iterations, there results a set of bins $\mathcal{B}^{(t)}$ with $b^{(t)} = b^{(0)} - t$. For $1 \leq i < b^{(t)}$, bin $B_i^{(t)}$ has capacity $c(B_j^{(t)}) > \gamma/2$. (Only the smallest capacity bin, $B_{b^{(t)}}^{(t)}$, may have capacity $c(B_{b^{(t)}}^{(t)}) \leq \gamma/2$.) Excluding $B_1^{(t)}$ and $B_{b^{(t)}}^{(t)}$, the total number of bins is at most $\lfloor (1-\gamma)/(\gamma/2) \rfloor$. Thus, $b^{(t)} \leq \lfloor (1-\gamma)/(\gamma/2) \rfloor + 2 = \lfloor 2/\gamma \rfloor$.

Let $\mathcal{B}^*$ be a bin set with $b^* = \lfloor 2/\gamma \rfloor$ bins of uniform capacity, i.e., such that bin B_i^* has capacity $c(B_i^*) = 1/b^*$. For $1 \leq i \leq b^{(t)}$, $c(B_i^{(t)}) \leq \gamma$ and $c(B_i^*) = 1/b^* \geq 1/\lfloor 2/\gamma \rfloor \geq \gamma/2$; thus $c(B_i^*) \geq c(B_i^{(t)})/2$.

For a bin $B_i^{(t)}$, with $1 \leq i \leq b^{(t)}$, given an experiment with a single thrown ball, $\Pr[X_{B_i^{(t)}} = 1] \leq \gamma$. For the corresponding bin B_i^*, given an experiment in which three balls are thrown, $\Pr[X_{B_i^*} \geq 1] \geq (1 - (1-\gamma/2)^3) \geq 3\gamma/2 - \gamma^2/2 + \gamma^3/8$. Algebraic manipulation shows that for these two ball-throwing events, $\Pr[X_{B_i^{(t)}} = 1] \leq \Pr[X_{B_i^*} \geq 1]$ for $0 \leq \gamma \leq 3 - \sqrt{5} \approx .76$. ∎

H Proof of Theorem 4

In Section 7, we gather together our results into comprehensive MR security bounds for the application of HE to various practical scenarios. Our main theorem, Theorem 4 (replicated below) treats the case of uniform-capacity bins and gives the tightest bounds when the number of bins is much larger than the number of balls. (Two corollaries in Section 7 treat cases of non-uniform bin capacities and cases where the number of balls is relatively small.)

Theorem 4 *Let $\mathsf{HE}[\mathsf{IS\text{-}DTE}, \mathsf{SE}]$ be an HE scheme with a suitable SE and DTE $\mathsf{IS\text{-}DTE}$ using an ℓ-bit representation. Let p_m be a uniform distribution over b messages and p_k be a key-distribution with maximum weight w. Let $\gamma = \lceil 1/w \rceil$. Then for any adversary $\mathcal{A}$,*

$$\mathbf{Adv}^{\mathrm{mr}}_{\mathsf{HE},p_m,p_k}(\mathcal{A}) \leq w(1+\delta) + \frac{1+\gamma}{2^\ell} ,$$

where $\delta = \frac{\gamma^2}{2b} + \frac{e\gamma^4}{27b^2}\left(1 - \frac{e\gamma^2}{b^2}\right)^{-1}$.

Proof: We apply Theorem 2, Theorem 6, and Lemma 1 to get that

$$\mathbf{Adv}^{\mathrm{mr}}_{\mathsf{HE},p_m,p_k}(\mathcal{A}) \leq \frac{1}{2^\ell} + \mathrm{E}\left[\, L_{p_k,p_d} \,\right]$$

We then apply majorization (Theorem 3) to get that $\mathrm{E}\left[\, L_{p_k,p_d} \,\right] \leq \mathrm{E}\left[L_{p'_k,p_d} \right]$ where $p'_k = (w, w, \ldots, w)$ with dimension $\lceil w^{-1} \rceil$. (Note that p'_k need not be a proper probability distribution, because this now represents the number of balls and their weights.) At this stage, we are analyzing load over bins selected according to p_d, which is (slightly) non-uniform due to representation error. However, we have that $\mathrm{E}\left[\, Lp'_k, p_d \,\right] \leq \mathrm{E}\left[\, Lp'_k, p_m \,\right] + \frac{\lceil w^{-1} \rceil}{2^\ell}$ by a union bound and $|p_m(M) - p_d(M)| \leq 1/2^\ell$ for all M (see Appendix B).

Now having uniform bins and balls, we can now apply Lemma 3 by setting $a = \lceil w^{-1} \rceil$, $b = |\mathcal{M}|$, and $c = b/a^2$ to get the bound

$$\mathrm{E}\left[L_{p'_k,p_m} \right] \leq w\left(1 + \frac{1}{2c} + r(c,b)\right) = w + w\delta$$

where $\delta = \frac{a^2}{2b} + \frac{ea^4}{27b^2}\left(1 - \frac{ea^2}{b^2}\right)^{-1}$. ∎