Fooling an Unbounded Adversary with a Short Key, Repeatedly: The Honey Encryption Perspective

Xinze Li
Tsinghua University, Beijing, China

Qiang Tang
The University of Sydney, Australia

Zhenfeng Zhang
Institute of Software, Chinese Academy of Sciences, Beijing, China

Abstract

This article is motivated by the classical results from Shannon that put the simple and elegant one-time pad away from practice: key length has to be as large as message length and the same key could not be used more than once. In particular, we consider encryption algorithm to be defined relative to specific message distributions in order to trade for unconditional security. Such a notion named honey encryption (HE) was originally proposed for achieving best possible security for password based encryption where secrete key may have very small amount of entropy.

Exploring message distributions as in HE indeed helps circumvent the classical restrictions on secret keys.We give a new and very simple honey encryption scheme satisfying the unconditional semantic security (for the targeted message distribution) in the standard model (all previous constructions are in the random oracle model, even for message recovery security only). Our new construction can be paired with an extremely simple yet "tighter" analysis, while all previous analyses (even for message recovery security only) were fairly complicated and require stronger assumptions. We also show a concrete instantiation further enables the secret key to be used for encrypting multiple messages.

2012 ACM Subject Classification Security and privacy → Cryptography; Theory of computation → Cryptographic primitives

Keywords and phrases unconditional security, information theoretic encryption, honey encryption

Digital Object Identifier 10.4230/LIPIcs.ITC.2021.23

Funding *Qiang Tang*: Supported by a Google Faculty Award, and a gift from Filecoin.
Zhenfeng Zhang: Supported by National Key R&D Program of China (2017YFB0802000)

Acknowledgements We thank anonymous reviewers for valuable comments, specifically the simplification of the key re-use analysis.

1 Introduction

The celebrated *one-time pad* is extremely simple and elegant, while satisfying perfect secrecy that can be against an even computationally unbounded attacker. It also has two well-known drawbacks that hinder its practical deployment: (1) the key length has to be as large as the message size as shown in Shannon's classical work [23] that perfect secrecy must incur such a cost; and (2) one key can only be used to encrypt one message. These two main drawbacks of one time pad have motivated cryptographers to introduce the concept of computational security, and invented corresponding tools. In particular, pseudo-random generators were developed to stretch a short key to be longer for the stream cipher, and pseudo-random functions were introduced to design a symmetric key encryption that can use the same short key to encrypt multiple messages.

Despite that encryption schemes based on computational hardness are commonly used in practice (as one of the greatest achievements of modern (computational) cryptography in general), various application scenarios exist where information-theoretic models and arguments still play a dominant or defining role: (i) one wishes to maintain security despite the use of "weak" keys, such as passwords [16], or short keys in resource constraint devices such as IoT nodes – the information-theoretic setting is forced upon by the fact that brute-force attacks on the key space become feasible; (ii) one wishes to maintain long-term or even everlasting security [8] for encrypted storage of highly confidential contents, such as user secret keys, passwords, or genomics data – the information theoretic setting is demanded as the validity of computational assumptions highly depends on existing cryptanalysis techniques and computing powers and infrastructures, (e.g., the emerging threats of quantum computers).

In this article, we ask whether we can have an information theoretic encryption in the plain model (without random oracles or extra common random sources) whose secret key can be used as "conveniently" as that in the computational encryption, i.e., with length at security parameter, and can be used to encrypt arbitrary number of messages (polynomially bounded). We consider how to circumvent the well-known obstacles by *trading generality of the encryption for unconditional security*, in particular, via the lens of honey encryption [15, 14], which works on message distributions that are known to the encryption algorithm.

Fooling an unbounded adversary with a short key. There have been interesting progresses regarding how to achieve information theoretic encryption with a short key. To circumvent the inherent barrier that perfect secrecy requires each message bit to "burn up" a bit of secret key during encryption, a thread of research considers to give encryptor more information about the plaintext (but no more than what is known to the adversary). Or to put it another way, instead of encrypting arbitrary message, *the encryption algorithm only works for specific types of messages to trade for unconditional security.*

It began with Russell and Wang 's 2002 article on Entropic Security [21], and the follow-up work of Dodis and Smith [9]. They showed how to sidestep the obstacle of key length, obtaining (information-theoretic) semantic security guarantees even with small keys, by assuming that the *message comes from a high-entropy distribution.* On the positive side, entropic security circumvents the classical entropy bound on the secret key. This model, however, still requires the key to be long enough and satisfies $\mu > n - \ell$, where n is the message length, and μ, ℓ are the key entropy and message entropy respectively. Such an entropy requirement (though relaxed than one-time pad) still puts a restriction on the key length, which is often much larger than $O(\lambda)$. To achieve best possible security of (weak) password based encryption facing an offline brute-force attack, Juels and Ristenpart proposed an interesting concept of honey encryption [15] that further explores the *message distribution.* The encryption algorithm there is provided with the details of the message distribution. The key insight is putting attacker to face all plausible messages after brute-force decryption. However, they considered only a security notion that prevents the adversary to recover the whole message, which is arguably insufficient in practice.

For this reason, Jaeger, Ristenpart and Tang (JRT) [14] did an in-depth study of honey encryption: they first defined a semantic security like notion [1] that requires the ciphertext to hide all partial information, when message is sampled from the distribution (for example, password, bio-metric data or secret key); and they demonstrated that a simple hash based encryption achieves the security assuming the hash is a random oracle (or ideal cipher).

[1] It is called targeted distribution semantic security, as the encryption algorithm now is designed relative to a specific message distribution.

More importantly, the analysis were done in the information theoretic setting. A particularly interesting message [14] about honey encryption was that designing an encryption algorithm for a particular message distribution may help to approach our goals: it seems possible in this case to have semantic security like notion against even unbounded attackers, but only requiring a secret key that could be as short as that in the computational setting.

Honey encryption with "semantic security" in the standard model. Relying on random oracles to establish security is certainly unsatisfying, especially in the information theoretic setting, as a random oracle can pump out unlimited amount of entropy as an idealized assumption that could not be instantiated in real-life. An immediate natural question to consider is *whether we can construct honey encryption that satisfies a unconditional semantic security in the standard model*, which was left as an open problem in [15, 14]. Thus this becomes the first problem for us to tackle in this article.

Our first standard model HE construction is a natural instantiation of the DTE-then-Encrypt construction from [14, 15]: previously, encrypting algorithm is simply the hash based encryption using hash as a random function (or even an ideal cipher) to generate a session key for the message. Now we would like to instantiate the random oracle. Let us first walk through the high-level intuition of the complicated security analysis from [14] (similar for [15] even though it was only against a weaker message recovery adversary):

Phase (1) of the security analysis in JRT [14] and JR [15] was to transform from the security game defining semantic security (for a targeted message distribution) to a game in which the ciphertext is chosen uniformly, and the secret key is sampled after the adversary has run. In the new game, one can show that the advantage of any adversary is no larger than that of an "optimal" adversary who decrypts the ciphertext using all possible keys, computes the predicate value on the resulting plaintext, and outputs the bit which has the higher cumulative mass of keys that resulted in this bit. Such a cumulative probability can be viewed as the total weight of the balls in a bin at the end of a balls-and-bins game.

Phase (2) is the complicated part that was to analyze the maximum load in the non-uniform bin selection with non-uniform balls. A majorization lemma [4] was applied to upper-bound the quantity that obtained from *uniformly* weighted balls (with the same weights). The latter thus can be simply be derived by bounding the number of balls in the experiment, for which we can use Chernoff inequality.

Now without the random oracle, the balls-into-bins experiment is no longer valid (at least not in the normal sense), since "ball throws" become correlated with each other. Furthermore, the majorization technique also requires independence. To get around the major challenges, we observe that a direct analysis might be possible without using the majorization techniques, if we leverage a generalized Chernoff-bound [22] to deal with correlated (to some extent) random variables that even may not be identically distributed. Such technique may also be applied to simplify the analysis of the JRT result regarding semantic security [14] and also the JR result [15] for even message recovery security. Since a generalized Chernoff bound dealt with q-wise independent random variables, it becomes natural to instantiate the random oracle with a q-wise independent hash. Moreover, after a careful analysis, we can set the parameter with just a small q such as the security parameter.

Our second standard model HE construction starts from an entropically secure encryption (ES). This construction and analysis turn out to be surprisingly simple. There is a seeming dilemma: what we would like to have is an HE scheme with a key of length at most $O(\lambda)$, where λ is the security parameter, as in the computational setting; while entropic security requires the key entropy (length) no smaller than message length minus message entropy. For most of the message distributions (say with entropy half of the length), it already requires the key length to be depending on message length.

But a closer look reveals the power of one important building block for all existing HE constructions, distribution transforming encoder (DTE for short). A distribution transforming encoder takes a message distribution, and encodes it to an almost uniform distribution over another space S. If S is close to $\{0,1\}^n$, the encoded distribution is already close to uniform. Applying an entropically secure encryption on "encoded message distribution" now offers opportunities to allow the key entropy to be minimal. Interestingly, it is very easy to see the security of the DTE-then-ES construction of HE, though previous concrete constructions of HE were all paired with very complicated analysis. [2]

It is also worth noting that we can consider the output of DTE to be a uniform distribution in our analysis, which only adds negligible error. Since the ES scheme is applied on this distribution, we can further lessen restrictions on the ES scheme: it only needs to work for uniform input (with full entropy). In this way, we can achieve an even better security bound than those previous ones, which was considered to be asymptotically optimal.

Fooling an unbounded adversary again with the same short key. It's well-known that in one-time pad (and many related constructions), if the secret key is used to encrypt two messages, some pattern (e.g., whether two messages are equal) is leaked immediately. Such vulnerability was widely exploited in practice, and lead to numerous highly impacting attacks [5, 25]. With the encouragement of HE in circumventing the entropy bound in the information theoretic setting, we are now more ambitious and would like to ask *whether one can further encrypt multiple messages using the same short key.* This becomes our second major question to address in this article.

Insecurity of ES when reusing a key. Neither previous works on entropic security, nor honey encryption discussed the key reuse issue. To see whether entropic security is helpful enabling key reuse, we start with the security notion which generalizes the conventional semantic security definition. Conventionally, the adversary who sees a ciphertext, tries to learn a predicate on the corresponding message. Now, adversary seeing a vector of ciphertext, can infer a predicate bit on the vector of messages (which were sampled independently).

Having this definition in place, (informal) intuition that entropic security does not seem to be promising enabling the key reuse can be seen as follows: One can view the vector of messages as one large piece of message. Plugging in the entropy bound from [21, 9], suppose there are t messages, each with entropy $n/2$ and length n. The entropy bound would require length of the key to be no smaller than $t \cdot n - t \cdot n/2 = tn/2$. Even a secret key with length n could be at most used twice. Such intuition can be easily generalized to the case that message entropy to be $n-1$ (in which the key could be reused at most μ times, where μ is the entropy of the key)! To formally prove the impossibility, we establish the lower bound on the key length if an ES scheme reuses the key for T times. This generalizes the original lower bound in the single ciphertext setting [21, 9]. An analysis on key length requirements is given in Sec. 4.1 to show that in order to reuse the key for T times, we must have the key length at least to be $(n-t)T$, where the messages are from distribution with entropy t.

[2] We remark that, as far as we know, such a simple construction has not been shown before. Previously, honey encryption was motivated by password based encryption, thus their main goal is to obtain security with a weak key. Thus comparison focused on the insufficiency of entropic security [15, 14] due to its key length requirement. When we consider honey encryption from the angle of unbounded security, entropic security becomes a natural candidate to leverage, which enables us to simplify the analysis for HE [15, 14] dramatically. We still presented the q-wise independent hash based construction to showcase (already simplified) existing analysis structure, which in turn demonstrates the simplicity of HE from entropic security. Ironically, with existing analysis of the q-wise independent hash based construction, we cannot even choose q to be as small as 2, while entropic security based construction can!

Honey encryption that allows key re-use in the standard model. The existence of entropy bounds in both one-time pad and entropic security hints that security in those two settings needs to burn key entropy for each ciphertext. However, the entropy bounds do not seem to appear in honey encryption when one encrypts only for a particular message distribution. We ask whether HE can remain secure when the secret key is reused.

We first remark that, key re-use does not come for free in honey encryption. The above impossibility in entropic security directly hints this, and also gives an example: the HE construction instantiating with the small-biased set based ES scheme from [21], does not allow key reuse, even when the message is uniform. There, the overall construction is basically in the form of $M \oplus s(K)$ for a function $s()$ defining the small biased set.

Fortunately, we show, the HE construction from the pairwise independent hash (trivially holds for the q-wise independent hash based construction as well) allows one to reuse key for arbitrary number of times t with only a security loss linear to t. We also give a definition of (targeted distribution) semantic security that allows key reuses. Interestingly, the view that leads to the entropy deficiency in entropic security explains the intuition why the second HE construction enables key reuses. As a high level intuition, now facing a vector of independently sampled messages, encoding them individually and then concatenating works as a good DTE; hashing them individually and then concatenating also works as a good pairwise independent hash! Those observations essentially reduce the security of key reuse to the security of HE on another message distribution.

Now we finally have an information theoretically secure encryption that can have and use a key as convenient as in the computational setting, just by giving the encryption algorithm details of the message distribution (which is no more than what the adversary knows).

Discussions. Honey encryption was originally proposed to deal with best possible security for password based encryption, where a very small amount of entropy in password is available. We find the underlying concept exploring the message distribution to "mute" the brute-force attack very inspiring and applicable to broader information theoretic setting, which motivated us to examine information theoretic encryption via the HE lens and vise versa.

As we demonstrate, exploring message distribution enables us to circumvent the major obstacle about secret key in information theoretic encryption; on the other hand, putting HE in the unbounded security domain also leads to a very natural and simple construction of HE from entropic security, which simplifies our understanding of HE itself. Indeed, all previous HE constructions (even in the random oracle model and only against a weaker message recovery adversary) require fairly complicated analysis.

Other related works. There exist several lines of exploration of relaxing information theoretic security or adding extra setup for smaller key length. One attempt is adding constraints to the adversary, e.g. restricting adversary's access to limited ciphertext bits[19], or constraining adversary's memory usage (bounded storage model)[6, 3, 8, 11, 18]. These schemes mostly focus on key expansion by leveraging honest guy's advantage. In bounded storage model, for example, a short secret key is expanded to a one-time pad key using the random sources, and the actual key in use should still satisfy Shannon's bound. In comparison, we do not give encryption algorithm any *extra* knowledge in honey encryption; what we give to the encryption algorithm (message distribution) is already known to the attacker. Another relaxed notation proposed by Calmon et al. is ε-symbol secrecy[10], in which it is hard for an adversary to recover message bits (but not functions of messages). Limited by underlying encoding scheme, though, the key size cannot be compressed to be

as small as security parameter. A third way of relaxation is to restrict input messages to concrete distribution, for example, maximal correlation secrecy[17] requires input to be uniformly distributed.

We note that, none of the constructions dealt with security when a key is reused, except that might be possible in the bounded storage model; but, essentially, the random source generate fresh randomness for each encryption, which implicitly uses a super large key.

Shikata[24] proposed formalization of several information theoretic security definitions and gave lower bounds of key length, which was later extended to multiple-use model. While our work restricts the input messages to follow certain distribution, in their work, however, a key is required to have $\Omega(m^T)$ length if it is reused T times, where m denotes the message length. Therefore, it is clear they cannot have a short key and enable the key reuse.

HE has a number of real-world applications, such as secure password vaults[7, 12], genomic data[13], and natural languages [20, 1]. Recently, a systematic study about various pseudorandom encodings including DTE was given in [2].

2 Preliminaries and Honey Encryption Background

Notations. Let $\mathcal{S}$ be a set, a distribution on $\mathcal{S}$ is defined to be a function $p : \mathcal{S} \to [0, 1]$ such that $\sum_{s\in\mathcal{S}} p(s) = 1$. Denote $U_{\mathcal{S}}$ to be the uniform distribution on $\mathcal{S}$. For a set $B \subseteq \mathcal{S}$, define $p(B) = \sum_{s\in B} p(s)$. By $s \leftarrow_p \mathcal{S}$ we mean sampling an element s from S according to distribution p, and by $s \leftarrow\$ \mathcal{S}$ we mean s is sampled uniformly from $\mathcal{S}$. Let $\mathcal{A}$ be a randomized algorithm, then by $y \leftarrow\$ \mathcal{A}(X)$ we mean y is the output of algorithm $\mathcal{A}$ running on input X. We use $y \leftarrow \mathcal{A}(X)$ if $\mathcal{A}$ is a deterministic algorithm instead. For a game G, we use $\Pr[G \Rightarrow \text{true}]$ to denote the probability that G outputs true.

Min-entropy. Let $X \leftarrow_p \mathcal{S}$ be a random variable with distribution p. The min-entropy of X is $\mathrm{H}_\infty(X) = -\log \max_{s\in\mathcal{S}} p(s)$. We also use notations $\mathrm{H}_\infty(p) = \mathrm{H}_\infty(X)$ for simplicity.

***q*-wise independent hash functions.** A family of hash functions $\{\mathsf{H}_i : \mathcal{M} \to \mathcal{S}\}_{i\in\mathcal{I}}$ is called universal hash family if for all $m_1, m_2 \in \mathcal{M}, m_1 \neq m_2, \Pr_{i\leftarrow\mathcal{I}}[\mathsf{H}_i(m_1) = \mathsf{H}_i(m_2)] \leq \frac{1}{|S|}$. Furthermore, the hash family $\{\mathsf{H}_i\}$ is called q-wise independent if for any distinct $m_1, m_2, \cdots, m_q \in \mathcal{M}$ and any $t_1, t_2, \cdots, t_q \in \mathcal{S}$,

$$\left|\Pr_{i\leftarrow\mathcal{I}}[\mathsf{H}_i(m_1) = t_1 \wedge \mathsf{H}_i(m_2) = t_2 \wedge \cdots \wedge \mathsf{H}_i(m_q) = t_q]\right| = \frac{1}{|\mathcal{S}|^q}$$

$\{\mathsf{H}_i\}$ is also called pairwise independent when $q = 2$.

Let $\mathbb{F}$ be a field. A (polynomial) q-wise independent hash family $\mathsf{H}_{(a_0,\cdots,a_{q-1})} : \mathbb{F} \to \mathbb{F}$, where each $a_i \in \mathbb{F}$, can be constructed [26] as

$$\mathsf{H}_{(a_0,\cdots,a_{q-1})}(m) = \sum_{i=0}^{q-1} a_i m^i$$

Entropic security. The definition of entropic security was first proposed by Russel and Wang in [21] and later studied by Dodis and Smith[9]. A probabilistic map Y is said to hide all functions of X with leakage ε if for every adversary $\mathcal{A}$, there exists some adversary $\mathcal{A}'$ such that for all functions f,

$$|\Pr[\mathcal{A}(Y(X)) = f(X)] - \Pr[\mathcal{A}'() = f(X)]| \leq \varepsilon$$

$\mathrm{TDSS0}^{\mathcal{A}_s,f}_{\mathsf{HE},p_m,p_k}$	$\mathrm{TDSS1}^{\mathcal{A},f}_{\mathsf{HE},p_m,p_k}$
$M \leftarrow_{p_m} \mathcal{M}$	$K \leftarrow_{p_k} \mathcal{K}$
$b \leftarrow\$ \mathcal{A}_s$	$M \leftarrow_{p_m} \mathcal{M}$
return $b = f(M)$	$C \leftarrow\$ \mathsf{HEnc}(K, M)$
	$b \leftarrow\$ \mathcal{A}(C)$
	return $b = f(M)$

Figure 1 TDSS Security Games.

The map $Y()$ is called (μ, ε)-entropically secure if $Y()$ hides all functions of X, whenever the min-entropy of X is at least μ. We say $Y()$ is (μ, ε)-entropically secure for predicates if $Y()$ hides all functions of X that take value in $\{0, 1\}$.

Honey encryption[15, 14]. A honey encryption (HE) scheme $\mathsf{HE} = (\mathsf{HEnc}, \mathsf{HDec})$ is designed for a specific input distribution. We use $\mathcal{K}$, $\mathcal{M}$ and $\mathcal{C}$ denote the key space, the message space and the ciphertext space, and p_k, p_m denote the key distribution on $\mathcal{K}$ and the message distribution on $\mathcal{M}$ respectively. The encryption algorithm will take p_m as input.

Target distribution semantic security (TDSS) [14]. Since honey encryption is designed only for each specific message distribution, the semantic security type of definition has to be adapted for a targeted distribution. For more detailed discussions, we refer to [14].

Let $f : \mathcal{M} \to \{0, 1\}$ be a predicate on $\mathcal{M}$, $p_f(b) = \Pr[f(M) = b \mid M \leftarrow_{p_m} \mathcal{M}]$, and $\omega_f = \max\{p_f(0), p_f(1)\}$. Define security games for HE with respect to distributions p_k, p_m in Figure 1: In game TDSS0 an adversary $\mathcal{A}_s$ called a simulator tries to guess the value of $f(M)$ with no access to ciphertexts, while in game TDSS1 the adversary $\mathcal{A}$ guess $f(M)$ given an encryption of M. The advantage of an adversary $\mathcal{A}$ with respect to HE, distributions p_m, p_k and predicate f is defined as:

$$\mathsf{Adv}^{\mathrm{tdss}}_{\mathsf{HE},p_m,p_k}(\mathcal{A}, f) = \Pr\left[\mathrm{TDSS1}^{\mathcal{A},f}_{\mathsf{HE},p_m,p_k} \Rightarrow \mathrm{true}\right] - \Pr\left[\mathrm{TDSS0}^{\mathcal{A}_s,f}_{\mathsf{HE},p_m,p_k} \Rightarrow \mathrm{true}\right]$$

The optimal strategy for $\mathcal{A}_s$ in TDSS0 is to output the most probable value of $f(M)$ given p_m, f, which gives $\Pr\left[\mathrm{TDSS0}^{\mathcal{A}_s,f}_{\mathsf{HE},p_m,p_k} \Rightarrow \mathrm{true}\right] = \omega_f$. Therefore we can rewrite $\mathsf{Adv}^{\mathrm{tdss}}_{\mathsf{HE},p_m,p_k}(\mathcal{A}, f) = \Pr\left[\mathrm{TDSS1}^{\mathcal{A},f}_{\mathsf{HE},p_m,p_k} \Rightarrow \mathrm{true}\right] - \omega_f$.

▶ **Definition 1.** *An HE scheme* HE *with respect to key distribution* p_k *and message distribution* p_m *is said to be* $\varepsilon-$TDSS *secure if*

$$\mathsf{Adv}^{\mathrm{tdss}}_{\mathsf{HE},p_m,p_k} = \max_{\mathcal{A},f} \mathsf{Adv}^{\mathrm{tdss}}_{\mathsf{HE},p_m,p_k}(\mathcal{A}, f) \le \varepsilon$$

where the max is over all (unbounded) adversary $\mathcal{A}$ *and arbitrary predicate* f*.*

Distribution-transforming encoder (DTE)[15]. A DTE is a pair of algorithms $\mathsf{DTE} = (\mathsf{encode}, \mathsf{decode})$ defined relative to sets $\mathcal{M}$ and $\mathcal{S}$, where randomized encoding algorithm encode takes as input a message $M \in \mathcal{M}$ and outputs $S \in \mathcal{S}$, and deterministic decoding algorithm decode takes as input $S \in \mathcal{S}$ and outputs $M \in \mathcal{M}$. DTE should always satisfy correctness: for all $M \in \mathcal{M}$, $\Pr[\mathsf{decode}(\mathsf{encode}(M)) = M] = 1$.

$\mathrm{SAMP0}^{\mathcal{D}}_{\mathsf{DTE}}$	$\mathrm{SAMP1}^{\mathcal{D}}_{\mathsf{DTE},p_m}$
$S \leftarrow\!\$\, \mathcal{S}$	$M \leftarrow_{p_m} \mathcal{M}$
$b \leftarrow\!\$\, \mathcal{D}(S)$	$S \leftarrow\!\$\, \mathsf{encode}(M)$
return $b = 1$	$b \leftarrow\!\$\, \mathcal{D}(S)$
	return $b = 1$

Figure 2 DTE Security Games.

$\mathsf{HEnc}(K, M)$	$\mathsf{HDec}(K, C)$
$S \leftarrow\!\$\, \mathsf{encode}(M)$	$S \leftarrow \mathsf{Dec}(K, C)$
$C \leftarrow\!\$\, \mathsf{Enc}(K, S)$	$M \leftarrow \mathsf{decode}(S)$
return C	**return** M

$\mathsf{HEnc}(K, M)$	$\mathsf{HDec}(K, C)$
$S \leftarrow\!\$\, \mathsf{encode}(M)$	$(R, \tilde{C}) \leftarrow C$
$R \leftarrow\!\$\, \{0,1\}^r$	$S \leftarrow \mathsf{H}_R(K) \oplus \tilde{C}$
$\tilde{C} \leftarrow \mathsf{H}_R(K) \oplus S$	$M \leftarrow \mathsf{decode}(S)$
$C \leftarrow (R, \tilde{C})$	**return** M
return C	

Figure 3 Left: DTE-then-Encrypt; Right: DTE-then-Hash.

The security property for DTE schemes is defined via the security games in Figure 2. The advantage of an adversary $\mathcal{D}$ against DTE and distribution p_m is:

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

Define DTE advantage as measurement for DTE security as follows:

▶ **Definition 2.** *The DTE advantage of a scheme* DTE *with respect to distribution* p_m *is defined to be* $\mathsf{Adv}^{\mathrm{dte}}_{\mathsf{DTE},p_m} = \max_{\mathcal{D}} \mathsf{Adv}^{\mathrm{dte}}_{\mathsf{DTE},p_m}(\mathcal{D})$*, where the maximization is over all (unbounded) adversary* $\mathcal{D}$*.*

Although it is mentioned in [2] that DTE does not exist for *all* distributions, a large number of distributions can be encoded using a DTE. For example, a distribution can be encoded using inverse-sampling DTE in [15] if values of all probability mass functions are explicitly given.

DTE-then-Encrypt[15, 14]. DTE-then-Encrypt serves as a framework to construct HE schemes with respect to a target distribution. A message is first encoded using DTE and then encrypted using a symmetric encryption. The framework is described in Fig 3. In [15, 14], the encryption is instantiated with a hash-based encryption scheme $\mathsf{Enc}(K, M) = (R, \mathsf{H}_R(K) \oplus M)$, where R is randomly drawn. We rename this DTE-then-Hash for clarity. The DTE-then-Hash construction is described in Fig 3.

3 HE Constructions Secure in the Standard Model

Recall that for an n-bit input, one-time pad requires the key to have at least n bits of min-entropy. Entropically secure encryption scheme (ES scheme) relaxes this entropy requirement to $n - t$ by restricting the input with at least t bits of min-entropy. If the ES scheme uses a key with min-entropy $n - t + \delta$, it would achieve a security around $2^{-\delta/2}$[9]. HE completely removes the key entropy requirement by working on a specific message distribution (called

target distribution). JRT showed that an HE scheme using a key with min-entropy δ would achieve ε-TDSS (targeted distribution semantic security) for ε slightly bigger than $2^{-\delta/2}$ (e.g. ε is around 2^{-13} when $\delta = 30$[14]), albeit in random oracle model.

Below we present our two attempts at HE constructions in the standard model as well as security analysis: the first using q-wise independent hash family, and the second using an ES scheme. The first construction, which is a natural extension of JR[15] and JRT[14]'s construction in standard model, gives a security bound comparable to the random oracle case, and requires q to be around key length. While this is acceptable for low-entropy key settings, it turns out that our second HE construction using ES scheme can achieve asymptotically optimal security bounds and only requires a pairwise independent hash family.

3.1 HE from q-wise independent hash

Our first attempt to construct HE in standard model follows from the previous works of [15] and [14]. Their construction applies DTE-then-Hash framework where the hash function is modeled as a random oracle. While we cannot use random oracle since we are working in standard model, it is natural to consider the case where the hash function is modeled as a q-wise independent hash family instead. As we will see, the overall structure of the analysis of [15] and [14] still applies with the use of new techniques, yet it results in a relatively weak bound. We give a high level overview in this section and leave the full proof in the appendix.

Let us recall the TDSS security analysis of DTE-then-Hash construction presented in [14, 15], which came in two main steps. In the first step, game transitions are performed on the original TDSS security game to another game where the adversary's best strategy is to decrypt the ciphertext using all possible keys, compute the predicate value on all decryptions, and output the bit which is supported by larger probability mass of keys. Note that such transition does not rely on random oracle, and incurs an error which is bounded by DTE advantage. In the second step, different balls-into-bins analysis are applied where each decryption attempt is considered as throwing a "ball" into a "bin". Each decryption result is considered to be independent if random oracle is used; however, we introduce correlations to the decryption results using q-wise independent hash family, and we cannot use majorization lemma to simplify the probability analysis. We overcome these problems by using a more general Chernoff-like bound on q-wise independent random variables[22]. An advantage of this bound is that it does not require the random variables to have the same distribution, so that we can even omit the majorization step in previous proofs.

Game transitions. Similar to previous results [14, 15] , the first part of the proof is summarized by the following lemma from [14], which transforms the estimation of the adversary's advantage $\mathsf{Adv}^{\mathrm{tdss}}_{\mathsf{HE},p_m,p_k}(\mathcal{A},f)$ in TDSS games to the expected outcome $\mathbb{E}\big[L^{\mathsf{H},\mathsf{DTE},f}_{p_k}\big]$ in an experiment $E^{\mathsf{H},\mathsf{DTE},f}_{p_k}$ defined in 4, via a sequence of games. The bias of the predicate is $\omega_f = \max\{p_f(0), p_f(1)\}$.

Note that experiment $E^{\mathsf{H},\mathsf{DTE},f}_{p_k}$ actually describes a brute-force attack, and the expectation $\mathbb{E}\big[L^{\mathsf{H},\mathsf{DTE},f}_{p_k}\big]$ denotes the success probability of this attack. In other words, we are transitioning to an experiment in which a brute-force attack is performed, and we are concerned with the success probability of such attack.

▶ **Lemma 3** ([14]). *Let* HE *be defined using DTE-then-Hash construction with respect to distributions* p_m, p_k*, q-wise independent hash family* $\{H_i\}$ *and DTE scheme* DTE*, f be a predicate on* $\mathcal{M}$*,* $\mathcal{A}$ *be any adversary, then*

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

where $L^{\mathsf{H},\mathsf{DTE},f}_{p_k}$ *is defined via experiment* $E^{\mathsf{H},\mathsf{DTE},f}_{p_k}$*.*

Experiment $E_{p_k}^{\mathsf{H},\mathsf{DTE},f}$

$R \leftarrow\$ \{0,1\}^r, \tilde{C} \leftarrow\$ \mathcal{S}$
$B_0 \leftarrow \emptyset, B_1 \leftarrow \emptyset$
for $i = 1, 2, \cdots, |\mathcal{K}|$ **do**
$S_i \leftarrow \mathsf{H}_R(K_i) \oplus \tilde{C}$
$M_i \leftarrow \mathsf{decode}(S_i)$
$b_i \leftarrow f(M_i)$
$B_{b_i} \leftarrow B_{b_i} \cup \{K_i\}$
endfor
$L_{p_k}^{\mathsf{H},\mathsf{DTE},f} \leftarrow \max_{b\in\{0,1\}} p_k(B_b)$

Figure 4 Experiment used in security analysis of DTE-then-Hash.

Proof for Lemma 3 is given in Appendix A.

Bounding success probability of predicting. In the second part of our proof, we give a bound on $\mathbb{E}\big[L_{p_k}^{\mathsf{H},\mathsf{DTE},f}\big]$, which represents the probability of a success brute-force attack. This is the different part of analysis we have to do without the luxury of relying on random oracle to get independence or majorization technique to simplify the balls-into-bins experiment. Our main tool is the following more general Chernoff-like result which is a special case of a theorem from [22]:

▶ **Lemma 4** ([22]). *Let $X_1, \cdots, X_n$ be q-wise independent random variables confined to the interval $[0,1]$, and $X = \sum_{i=1}^n X_i$ with $\mu = \mathbb{E}[X]$, then*
1) For $\delta \le 1$ satisfying $q \le \lfloor \delta^2 \mu e^{-1/3} \rfloor$, $\Pr[|X-\mu| \ge \delta\mu] \le e^{-\lfloor q/2 \rfloor}$;
2) For $\delta \ge 1$ satisfying $q \le \lfloor \delta\mu e^{-1/3} \rfloor$, $\Pr[|X-\mu| \ge \delta\mu] \le e^{-\lfloor q/2 \rfloor}$.
Define p_d to be the probability distribution of $\mathcal{M}$ given by sampling a uniformly random seed from $\mathcal{S}$ and then applying decode, i.e.

$$p_d(M) = \Pr[M^* = M \mid S \leftarrow\$ \mathcal{S}, M^* \leftarrow \mathsf{decode}(S)]$$

The following lemma gives a bound on $\mathbb{E}\big[L_{p_k}^{\mathsf{H},\mathsf{DTE},f}\big]$; we defer detailed proof to Appendix B:

▶ **Lemma 5.** *Let $p_t(b) = \Pr[f(M) = b \mid M \leftarrow_{p_d} \mathcal{M}]$ and $\omega_t = \max\{p_t(0), p_t(1)\}$. Let ω_k denote the maximum key probability. Then for all $q \le e^{-1/3}/2\omega_k$,*

$$\mathbb{E}\big[L_{p_k}^{\mathsf{H},\mathsf{DTE},f}\big] \le \omega_t + e^{-\lfloor q/2 \rfloor} + (1 - 2e^{-\lfloor q/2 \rfloor})(q\omega_k)^{1/2} e^{1/6}$$

Finalizing the bound. The following theorem sums up the two steps above and gives a TDSS security bound:

▶ **Theorem 6.** *Let HE be constructed using DTE-then-Hash with respect to distributions p_m, p_k and q-wise independent hash family $\{\mathsf{H}_i\}_{i\in\{0,1\}^r}$, where ω_k denotes the maximum key probability. Then for all $q \le e^{-1/3}/2\omega_k$,*

$$\mathsf{Adv}^{\mathrm{tdss}}_{\mathsf{HE},p_m,p_k} \le 2\mathsf{Adv}^{\mathrm{dte}}_{\mathsf{DTE},p_m} + e^{-\lfloor q/2 \rfloor} + (1 - 2e^{-\lfloor q/2 \rfloor})(q\omega_k)^{1/2} e^{1/6}$$

In other words, HE is $\varepsilon-$TDSS secure for

$$\varepsilon = 2\mathsf{Adv}^{\mathrm{dte}}_{\mathsf{DTE},p_m} + e^{-\lfloor q/2 \rfloor} + (1 - 2e^{-\lfloor q/2 \rfloor})(q\omega_k)^{1/2} e^{1/6}$$

The proof for Lemma 5 and Theorem 6 are given in Appendix B.

Note that if we choose $q = \log(1/\omega_k) = \mathrm{H}_\infty(p_k)$ in Theorem 6, the security bound is asymptotic to $O((\omega_k \log(1/\omega_k))^{1/2})$, which is close to $O(\omega_k^{1/2})$ in low min-entropy key settings. This implies that we can choose $q \approx \lambda$ if we are using a key of length λ. However, we cannot choose $q = O(1)$ because of the results in Lemma 4. Furthermore, although the $O(\log(1/\omega_k)))$ hardly affects the security bound, we can actually push harder towards the $O(\omega_k^{1/2})$ bound, which is considered to be optimal in [14].

3.2 HE from entropically secure encryption

We now give an HE construction satisfying TDSS in the standard model via entropic security. This idea arises when we view both of them as information theoretic encryption candidates. An observation is that entropic security notion and TDSS notion are similar in some way: they both capture the hardness for an unbounded adversary to learn any predicate of the input message given an encryption of this message. The difference is that entropic security expects entropy from the message, while HE scheme further explores the message distribution. Intuitively, we would like to "modify" input message to gain enough entropy.

In order to match the entropy requirement in entropic security, we first encode input messages using DTE (constructed specifically for message distribution, see above definition in Sec.2). It should be pointed out, however, that DTE actually outputs a near-uniform distribution which has almost full entropy. We can think of the DTE output as a uniform distribution in our analysis, which only incurs negligible error. Such ES schemes are easy to find; in fact, any $(n-\alpha, \varepsilon)$-ES scheme for $\alpha \geq 0$ supports uniform input, since it supports any input with min-entropy at least $n - \alpha$. In this way, the entropy requirements in the ES scheme becomes unimportant since we are using uniform distribution as input; We can even use an ES scheme which only supports uniform input, which leads to better parameters. In fact, according to an observation in JRT[14], our construction achieves asymptotically best TDSS security bound $O(\omega_k^{1/2})$.

More interestingly, our analysis is much simpler than that in JRT[14] and even the message recovery security analysis in JR[15], which is a strictly weaker security notion than TDSS; while at the same time, this simpler analysis is "tighter": the entropic security path gives an instantiation from *pairwise*-independent hash, but following the more complicated analysis structure in Sec. 3.1, the resulting bound does not allow us to choose q to be as small as 2 in the q-wise independent hash based construction.

▶ **Theorem 7.** *Let p_m be a distribution on a set $\mathcal{M}$, p_k be a distribution on a set $\mathcal{K}$, and n be an integer. Let $\mathsf{e} = (\mathsf{EEnc}, \mathsf{EDec})$ be an $(n-\alpha, \varepsilon)$-ES scheme for arbitrary $0 \leq \alpha < n$ with key space $\mathcal{K}$, key distribution p_k and message space $\{0,1\}^n$, and $\mathsf{DTE} = (\mathsf{encode}, \mathsf{decode})$ a DTE scheme with respect to p_m that outputs an n-bit binary string. Then the DTE-then-Encrypt construction using e and DTE is an $\varepsilon'-$TDSS secure HE scheme with respect to key distribution p_k and message distribution p_m, where $\varepsilon' = \varepsilon + \mathsf{Adv}^{\mathrm{dte}}_{\mathsf{DTE},p_m}$.*

Proof. It is easy to check that the HE construction is well defined and satisfies correctness. We show that HE satisfies $\varepsilon-$TDSS security. For every adversary $\mathcal{A}$ and arbitrary predicate f, consider the following sequence of games:

Now game G_0 is exactly the same as game $\mathrm{TDSS1}^{\mathcal{A},f}_{\mathsf{HE},p_m,p_k}$. Consider the following adversary $\mathcal{D}(S)$ against DTE security: One can check that $\mathcal{D}$ simulates G_1 when $S \leftarrow\!\!\$\, \mathcal{S}$ and simulates G_0 when S is an encoding of $M \leftarrow_{pm} \mathcal{M}$. Furthermore, $\mathcal{D}$ returns 1 if and only if $\mathcal{A}$ wins in corresponding games. It follows from DTE advantage definition that

$$\Pr[G_0 \Rightarrow \text{true}] - \Pr[G_1 \Rightarrow \text{true}] \leq \mathsf{Adv}^{\mathrm{dte}}_{\mathsf{DTE},p_m} \tag{1}$$

Game G_0	Game G_1
$K \leftarrow_{p_k} \mathcal{K}$	$K \leftarrow_{p_k} \mathcal{K}$
$M \leftarrow_{p_m} \mathcal{M}$	$S \leftarrow\$ \{0,1\}^n$
$S \leftarrow\$ \mathsf{encode}(M)$	$M \leftarrow \mathsf{decode}(S)$
$C \leftarrow\$ \mathsf{EEnc}(K, S)$	$C \leftarrow\$ \mathsf{EEnc}(K, S)$
$b \leftarrow\$ \mathcal{A}(C)$	$b \leftarrow\$ \mathcal{A}(C)$
return $b = f(M)$	**return** $b = f(M)$

Figure 5 Sequence of games used in Theorem 7.

Adversary $\mathcal{D}(S)$
$K \leftarrow_{p_k} \mathcal{K}$
$M \leftarrow \mathsf{decode}(S)$
$C \leftarrow\$ \mathsf{EEnc}(K, S)$
$b \leftarrow\$ \mathcal{A}(C)$
if $b = f(M)$ **return** 1
else return 0

Figure 6 Adversary $\mathcal{D}(S)$ against DTE security.

We now work in game G_1. Note that the random variable S is uniformly sampled from $\{0,1\}^n$, therefore S has min-entropy n. By the definition of entropic security, there exists some adversary $\mathcal{A}'$ such that for all functions $\tilde{f}$,

$$\left|\Pr\left[\mathcal{A}(\mathsf{EEnc}(K,S)) = \tilde{f}(S)\right] - \Pr\left[\mathcal{A}'() = \tilde{f}(S)\right]\right| \leq \varepsilon$$

Setting $\tilde{f}(S) = f(\mathsf{decode}(S))$ we get

$$|\Pr[\mathcal{A}(\mathsf{EEnc}(K,S)) = f(M)] - \Pr[\mathcal{A}'() = f(M)]| \leq \varepsilon$$

Now $\Pr[\mathcal{A}(\mathsf{EEnc}(K,S)) = f(M)]$ is exactly the probability that $\mathcal{A}$ returns true in game G_1, in other words $\Pr[\mathcal{A}(\mathsf{EEnc}(K,S)) = f(M)] = \Pr[G_1 \Rightarrow \text{true}]$. On the other hand, we have $\Pr[\mathcal{A}'() = f(M)] \leq \Pr\left[\text{TDSS0}_{p_m}^{\mathcal{A}_s, f} \Rightarrow \text{true}\right]$ since the simulator $\mathcal{A}_s$ can simply run $\mathcal{A}'$ and return the same value as $\mathcal{A}'$ does. In other words,

$$\left|\Pr[G_1 \Rightarrow \text{true}] - \Pr\left[\text{TDSS0}_{p_m}^{\mathcal{A}_s, f} \Rightarrow \text{true}\right]\right| \leq \varepsilon \tag{2}$$

Combining 1 and 2 we have

$$\begin{aligned}\mathsf{Adv}^{\text{tdss}}_{\mathsf{HE},p_m,p_k}(\mathcal{A}, f) &= \Pr\left[\text{TDSS1}^{\mathcal{A},f}_{\mathsf{HE},p_m,p_k} \Rightarrow \text{true}\right] - \Pr\left[\text{TDSS0}_{p_m}^{\mathcal{A}_s,f} \Rightarrow \text{true}\right] \\ &\leq \varepsilon + \mathsf{Adv}^{\text{dte}}_{\mathsf{DTE},p_m} = \varepsilon'\end{aligned}$$

Since this holds for all $\mathcal{A}$ and f, we have

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

In other words, HE is $\varepsilon'-$TDSS secure. ◀

As a special case, let e be the random hashing construction in [9]:

▶ **Lemma 8** ([9]). *Let* $\{\mathsf{H}_i\}_{i\in\mathcal{I}}$ *be a pairwise independent hash family from* $\{0,1\}^n$ *to* $\{0,1\}^n$*, and* K *is sampled according to* p_k*, then the encryption scheme* $\mathsf{Enc}(K,M;i) = (i, \mathsf{H}_i(K)\oplus M)$ *is* (μ,ε)*-entropically secure for* $\mu = n - \mathrm{H}_\infty(p_k) + 2\log(1/\varepsilon) + 2$*. Specifically,* $\mathsf{Enc}(K,M;i)$ *is* $(n-\delta,\varepsilon)$*-entropically secure for* $\mathrm{H}_\infty(p_k) = \delta + 2\log(1/\varepsilon) + 2$*.*

In our HE construction we only require the entropically secure encryption scheme to support uniform distribution, therefore we can choose $\delta = 0$ for optimal parameters. This leads to the following corollary:

▶ **Corollary 9.** *Let* HE *with respect to key distribution* p_k *and message distribution* p_m *be constructed using the DTE-then-Hash construction, in which a pairwise independent hash family and an DTE scheme with advantage* $\mathsf{Adv}^{\mathrm{dte}}_{\mathsf{DTE},p_m}$ *are applied. Then* HE *is* $\varepsilon-$TDSS *secure for* $\varepsilon = 2^{(2-\mathrm{H}_\infty(p_k))/2} + \mathsf{Adv}^{\mathrm{dte}}_{\mathsf{DTE},p_m} = 2\omega_k^{1/2} + \mathsf{Adv}^{\mathrm{dte}}_{\mathsf{DTE},p_m}$*, where* ω_k *denotes the maximum key probability.*

Corollary 9 shows that we can achieve $O(\omega_k^{1/2})$ TDSS security using a *pairwise* independent hash family. Comparing this to Remark 5.6 in JRT[14] which states that TDSS security is at least at the order of $\omega_k^{1/2}$, we conclude that our construction achieves asymptotically best security bound while only requiring pairwise independent hash. This is an improvement over JRT[14]'s results, especially since we are working in standard model compared to their random oracle assumption (and also to JR[15] which only considered a weaker message recovery attack with RO).

4 Multi-Message Security

In this section, we are concerned with another drawback of information theoretic encryption besides the key length: the same key must not be used to encrypt multiple messages. Indeed, using one-time pad to encrypt two messages m_1, m_2 with the same key k yields two ciphertexts $m_1 \oplus k, m_2 \oplus k$, from which one can easily recover the value of $m_1 \oplus m_2$. We first analyze the (in)security of key reuse in entropic security, which also has implication that a honey encryption which is not carefully designed for key reuse will also be facing attacks when re-using the same key. Nevertheless, we prove that our HE construction in the standard model using pairwise independent hash further allows one to re-use a short key: this HE construction finally addresses both issues.

4.1 Insecurity of key re-use in entropic security

Entropic security leveraging message entropy helps decrease the key length, however, entropy security does not give a solution to this problem: ES schemes become insecure when a single key is used to encrypt multiple messages, even if these messages are independently sampled. Informally speaking, each encryption requires a slice of fresh randomness from the key, thus the key has to be long enough in order to provide sufficient randomness. We give an analysis on lower bound of the key needed for reuse: generalizing the analysis from single-message settings in [9], first we show that entropic security for multiple messages implies indistinguishability of multiple ciphertexts; then the lower bound can be derived from a Shannon-style bound (when we choose a special representative message distribution). This lower bound implies that a secret key in an ES scheme can only be used to encrypt very limited number of messages.

Key reuse for independent messages in entropy security. We first give a formal definition of entropic security in key reuse scenario, where a single key is used to encrypt multiple independently sampled messages. Note that the security definition becomes stronger if we remove the independence restriction; since we are after a negative result, it suffices to consider this weaker variant.

▶ **Definition 10.** *A probabilistic map $Y()$ is called (t, ε, T)-entropically secure if for all independent random variables $X_1, \cdots, X_T$ where each X_i has min-entropy at least t, and for all adversary $\mathcal{A}$, there exists some adversary $\mathcal{A}'$ such that for all functions f,*

$$|\Pr[\mathcal{A}(Y(X_1), \cdots, Y(X_T)) = f(X_1, \cdots, X_T)] - \Pr[\mathcal{A}'() = f(X_1, \cdots, X_T)]| \leq \varepsilon$$

In the first part of the proof, we will show that for a (t, ε, T)-entropically secure encryption scheme, the joint distribution of T ciphertexts (using the same key) satisfies indistinguishability definition. The latter basically requires that for any two message distributions with the same entropy, the ciphertext distribution would be indistinguishable. An alternative (and equivalent) definition that makes the following easier is that there exists one particular distribution G (that is irrelevant to the system), for all message distributions, the resulting ciphertext is indistinguishable with G. We first generalize those definitions to fit our setting of multiple messages:

▶ **Definition 11.** *A randomized map $Y()$ is (t, ε, T)-indistinguishable, if there is a random variable G, such that for every independent random variables $X_t, \cdots, X_T$ over $\{0,1\}^n$ where each X_i has min-entropy at least t, we have*

$$\mathrm{SD}((Y(X_1), \cdots, Y(X_T)), G) \leq \varepsilon$$

We prove the following lemma: an entropic secure encryption that can re-use the key for T times implies a form of indistinguishability.

▶ **Lemma 12.** *(t, ε, T)-entropic security for predicates implies $(t-1, 4T\varepsilon, T)$ -indistinguishability.*

Proof. Let $(X_1, \cdots, X_T)$ and $(X'_1, \cdots, X'_T)$ be two vectors of random variables where each X_i is independent from each X_j, each X'_i is independent from each X'_j, and each X_i, X'_i has min-entropy at least $t-1$.

First of all, it suffices to prove the indistinguishability of $(X_1, \cdots, X_T)$ and $(X'_1, \cdots, X'_T)$ when each X_i and each X'_i is a flat distribution on some set of 2^{t-1} points. Otherwise we can rewrite $(X_1, \cdots, X_T)$ and $(X'_1, \cdots, X'_T)$ as sum of distributions

$$(X_1, \cdots, X_T) = \sum_i a_i (X_{i_1}, \cdots, X_{i_T}), (X'_1, \cdots, X'_T) = \sum_j b_j (X'_{j_1}, \cdots, X'_{j_T})$$

where each coordinate X_{i_k}, X'_{j_l} is a flat distribution. $\mathrm{SD}((X_1, \cdots, X_T), (X'_1, \cdots, X'_T))$ can then be upper bounded by $\sum_{i,j} a_i b_j \mathrm{SD}(X_{i_1}, \cdots, X_{i_T}), (X'_{j_1}, \cdots, X'_{j_T})$. Therefore it suffices to show that for every pair of $(X_{i_1}, \cdots, X_{i_T}), (X'_{j_1}, \cdots, X'_{j_T})$,

$$\mathrm{SD}((X_{i_1}, \cdots, X_{i_T}), (X'_{j_1}, \cdots, X'_{j_T})) \leq 4T\varepsilon.$$

Now assume $(X_1, \cdots, X_T)$ and $(X'_1, \cdots, X'_T)$ satisfy that: for each i, X_i and X'_i are two flat distributions over *disjoint* sets of 2^{t-1} points each. Let $\tilde{X} = (\tilde{X}_1, \cdots, \tilde{X}_T)$ be sampled as follows: to sample from $\tilde{X}_i$, first sample a random bit b_i uniformly; if $b_i = 0$, sample $\tilde{X}_i$ according to X_i, and otherwise sample $\tilde{X}_i$ according to X'_i. In this way, each $\tilde{X}_i$ has

min-entropy t, and $\tilde{X}_1, \cdots, \tilde{X}_T$ are independent from each other. For every i, let f_i be the predicate which outputs 0 if $\tilde{X}_i$ is sampled according to X_i, and 1 if $\tilde{X}_i$ is sampled according to X'_i, regardless of the choices of other coordinates.

For each i, define an adversary $\mathcal{A}_i$ which, given inputs $y = (Y(\tilde{X}_1), \cdots, Y(\tilde{X}_T))$, outputs 0 if $Y(\tilde{X}_i)$ is more likely under the distribution $Y(X_i)$ than $Y(X'_i)$, and 1 otherwise. Note that the output of $\mathcal{A}_i$ is independent of the choices of $\tilde{X}_1, \cdots, \tilde{X}_{i-1}, \tilde{X}_{i+1}, \cdots, \tilde{X}_T$. Therefore the probability that $\mathcal{A}_i$ successfully predicts f_i is

$$\Pr\left[\mathcal{A}_i(Y(\tilde{X}_1), \cdots, Y(\tilde{X}_T)) = f_i(\tilde{X}_1, \cdots, \tilde{X}_T)\right] = \frac{1}{2} + \frac{1}{2}\mathrm{SD}(Y(X_i), Y(X'_i))$$

On the other hand, for any random variable G over $\{0,1\}$ independent of $\tilde{X}_i$, the probability that $G = f_i(\tilde{X}_1, \cdots, \tilde{X}_T)$ is exactly $\frac{1}{2}$. By (t, ε, T)-entropic security we get

$$\Pr\left[\mathcal{A}_i(Y(\tilde{X}_1), \cdots, Y(\tilde{X}_T)) = f_i(\tilde{X}_1, \cdots, \tilde{X}_T)\right] \leq \max_G \Pr\left[G = f_i(\tilde{X}_1, \cdots, \tilde{X}_T)\right] + \varepsilon = \frac{1}{2} + \varepsilon$$

From two inequalities above we get $\mathrm{SD}((Y(X_i), Y(X'_i))) \leq 2\varepsilon$ for every $i \in [1, T]$. Therefore,

$$\mathrm{SD}((Y(X_1), \cdots, Y(X_T)), (Y(X'_1), \cdots, Y(X'_T))) \leq \sum_{i=1}^{T} \mathrm{SD}((Y(X_i), Y(X'_i))) \leq 2T\varepsilon$$

For the case where X_i and X'_i are not disjoint, we can find a third vector $(X''_1, \cdots, X''_T)$ where each X''_i is a flat distribution on 2^{t-1} points disjoint from both X_i and X'_i. In this way,

$$\mathrm{SD}((Y(X_1), \cdots, Y(X_T)), (Y(X''_1), \cdots, Y(X''_T))) \leq 2T\varepsilon$$

$$\mathrm{SD}((Y(X'_1), \cdots, Y(X'_T)), (Y(X''_1), \cdots, Y(X''_T))) \leq 2T\varepsilon$$

We then use the triangle inequality to show that

$$\mathrm{SD}((Y(X_1), \cdots, Y(X_T)), (Y(X'_1), \cdots, Y(X'_T))) \leq 4T\varepsilon \quad \blacktriangleleft$$

The indistinguishability result can be used to bound the key size: essentially, we will choose a special distribution of vector such that each coordinate has a fixed prefix w_i, while the remaining parts are sampled uniformly. ES ciphertexts can be seen as a statistically secure encryption scheme with all the w_i as input, which gives us the desired bound.

▶ **Lemma 13.** *Any encryption scheme which is (t, ε, T)-entropically secure for inputs of length n requires a key of length at least $(n - t + 1)T - 1$.*

Proof. For every $w = (w_1, \cdots, w_T) \in \{0,1\}^{(n-t+1)T}$, where each $w_i \in \{0,1\}^{n-t+1}$, let M_{w_i} be uniformly chosen from $\{w_i\} \times \{0,1\}^{t-1}$ and $M_w = (M_{w_1}, \cdots, M_{w_T})$. Then each M_{w_i} has min-entropy $t - 1$, and any (t, ε, T)-entropically secure encryption scheme Enc produces indistinguishable distributions $(\mathsf{Enc}(M_{w_1}), \cdots, \mathsf{Enc}(M_{w_T}))$, and $(\mathsf{Enc}(M_{w'_1}), \cdots, \mathsf{Enc}(M_{w'_T}))$ for any pair (w, w'). Therefore $(\mathsf{Enc}(M_{w_1}), \cdots, \mathsf{Enc}(M_{w_T}))$ can be seen as an encryption scheme for $(n - t + 1)T$-bit strings, and thus Enc must have key length $(n - t + 1)T - 1$. ◀

▶ Remark 14. Lemma 13 implies that in an ES scheme, a key of length μ can only be used to encrypt $O(\mu)$ messages even if these messages have entropy $n - O(1)$. Therefore one cannot expect ES to remain secure in multi-message settings, especially with the use of a short key.

5 Conclusions and Future Works

In this paper, we investigate the following problem: is it possible to have an encryption scheme (for a class of messages) that satisfies unbounded semantic type of security, but using only a short key and the key can be re-used. We give an affirmative answer with a construction of honey encryption from pair-wise independent hash that satisfies both.

Approaching the problem via the lens of honey encryption inspires us to explore a nice trade-off between security and generality. We hope our initial positive results can motivate more researches on exploring message distribution for better information theoretic encryption: more general encoding mechanisms, relaxing message independence requirement in key reuse, considering integrity and more.

References

1 Esther Omolara Abiodun and Aman Jantan. Modified honey encryption scheme for encoding natural language message. *Int. J. Electr. Comput. Eng.*, 9(3):1871, 2019.

2 Thomas Agrikola, Geoffroy Couteau, Yuval Ishai, Stanislaw Jarecki, and Amit Sahai. On pseudorandom encodings. In Rafael Pass and Krzysztof Pietrzak, editors, *Theory of Cryptography - 18th International Conference, TCC 2020, Durham, NC, USA, November 16-19, 2020, Proceedings, Part III*, volume 12552 of *Lecture Notes in Computer Science*, pages 639–669. Springer, 2020. doi:10.1007/978-3-030-64381-2_23.

3 Yonatan Aumann, Yan Zong Ding, and Michael O. Rabin. Everlasting security in the bounded storage model. *IEEE Trans. Inf. Theory*, 48(6):1668–1680, 2002. doi:10.1109/TIT.2002.1003845.

4 Petra Berenbrink, Tom Friedetzky, Zengjian Hu, and Russell A. Martin. On weighted balls-into-bins games. *Theor. Comput. Sci.*, 409(3):511–520, 2008. doi:10.1016/j.tcs.2008.09.023.

5 Nikita Borisov, Ian Goldberg, and David A. Wagner. Intercepting mobile communications: the insecurity of 802.11. In Christopher Rose, editor, *MOBICOM 2001, Proceedings of the seventh annual international conference on Mobile computing and networking, Rome, Italy, July 16-21, 2001*, pages 180–189. ACM, 2001. doi:10.1145/381677.381695.

6 Christian Cachin and Ueli M. Maurer. Unconditional security against memory-bounded adversaries. In Burton S. Kaliski Jr., editor, *Advances in Cryptology - CRYPTO '97, 17th Annual International Cryptology Conference, Santa Barbara, California, USA, August 17-21, 1997, Proceedings*, volume 1294 of *Lecture Notes in Computer Science*, pages 292–306. Springer, 1997. doi:10.1007/BFb0052243.

7 Rahul Chatterjee, Joseph Bonneau, Ari Juels, and Thomas Ristenpart. Cracking-resistant password vaults using natural language encoders. In *2015 IEEE Symposium on Security and Privacy, SP 2015, San Jose, CA, USA, May 17-21, 2015*, pages 481–498. IEEE Computer Society, 2015. doi:10.1109/SP.2015.36.

8 Yan Zong Ding and Michael O. Rabin. Hyper-encryption and everlasting security. In Helmut Alt and Afonso Ferreira, editors, *STACS 2002, 19th Annual Symposium on Theoretical Aspects of Computer Science, Antibes - Juan les Pins, France, March 14-16, 2002, Proceedings*, volume 2285 of *Lecture Notes in Computer Science*, pages 1–26. Springer, 2002. doi:10.1007/3-540-45841-7_1.

9 Yevgeniy Dodis and Adam D. Smith. Entropic security and the encryption of high entropy messages. In Joe Kilian, editor, *Theory of Cryptography, Second Theory of Cryptography Conference, TCC 2005, Cambridge, MA, USA, February 10-12, 2005, Proceedings*, volume 3378 of *Lecture Notes in Computer Science*, pages 556–577. Springer, 2005. doi:10.1007/978-3-540-30576-7_30.

10 Flávio du Pin Calmon, Muriel Médard, Linda M. Zeger, João Barros, Mark M. Christiansen, and Ken R. Duffy. Lists that are smaller than their parts: A coding approach to tunable secrecy. In *50th Annual Allerton Conference on Communication, Control, and Computing,*

Allerton 2012, Allerton Park & Retreat Center, Monticello, IL, USA, October 1-5, 2012, pages 1387–1394. IEEE, 2012. doi:10.1109/Allerton.2012.6483380.

11 Stefan Dziembowski and Ueli M. Maurer. Optimal randomizer efficiency in the bounded-storage model. *J. Cryptol.*, 17(1):5–26, 2004. doi:10.1007/s00145-003-0309-y.

12 Maximilian Golla, Benedict Beuscher, and Markus Dürmuth. On the security of cracking-resistant password vaults. In Edgar R. Weippl, Stefan Katzenbeisser, Christopher Kruegel, Andrew C. Myers, and Shai Halevi, editors, *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, Vienna, Austria, October 24-28, 2016*, pages 1230–1241. ACM, 2016. doi:10.1145/2976749.2978416.

13 Zhicong Huang, Erman Ayday, Jacques Fellay, Jean-Pierre Hubaux, and Ari Juels. Genoguard: Protecting genomic data against brute-force attacks. In *2015 IEEE Symposium on Security and Privacy, SP 2015, San Jose, CA, USA, May 17-21, 2015*, pages 447–462. IEEE Computer Society, 2015. doi:10.1109/SP.2015.34.

14 Joseph Jaeger, Thomas Ristenpart, and Qiang Tang. Honey encryption beyond message recovery security. In Marc Fischlin and Jean-Sébastien Coron, editors, *Advances in Cryptology - EUROCRYPT 2016 - 35th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Vienna, Austria, May 8-12, 2016, Proceedings, Part I*, volume 9665 of *Lecture Notes in Computer Science*, pages 758–788. Springer, 2016. doi:10.1007/978-3-662-49890-3_29.

15 Ari Juels and Thomas Ristenpart. Honey encryption: Security beyond the brute-force bound. In Phong Q. Nguyen and Elisabeth Oswald, editors, *Advances in Cryptology - EUROCRYPT 2014 - 33rd Annual International Conference on the Theory and Applications of Cryptographic Techniques, Copenhagen, Denmark, May 11-15, 2014. Proceedings*, volume 8441 of *Lecture Notes in Computer Science*, pages 293–310. Springer, 2014. doi:10.1007/978-3-642-55220-5_17.

16 Burt Kaliski. PKCS #5: Password-based cryptography specification version 2.0. *RFC*, 2898:1–34, 2000. doi:10.17487/RFC2898.

17 Cheuk Ting Li and Abbas El Gamal. Maximal correlation secrecy. *IEEE Trans. Inf. Theory*, 64(5):3916–3926, 2018. doi:10.1109/TIT.2018.2816066.

18 Ueli M. Maurer. Conditionally-perfect secrecy and a provably-secure randomized cipher. *J. Cryptol.*, 5(1):53–66, 1992. doi:10.1007/BF00191321.

19 Ueli M. Maurer and James L. Massey. Local randomness in pseudorandom sequences. *J. Cryptol.*, 4(2):135–149, 1991. doi:10.1007/BF00196773.

20 Abiodun Esther Omolara, Aman Jantan, Oludare Isaac Abiodun, and Howard Eldon Poston. A novel approach for the adaptation of honey encryption to support natural language message. In *Proceedings of the International MultiConference of Engineers and Computer Scientists*, volume 1, 2018.

21 Alexander Russell and Hong Wang. How to fool an unbounded adversary with a short key. In Lars R. Knudsen, editor, *Advances in Cryptology - EUROCRYPT 2002, International Conference on the Theory and Applications of Cryptographic Techniques, Amsterdam, The Netherlands, April 28 - May 2, 2002, Proceedings*, volume 2332 of *Lecture Notes in Computer Science*, pages 133–148. Springer, 2002. doi:10.1007/3-540-46035-7_9.

22 Jeanette P. Schmidt, Alan Siegel, and Aravind Srinivasan. Chernoff-hoeffding bounds for applications with limited independence. *SIAM J. Discret. Math.*, 8(2):223–250, 1995. doi:10.1137/S089548019223872X.

23 Claude E. Shannon. Communication theory of secrecy systems. *Bell Syst. Tech. J.*, 28(4):656–715, 1949. doi:10.1002/j.1538-7305.1949.tb00928.x.

24 Junji Shikata. Formalization of information-theoretic security for key agreement, revisited. In *Proceedings of the 2013 IEEE International Symposium on Information Theory, Istanbul, Turkey, July 7-12, 2013*, pages 2720–2724. IEEE, 2013. doi:10.1109/ISIT.2013.6620721.

25 Mathy Vanhoef and Frank Piessens. Key reinstallation attacks: Forcing nonce reuse in WPA2. In Bhavani M. Thuraisingham, David Evans, Tal Malkin, and Dongyan Xu, editors, *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, CCS*

2017, Dallas, TX, USA, October 30 - November 03, 2017, pages 1313–1328. ACM, 2017. doi:10.1145/3133956.3134027.

26 Mark N. Wegman and Larry Carter. New classes and applications of hash functions. In *20th Annual Symposium on Foundations of Computer Science, San Juan, Puerto Rico, 29-31 October 1979*, pages 175–182. IEEE Computer Society, 1979. doi:10.1109/SFCS.1979.26.

A Proof for Lemma 3

Proof. We use the sequence of games in Figure 7:

Game G_0		Game G_1		Game G_2	
1 :	$K \leftarrow_{p_k} \mathcal{K}$		$K \leftarrow_{p_k} \mathcal{K}$		$R \leftarrow\$ \{0,1\}^r$
2 :	$M \leftarrow_{p_m} \mathcal{M}$		$S \leftarrow\$ \mathcal{S}$		$\tilde{C} \leftarrow\$ \mathcal{S}$
3 :	$S \leftarrow\$ \mathsf{encode}(M)$		$M \leftarrow \mathsf{decode}(S)$		$C \leftarrow (R, \tilde{C})$
4 :	$R \leftarrow\$ \{0,1\}^r$		$R \leftarrow\$ \{0,1\}^r$		$b \leftarrow\$ \mathcal{A}(C)$
5 :	$\tilde{C} \leftarrow \mathsf{H}_R(K) \oplus S$		$\tilde{C} \leftarrow \mathsf{H}_R(K) \oplus S$		$K \leftarrow_{p_k} \mathcal{K}$
6 :	$C \leftarrow (R, \tilde{C})$		$C \leftarrow (R, \tilde{C})$		$S \leftarrow \mathsf{H}_R(K) \oplus \tilde{C}$
7 :	$b \leftarrow\$ \mathcal{A}(C)$		$b \leftarrow\$ \mathcal{A}(C)$		$M \leftarrow \mathsf{decode}(S)$
8 :	**return** $b = f(M)$		**return** $b = f(M)$		**return** $b = f(M)$

Figure 7 Sequence of games used in Lemma 3.

First of all, game G_0 is exactly game TDSS1 with the HEnc part written in details. Therefore, we have $\Pr[G_0 \Rightarrow \text{true}] = \Pr\left[\text{TDSS1}^{\mathcal{A},f}_{\mathsf{HE},p_m,p_k} \Rightarrow \text{true}\right]$. By definition we have

$$\mathsf{Adv}^{\text{tdss}}_{\mathsf{HE},p_m,p_k}(\mathcal{A}, f) = \Pr[G_0 \Rightarrow \text{true}] - \omega_f \tag{3}$$

Next, the gap between G_0 and G_1 can be reduced to DTE security. Note that the only difference between G_0 and G_1 appears in line 2 and line 3. For any adversary $\mathcal{A}$, consider the following adversary $\mathcal{D}$ (in Fig 8) against DTE security: It is clear that $\mathcal{D}$ outputs 1 if and only if $\mathcal{A}$ outputs the correct bit. Furthermore, $\mathcal{D}$ simulates G_1 when $S \leftarrow\$ \mathcal{S}$, and $\mathcal{D}$ simulates G_0 when S is the DTE encoding of message $M \leftarrow_{p_m} \mathcal{M}$. It follows that:

$$\Pr[G_0 \Rightarrow \text{true}] - \Pr[G_1 \Rightarrow \text{true}] \leq \mathsf{Adv}^{\text{dte}}_{\mathsf{DTE},p_m} \tag{4}$$

Adversary $\mathcal{D}$(S)
$K \leftarrow_{p_k} \mathcal{K}$
$M \leftarrow \mathsf{decode}(S)$
$R \leftarrow\$ \{0,1\}^r$
$\tilde{C} \leftarrow \mathsf{H}_R(K) \oplus S$
$C \leftarrow (R, \tilde{C})$
$b \leftarrow\$ \mathcal{A}(C)$
if $b = f(M)$ **return** 1
else return 0

Figure 8 Adversary $\mathcal{D}(S)$ against DTE security games.

Adversary $\mathcal{A}^*(C)$

$(R, \tilde{C}) \leftarrow C$
$L_0 \leftarrow 0, L_1 \leftarrow 0$
for $K \in \mathcal{K}$ **do**
$S \leftarrow \mathsf{H}_R(K) \oplus \tilde{C}$
$M \leftarrow \mathsf{decode}(S)$
$L_{f(M)} \leftarrow L_{f(M)} + p_k(K)$
endfor
$b^* \leftarrow \mathrm{argmax}_{b\in\{0,1\}} L_b$
return b^*

Figure 9 Adversary $\mathcal{A}^*(C)$ in game G_2.

The next step is to show that G_2 is equivalent to G_1. Note that in G_1 we first sample S uniformly from $\mathcal{S}$ and independently from K, which guarantees $\tilde{C}$ also to be a uniform sample from $\mathcal{S}$ independent from K. Therefore, we can first sample $\tilde{C}$ uniformly and choose K after the execution of $\mathcal{A}$, which is exactly the case in G_2. Therefore,

$$\Pr[G_1 \Rightarrow \text{true}] = \Pr[G_2 \Rightarrow \text{true}] \tag{5}$$

Now consider the following adversary $\mathcal{A}^*$ in G_2: Adversary $\mathcal{A}^*$ adds up the probability mass of all the keys resulting in $f(M) = 0$ and $f(M) = 1$ respectively; therefore, we have $L_0 = \Pr[f(M) = 0]$ and $L_1 = \Pr[f(M) = 1]$. This implies that $\mathcal{A}^*$ is the best possible adversary in game G_2. If we denote $\Pr[G_2^* \Rightarrow \text{true}]$ to be the probability that $\mathcal{A}^*$ succeeds in G_2, we have

$$\Pr[G_2 \Rightarrow \text{true}] \leq \Pr[G_2^* \Rightarrow \text{true}] \tag{6}$$

Finally, consider Experiment $E_{p_k}^{\mathsf{H},\mathsf{DTE},f}$. For fixed choice of $(R, \tilde{C})$, the value $L_{p_k}^{\mathsf{H},\mathsf{DTE},f}$ is exactly L_{b^*} in adversary $\mathcal{A}^*$, which is the probability that $\mathcal{A}^*$ succeeds conditioned on the choice of $(R, \tilde{C})$. Taking expectation over all $(R, \tilde{C})$ gives

$$\Pr[G_2^* \Rightarrow \text{true}] = \mathbb{E}\left[L_{p_k}^{\mathsf{H},\mathsf{DTE},f}\right] \tag{7}$$

Combining 3, 4, 5, 6 and 7 gives the proof for Lemma 3. ◀

B Proof for Lemma 5 and Theorem 6

Proof. In order to bound $\mathbb{E}\left[L_{p_k}^{\mathsf{H},\mathsf{DTE},f}\right]$, we would like to give an upper bound of $\Pr\left[L_{p_k}^{\mathsf{H},\mathsf{DTE},f} \geq \alpha\right]$ for some $\alpha \in (0, 1)$. Recall that $L_{p_k}^{\mathsf{H},\mathsf{DTE},f} = \max\{p_k(B_0), p_k(B_1)\}$, where $p_k(B_0), p_k(B_1)$ represents the probability that predicate f returns 0 or 1 respectively, under random choices of K. by union bound

$$\begin{aligned}\Pr\left[L_{p_k}^{\mathsf{H},\mathsf{DTE},f} \geq \alpha\right] &= \Pr[\max\{p_k(B_0), p_k(B_1)\} \geq \alpha] \\ &\leq \Pr[p_k(B_0) \geq \alpha] + \Pr[p_k(B_1) \geq \alpha]\end{aligned} \tag{8}$$

It turns out that we only need to bound $\Pr[p_k(B_0) \geq \alpha]$ and $\Pr[p_k(B_1) \geq \alpha]$, where $p_k(B_1) = \sum_{f(M_i)=1} p_k(K_i) = \sum_{i=1}^{|\mathcal{K}|} f(M_i)p_k(K_i), p_k(B_0) = 1 - p_k(B_1) = \sum_{i=1}^{|\mathcal{K}|}(1 - f(M_i))p_k(K_i)$. At this point, the value of each $p_k(K_i)$ is fixed given p_k, therefore we are only concerned with the distributions of $f(M_i)$.

▶ **Lemma 15.** *The random variables $M_1, \cdots, M_{|\mathcal{K}|}$ are q-wise independent, and each M_i has the same distribution p_d.*

Here the q-wise independence follows from the fact that $\mathsf{H}_R(K_1), \cdots, \mathsf{H}_R(K_{|\mathcal{K}|})$ are q-wise independent (for randomly chosen R), and that each M_i is a function of $\mathsf{H}_R(K_i)$. Each M_i has distribution p_d since $\tilde{C}$ is uniformly chosen from $\mathcal{S}$ and independent from R, and therefore each S_i is uniformly chosen from $\mathcal{S}$.

Now let $p_t(b) = \Pr[f(M) = b \mid M \leftarrow_{p_d} \mathcal{M}]$ and $\omega_t = \max\{p_t(0), p_t(1)\}$. We can assume without loss of generality that $\omega_t = p_t(1) \geq p_t(0)$. It follows that $1/2 \leq \omega_t \leq 1$. In this way, $f(M_1), \cdots, f(M_{|\mathcal{K}|})$ are q-wise independent random variables satisfying for every i, $\Pr[f(M_i) = 1] = \omega_t, \Pr[f(M_i) = 0] = 1 - \omega_t$.

We can apply Lemma 4 to prove the following proposition:

▶ **Proposition 16.** *Let $\omega_k = \max_{1 \leq i \leq |\mathcal{K}|} p_k(K_i)$, and $\alpha = \omega_t + (q\omega_k\omega_t)^{1/2}e^{1/6}$. For $q \leq \omega_t e^{-1/3}/\omega_k$,*

$$\Pr[p_k(B_1) \geq \alpha] \leq e^{-\lfloor q/2 \rfloor}, \Pr[p_k(B_0) \geq \alpha] \leq e^{-\lfloor q/2 \rfloor}$$

Proof. We first prove the proposition for $p_k(B_1)$. Define $X_i = f(M_i)p_k(K_i)/\omega_k$ for $1 \leq i \leq |\mathcal{K}|$, $X = \sum_{i=1}^{|\mathcal{K}|} X_i$ and $\mu = \mathbb{E}[X]$. Since for each i, $p_k(K_i)/\omega_k$ is a constant value independent of M_i, the random variables $X_1, X_2, \cdots, X_{|\mathcal{K}|}$ are q-wise independent satisfying $\Pr[X_i = p_k(K_i)/\omega_k] = \omega_t, \Pr[X_i = 0] = 1 - \omega_t$ for all $1 \leq i \leq |\mathcal{K}|$. Therefore

$$\mu = \mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{|\mathcal{K}|} X_i\right] = \sum_{i=1}^{|\mathcal{K}|} \mathbb{E}[X_i] = \sum_{i=1}^{|\mathcal{K}|} \frac{p_k(K_i)}{\omega_k}\omega_t = \frac{\omega_t}{\omega_k}$$

Furthermore, $p_k(B_1) = \sum_{i=1}^{|\mathcal{K}|} f(M_i)p_k(K_i) = \sum_{i=1}^{|\mathcal{K}|} X_i\omega_k = \omega_k X$.

We apply Lemma 4 on X by choosing $\delta = (q/\mu e^{-1/3})^{1/2} = (q\omega_k/\omega_t)^{1/2}e^{1/6}$ and assuming that $\delta \leq 1$, which is equivalent to $q \leq \omega_t e^{-1/3}/\omega_k$. One can check that $\alpha = (1+\delta)\omega_t$. Lemma 4 states that $\Pr[X \geq (1+\delta)\mu] \leq e^{-\lfloor q/2 \rfloor}$. On the other hand,

$$\begin{aligned}\Pr[X \geq (1+\delta)\mu] &= \Pr\left[X \geq (1+\delta)\frac{\omega_t}{\omega_k}\right] \\ &= \Pr[\omega_k X \geq (1+\delta)\omega_t] = \Pr[p_k(B_1) \geq \alpha]\end{aligned}$$

Therefore $\Pr[p_k(B_1) \geq \alpha] \leq e^{-\lfloor q/2 \rfloor}$. This proves the first part of the proposition.

The second part of the proof for $p_k(B_0)$ comes in a similar fashion. This time we redefine $X_i = (1 - f(M_i))p_k(K_i)/\omega_k, X = \sum_{i=1}^{|\mathcal{K}|} X_i$ and $\mu = \mathbb{E}[X] = (1 - \omega_t)/\omega_k$. Note that $p_k(B_0) = \omega_k X$. Consider both cases of Lemma 4:

1) If $q \leq \mu e^{-1/3} = (1 - \omega_t)e^{-1/3}/\omega_k$, we choose $\delta = (q/\mu e^{-1/3})^{1/2} \leq 1$, ensuring that $q = \delta^2\mu e^{-1/3}$. Conditions of the first inequality of Lemma 4 are satisfied since q is always an integer. Therefore,

$$\Pr[p_k(B_0) \geq (1+\delta)\mu\omega_k] = \Pr[X \geq (1+\delta)\mu] \leq e^{-\lfloor q/2 \rfloor}$$

One can check that $(1+\delta)\mu\omega_k = (1-\omega_t) + (q\omega_k(1-\omega_t))^{1/2}e^{1/6} \leq \omega_t + (q\omega_k\omega_t)^{1/2}e^{1/6} = \alpha$. (inequality follows from $1/2 \leq \omega_t \leq 1$) Thus

$$\Pr[p_k(B_0) \geq \alpha] \leq \Pr[p_k(B_0) \geq (1+\delta)\mu\omega_k] \leq e^{-\lfloor q/2 \rfloor}$$

2) If $q \geq \mu e^{-1/3} = (1-\omega_t)e^{-1/3}/\omega_k$, we choose $\delta = q/\mu e^{-1/3} \geq 1$, ensuring that $q = \delta\mu e^{-1/3}$. Again, conditions of the second inequality of Lemma 4 are satisfied since q is an integer. Therefore,

$$\Pr[p_k(B_0) \geq (1+\delta)\mu\omega_k] = \Pr[X \geq (1+\delta)\mu] \leq e^{-\lfloor q/2 \rfloor}$$

This time we have $(1+\delta)\mu\omega_k = (1-\omega_t) + q\omega_k e^{1/3} \leq \omega_t + (q\omega_k\omega_t)^{1/2}e^{1/6} = \alpha$. (inequality follows from the assumption $q \leq \omega_t e^{-1/3}/\omega_k$ and $1/2 \leq \omega_t \leq 1$) Thus

$$\Pr[p_k(B_0) \geq \alpha] \leq \Pr[p_k(B_0) \geq (1+\delta)\mu\omega_k] \leq e^{-\lfloor q/2 \rfloor}$$

Combining both cases, we conclude that for $q \leq \omega_t e^{-1/3}/\omega_k$, $\Pr[p_k(B_0) \geq \alpha] \leq e^{-\lfloor q/2 \rfloor}$. This ends the proof for the second part of the proposition. ◀

For the rest of the proof assume $q \leq \omega_t e^{-1/3}/\omega_k$. From Proposition 16 and eq.8, $\Pr\left[L_{p_k}^{\mathsf{H},\mathsf{DTE},f} \geq \alpha\right] \leq 2e^{-\lfloor q/2 \rfloor}$. In this way

$$\begin{aligned}
\mathbb{E}\left[L_{p_k}^{\mathsf{H},\mathsf{DTE},f}\right] &\leq \alpha(1 - \Pr\left[L_{p_k}^{\mathsf{H},\mathsf{DTE},f} \geq \alpha\right]) + \Pr\left[L_{p_k}^{\mathsf{H},\mathsf{DTE},f} \geq \alpha\right] \\
&\leq \alpha + 2e^{-\lfloor q/2 \rfloor}(1-\alpha) \\
&= \omega_t + 2e^{-\lfloor q/2 \rfloor}(1-\omega_t) + (1 - 2e^{-\lfloor q/2 \rfloor})(q\omega_k\omega_t)^{1/2}e^{1/6}
\end{aligned}$$

Since $1/2 \leq \omega_t \leq 1$, $\mathbb{E}\left[L_{p_k}^{\mathsf{H},\mathsf{DTE},f}\right]$ is further bounded by

$$\mathbb{E}\left[L_{p_k}^{\mathsf{H},\mathsf{DTE},f}\right] \leq \omega_t + e^{-\lfloor q/2 \rfloor} + (1 - 2e^{-\lfloor q/2 \rfloor})(q\omega_k)^{1/2}e^{1/6} \tag{9}$$

This finishes the proof for Lemma 5.

From Lemma 3 and 9,

$$\begin{aligned}
\mathsf{Adv}^{\mathrm{tdss}}_{\mathsf{HE},p_m,p_k}(\mathcal{A},f) &\leq \mathsf{Adv}^{\mathrm{dte}}_{\mathsf{DTE},p_m} + \mathbb{E}\left[L_{p_k}^{\mathsf{H},\mathsf{DTE},f}\right] - \omega_f \\
&\leq \mathsf{Adv}^{\mathrm{dte}}_{\mathsf{DTE},p_m} + \omega_t + e^{-\lfloor q/2 \rfloor} + (1 - 2e^{-\lfloor q/2 \rfloor})(q\omega_k)^{1/2}e^{1/6} - \omega_f
\end{aligned} \tag{10}$$

for all $\mathcal{A}, f$ and $q \leq e^{-1/3}/2\omega_k$.

Now consider the following adversary $\mathcal{D}_f$ against the DTE security game: on input S, $\mathcal{D}_f$ decodes S, applies f to the decoded message and outputs the f-value obtained. One can check that $\Pr\left[\mathrm{SAMP1}^{\mathcal{D}_f}_{\mathsf{DTE},p_m} \Rightarrow \mathrm{true}\right] = \omega_f$ and $\Pr\left[\mathrm{SAMP0}^{\mathcal{D}_f}_{\mathsf{DTE}} \Rightarrow \mathrm{true}\right] = \omega_t$. By the definition of DTE advantage $|\omega_f - \omega_t| \leq \mathsf{Adv}^{\mathrm{dte}}_{\mathsf{DTE},p_m}$. This combining with inequality (10) gives

$$\mathsf{Adv}^{\mathrm{tdss}}_{\mathsf{HE},p_m,p_k}(\mathcal{A},f) \leq 2\mathsf{Adv}^{\mathrm{dte}}_{\mathsf{DTE},p_m} + e^{-\lfloor q/2 \rfloor} + (1 - 2e^{-\lfloor q/2 \rfloor})(q\omega_k)^{1/2}e^{1/6} \tag{11}$$

Notice that the right hand side of inequality (11) is independent of the choice of $(\mathcal{A}, f)$. This finishes the proof of Theorem 6. ◀