Key Derivation for Squared-Friendly Applications: Lower Bounds

Maciej Skorski
maciej.skorski@mimuw.edu.pl

Abstract

Security of cryptographic applications is typically defined by security games. The adversary, within certain resources, cannot win with probability much better than 0 (for unpredictability applications, like one-way functions) or much better than $\frac{1}{2}$ (indistinguishability applications for instance encryption schemes). In so called *squared-friendly applications* the winning probability of the adversary, for different values of the application secret randomness, is not only close to 0 or $\frac{1}{2}$ on average, but also concentrated in the sense that its second central moment is small. The class of squared-friendly applications, which contains all unpredictability applications and many indistinguishability applications, is particularly important in the context of key derivation. Barak et al. observed that for square-friendly applications one can beat the "RT-bound", extracting secure keys with significantly smaller entropy loss. In turn Dodis and Yu showed that in squared-friendly applications one can directly use a "weak" key, which has only high entropy, as a secure key.

In this paper we give sharp lower bounds on square security assuming security for "weak" keys. We show that *any* application which is either (a) secure with weak keys or (b) allows for saving entropy in a key derived by hashing, *must* be square-friendly. Quantitatively, our lower bounds match the positive results of Dodis and Yu and Barak et al. (TCC'13, CRYPTO'11) Hence, they can be understood as a general characterization of squared-friendly applications.

While the positive results on squared-friendly applications where derived by one clever application of the Cauchy-Schwarz Inequality, for tight lower bounds we need more machinery. In our approach we use convex optimization techniques and some theory of circular matrices.

Keywords and phrases key derivation, square-friendly applications, lower bounds

1 Introduction

When analyzing the security of cryptographic primitives one typically assumes the access to *perfect* randomness. In practice, we are often limited to *imperfect* sources of randomness.

1.1 Key derivation

IDEAL AND REAL SETTINGS. For any cryptographic primitive (like encryption or signatures), which needs a "random" m-bit string R to sample the secure key[1], we compare two different settings:

(a) ideal setting: R is *perfectly* random: uniform and independent of any side information available to the attacker
(b) real settings: there is only an *imperfect* entropy source X and the secure key R needs to be derived from X. The attacker may have some *side information* about X, in particular the additional randomness used to derive R from X.

[1] In applications like block-ciphers R is simply the key. In other applications like RSA R is used to sample public or secret keys. We will simply refer to R as the key.

The security of the primitive is parametrized by ϵ, which is the success probability (for so called unpredictability applications) or the advantage (for so called indistinguishability applications) of an attacker with certain resources[2].

Generic approach and the entropy loss. The general way to derive a secure key is to "extract" the randomness from X by a seeded extractor. In particular, the Leftover Hash Lemma implies that if the min-entropy of X is at least $m + L$ then $H(X), H$, where H is randomly chosen function from a universal family, is δ-close to uniform with $\delta = \sqrt{2^{-L}}$. This means that if an application is ϵ-secure for uniform R, then the same application keyed with $R = H(X)$, and even published H, is ϵ'-secure with

$$\epsilon' \leqslant \epsilon + \sqrt{2^{-L}}. \tag{1}$$

where the entropy loss L is the difference between the entropy of X and m. Note that from Equation (1) it follows that we need $L = 2\log(1/\epsilon)$ to obtain (roughly) the same security $\epsilon' = 2\epsilon$. Unfortunately, if we want the security against *all statistical tests*, this loss is necessary for *any* extractor, as implied by the so called "RT-bound" [8].

Need for better techniques for cryptographic applications. The RT-bound does not exclude the possibility of deriving a secure key wasting *much less* than $2\log(1/\epsilon)$ bits of entropy for *particular* applications. Saving the entropy, apart from scientific curiosity, is a problem of real-word applications. Minimizing the entropy loss is of crucial importance for some applications where it affects efficiency (for example in the elliptic-curve Diffie-Hellman key exchange) and sometimes the entropy amount we have is bounded (e.g. biometric data) than the required length of a key; see also the discussion in [2]. Hence, better techniques than simple extracting are desired. Below we discuss what is known about possible improvements in key derivation for cryptographic applications.

Key derivation for unpredictability applications. It is known that unpredictability applications directly tolerate weak keys, provided that the entropy deficiency is not too big. More precisely, any unpredictability application which is ϵ-secure with the uniform m-bit key, is also $\epsilon' = 2^d\epsilon$-secure for any key of entropy $m - d$. If we have a source X that has "enough" entropy but its length is too big, we can condense it to a string of length m with almost full entropy. Essentially, since we achieve a very good condensing rate: any X of $m + \log\log(1/\epsilon)$ bits of entropy can be condensed to an m bit string with the entropy deficiency $d = 3$ which is $\epsilon' = 2^3\epsilon$-close to uniform[3], we are able to derive a key (roughly) equally secure as the uniform key, with the entropy loss only $L = \mathcal{O}(\log\log(1/\epsilon))$, i.e. actually without entropy waste [5].

Key derivation for indistinguishability applications. The situation for indistinguishability applications is completely different. For the one-time pad which needs an m-bit uniform key, a key of even $m-1$ bits of entropy might be insecure [2]. For some applications we can overcome this difficulty if the winning probability of the adversary, as a function of the key, is not only close to $1/2$ on average, but also *concentrated* around $1/2$. Recall that the advantage of an attacker, for a particular key, is defined as the difference[4] between the winning probability and $1/2$. One introduces the following two interesting properties:

(a) strong security: the *absolute* advantage is small on average (close to the advantage)

[2] For example bounded running time, circuit size or the number of oracle queries.
[3] Thus, for condensing we lose incomparably less in the amount than for extracting.
[4] In indistinguishability games an adversary needs to guess a bit at the end of the game. Since he can flip his answer, any bias indicates that his guess is better than a random answer.

(b) square security: the *squared* advantage is small on average (close to the advantage)
Property (a) provides basically the same bounds as for the unpredictability applications. Namely, we can apply a weak key directly, losing a factor 2^d in the security where d is the entropy deficiency. Unfortunately, this holds only for a very limited class of applications. Property (b) offers slightly worse bounds but is satisfied for a wide class of indistinguishability applications, called "squared-friendly". One can use a "weak key" *directly* with a squared-friendly application achieving security of roughly $\sqrt{2^d\epsilon}$ where ϵ is the security with the uniform key and d is the entropy deficiency [6]. Alternatively, if we want to obtain security $\mathcal{O}(\epsilon)$ instead of $\mathcal{O}(\sqrt{\epsilon})$, one can use universal hashing to extract an ϵ-secure key with the entropy loss reduced by half [2] , i.e. up to $L=\log(1/\epsilon)$. The improvement in the security analysis over the "standard" Leftover Hash Lemma (LHL) comes from restrictions on the class of the test functions, imposed by the squared-friendly assumption.

1.2 Our results

In what follows we assume that P is an arbitrary indistinguishability application which needs an m-bit uniform key. We give *tight* lower bounds on the amount of square security (the expected square of the attacker's advantage) or strong-security (the expected absolute average of the attacker's advantage) that is necessary for an application to be secure with weak keys, that is keys with entropy deficiency. The notion of entropy here is either the min-entropy or the collision entropy. Collision entropy is less restrictive than min-entropy and is a natural choice to applications involving hash functions, like the LHL[5]. It is equally good for squared-friendly applications as min-entropy. Therefore, as remarked in [6], results for collision entropy are more desired. Nevertheless, we provide bounds for both entropy notions[6].

Summary of our contribution. We characterize squared-friendly applications by their "nice" features. Namely, we show that square-friendly applications are *precisely* those applications which are secure with weak keys or offers improvements in the entropy loss for a key derived by the LHL. Hence the current state of art is optimal: we cannot do better key derivation than for squared-friendly applications unless we build a theory on stronger than collision entropy requirements for weak keys (which would be in some sense inconvenient because of a natural connection between collision entropy and hash functions).

Any application secure with weak keys has large square-security. The following results was proved by Dodis and Yu:

▶ **Theorem** ([6]). *Applications which are σ-square-secure with the uniform key, i.e. when the averaged squared advantage of any attacker is less than σ, are $\epsilon=\sqrt{2^d\sigma}$-secure with any key of collision entropy at least $m-d$.*

The following question is therefore natural

> **Q**: If P is secure for all keys of *high (collision or min-) entropy*, how much square-security does it have?

We give an answer in the following two theorems. The first is actually trivial and perhaps known in folklore.

---

[5] For some applications we really need the LHL because of its simplicity, efficiency and nice algebraic features [2].

[6] Actually collision entropy is more challenging and our observations on strong security are known in folklore, but we study also the min-entropy case for the sake of completeness.

▶ **Theorem.** *Let $d \geqslant 1$. Suppose that P is ϵ-secure with any key of min-entropy at least $m - d$. Then P is ϵ'-strongly secure with $\epsilon' = \mathcal{O}(\epsilon)$.*

The second one is more interesting

▶ **Theorem** (Informal)**.** *Let $d \geqslant 1$. If P is ϵ-secure with any key of collision entropy at least $m - d$, then P is σ-square-secure with $\sigma = \mathcal{O}(\epsilon^2)$.*

The bounds in both cases are tight. Note that if the entropy deficiency d is bounded then our lower bound perfectly (up to a constant factor) matches the result of Dodis and Yu for *any* P.

Square Security is necessary to improve key derivation by condensing collision entropy. In the previous paragraph, we discussed the case when the entropy deficiency d is bounded away from 0. However, sometimes we intentionally extremely condense collision entropy so that this gap is close to 0, to achieve better than $\mathcal{O}(\sqrt{\epsilon})$ security at the price of starting with more than m-bits of entropy. For ϵ-secure square friendly applications one can derive by universal hashing a (roughly) ϵ-secure key from any source having $m + \log(1/\epsilon)$ bits of min-entropy (or even collision entropy) [2]. Let us briefly discuss this result. The proof of the classical Leftover Hash lemma consists of two separate claims:

(a) *Universal hash functions can extremely condense collision entropy.*
(b) *Distributions of extremely high collision entropy are close to uniform.*
More precisely, in the first step one applies a random function from a universal family to "condense" the collision entropy of X from $m+L$ bits, where $L \gg 0$, to an m-bit string with $m - \log(1 + 2^{-L}) \approx m - 2^{-L}$ bits of collision entropy [7]. In the next step one shows that any m-bit random variable with collision entropy at least $m - \epsilon^2$ is ϵ-close to uniform. Thus, $L = 2\log(1/\epsilon)$ is enough to obtain ϵ-security. As observed by Barak et al. [2], for ϵ-secure applications which are *in addition* ϵ-square-secure, it suffices to start with $m - \epsilon$ bits of the collision entropy in step (b), which reduces by half, i.e. up to $L = \log(1/\epsilon)$ (comparing to the RT-bound), the entropy loss needed to achieve ϵ-closeness.

▶ **Theorem** ([2])**.** *Suppose that P with a uniform m-bit key is ϵ-secure and σ-square secure. Let R be any key of collision entropy at least $m - d$ (possibly given some side information). Then P keyed with R is ϵ'-secure with $\epsilon' = \epsilon + \sqrt{\sigma(2^d - 1)}$, even if the used hash function is published. In particular, for $\sigma = \mathcal{O}(\epsilon)$ and $d = \mathcal{O}(\epsilon)$ we obtain $\epsilon' = \mathcal{O}(\epsilon)$.*

Applying this to R being X condensed by universal hashing we get

▶ **Corollary** ([2])**.** *Suppose that P with a uniform m-bit key is ϵ-secure and σ-square secure (that is, average squared advantage of attackers is not bigger than σ). Suppose that X has min-entropy (or collision entropy) at least $m + L$ and let R be an m-bit key derived by universal hashing. Then P keyed with R is ϵ'-secure with $\epsilon' = \epsilon + \sqrt{2^{-L}\sigma}$, even if the used hash function is published. In particular, $\epsilon' = \mathcal{O}(\epsilon)$ for $\sigma = \mathcal{O}(\epsilon)$ and $L = \log(1/\epsilon)$.*

The first result motivates the following question about weak keys with the entropy deficiency close to 0.

> **Q**: Suppose that an application P is secure for all keys of *extremely condensed collision entropy*, possibly given side information. How much square-security does P have?

[7] conditioned on the choice of the function, which can be thought as a *seed* for the condenser.

We give an answer in the following theorem

▶ **Theorem** (Informal). *Let $d \ll 1$ and suppose that P is ϵ-secure with all keys of collision entropy at least $m-d$ (possibly given side information, like the condenser's seed). Then P is σ-square secure with $\sigma = \mathcal{O}(\max(d, \epsilon^2/d))$.*

Our theorem, applied for $d = \epsilon$, shows the full converse of the observation of Barak et al. A good illustrative example is the case of the Leftover Hash Lemma. As mentioned, universal hash functions condense $m + \log(1/\epsilon)$ bits of entropy into an m-bit string with $m - \epsilon$ bits of entropy. If we use universal hash functions *only as a condenser* (which is exactly how we use them in the LHL), then we have a "black-box" equivalence between distributions of collision entropy at least $m-\epsilon$ and hashes of distributions having at least $m+\log(1/\epsilon)$ bits of entropy[8]. If follows then that we want to reduce the entropy loss by half to $L = \log(1/\epsilon)$ and achieve ϵ-security, then our application must be ϵ-square secure. This lower bound matches the positive result of Barak et al. [2] and, since is holds for *any* application, can be viewed as a *general* characterization of squared-friendly applications.

Square security is necessary for reducing the entropy loss in the LHL. As remarked in the discussion in the previous paragraph, we can *heuristically* identify the set of randomly "hashed" high entropy distributions with the set of distributions of extremely high collision entropy (conditioned on the choice of a hash function as the uniform "seed"). This "equivalence", is reasonable for the "black-box" use of hash functions. However, it is natural to ask if we can prove it formally. That is, we ask if square security is necessary for improvements in the entropy loss for key derived not by a general "black-box" collision entropy condenser but precisely by *hashing*.

> **Q**: Suppose that an application P is secure for any key derived by applying a randomly chosen (almost) universal hash function to a min-entropy source, even if the hash function is published. Suppose that the entropy loss vs security trade-off is significantly better than (pessimistic) RT-bound. Is P square-secure?

We give an affirmative answer and a lower bound that (almost) matches the results of [2] for *any* application.

▶ **Theorem** (The improved LHL [2] is tight for any application, informal). *Let $\epsilon = 2^{-(1-\beta)m}$ where β is some small positive number. Then there exists an ϵ-universal family $\mathcal{H}$ of hash functions from n to m bits, efficiently commutable and samplable with the use of n^2 uniformly random bits, with the following property: for any application P, if for every source X of min-entropy at least $k = m + \log(1/\epsilon)$ and H chosen randomly from $\mathcal{H}$ we have that P is secure with the key $H(X)$ and published H, then P must be σ-square-secure with $\sigma = \epsilon^{\frac{1-3\beta/2}{1-\beta}}$.*

This theorem for β close to 1 (exponential but meaningful security) shows that $\epsilon^{1-o(1)}$-square-security is *necessary* for saving $\log(1/\epsilon)$ bits of entropy in deriving an ϵ-secure key by universal hashing (which is almost tight since ϵ-square-security is enough.

Square-security bounds are generally optimal. The improved bound for applications which are square-secure is generally optimal, as observed in [5]. We provide an alternative proof of this result, using our techniques.

[8] Because the only information that a general condenser provides is about the entropy in its output.

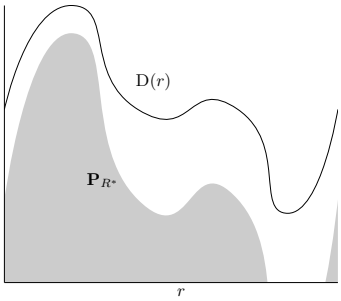

Figure 1 The shape of the best-advantage key distribution (under collision entropy constraints). In application the function $\mathrm{D}(r)$ equals the advantage of an attacker on the key r.

▶ **Theorem 1** ([5]). *For any n, $k \leqslant n$, and $\delta \in (0,1)$ there exists an application which has δ-square security but for some key of Renyi entropy at least k it achieves only security $\epsilon = \Omega\left(\sqrt{2^{n-k}\delta}\right)$.*

The proof appears in Appendix G. The advantage of our proof is that it abstracts a general condition for this bound to be satisfied. In particular, similar bounds can obtained for all so called "strongly secure" applications, where attacker's advantage is nearly zero except a tiny subset of "weak" keys where the attacker wins with heavy advantage.

1.3 Our techniques

Our main technical contribution is an explicit characterization of a distribution which maximizes the expectation of a given function, subject to collision entropy constraints. We show that the worst-case distribution has the shape similar to the function, up to a transform which involves taking a threshold and scaling, as illustrated in Figure 1. We apply this characterization to settings where we want to find the distribution of keys which maximizes the attacker advantage. We stress that with our characterization one can compute *optimal* security with weak keys. Previous works [2, 6] obtained good bounds with the Cauchy-Schwarz inequality only, however these techniques cannot be extended to obtain optimal or lower bounds, as we do.

1.4 Organization of the paper

In Section 2 we provide the basic notations and definitions for security, square-security and entropy. In Section 3 we state the known positive result. Our main auxiliary result on optimization problems with collision entropy constraints is presented in Section 4. The lower bounds are given in Section 5.

2 Preliminaries

Basic notions. The min entropy of X is $\mathbf{H}_\infty(X) = \log(1/\max_x \Pr[X=x])$. The collision probability of X is $\mathrm{CP}(X) = \sum_x \Pr[X=x]^2$, that is $\mathrm{CP}(X) = \Pr[X = X']$ where X' is an independent copy of X. The collision entropy of X is $\mathbf{H}_2 = \mathrm{CP}(X)$ and the conditional collision entropy $\mathbf{H}_2(X|Z)$ equals $-\log\left(\mathbf{E}_{z\leftarrow Z}\,\mathrm{CP}(X|_{Z=z})\right)$. The statistical distance of X and Y (taking values in the same space) is $\Delta(X;Y) = \sum_x |\Pr[X=x] - \Pr[Y=x]|$.

Security of indistinugishability and unpredictability applications. Consider

any application whose security is defined by a *security game* between an attacker A and a challenger $\mathsf{C}(r)$, where r is an m-bit key derived from U_m in the "ideal" setting and from some distribution R in the "real" setting. For every key r we denote by $\mathsf{Win}_{\mathsf{A}}(r)$ the probability (over the randomness used by A and C) that the adversary A wins the game when challenged on the key r. The advantage of the adversary A on the key r is defined, depending on the type of the application (unpredictability, indistinugishability) as follows:

$$\mathsf{Adv}_{\mathsf{A}}(r) \stackrel{\text{def}}{=} \mathsf{Win}_{\mathsf{A}}(r), \qquad \text{(unpredictability)} \tag{2}$$

$$\mathsf{Adv}_{\mathsf{A}}(r) \stackrel{\text{def}}{=} \mathsf{Win}_{\mathsf{A}}(r) - \frac{1}{2}, \qquad \text{(indistingusihability)} \tag{3}$$

Now we define the security in the ideal and real models as follows:

▶ **Definition 2** (Security in the ideal and real model)**.** An application $\mathcal{P}$ is (T,ϵ)-secure in the ideal model if

$$|\mathop{\mathbf{E}}_{r\leftarrow U_m} \mathsf{Adv}_{\mathsf{A}}(r)| \leqslant \epsilon \tag{4}$$

for all attackers A with resources less than T. We say that $\mathcal{P}$ is (T,ϵ)-secure in the $(m-d)$-real$_2$ if for every distribution R of collision entropy at least $m-d$

$$|\mathop{\mathbf{E}}_{r\leftarrow R} \mathsf{Adv}_{\mathsf{A}}(r)| \leqslant \epsilon, \tag{5}$$

for all attackers A with resources less than T.

▶ Remark 1 (Strong security in the ideal model). If $\mathbf{E}_{r\leftarrow U_m} |\mathsf{Adv}_{\mathsf{A}}(r)| \leqslant \epsilon$ in the above setting then we say that $\mathcal{P}$ is (T,ϵ)-strongly secure (in the ideal model).

Square security. Finally, we define the notion of square-security (in the ideal model)

▶ **Definition 3** (Square security)**.** An application $\mathcal{P}$ is (T,ϵ)-square-secure if

$$\mathop{\mathbf{E}}_{r\leftarrow U_m} \mathsf{Adv}_{\mathsf{A}}(r)^2 \leqslant \epsilon, \tag{6}$$

for all attackers A with resources less than T.

Security in the presence of side information. Sometimes we need to consider stronger adversaries, which has additional information S. For example, this is always the case where the weak key has been derived from an entropy source using *public* randomness.

▶ **Definition 4** (Security in the presence of side information)**.** Given a side information $S\in\mathcal{S}$, an application $\mathcal{P}$ is (T,ϵ)-secure in the $(m-d)$-real$_2$ model if for every distribution R such that $\mathbf{H}_2(R|S) \geqslant m-d$ we have

$$\max_{s\in\mathcal{S}} |\mathop{\mathbf{E}}_{r\leftarrow R} \mathsf{Adv}_{\mathsf{A}}(r,s)| \leqslant \epsilon, \tag{7}$$

for all attackers A with resources less than T. In the ideal model $\mathcal{P}$ is (T,ϵ)-secure if $\max_{s\in\mathcal{S}} |\mathbf{E}_{r\leftarrow U_m} \mathsf{Adv}_{\mathsf{A}}(r,s)| \leqslant \epsilon$ and respectively (T,ϵ)-square-secure if $\max_{s\in\mathcal{S}} |\mathbf{E}_{r\leftarrow U_m} \mathsf{Adv}_{\mathsf{A}}(r,s)^2| \leqslant \epsilon$ for all attackers A with resources less than T.

▶ Remark 2. Note that in the nonuniform setting, security and square security in the ideal model with and without side information coincide.

3 Square security- positive results

Improved key derivation for square-secure applications. Let D be an arbitrary real-valued function on $\{0,1\}^m$ and let Y be an arbitrary m-bit random variable with collision entropy $\mathbf{H}_2(X) \geqslant m - d$. By the Cauchy Schwarz Inequality one obtains [2, 6] the following inequalities

$$\mathbf{E}\,\mathrm{D}(Y) \leqslant \sqrt{\mathbf{E}\,\mathrm{D}(U_m)^2} \cdot \sqrt{2^d}, \tag{8}$$

$$\mathbf{E}\,\mathrm{D}(Y) - \mathbf{E}\,\mathrm{D}(U_m) \leqslant \sqrt{\mathrm{VarD}(U_m)} \cdot \sqrt{2^d - 1}. \tag{9}$$

When the side information S is present, and $\mathbf{H}_2(Y|S) \geqslant m - d$, we get

$$\mathbf{E}\,\mathrm{D}(Y,S) \leqslant \sqrt{\mathbf{E}\,\mathrm{D}(U_m,S)^2} \cdot \sqrt{2^d}, \tag{10}$$

$$\mathbf{E}\,\mathrm{D}(Y) - \mathbf{E}\,\mathrm{D}(U_\mathcal{Y}) \leqslant \sqrt{\mathop{\mathbf{E}}_{s\leftarrow S} \mathrm{VarD}(U_m,s)} \cdot \sqrt{2^d - 1}. \tag{11}$$

These inequalities, applied to $\mathrm{D} = \mathsf{Adv}_\mathsf{A}$ link the security in the real model with the entropy deficiency of a weak key and the security in the ideal model. In particular, one obtains the following results, already mentioned in Section 1.1

▶ **Theorem 5** ([6]). *Suppose that $\mathcal{P}$ is (T,σ)-square secure in the ideal model. Then it is (T,ϵ) secure in the $(m-d)$-real$_2$ model with $\epsilon = \sqrt{2^d\sigma}$.*

▶ **Theorem 6** ([2, 6]). *Suppose that an application $\mathcal{P}$ in the ideal model is (T,ϵ)-secure and (T,σ)-square-secure. Then it is (T,δ) secure in the real $(m-d)$-model with $\delta = \epsilon + \sqrt{(2^d-1)\sigma}$.*

Theorem 5 states that a weak key can be used directly in a square-secure application provided that the entropy deficiency is not too big. The second theorem deals with the case where the deficiency is *extremely* small. It is essentially important when one notices that *universal hash functions* condense collision entropy at a very good rate. Theorem 6 yields the following important corollary

▶ Corollary 1 (Improved LHL, [2]). Suppose that $\mathcal{P}$ is as above. Let X be an n-bit random variable of collision entropy at least $m+L$, let $\mathcal{H}$ be a $\frac{1+\gamma}{2^m}$-universal family of functions from n to m bits and let H be a random member of $\mathcal{H}$. Then $\mathcal{P}$ keyed with $H(X)$ is ϵ'-secure with $\epsilon' \leqslant \epsilon + \sqrt{\sigma\,(\gamma + 2^{-L})}$ against all adversaries with resources T and given H. In other words, for all A with resources T we have

$$\mathop{\mathbf{E}}_{(r,h)\leftarrow H(X),H} \mathsf{Adv}_\mathsf{A}(r,h) \leqslant \epsilon + \sqrt{\sigma\,(\gamma + 2^{-L})}. \tag{12}$$

In particular, for $\gamma \leqslant \epsilon$ and $\sigma \leqslant 4\epsilon$ we achieve security $\epsilon' \leqslant 3\epsilon$ with only $Ł = \log(1/\epsilon)$ bits of the entropy loss.

Summing up, when we want to derive a secure key for an ϵ-square-secure application from a source X, we have two options

(a) We condense (if necessary) X by hashing into a string with small entropy deficiency. From a source which has $m - \mathcal{O}(1)$ bits of entropy we derive a $\mathcal{O}(\sqrt{\epsilon})$-secure key.
(b) If we want more security, we can condense X even stronger, with deficiency extremely close to 0, sacrificing some entropy amount. From a source which has $m+\log(1/\epsilon)-\mathcal{O}(1)$ bits of entropy we derive a $\mathcal{O}(\epsilon)$-secure key.

In every case we obtain the meaningful security, in particular even if entropy amount we start with is *smaller* than the length of the key we need. The application of a generic extractor in such a case gives *no* security guarantee! For more examples and applications we refer the reader to [6] and [2].

SECURITY AND SQUARE SECURITY - MATHEMATICAL INSIGHT. It is worth of mentioning that the idea behind square security is, conceptually, simple and natural. All we need is the *concentration* of the adversary's winning probability, which is guaranteed by the small first central or second central moment.

WHAT APPLICATIONS ARE SQUARE-SECURE? It is known that PRGs, PRFs and one-time pads cannot have good square security [3]. In turn, many applications such as such as stateless chosen plaintext attack (CPA) secure encryption and weak pseudo-random functions (weak PRFs), can be proven to be "square-friendly" that is they have square-security roughly the same as the standard security. The general method to prove that security implies square-security is the so called "double run trick" [2, 6].

4 Optimization: auxiliary results

Our main technical tool is a characterization of a distribution that maximizes the expectation of a given function under the collision entropy constraints. It has a nice geometrical interpretation, as the best shape is simply a combination of a threshold and scaling transformation, see Figure 1.

▶ Lemma 1 (Maximizing the expectation subject to collision entropy constraints). Let $\mathrm{D} : S \to [0,1]$ be a function on a finite set S and let Y^* be any optimal solution to the following problem

$$\begin{aligned} \underset{Y}{\text{maximize}} \quad & \mathbf{E}\,\mathrm{D}(Y) \\ \text{subject to} \quad & \mathbf{H}_2(Y) \geqslant k \end{aligned} \tag{13}$$

where the maximum is taken over all random variables Y taking values in S. Then there exist numbers $\lambda \geqslant 0$ and $t \in \mathbb{R}$ such that Y^* satisfies the following condition

$$\max(\mathrm{D}(x) - t, 0) = \lambda \mathbf{P}_{Y^*}(x) \quad \text{for all } x \in S. \tag{14}$$

In particular $\mathrm{Var}\mathrm{D}'(U) = \frac{\lambda^2}{|S|^2}$ where $\mathrm{D}'(x) = \max(\mathrm{D}(x) - t, 0)$. Moreover, if values of $\mathrm{D}(\cdot)$ are all different, then we have $\lambda > 0$ and λ, t are unique.

▶ Remark 3. If values of $\mathrm{D}(\cdot)$ are all different, then $\lambda > 0$ (see Appendix B).

▶ Corollary 2. We have the following identities $\mathbf{E}\,\mathrm{D}'(U_m) = \frac{\lambda}{|S|}$, $\mathbf{E}\,\mathrm{D}'(U_m)^2 = \frac{1+\theta}{|S|^2\lambda^2}$, and $\mathbf{E}\,\mathrm{D}'(Y^*) = \frac{(1+\theta)\lambda}{|S|}$ (D, θ, Y^*, λ, t and D' are as in Lemma 1).

5 Square security- lower bounds

5.1 Weak keys with the entropy deficiency bounded away from 0

We start with the following results, which states that every indistinguishability application which is secure with all keys of high min-entropy must be *strongly secure*. The proof is straightforward but for the sake of the completeness we give it in Appendix A.

▶ **Theorem 7.** *Suppose that an indistinguishability application P, which needs an m-bit key, is (T, ϵ)-secure in the $(m-d)$-real$_\infty$ model, for some $d \geqslant 1$. Then P is $(T, 2\epsilon)$-strongly secure. The bound 2ϵ here is tight.*

More challenging and more interesting is the case of an application secure with all keys of high collision entropy.

▶ **Theorem 8.** *Suppose that an indistinguishability application P, which needs an m-bit key, is (T, ϵ)-secure in the $(m-d)$-real$_2$ model, for some $d \geqslant 1$. Then P is (T, σ)-square-secure with $\sigma = 4\epsilon^2$*

Note that for bounded d the level of square security perfectly matches the positive result of Dodis and Yu (Theorem 5). We also show (see the end remark in the proof) that this bound is tight up to a constant factor and thus we cannot get the bound $\mathcal{O}\left(\epsilon^2/2^d\right)$, which would exactly match to the positive result in Theorem 5 for all d. The proof is heavily based on Lemma 1 and appears in Appendix C.

5.2 Weak keys with the entropy deficiency close to 0

Below we provide a lower bound when the entropy deficiency is close to 0.

▶ **Theorem 9.** *Suppose that P, which uses an m-bit key, is (T, ϵ)-secure in the $(m-d)$-real$_2$ model (possibly with side information). Then P is σ-square-secure with $\sigma \leqslant \epsilon^2 + \max\left(2^d - 1, \frac{4\epsilon^2}{2^d - 1}\right)$. In particular, if $d \ll 1$ then $\sigma \leqslant 2\max(d, \frac{\epsilon^2}{d})$.*

The proof is based on Lemma 1 and is given in Appendix D. From this we see that Theorem 6 for $d = \epsilon$ is tight.

5.3 Leftover Hash Lemma as a Key Derivation Function

Finally, we consider the case of a key derived by hashing.

▶ **Theorem 10.** *Let $\alpha \in [1, 2]$ and let $\epsilon > 0$. Suppose that an application P, which uses an m-bit secure key, has the following property: for every n-bit source X of min-entropy $k \geqslant m + \alpha \log(1/\epsilon)$, and every efficient ϵ^α-universal family $\mathcal{H}$ of hash functions from n to m bits, we have*

$$\mathbf{E}\,\mathsf{Adv}_{\mathsf{A}}(H(X), H) \leqslant C\epsilon,$$

for some constant C and all adversaries A with resources at most T. Then P is (T, σ)-square-secure with

$$\sigma \leqslant \frac{3}{2} \cdot \max\left(2^{-m/2}\epsilon^{\alpha/2}, 4(C+1)^2 2^{m/2}\epsilon^{2-\alpha/2}\right). \tag{15}$$

For $\epsilon > 2^{-m}$ we get $\sigma = \mathcal{O}\left(2^{m/2}\epsilon^{2-\alpha/2}\right)$. In particular, if $\alpha = 1$ and $\epsilon = 2^{-(1-\beta)m}$ for some $\beta > 0$ then $\sigma = \mathcal{O}\left(2^{-(1-3\beta/2)m}\right) = \mathcal{O}\left(\epsilon^{\frac{1-3\beta/2}{1-\beta}}\right)$. Thus, *any* application P which allows deriving an ϵ'-secure key with $\epsilon' = \mathcal{O}(\epsilon)$ and entropy loss $L = \log(1/\epsilon)$ must be $\sigma = \mathcal{O}\left(\epsilon^{1-o(1)}\right)$-square-secure. On the positive side we know that σ-square-security with $\sigma = \epsilon$ is enough (Corollary 1).

▶ Corollary 3 (The Improved LHL is tight for any application). For *any* application P, the security guarantee in the improved Leftover Hash Lemma (Corollary 1) cannot be improved by more than a factor $\epsilon^{o(1)}$. Note that we require $\mathcal{H}$ to be efficiently computable and samplable, in order to exclude some (possible) "pathological" counterexamples.

The proof of Theorem 10 relies on some advanced facts from matrices theory. We briefly sketch our approach, the full proof is given in Appendix E. The key technical fact we prove is that the hashes of high-min-entropy distributions are really mapped *onto* high collision entropy distributions (with quantitative parameters good enough for our purposes). Once we have a such a correspondence, we reduce the problem to Theorem 9. To this end, we consider the probability $\Pr[H(x)=y]$ that x is hashed into y as a *matrix* with rows y and columns x and observe use this matrix to obtain a linear map which realizes that correspondence. To obtain a map with a good behavior, we fill it using some special "pattern" which ensures nice algebraic properties and simplifies inverting.

6 Conclusion

We show that the technical condition called *square security* introduced in previous works of Dodis,Yu and Barak et al.is not only sufficient but also necessary for better security with weak keys used directly.

References

1 G. Allaire, S. M. Kaber, and K. Trabelsi. *Numerical linear algebra*. Texts in applied mathematics. Springer, 2008.

2 B. Barak, Y. Dodis, H. Krawczyk, O. Pereira, K. Pietrzak, F. Standaert, and Y. Yu. Leftover hash lemma, revisited. In *Proc. 31th CRYPTO*, 2011.

3 B. Barak, R. Shaltiel, and A. Wigderson. In *RANDOM-APPROX*, 2003.

4 S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, NY, USA, 2004.

5 Y. Dodis, K. Pietrzak, and D. Wichs. Key derivation without entropy waste. In *EUROCRYPT*, pages 93–110. Springer Berlin Heidelberg, 2014.

6 Y. Dodis and Y. Yu. Overcoming weak expectations. In *Theory of Cryptography*, volume 7785 of *Lecture Notes in Computer Science*. 2013.

7 R. M. Gray. Toeplitz and circulant matrices: A review. *Commun. Inf. Theory*, 2(3):155–239, Aug. 2005.

8 J. Radhakrishnan and A. Ta-Shma. Bounds for dispersers, extractors, and depth-two superconcentrators. *SIAM JOURNAL ON DISCRETE MATHEMATICS*, 13:2000, 2000.

A Proof of Theorem 7

Fix an attacker $\mathcal{A}$. Let S^+ be the subset of keys where $\mathcal{A}$ achieves positive advantage and S^- be the remaining subset where the advantage is negative. Suppose that $|S^+| \geqslant 2^{m-1}$ (without loosing generality, consider the adversary with the flipped output otherwise).

Define the following distribution R: for $x \in S^-$ it puts the weight $\frac{1}{2^{m-1}}$ and for $x \in S^+$ it puts the weight $\frac{1-\frac{|S^-|}{2^{m-1}}}{|S^+|}$. Note that the min-entropy is at least $m-1$. Let $\epsilon^- = \frac{\sum_{r\in S^-}\mathsf{Adv}(r)}{|S^-|}$ and $\epsilon^+ = \frac{\sum_{r\in S^+}\mathsf{Adv}(r)}{|S^+|}$ be the average advantages on $S^+, S-$ respectively. Applying the security assumption to the distribution R, we obtain $\frac{|S^-|}{2^{m-1}}\epsilon^- + \left(1-\frac{|S^-|}{2^{m-1}}\right)\epsilon^+ \geqslant -\epsilon$ and hence (since $\epsilon^- < 0$) we have $\frac{|S^-|}{2^m}\cdot|\epsilon^-| \leqslant \frac{\epsilon}{2} + \left(\frac{1}{2}-\frac{|S^-|}{2^m}\right)\epsilon^+$. In turn, by choosing R to be uniform over S^+ (again the min-entropy is at least $m-1$) we obtain $\epsilon^+ \leqslant \epsilon$. Combining the

last two inequalities yields

$$\begin{aligned}\frac{|S^-|}{2^m}\cdot|\epsilon^-|+\frac{|S^+|}{2^m}\cdot\epsilon^+ &\leqslant \frac{\epsilon}{2}+\left(\frac{1}{2}-\frac{|S^-|}{2^m}\right)\epsilon^+ + \frac{|S^+|}{2^m}\cdot\epsilon^+\\ &= \frac{\epsilon+\epsilon^+}{2}+\frac{|S^+|-|S^-|}{2^m}\epsilon^+\\ &\leqslant 2\epsilon.\end{aligned}$$

But this means precisely that $\sum_{r\in\{0,1\}^n}|\mathsf{Adv}(r)|\leqslant 2\epsilon$. The bound is optimal is when the attacker fails on the half of the key with advantage (negative) $-\epsilon$ but performs well on the remaining part with advantage (positive) ϵ.

B Proof of Lemma 1

Proof. Our problem is equivalent to the following constrained maximization problem over $\mathbb{R}^{|S|}$

$$\begin{aligned}\underset{(p(x))_x\in\mathbb{R}^{|S|}}{\text{maximize}} \quad & \sum_x \mathrm{D}(x)p(x)\\ \text{subject to} \quad & -p(x)\leqslant 0 \quad \text{for all } x\in S\\ & \sum_x p(x)=1\\ & \sum_x p(x)^2\leqslant 2^{-k}\end{aligned} \tag{16}$$

The corresponding Lagrangian is given by

$$\begin{aligned}L((p(x))_x;(\lambda_1(x))_x,\lambda_2,\lambda_3) = &\sum_x \mathrm{D}(x)p(x)+\sum_x\lambda_1(x)p(x)-\lambda_2\left(\sum_x p(x)-1\right)\\ &-\lambda_3\left(\sum_x p(x)^2-2^{-k}\right)\end{aligned} \tag{17}$$

Note that the equality constraint is linear, the inequality constraints are convex and, since $k<n$, there exists a vector $p=p(x)$ such that $p(x)\geqslant 0$ for all x, $\sum_x p(x)=1$ and $\sum_x p(x)^2<2^k$. This means that Slater's Constraint Qualification is satisfied and the strong duality holds [4]. In this case the Karush-Kuhn-Tucker (KKT) conditions imply that for the optimal solution $p=p^*$ we have

$$\mathrm{D}(x)=-\lambda_1(x)+\lambda_2+\lambda_3 p^*(x) \tag{18}$$

where $\lambda_1(x)\geqslant 0$ for all x, $\lambda_3\geqslant 0$ and $\lambda_2\in\mathbb{R}$ are the Lagrange Multipliers, satisfying the following so called "complementary slackness" condition

$$\begin{aligned}\forall x\quad \lambda_1(x)&=0 \qquad \text{if } p^*(x)>0,\\ \lambda_3&=0 \qquad \text{if } \textstyle\sum_x(p^*(x))^2<2^{-k}.\end{aligned} \tag{19}$$

The characterization in Equation (14) follows now by setting $\lambda=\lambda_3$ and $t=\lambda_2$. Indeed, by Equation (18), Equation (19) and $\lambda_1(x)\geqslant 0$ we get

$$\max(\mathrm{D}(x)-\lambda_2,0)=\max(-\lambda_1(x)+\lambda_3 p^*(x),0)=\lambda_3 p^*(x).$$

Finally, note that if all values of $\mathrm{D}(\cdot)$ are different then in Equation (18) we cannot have $\lambda_3 = 0$, because then Equation (19) implies that D is constant on the support of p^* (which has at least two points provided that $k > 0$). To proof the uniqueness part, observe that if there exists a different pair (t', λ') for the same optimal solution p^*, then for all x such that $p^*(x) > 0$ we have

$$\forall x \in \mathrm{supp}(p^*) \quad \mathrm{D}(x) = t + \lambda p^*(x) + t' = \lambda' p^*(x), \tag{20}$$

and, since the case $\lambda = \lambda'$ cannot happen because it implies $t = t'$, we get $p^*(x) = \frac{t-t'}{\lambda-\lambda'}$. ◀ ◀

C Proof of Theorem 8

Proof. We show that the following implication is true:

(I) If for $\mathrm{D}_1 = \mathrm{D}$ and $\mathrm{D}_2 = \mathbf{1} - \mathrm{D}$ we have $\mathbf{E}\,\mathrm{D}_1(Y) \leqslant \frac{1}{2} + \epsilon$, $\mathbf{E}\,\mathrm{D}_2(Y) \leqslant \frac{1}{2} + \epsilon$ for every Y of collision entropy at least $m - d$, then $\mathrm{Var}\mathrm{D}(U) \leqslant 3\epsilon^2$.

Having this, we easily derive the result of Theorem 8.

▶ Claim 10.1. If **(I)** is satisfied then Theorem 8 holds.

Proof of Claim 10.1. First, we observe that

$$\mathbf{E}\left(\mathsf{Win}_{\mathsf{A}}(U) - 1/2\right)^2 = \mathrm{Var}\left(\mathsf{Win}_{\mathsf{A}}(U)\right) + \left(\mathbf{E}\,\mathsf{Win}_{\mathsf{A}}(U) - 1/2\right)^2, \tag{21}$$

hence it is enough to bound the variance of the winning probability. Next, we fix any adversary A with resources T, and set $\mathrm{D}(x) = \mathsf{Win}_{\mathsf{A}}(x)$, $\mathrm{D}_1(x) = \mathrm{D}(x)$ and $\mathrm{D}_2(x) = 1 - \mathrm{D}(x)$; by Theorem 2 we get $\max_Y \mathbf{E}\,\mathrm{D}_1(Y) \leqslant \frac{1}{2} + \epsilon$ where Y runs over al m-bit random variables of collision entropy at least $m - d$. Second, we notice that for the indistinguishability case, there exists adversary A', with the same running time, which flips the success probability of A, i.e. such that $\mathsf{Win}_{\mathsf{A}'}(x) = 1 - \mathsf{Win}_{\mathsf{A}}(x)$. Thus, for $\mathrm{D}_2(x) = 1 - \mathrm{D}(x)$ we obtain $\max_Y \mathbf{E}\,\mathrm{D}_2(Y) \leqslant \frac{1}{2} + \epsilon$. It follows that for every Y we have $\frac{1}{2} - \epsilon \leqslant \mathbf{E}\,\mathrm{D}(Y) \leqslant \frac{1}{2} + \epsilon$. Hence the assumption of **(I)** is satisfied and we get $\mathrm{Var}\left(\mathsf{Win}_{\mathsf{A}}(U)\right) \leqslant 2\epsilon^2$ which, combined with Equation (21), finishes the proof. ◀ ◀

It remains to prove that **(I)** holds under the assumptions in Theorem 8. Let $\mathrm{D}_1 = \mathrm{D}$, $\mathrm{D}_2 = \mathbf{1} - \mathrm{D}$ and let Y_i^*, for $i = 1, 2$, be optimal for maximizing $\mathbf{E}\,\mathrm{D}(Y)$ under restrictions $\mathrm{CP}(Y) \geqslant \frac{1+\theta}{2^m}$ where $\theta = 2^d - 1$, and let (Y_1^*, t_1, λ_1), (Y_2^*, t_2, λ_2) be the corresponding maximizing distributions and numbers from the characterization in Lemma 1 (by the approximation argument, we can assume that all values of D are different, perturbing them slightly if necessary; this, according to Remark 3, implies that the numbers t_i, λ_i for $i = 1, 2$ are unique and $\lambda_i > 0$). According to (**I**), we assume that

$$\mathbf{E}\,\mathrm{D}_i(Y_i^*) \leqslant \frac{1}{2} + \epsilon, \quad i = 1, 2 \tag{22}$$

Consider now the following cases:
Case 1: $t_1 < 0$ or Case 2: $t_2 < 0$. Suppose that $t_i < 0$ for $i = 1$ or $i = 2$. Since $\mathrm{D}_i \geqslant 0$, this implies that $\mathrm{D}_i'(x) = \max(\mathrm{D}(x) - t, 0) = \mathrm{D}(x) - t$ and thus $\mathrm{Var}\mathrm{D}_i = \mathrm{Var}\mathrm{D}_i'$. By Corollary 2

we get

$$\begin{aligned}\mathbf{E}\,\mathrm{D}_i(Y^*) - \mathbf{E}\,\mathrm{D}_i(U_m) &= \mathbf{E}\,\mathrm{D}'_i(Y^*) - \mathbf{E}\,\mathrm{D}'_i(U_m)\\ &= 2^{-m}\theta\lambda_i\\ &= \sqrt{\theta}\cdot\sqrt{2^{-2m}(1+\theta)\lambda_i^2 - (2^{-m}\lambda_i)^2}\\ &= \sqrt{\theta\cdot\mathrm{VarD}'_i(U_m)}\end{aligned}$$

Since we have $\mathrm{VarD} = \mathrm{VarD}_i$ and $|\mathbf{E}\,\mathrm{D}_i(Y^*) - \mathbf{E}\,\mathrm{D}_i(U_m)| \leqslant 2\epsilon$ by Equation (22), we obtain

$$\mathrm{VarD}_i(U_m) \leqslant 4\epsilon^2/\theta \tag{23}$$

We assume further that $t_1, t_2 \geqslant 0$ and split the analysis into two cases, illustrated in Figure 4.

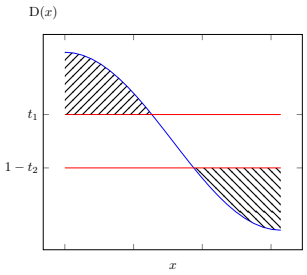

Figure 2 $t_1, t_2 \geqslant 0, t_1 + t_2 \geqslant 1$

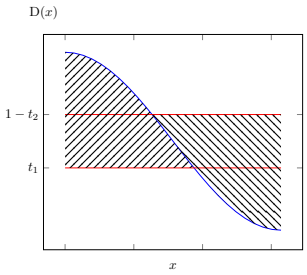

Figure 3 $t_1, t_2 \geqslant 0,\ 1 > t_1 + t_2$

Figure 4 Cases $t_1, t_2 \geqslant 0$

Case 3: $t_1 \geqslant 0, t_2 \geqslant 0, t_1 + t_2 \geqslant 1$ (see Figure 2). In this and the next case we will use the following simple lemma:

▶ Lemma 2. Let V, V_1, V_2 be r.v.'s such that $V_1, V_2 \geqslant 0$, $V_1V_2 = 0$ and $V = V_1 + V_2$. Then $\mathrm{Var}V \leqslant \mathrm{Var}V_1 + \mathrm{Var}V_2$.

▶ Corollary 4. For any discrete real random variable V, the expression $\mathrm{Var}\,(\max(V - t, 0))$ decreases with t.

Define $S_1 = \{x :\ \mathrm{D}(x) \geqslant t_1\}$ and $S_2 = \{x :\ \mathrm{D}(x) \leqslant 1 - t_2\}$. By the assumption on t_1 and t_2, the sets S_1 and S_2 are disjoint. Note that Y_1^* and Y_2^* are supported, respectively, on the sets S_1 and S_2 by the characterization in Lemma 1. From Equation (22) it follows now that

$$1/2 - \epsilon \leqslant 1 - t_2 \leqslant t_1 \leqslant 1/2 + \epsilon. \tag{24}$$

By Corollary 2 this implies that $\mathrm{D}'_i(x) = \max(\mathrm{D}_i(x) - t_i, 0)$ satisfy

$$(1+\theta)\,\mathbf{E}\,\mathrm{D}'(U) = \mathbf{E}\,\mathrm{D}'(Y_i^*) = \mathbf{E}\,\mathrm{D}(Y_i^*) - t_i \leqslant 2\epsilon \tag{25}$$

Hence, again by Corollary 2 we get

$$\mathrm{VarD}'_i(U) = \theta(\mathbf{E}\,\mathrm{D}'_i(U))^2 \leqslant \frac{4\theta\epsilon^2}{(1+\theta)^2} \tag{26}$$

Define $V_1 = \max(D_1, t_1)$, $V_2 = D_1$ if $1-t_2 \leqslant D_1 \leqslant t_1$ and 0 otherwise, and $V_3 = \min(D_1, 1-t_2) = 1 - \max(D_2, t_2)$. Since $D_1 = V_1 + V_2 + V_3$, applying Lemma 2 twice, we get

$$\begin{aligned}\mathrm{Var} D_1 &\leqslant \mathrm{Var} V_1 + \mathrm{Var} V_2 + \mathrm{Var} V_3 \\ &\leqslant \frac{8\theta\epsilon^2}{(1+\theta)^2} + \epsilon^2 \\ &\leqslant 3\epsilon^2 \end{aligned} \tag{27}$$

Case 4: $t_1 \geqslant 0, t_2 \geqslant 0, t_1 + t_2 < 1$ (see Figure 3). We show the following estimate

▶ Claim 10.2. We have $\mathrm{VarD} \leqslant (1 + 1/\theta)\left(\mathrm{VarD}'_1(U) + \mathrm{VarD}'_2(U)\right)$

Proof of Claim 10.2 . Observe that we have

$$\begin{aligned}\mathrm{VarD}(U) &= \mathrm{Var}(\mathrm{D}(U) - t_1) \\ &= \mathrm{Var}\left(\max(\mathrm{D}(U) - t_1, 0)\right) + \mathrm{Var}\left(\max(t_1 - \mathrm{D}(U), 0)\right) \\ &\quad + 2\,\mathbf{E}\max(\mathrm{D}(U) - t_1, 0) \cdot \mathbf{E}\max(t_1 - \mathrm{D}(U), 0).\end{aligned}$$

Since $t_2 \leqslant 1 - t_1$ we get $\mathrm{Var}\left(\max(t_1 - \mathrm{D}(U), 0)\right) \leqslant \mathrm{Var}\left(\max(1 - t_2 - \mathrm{D}(U), 0)\right)$, by Corollary 4 applied to $V = 1 - \mathrm{D}(U)$. Clearly we have $\mathbf{E}\max(t_1 - \mathrm{D}(U), 0) \leqslant \mathbf{E}\max(1 - t_2 - \mathrm{D}(U), 0)$. Therefore

$$\mathrm{VarD}(U) \leqslant \mathrm{VarD}'_1(U) + \mathrm{VarD}'_2(U) + 2\,\mathbf{E}\,\mathrm{D}'_1(U)\,\mathbf{E}\,\mathrm{D}'_2(U)$$

By Corollary 2 we get $\mathrm{VarD}'_i(U) = \theta \cdot (\mathbf{E}\,\mathrm{D}'_i(U))^2$ for $i = 1, 2$. Plugging this into the above equation, we get

$$\mathrm{VarD}(U) = (1 + \theta^{-1})\left(\mathrm{VarD}'_1(U) + \mathrm{VarD}'_2(U)\right) - \theta\left(\sqrt{\mathrm{VarD}'_1(U)} - \sqrt{\mathrm{VarD}'_2(U)}\right)^2$$

and the claim follows. ◀ ◀

▶ Claim 10.3. We have $\sqrt{\mathrm{VarD}'_1(U)} + \sqrt{\mathrm{VarD}'_1(U)} \leqslant 2\epsilon/\sqrt{\theta}$.

Proof of Claim. By Corollary 2 we have $\sqrt{\mathrm{VarD}'_i(U)} = \left(\mathbf{E}\,\mathrm{D}'_i(Y_i^*) - \mathbf{E}\,\mathrm{D}'_i(U)\right)/\sqrt{\theta}$ for $i = 1, 2$. This implies

$$\sqrt{\mathrm{VarD}'_1(U)} + \sqrt{\mathrm{VarD}'_2(U)} \leqslant \left(\mathbf{E}\,\mathrm{D}'_1(Y_i^*) - \mathbf{E}\,\mathrm{D}'_1(U) + \mathbf{E}\,\mathrm{D}'_2(Y_i^*) - \mathbf{E}\,\mathrm{D}'_2(U)\right)/\sqrt{\theta}.$$

Since $\mathbf{E}\,\mathrm{D}'(Y_i^*) = \mathbf{E}\,\mathrm{D}(Y_i^*) - t_i$, which follows from Equation (14), and since $\max(\mathrm{D}_i(x) - t_i, 0) + t_i = \max(\mathrm{D}_i(x), t_i)$ we can rewrite the above equation as

$$\sqrt{\mathrm{VarD}'_1(U)} + \sqrt{\mathrm{VarD}'_2(U)} \leqslant \left(\mathbf{E}\,\mathrm{D}_1(Y_i^*) + \mathbf{E}\,\mathrm{D}_2(Y_i^*) - \mathbf{E}(\max(\mathrm{D}_1(U), t_1) + \max(\mathrm{D}_2(U), t_2))\right)/\sqrt{\theta}.$$

Since $\max(\mathrm{D}_1(x), t_1) + \max(\mathrm{D}_2(x), t_2) \geqslant \mathrm{D}_1(x) + \mathrm{D}_2(x) = 1$, the result follows. ◀ ◀

By combining the inequality $\sqrt{\mathrm{VarD}'_1(U) + \mathrm{VarD}'_2(U)} \leqslant \sqrt{\mathrm{VarD}'_1(U)} + \sqrt{\mathrm{VarD}'_2(U)}$ with Claim 10.2 and Claim 10.3 we finally obtain

$$\mathrm{VarD} \leqslant \left(\theta^{-1} + \theta^{-2}\right) \cdot 4\epsilon^2 \tag{28}$$

Summarizing. The result follows by combining the estimates Equation (23), Equation (27) and Equation (28).

Tightness. To see that the lower bound is sharp, suppose that the class of adversaries consists of A and A' where A' is obtained by "flipping" the guess A. Let S_1, S_2, S_3, S_4 be any disjoint subsets of $\{0,1\}^m$ such that $|S_1| = |S_4| = 2^{m-d}$ and $|S_2| = |S_3| = 2^{m-1} - 2^{m-1-d}$. Let $\gamma \in (0,1)$ and suppose that the distribution of $\mathsf{Win}_{\mathsf{A}}(r)$ is as follows: $\frac{1}{2}+\epsilon$ for $r \in S_1$, $\frac{1}{2}+\gamma\epsilon$ for $r \in S_2$, $\frac{1}{2}-\gamma\epsilon$ for $r \in S_3$ and $\frac{1}{2}-\epsilon$ on $r \in S_4$. It is easy to see that $\max_Y \mathbf{E}_{r\leftarrow Y} \mathsf{Win}_{\mathsf{A}}(r) = \frac{1}{2}+\epsilon$ and $\max_Y \mathbf{E}_{r\leftarrow Y} \mathsf{Win}_{\mathsf{A}'}(r) = \frac{1}{2}+\epsilon$, where the maximum is over all distributions of collision entropy $m-d$. However

$$\begin{aligned}\mathop{\mathbf{E}}_{r\leftarrow U}\left(\mathsf{Win}_{\mathsf{A}}(r) - 1/2\right)^2 &= \mathop{\mathbf{E}}_{r\leftarrow U}\left(\mathsf{Win}_{\mathsf{A}'}(r) - 1/2\right)^2 \\ &= 2\cdot 2^{-d}\epsilon^2 + 2\cdot\left(2^{-1} - 2^{-1-d}\right)\cdot\gamma^2\epsilon^2,\end{aligned}$$

which for $\gamma \to 1$ becomes arbitrarily close to $2\epsilon^2$. ◀ ◀

D Proof of Theorem 9

Proof. The observation in Remark 2 implies that we can assume, without losing generality, that there is no side information. Using Lemma 1 we derive an estimate on the variance of a function D, which has small advantage in distinguishing between distributions of high collision entropy and the uniform distribution.

▶ Lemma 3. Let D be a real-valued function on m-bit strings. Suppose that $\mathbf{E}\,\mathrm{D}(Y) - \mathbf{E}\,\mathrm{D}(U_m) \leqslant \delta$ for all distributions Y over $\{0,1\}^m$ such that $\mathrm{CP}(Y) \leqslant \frac{1+\theta}{2^m}$. Then $\mathrm{VarD} \leqslant \max(\theta(\mathbf{E}\,\mathrm{D}(U_m))^2, \theta^{-1}\delta^2)$.

Proof of Lemma 3. We can assume that the values of D are all different (by the approximation argument we can assume this by perturbing the values slightly and once we prove the theorem under that restriction, we can pass with the changes to 0). From Lemma 1 and Remark 3 it follows then that the expectation $\mathbf{E}\,\mathrm{D}(Y)$ is maximized by a distribution $Y = Y^*$ satisfying $\mathrm{CP}(Y^*) = \frac{1+\theta}{2^m}$ and $\lambda\mathbf{P}_{Y^*}(x) = \max(\mathrm{D}(x)-t, 0)$ for some *unique* numbers $\lambda > 0$, $t \in \mathbb{R}$ and all x. Let $\mathrm{D}'(x) = \max(\mathrm{D}(x) - t, 0)$. By Corollary 2 we get

$$\mathbf{E}\,\mathrm{D}'(U_m)^2 - (\mathbf{E}\,\mathrm{D}'(U_m))^2 = \frac{\lambda^2\theta}{2^{2m}} = \theta\cdot(\mathbf{E}\,\mathrm{D}'(U_m))^2. \tag{29}$$

This equation and the uniqueness of t show that $t < 0$ if and only if $\mathrm{VarD}(U_m) > \theta(\mathbf{E}\,\mathrm{D}(U_m))^2$. Consider now two cases:
Case 1: $t \geqslant 0$. Then $\mathrm{VarD}(U_m) \leqslant \theta(\mathbf{E}\,\mathrm{D}(U_m))^2$.
Case 2: $t < 0$. We have $\mathrm{D}'(x) = \mathrm{D}(x) - t$ and therefore, as in the proof of Theorem 8, Case 1 and Case 2, we obtain

$$\mathbf{E}\,\mathrm{D}(Y^*) - \mathbf{E}\,\mathrm{D}(U_m) = \sqrt{\theta\cdot\mathrm{VarD}(U_m)}, \tag{30}$$

which implies that $\mathrm{VarD}(U_m) \leqslant \theta^{-1}\delta^2$. The result follows by taking the maximum over both estimates. ◀ ◀

Let $\theta = 2^d - 1$ and let $\mathrm{D}(r) = \mathsf{Win}_{\mathsf{A}}$, where A is any adversary with resources T. We observe that for every Y such that $\mathrm{CP}(Y) \leqslant \frac{1+\theta}{2^m}$, by the triangle inequality and Theorem 2 we obtain

$$\mathbf{E}\,\mathrm{D}(Y) - \mathbf{E}\,\mathrm{D}(U) \leqslant |\,\mathbf{E}\,\mathrm{D}(Y) - 1/2| + |\,\mathbf{E}\,\mathrm{D}(U) - 1/2| \leqslant 2\epsilon.$$

Applying Lemma 3 we get $\mathrm{Var}(\mathsf{Win}_{\mathsf{A}}) \leqslant \max(\theta, 4\epsilon^2/\theta)$. Recalling that we have $|\mathbf{E}\,\mathrm{D}(U) - 1/2| \leqslant \epsilon$, by the identity in Equation (21) we obtain

$$\mathbf{E}(\mathrm{D}(U) - 1/2)^2 \leqslant \max(\theta, 4\epsilon^2/\theta) + \epsilon^2,$$

which finishes the proof. ◀ ◀

E Proof of Theorem 10

of Theorem 10. The proof is based on the following lemma

▶ Lemma 4 (Distributions of high collision entropy as random hashes). Let $\mathcal{X} = \{x_1, \ldots, x_N\}$ and $\mathcal{Y} = \{y_1, \ldots, y_M\}$ be any sets and let η be a number such that $0 < \eta < 1/(2\sqrt{M})$. Then there exists an efficient family $\mathcal{H} = \{h_\xi\}_\xi$ of functions from $\mathcal{X}$ to $\mathcal{Y}$, which is γ-universal with $\gamma = M\eta^2$, and has the following property: for every K and every distribution Y over $\mathcal{Y}$ such that $\mathrm{CP}(Y) \leqslant (1+\eta)/M$ there exists a set $S \subset \mathcal{X}$ of size $|S| = K$ such that (a) $\mathrm{SD}(H(U_S); Y) \leqslant \frac{M}{K}$ and (b) $\mathrm{CP}(H(U_S)) \leqslant \mathrm{CP}(Y) + \frac{2M}{K}$.

We show how Lemma 4 implies Theorem 10. Set $M = 2^m$, $N = 2^n$, $M = 2^k$, and $\eta = 2^{-m/2}\epsilon^{\alpha/2}$ so that $\gamma = \epsilon^\alpha$. Note that $\epsilon < 1/4$ implies $\eta < 1/(2\sqrt{M})$. Fix any adversary A' which doesn't use side information and let A be the adversary which challenged on r and given h as advise, runs A'. Clearly, we have $\mathsf{Win}_{\mathsf{A}}(r,h) = \mathsf{Win}_{\mathsf{A}'}(r)$. By the assumption we have $\mathbf{E}\,\mathsf{Win}_{\mathsf{A}}(H(X), H) \leqslant 1/2 + C\epsilon$ for all X of min-entropy at least k. Therefore, we obtain

$$\mathbf{E}\,\mathsf{Win}_{\mathsf{A}'}(H(X)) \leqslant 1/2 + C\epsilon, \quad \text{for every } X \text{ such that } \mathbf{H}_\infty(X) \geqslant k.$$

This inequality, combined with Lemma 4 where $M = 2^m$ and $\eta =$, yields

$$\mathbf{E}\,\mathsf{Win}_{\mathsf{A}'}(Y) \leqslant 1/2 + C\epsilon + 2^{m-k}, \quad \text{for every } Y \text{ such that } \mathrm{CP}(Y) \leqslant \frac{1+\eta}{2^m}.$$

In other words, our application is (T, ϵ')-secure in the $(m - d')$-real$_2$ model where $\epsilon' = C\epsilon + 2^{m-k} \leqslant (C+1)\epsilon$ and $2^{d'} = 1 + \eta$. Applying the lower bound in Theorem 9, we obtain (T, σ)-square-security with $\sigma = \epsilon'^2 + \max\left(\eta, 4\epsilon'^2\eta^{-1}\right)$. This yields Equation (15).

Proof. Proof of Lemma 4 Let $\{\xi_j\}_{j\in\{1,\ldots,N\}}$ be pairwise independent random variables taking values in $\{y_1, \ldots, y_M\}$ such that

$$\Pr[\xi_j = y_i] = \begin{cases} \frac{1}{M} + \delta & \text{if } i \equiv j \mod M \\ \frac{1}{M} & \text{if } i \equiv j+1, \ldots, j+\frac{M}{2}-1 \mod M \\ \frac{1}{M} - \frac{2\delta}{M} & \text{if } i \equiv j+\frac{M}{2}, \ldots, j+M-1 \mod M \end{cases} \tag{31}$$

The distribution of $\{\xi_j\}_j$ has a circular "pattern", illustrated in the table below.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
y_1	$\frac{1}{4}+\delta$	$\frac{1}{4}$	$\frac{1}{4}-\frac{\delta}{2}$	$\frac{1}{4}-\frac{\delta}{2}$	$\frac{1}{4}+\delta$	$\frac{1}{4}$	$\frac{1}{4}-\frac{\delta}{2}$	$\frac{1}{4}-\frac{\delta}{2}$	$\frac{1}{4}+\delta$	$\frac{1}{4}$	$\frac{1}{4}-\frac{\delta}{2}$	$\frac{1}{4}-\frac{\delta}{2}$
y_2	$\frac{1}{4}-\frac{\delta}{2}$	$\frac{1}{4}+\delta$	$\frac{1}{4}$	$\frac{1}{4}-\frac{\delta}{2}$	$\frac{1}{4}-\frac{\delta}{2}$	$\frac{1}{4}+\delta$	$\frac{1}{4}$	$\frac{1}{4}-\frac{\delta}{2}$	$\frac{1}{4}-\frac{\delta}{2}$	$\frac{1}{4}+\delta$	$\frac{1}{4}$	$\frac{1}{4}-\frac{\delta}{2}$
y_3	$\frac{1}{4}-\frac{\delta}{2}$	$\frac{1}{4}-\frac{\delta}{2}$	$\frac{1}{4}+\delta$	$\frac{1}{4}$	$\frac{1}{4}-\frac{\delta}{2}$	$\frac{1}{4}-\frac{\delta}{2}$	$\frac{1}{4}+\delta$	$\frac{1}{4}$	$\frac{1}{4}-\frac{\delta}{2}$	$\frac{1}{4}-\frac{\delta}{2}$	$\frac{1}{4}+\delta$	$\frac{1}{4}$
y_4	$\frac{1}{4}$	$\frac{1}{4}-\frac{\delta}{2}$	$\frac{1}{4}-\frac{\delta}{2}$	$\frac{1}{4}+\delta$	$\frac{1}{4}$	$\frac{1}{4}-\frac{\delta}{2}$	$\frac{1}{4}-\frac{\delta}{2}$	$\frac{1}{4}+\delta$	$\frac{1}{4}$	$\frac{1}{4}-\frac{\delta}{2}$	$\frac{1}{4}-\frac{\delta}{2}$	$\frac{1}{4}+\delta$

Table 1 The distribution of $\{\xi_j\}_j$ for $M = 4$ and $N = 12$

Let $A = [a_{i,j}]$ be the $M \times M$ matrix with the entries

$$a_{i,j} = \Pr[\xi_j = y_i]. \tag{32}$$

The matrix A is the "base" pattern of the table $(\Pr[\xi_j = y_i])_{i,j}$. For the special case $M = 4$ this matrix is

$$A = \begin{pmatrix} \frac{1}{4}+\delta & \frac{1}{4} & \frac{1}{4}-\frac{\delta}{2} & \frac{1}{4}-\frac{\delta}{2} \\ \frac{1}{4} & \frac{1}{4}+\delta & \frac{1}{4}-\frac{\delta}{2} & \frac{1}{4}-\frac{\delta}{2} \\ \frac{1}{4}-\frac{\delta}{2} & \frac{1}{4} & \frac{1}{4}+\delta & \frac{1}{4}-\frac{\delta}{2} \\ \frac{1}{4}-\frac{\delta}{2} & \frac{1}{4}-\frac{\delta}{2} & \frac{1}{4} & \frac{1}{4}+\delta \end{pmatrix}.$$

Define h to be a function such that

$$h(x_j) = \xi_j \tag{33}$$

We construct $\mathcal{H}$ X as follows: for every $i = 1, \ldots, M$ let S_i be a set such that

$$S_i \subset \{x_j : j \equiv i \mod M\}. \tag{34}$$

Let $S = \bigcup_i S_i$. Observe that $\Pr_{x \leftarrow S}[H(x) = y_i] = \sum_j a_{i,j} \cdot \frac{|S_j|}{|S|}$. Hence we obtain

$$\left(\Pr_{x \leftarrow S}[H(x) = y_1], \ldots, \Pr_{x \leftarrow S}[H(x) = y_M]\right)^T = A \cdot \left(\frac{|S_1|}{|S|}, \frac{|S_2|}{|S|} \ldots, \frac{|S_M|}{|S|}\right)^T \tag{35}$$

From Equation (35) we immediately obtain the following corollary

▶ Claim 10.4. Let Y be a random variable with values in $\{y_1, \ldots, y_M\}$ and let S_j be sets satisfying (34). Denote $s_j = \frac{|S_j|}{|S|}$, $p_j = \Pr[Y = y_j]$ and suppose that

$$A \cdot (s_1, s_2, \ldots, s_M)^T = (p_1, p_2, \ldots, p_M)^T \tag{36}$$

where A is the matrix defined in Equation (32). Let $S = \bigcup_j S_j$. Then we have $Y \stackrel{d}{=} H(U_S)$.

This claim reduces the problem of representing a distribution Y as a random hash of a high-entropy distribution X to the problem of solving system of linear equations. A natural way is to solve the system in Equation (36) for s and then find sets S_j such that $s_j = \frac{|S_j|}{|S|}$ or at least $s_j \approx \frac{|S_j|}{|S|}$. Note that there are two issues in this approach.

(a) The solution s of the equation $A \cdot s = p$ is not necessarily positive.
(b) The value $|S|$, necessary to obtain a good approximation $s_j \approx \frac{|S_j|}{|S|}$, could be much bigger than 2^m.

We deal with issue (a) by observing that from Claim 10.4 it follows that *uniform* Y can be represented by X uniform over a 2^m-element set, with the corresponding weights $s_1 = \ldots = s_M = \frac{1}{M}$. Intuitively, we expect that the same is true for every distribution Y close to U_m and that the corresponding weights s_j are also close to $\frac{1}{M}$. To state this in a quantitative form, we estimate the so called *matrix condition number*, a toll widely used in numerical analysis to estimate changes of a solution caused by a perturbation of linear systems. Problem of computing the condition number is extremely simplified due to the fact that the matrix A has a "cyclic" pattern. To deal with problem (b) we observe that under the condition $\log|S| \leqslant m + \log(1/\epsilon)$ we approximate the distribution Y well enough.

We return to the proof. Below we give the standard result on perturbed linear system.

▶ Lemma 5 (Perturbation of a linear system, [1]). Let $\|\cdot\|$ be a vector norm on the space $\mathbb{C}^M$ and let $\|A\|$, for any $M \times M$ matrix A, be the associated matrix norm defined by $\|A\| = \sup_{\|x\|=1} \|Ax\|$. Define the condition number of a non-singular $M \times M$ matrix A, relative to the norm $\|\cdot\|$, as $\mathrm{cond}(A) = \|A\| \cdot \|A^{-1}\|$. Suppose that s_0 and s are solutions of the systems

$$As_0 = p_0 \quad \text{and} \quad As = p$$

Then we have

$$\frac{\|s - s_0\|}{\|s_0\|} \leqslant \mathrm{cond}(A) \cdot \frac{\|b' - b\|}{\|b\|} \tag{37}$$

In our case, the convenient norm is the matrix norm induced by the second norm. This norm is also called the *spectral norm*, because it can be computed directly from the singular values of the matrix.

▶ Lemma 6 (Computing spectral norm [1]). For any matrix A with complex elements we have

$$\|A\|_2 = \frac{\sigma_{\max}}{\sigma_{\min}} \tag{38}$$

where $\sigma_{\max}$ and $\sigma_{\min}$ are, respectively, the biggest and the smallest singular values of A, i.e. the biggest and the smallest eigenvalues of the (symmetric) matrix AA^*.

By Equation (31) we obtain the following corollary

▶ Claim 10.5. The matrix A defined in Equation (32) is a doubly stochastic and circulant matrix. That is, rows and columns sum to 1 and every row is a right cyclic shift of the row above.

Circular matrices appear in many problems in physics, signal and image processing, probability, statistics, numerical analysis, and coding theory. Their theory is relatively simple comparing to the case of general matrices, and many basic questions can be answered in a simple closed form. For circulant matrices the eigenvalues and eigenvectors could be directly computed. In particular, we have the following decomposition theorem

▶ Lemma 7 (Spectral decomposition theorem for circulant matrices [7]). Let C be a circulant $M \times M$ matrix with complex elements whose first row is $(c_0, c_1, \ldots, c_{M-1})$. Then we have

$$C = U\Lambda U^*, \tag{39}$$

for the diagonal matrix $\Lambda = \mathrm{diag}\,(\lambda_0, \ldots, \lambda_{M-1})$ which consists of the elements (eigenvalues of C)

$$\lambda_j = \sum_{i=0}^{M} c_i \rho_j^i \tag{40}$$

and the unitary matrix $U = (u_0, \ldots, u_{M-1})$ with the columns (eigenvectors of C)

$$u_j = \frac{1}{\sqrt{M}} \left(1, \rho_j, \rho_k^2, \ldots, \rho_j^{(M-1)}\right)^T \tag{41}$$

where $\rho_j = \exp\left(-\frac{2k\pi}{M}\right)$ is the j-th of the M-th roots of unity.

Combining Lemma 5 and Lemma 6 we obtain the following well known fact

▶ Lemma 8. Let C be a normal matrix, i.e. matrix unitarily similar to a diagonal matrix: $C = U\mathrm{diag}(\lambda_1, \ldots, \lambda_M)U^*$. Then $\|C\|_2 = \max_i |\lambda_i|$.

From Lemma 8 as a corollary we immediately obtain

▶ Lemma 9. Let C be a $M \times M$ normal matrix with eigenvalues $\lambda_1, \ldots, \lambda_M$. Then the condition number with respect to the spectral norm is given by $\mathrm{cond}_2(C) = \frac{\max_i |\lambda_i|}{\min_i |\lambda_i|}$.

The last claim together with Lemma 7 allows us computing the condition number of a circulant matrix.

▶ Claim 10.6. Let $\delta \leqslant \frac{1}{2}$ and let A be the matrix defined in Equation (32). Then $\mathrm{cond}_2(A) \leqslant 1/\delta$.

Proof of Claim. Denote $\rho = \exp\left(-\frac{2\pi \mathrm{i}}{M}\right)$. By Claim 10.5 and Lemma 7 we know that the eigenvalues $\lambda_0, \lambda_1, \ldots, \lambda_{M-1}$ are given by $\lambda_i = \sum_{j=0}^{M-1} A_{1,j+1}\rho^{ij}$. By Equation (32) we have $A_{1,j+1} = \frac{1}{M} + \delta$ for $j = 0$, $A_{1,j+1} = \frac{1}{M}$ for $j = 1, \ldots, \frac{M}{2} - 1$ and $A_{1,j+1} = \frac{1}{M} - \frac{2\delta}{M}$. Since $1 + \rho^i + \rho^{2i} + \ldots + \rho^{(M-1)i} = 0$ for $i \not\equiv 0 \mod M$, we obtain

$$\lambda_i = \begin{cases} 1 & \text{if } i = 0 \\ \delta + \frac{2\delta}{M} \cdot \frac{1-(-1)^i}{1-\rho^i} & \text{if } i \in \{1, \ldots, M-1\} \end{cases} \tag{42}$$

Having established the explicit formula on the eigenvalues of A, we can easily identify the smallest and the biggest one in absolute value. We start by observing that the homographic mapping of the complex plane $z \to \frac{1}{1-z}$, maps the unit circle $|z| = 1$ into the line $\Re(z) = \frac{1}{2}$. Indeed,

$$\frac{1}{1-\exp(\theta \mathrm{i})} = \frac{\exp(-\theta \mathrm{i}/2)}{\exp(-\theta \mathrm{i}/2) - \exp(\theta \mathrm{i}/2)} = \frac{\cos(\theta/2) - \mathrm{i}\sin(\theta/2)}{-2\mathrm{i}\sin(\theta/2)} = \frac{1}{2} + \frac{\mathrm{i}}{2\tan(\theta/2)}. \tag{43}$$

From Equation (43) it follows that if $z = \exp(\theta \mathrm{i})$ runs around the unit circle in the counter-clockwise direction, starting from $z = 0$, then the point $\frac{1}{1-z}$ runs along the line $\Re(z) = \frac{1}{2}$ from $\left(\frac{1}{2}, +\infty\right)$ to $\left(\frac{1}{2}, -\infty\right)$. Applying this observation to the points $z = \rho^i$ with odd i and the formula in Equation (42), which for odd i becomes $\lambda = \delta\left(1 + \frac{4}{M} \cdot \frac{1}{1-z}\right)$, we see that

$$|\lambda_{M/2}| \leqslant |\lambda_i| \leqslant |\lambda_1|, \quad i \equiv 1 \mod 2$$

Since we have $\lambda_{M/2} = \delta\left(1 + \frac{2}{M}\right)$ and, by Equation (43), $\lambda_1 = \delta\left(1 + \frac{2}{M} + \frac{2\mathrm{i}}{M\tan(\pi/M)}\right)$ the above inequality implies

$$\delta\left(1 + \frac{2}{M}\right) \leqslant |\lambda_i| \leqslant \delta \cdot \sqrt{1 + \frac{4}{M} + \frac{4}{M^2 \sin^2\left(\frac{\pi}{M}\right)}}, \quad \text{for } i \equiv 1 \mod 2 \tag{44}$$

Noticing that $\lambda_i = 1$ for $i = 0$ and $\lambda_i = \delta$ for positive even i, we finally obtain

$$\begin{aligned} |\lambda_{\min}| &= \delta \\ |\lambda_{\max}| &= \max\left(1, \delta\left(1 + \frac{4}{M} + \frac{4}{M^2 \sin^2\left(\frac{\pi}{M}\right)}\right)^{\frac{1}{2}}\right) . \end{aligned} \tag{45}$$

Recall the inequality $\sin t \geqslant \frac{2t}{\pi}$ for $t \in \left(0, \frac{\pi}{2}\right)$. Applying this to Equation (45) we get $|\lambda_{\max}| \leqslant \max\left(1, \delta \cdot \sqrt{2 + \frac{4}{M}}\right) \leqslant \max(1, 2\delta)$, hence $|\lambda_{\max}| = 1$ for $\delta \leqslant \frac{1}{2}$. Thus, the result follows. ◀ ◀

Now we are ready to make the first step towards the actual proof. Namely, we prove that every distribution of sufficiently low collision probability can be (exactly) represented as a random hash of a distribution uniform over some set. The problem is the size of this set, which is not guaranteed to be sufficiently small.

▶ Claim 10.7. Let A be the matrix defined in Equation (32) and let Y be a distribution over $\{y_1, \ldots, y_M\}$ such that $\mathrm{CP}(Y) \leqslant \frac{1+\eta}{M}$. Define the vector p as $p_j = \Pr[Y = y_j]$. Then the solution s of the equation $As = p$ is non-negative, provided that $\delta \geqslant M^{1/2}\eta$.

Proof of Claim. Let p^0 and p be the vectors of probability masses of U and Y respectively, that is $p^0_j = \frac{1}{M}$ and $p_j = \Pr[Y = y_j]$. Let s and s^0 be the solutions of the systems $As^0 = p^0$ and $As = p$. Using Lemma 5, Claim 10.6 and the fact that $s^0 = p^0$, we get

$$\|s - s^0\|_2 \leqslant \delta^{-1} \cdot \|p - p^0\|_2 \tag{46}$$

Note that ,by assumption, we have

$$\|p - p^0\|_2 = \sum_y |\Pr[Y = y] - 1/M|^2 = \mathrm{CP}(Y) - \frac{1}{M} = \eta/M, \tag{47}$$

and recall the basic inequalities between the vector norms

$$\left\|s - s^0\right\|_\infty \leqslant \left\|s - s^0\right\|_1 \leqslant \sqrt{M} \cdot \left\|s - s^0\right\|_2. \tag{48}$$

Note that if the condition $\|s - s^0\|_\infty \leqslant \frac{1}{M}$ is satisfied, then $s_j \geqslant 0$ for all j. The claim follows now easily by combining Equation (46), Equation (47) and Equation (48). ◀ ◀

In the next step we relax the assumption on the equality of the distributions in Claim 10.7 and show how to *approximately* obtain any low collision entropy distribution by hashing a *bounded* set.

▶ Claim 10.8. Let A be the matrix defined in Equation (32) and let Y be a distribution over $\{y_1, \ldots, y_M\}$. Then for any K there exists a distribution Y' such that

(a) The solution s' of the equation $As' = p'$, where $p'_j = \Pr[Y' = y_j]$, is a vector whose entries are positive and are integer multiples of $\frac{1}{K}$
(b) We have $\mathrm{SD}(Y'; Y) \leqslant \frac{M}{K}$ and $\mathrm{CP}(Y') \leqslant \mathrm{CP}(Y) + \frac{2M}{K}$.

Proof of Claim. Define the probability vector s' as follows

$$s'_j = \gamma/K + \lfloor Ks_j \rfloor / K \tag{49}$$

where $\gamma = 1 - \sum_{j=1}^M \lfloor Ks_j \rfloor$. Observe that we have

$$0 \leqslant \gamma \leqslant 1 - M/K \tag{50}$$

and

$$s_j + (\gamma - 1)/K \leqslant s'_j \leqslant s_j + \gamma/K. \tag{51}$$

Since the matrix A has positive entries and since the rows of A sum to 1, we have

$$A_j s + (\gamma - 1)/K \leqslant A_j s' \leqslant A_j s + \gamma/K. \tag{52}$$

for every row A_j of the matrix A. By definition we have $A_j s = p_j$ and $A_j s' = p'_j$. Hence Equation (52) implies

$$\|p' - p\|_1 \leqslant M/K \tag{53}$$

and

$$\|p\|_2^2 - \frac{M(1-\gamma)}{K} + \frac{M(1-\gamma)^2}{K^2} \leqslant \|p'\|_2^2 \leqslant \|p\|_2^2 + \frac{2M\gamma}{K} + \frac{M\gamma^2}{K^2}. \tag{54}$$

From Equation (50) it follows that $\gamma\left(1 + \frac{1}{K}\right) < 1$. Thus, from Equation (54) we have $\|p'\|_2^2 \leqslant \|p\|_2^2 + \frac{2M}{K}$. This, combined with Equation (53) finishes the proof. ◀ ◀

The lemma follows now directly from the last claim. ◀ ◀

◀ ◀

F Proof of Lemma 8

Proof of Lemma. It is well known that multiplying by a unitary matrix preserves the spectral norm, that is $\|UA\|_2 = \|AU\|_2 = \|A\|$ for any square matrix A and unitary U of the same dimension. To see this, note that for any vector $v \in \mathbb{C}^M$ we have

$$\|UAv\|_2^2 = (UAv)^* UAv = v^* A^* U^* UAv = v^* A^* Av = \|Av\|_2^2.$$

The equality $\|UA\|_2 = \|A\|_2$ follows from this directly, and the second equality $\|AU\|_2 = \|A\|_2$ follows by observing that $|v\|_2 = 1$ if and only if $\|Uv\| = 1$ and combining this with the definition of the spectral norm. Therefore we have $\|C\|_2 = \|\mathrm{diag}(\lambda_1, \ldots, \lambda_M)\|_2$ and the claim follows. ◀ ◀

G Proof of Theorem 1

Proof. Let $d = n - k$. Consider an application with the following advantage profile: the adversary achieves advantage $\sqrt{2^{d+1}\delta}$ on a $\frac{1}{2^{d+1}}$-fraction of the keys, and zero advantage on the remaining keys. Let S be as in the proof of Lemma 1 In particular we have $S = \{r : \mathsf{Adv}(r) \geqslant t\}$ for some t. Since $\mathsf{Adv}(\cdot)$ is non-negative, and positive on $2^{n-k-1} < |S|$ elements, we conclude that $t \leqslant 0$ and $S = \{0,1\}^n$. By definition, we have

$$\mathop{\mathbf{E}}_{r \leftarrow U_n} \mathsf{Adv}^{\mathcal{A}}(r)^2 = \delta.$$

On the other hand, according to Lemma 1 we have (taking the maximal advantage over keys R with Renyi entropy at least k)

$$\max_R \mathbf{E}\, \mathsf{Adv}^{\mathcal{A}}(R) = \sqrt{\frac{\delta}{2^{d+1}}} + \sqrt{2^d - 1} \cdot \sqrt{\delta - \frac{\delta}{2^{d+1}}} = \Omega\left(\sqrt{2^{d+1}\delta}\right).$$

◀ ◀