Reducing Complexity Assumptions for Statistically-Hiding Commitment

Iftach Haitner[1⋆], Omer Horvitz[2⋆⋆], Jonathan Katz[2⋆⋆⋆], Chiu-Yuen Koo[2], Ruggero Morselli[2], and Ronen Shaltiel[3]

[1] Department of Computer Science, Weizmann Institute of Science
iftach.haitner@weizmann.ac.il
[2] Department of Computer Science, University of Maryland
{horvitz,jkatz,cykoo,ruggero}@cs.umd.edu
[3] Department of Computer Science, University of Haifa
ronen@cs.haifa.ac.il

Abstract. Determining the minimal assumptions needed to construct various cryptographic building blocks has been a focal point of research in theoretical cryptography. Here, we revisit the following question: *what are the minimal assumptions needed to construct statistically-hiding commitment schemes*? Previously, it was known how to construct such schemes based on one-way permutations. We improve upon this by constructing statistically-hiding commitment schemes based on *approximable-preimage-size* one-way functions. These are one-way functions for which there is an efficient way to approximate the number of preimages of a given output. A special case (for which we show a somewhat simpler construction) is that of *regular* one-way functions where all outputs have the same number of preimages.

We utilize two different approaches in constructing statistically-hiding commitment schemes. Our first approach proceeds by showing that the scheme of Naor et al. can be implemented using any one-way function having an output distribution which is "sufficiently similar" to uniform. We then construct one-way functions with this property from approximable-preimage-size one-way functions. Our second approach begins by constructing a commitment scheme which is statistically hiding against an honest-but-curious receiver. We then demonstrate a *compiler* which transforms any such commitment scheme into one which is statistically hiding even against a malicious receiver. This compiler and its analysis may be of independent interest.

1 Introduction

A central focus of modern cryptography has been to investigate the weakest possible assumptions under which various cryptographic primitives exist. This

⋆ Research supported by U.S.-Israel Binational Science Foundation grant 2002246.
⋆⋆ Research supported by U.S. Army Research Office award DAAD19-01-1-0494.
⋆⋆⋆ Supported by NSF CAREER award 0447075.

direction of research has been quite fruitful, and minimal assumptions are known for a wide variety of primitives: e.g., pseudorandom generators, pseudorandom functions, symmetric-key encryption/message authentication, and digital signatures [21, 12, 13, 20, 24, 26, 29]. In other cases, black-box separation results exist which indicate the difficulty — if not impossibility — of constructing "strong" cryptographic protocols (say, key-exchange) from "weak" building blocks (say, one-way permutations; see [22]).

The above may give the impression that exact characterizations for all primitives of interest (at least in terms of equivalent complexity-theoretic assumptions) are known; however, this is not the case. Questions that remain open (to choose two examples) include the possibility of constructing efficient-prover non-interactive zero-knowledge proofs [4] based on assumptions weaker than trapdoor permutations [9], as well as determining whether constant-round ZK proofs exist based only on the assumption of one-way functions (see [10, Chap. 4]).

Another key cryptographic primitive in which a weakest possible assumption is not known is *statistically-hiding commitment*. Informally, a commitment scheme defines a two-phase interactive protocol between a sender $\mathcal{S}$ and a receiver $\mathcal{R}$; after the *commitment phase*, $\mathcal{S}$ is uniquely bound to (at most) one value which is not yet revealed to $\mathcal{R}$, and in the *decommitment phase* $\mathcal{R}$ finally learns this value. The two security properties hinted at in this informal description are known as *binding* (namely, that $\mathcal{S}$ is bound to at most one value after the commitment phase) and *hiding* (namely, that $\mathcal{R}$ does not learn the value to which $\mathcal{S}$ commits before the decommitment phase). In a statistically-hiding commitment scheme the hiding property holds *even against all-powerful receivers* (i.e., hiding holds information-theoretically), while the binding property is required to hold only for computationally-bounded (say, polynomial-time) senders.

Statistically-hiding commitment schemes can be used as a building block in constructions of statistical zero-knowledge arguments [6, 25] or certain coin-tossing protocols [2, 23]. They are also advantageous when used within protocols in which certain commitments are never revealed; in this case, it need only be infeasible to violate the binding property *during the period of time the protocol is run*, whereas the committed values will remain hidden *forever* (i.e., regardless of how much time the receiver invests after completion of the protocol). Indeed, this is part of the motivation for statistical zero-knowledge as well. For further discussion, the reader is referred to [27, 28, 25].

Perfectly-hiding[1] commitment schemes were first shown to exist based on specific number-theoretic assumptions [6, 5] or, more generally, based on any collection of claw-free permutations [18] with an efficiently-recognizable index set [15] (see [15] for a definition of a weaker variant of statistically-hiding commitment which suffices for some applications and for which an efficiently-recognizable index set is not needed). Naor, et al. [25], using techniques developed earlier

[1] Very informally, in a statistically-hiding commitment scheme the receiver learns only a negligible amount of information about the sender's committed value, whereas in a perfectly-hiding commitment scheme the receiver learns *nothing*. Note that any perfectly-hiding scheme is also statistically-hiding.

by Ostrovsky, et al. [27, 28], later showed a construction of a perfectly-hiding commitment scheme based on one-way permutations. Statistically-hiding commitment schemes can also be constructed from collision-resistant hash functions [8, 19] (see [30] for minimal assumptions for the existence of the latter).

1.1 Our Results

We show how to construct a statistically-hiding commitment scheme given any *approximable-preimage-size* one-way function. Informally, this is a one-way function f satisfying the additional property that, given any y in the image of f, the value $|\{x : f(x) = y\}|$ (i.e., the number of points mapping to y) can be efficiently estimated. An interesting special case, for which our construction may be somewhat simplified, is that of *regular* one-way functions for which every point in the image of f has the same number of preimages. (We still require that it be feasible to approximate the number of preimages.) A variety of conjectured one-way functions are regular; we refer the reader to [16] for examples.

We show two different approaches to constructing statistically-hiding commitment schemes: the first is more direct and achieves better computational efficiency, while the second achieves better round complexity (in fact, it achieves round complexity identical to [25]). As part of our second approach, we show a *compiler* transforming any commitment scheme which is statistically-hiding against an honest-but-curious (a.k.a. semi-honest) receiver into one which is statistically-hiding against an arbitrarily-malicious receiver. Since our compiler requires only the existence of one-way functions, our result implies an equivalence between the two formulations of the problem. (Due to space limitations the details of our second approach do not appear in this version.)

Our results may be viewed as an example of the paradigm in which a sequence of works constructs a given primitive from ever-weaker assumptions; e.g., in the cases of pseudorandom generators and universal one-way hash functions/signature schemes (see [10, Chap. 2] and [11, Chap. 6]), constructions were first based on specific, number-theoretic assumptions [3, 18], and then the minimal assumptions were gradually reduced to trapdoor permutations [1] (in the case of signatures), one-way permutations [17, 26], regular one-way functions [16, 31], and (finally) one-way functions [20, 29]. We hope our work will similarly serve as a step toward resolving the question of the minimal assumptions required for statistically-hiding commitment.

1.2 Overview of Our Techniques

Our constructions are based on the protocol of Naor et al. [25], which is shown there to be perfectly hiding (as well as computationally binding) when applied using a one-way permutation. It is natural to ask what happens when this protocol is applied using some other function $f : \{0,1\}^n \to \{0,1\}^\ell$. We first observe that the main argument of [25] shows that the protocol is computationally binding as long as f cannot be efficiently inverted with respect to the uniform distribution U_ℓ (more formally, no efficient algorithm can compute $f^{-1}(y)$, for uniformly-chosen y, with non-negligible probability). We call a function with this property *one-way over its range*. Note that a function with this property is not necessarily one-way.

As our first main technical result, we then show that the protocol of Naor et al. is "somewhat hiding" when applied using a function f for which the distribution $f(U_n)$ is *balanced*. (By "somewhat hiding" we mean that the receiver cannot guess the committed bit with probability better than some constant $\rho < 1$. Such a protocol can be "amplified" using repetition to give a statistically-hiding protocol.) Loosely speaking, a distribution over $\{0,1\}^\ell$ is balanced if it assigns to "most" elements $y \in \{0,1\}^\ell$ a probability that is close to $2^{-\ell}$ (say between $(99/100) \cdot 2^{-\ell}$ and $(101/100) \cdot 2^{-\ell}$). (In the precise definition we allow some elements to have probability outside this range as long as both the number of such elements and their total weight are small.)

The remainder of the paper is devoted to constructing functions that are both balanced and one-way over their range.[2] Intuitively, both these properties require the output distribution $f(U_n)$ to be "somewhat similar" to uniform. While we do not know how to construct such a function given a general one-way function, we show how to construct such functions given regular or approximable-preimage-size one-way functions. We achieve this goal using poly-wise independent hashing, inspired by [20, 29]. More precisely, given a regular one-way function f (the case of approximable-preimage-size one-way functions is more complex), we define $f'(h, x) = (h, h(f(x)))$ where h is selected from a family of $O(k)$-wise independent hash functions (here, k is the security parameter). This hashing "smoothes" the output distribution, and we show that by choosing the output length of h appropriately we obtain an f' which is both balanced and one-way over its range. Note that making the output length of h "too small" makes f' more balanced, but possibly no longer one-way over its range (and vice versa); we use the fact that f is regular (and that the number of preimages is known) when setting the output length of h. This is why our approach does not extend for general one-way functions.

Due to space limitations, some proofs have been omitted or shortened.

2 Preliminaries

Throughout this paper, we let k denote the security parameter. If X_1 and X_2 are two distributions over a set $\mathcal{X}$, their statistical difference (written $\mathsf{SD}(X_1, X_2)$) is defined as:

$$\mathsf{SD}(X_1, X_2) \stackrel{\text{def}}{=} \frac{1}{2} \sum_{x \in \mathcal{X}} \left| \Pr_{X_1}[x] - \Pr_{X_2}[x] \right| .$$

Two distribution ensembles $\mathcal{X}_1 = \{X_1(k)\}_{k \in \mathbb{N}}$ and $\mathcal{X}_2 = \{X_2(k)\}_{k \in \mathbb{N}}$ have statistical difference ρ (for ρ a function of k) if $\mathsf{SD}(X_1(k), X_2(k)) \leq \rho(k)$ for all k large

[2] We remark that known constructions of "almost-everywhere one-to-one" one-way functions [14], "almost one-to-one" one-way functions [10, Sect. 3.5], and the constructions of [20] do not suffice for our purposes.

enough. If ρ is negligible, we say the ensembles are *statistically indistinguishable*. For a function $f : \{0,1\}^n \to \{0,1\}^\ell$, we let $\mathsf{image}(f) \stackrel{\text{def}}{=} \{f(x) \mid x \in \{0,1\}^n\}$.

2.1 Commitment schemes

An interactive bit commitment scheme is defined via a triple of PPT algorithms $(\mathcal{S}, \mathcal{R}_1, \mathcal{R}_2)$. Looking ahead, $\mathcal{S}$ and $\mathcal{R}_1$ will interact during what is called a *commitment phase*, while $\mathcal{R}_2$ will be used during the (non-interactive) *decommitment phase*. More formally:

- $\mathcal{S}$ (the *sender*) is an interactive Turing machine (ITM) which receives as initial input the security parameter 1^k and a bit b. Following its interaction, it outputs some information decom (the *decommitment*).
- $\mathcal{R}_1$ (the *receiver*) is an ITM which receives the security parameter 1^k as initial input. Following its interaction, it outputs some state information s.
- $\mathcal{R}_2$ (acting as a receiver, in the decommitment phase) is a deterministic algorithm which receives as input state information s and a decommitment decom; it outputs either a bit b or the distinguished value $\perp$.

Denote by $(\mathsf{decom} \mid s) \leftarrow \langle \mathcal{S}(1^k, b), \mathcal{R}_1(1^k)\rangle$ the experiment in which $\mathcal{S}$ and $\mathcal{R}_1$ interact (using the given inputs and uniformly random coins), and then $\mathcal{S}$ outputs decom while $\mathcal{R}_1$ outputs s. We make the following correctness requirement: for all k, all b, and every pair $(\mathsf{decom} \mid s)$ that may be output by $\langle \mathcal{S}(1^k, b), \mathcal{R}_1(1^k)\rangle$, it is the case that $\mathcal{R}_2(s, \mathsf{decom}) = b$.

The security of a commitment scheme can be defined in two complementary ways, protecting against either an all-powerful sender or an all-powerful receiver. Since we are interested in the case of statistically-hiding commitment (i.e., the latter case), we only provide the definition for this case.

Definition 1. Commitment scheme $(\mathcal{S}, \mathcal{R}_1, \mathcal{R}_2)$ is *ρ-hiding* (for ρ a function of k) if the following holds: Given a deterministic ITM $\mathcal{R}_1^*$, let $\mathsf{view}_{\langle \mathcal{S}(b), \mathcal{R}_1^*\rangle}(k)$ denote the distribution on the view of $\mathcal{R}_1^*$ when interacting with $\mathcal{S}(1^k, b)$ (this view simply consists of the sequence of messages it receives from $\mathcal{S}$), where this distribution is taken over the random coins of $\mathcal{S}$. Then we require that for any (even all-powerful) $\mathcal{R}_1^*$ the ensembles $\{\mathsf{view}_{\langle \mathcal{S}(0), \mathcal{R}_1^*\rangle}(k)\}$ and $\{\mathsf{view}_{\langle \mathcal{S}(1), \mathcal{R}_1^*\rangle}(k)\}$ have statistical difference at most ρ.

Note that in the above, considering a deterministic $\mathcal{R}_1^*$ is without loss of generality. We say a scheme is *statistically hiding* if it is ρ-hiding for negligible ρ. A 0-hiding scheme is called *perfectly hiding*.

Definition 2. Commitment scheme $(\mathcal{S}, \mathcal{R}_1, \mathcal{R}_2)$ is *computationally-binding* if the following is negligible for all PPT $\mathcal{S}^*$:

$$\Pr\left[((\mathsf{decom}, \mathsf{decom}') \mid s) \leftarrow \langle \mathcal{S}^*(1^k), \mathcal{R}_1(1^k)\rangle : \begin{array}{c}\mathcal{R}_2(s, \mathsf{decom}), \mathcal{R}_2(s, \mathsf{decom}') \in \{0,1\} \\ \bigwedge \mathcal{R}_2(s, \mathsf{decom}) \neq \mathcal{R}_2(s, \mathsf{decom}')\end{array}\right],$$

where the probability is taken over the random coins of both $\mathcal{S}^*$ and $\mathcal{R}_1$.

Given the above, we now define a statistically-secure commitment scheme:

Definition 3. Commitment scheme $(\mathcal{S}, \mathcal{R}_1, \mathcal{R}_2)$ is *ρ-secure* (resp., *statistically secure*, *perfectly secure*) if it is computationally binding and ρ-hiding (resp., statistically hiding, perfectly hiding).

2.2 One-Way Function Families and Variants

Let $n, \ell = \mathsf{poly}(k)$ be poly-time computable and let $\mathcal{F} = \{f_k : \{0,1\}^{n(k)} \to \{0,1\}^{\ell(k)}\}_{k \in \mathbb{N}}$ be a function family. We say $\mathcal{F}$ is *one-way* if the following hold:

– (**efficiently computable**) There exists a (deterministic) polynomial-time algorithm E such that, for all k and all $x \in \{0,1\}^{n(k)}$, $E(1^k, x) = f_k(x)$.
– (**one-way**) For all PPT algorithms A, the following is negligible (in k):

$$\Pr_{x \leftarrow \{0,1\}^{n(k)}}[f_k(A(1^k, f_k(x))) = f_k(x)].$$

We consider two additional properties of function families:

– $\mathcal{F}$ is $r(k)$**-regular** if for every k and every $x \in \{0,1\}^{n(k)}$ we have

$$\left|\{x' \in \{0,1\}^{n(k)} \mid f_k(x') = f_k(x)\}\right| = 2^{r(k)}$$

and $r(k)$ is poly-time computable.[3] In other words, for each $x \in \{0,1\}^{n(k)}$ there are exactly $2^{r(k)}$ elements (including x itself) which f_k maps to the same value.
– $\mathcal{F}$ is **approximable-preimage-size** if the function $\tilde{D}_{\mathcal{F}}(y,k) \stackrel{\text{def}}{=} \lceil \log(|f_k^{-1}(y)|\ \rceil$ is polynomial-time computable.[4]

For simplicity, we drop the explicit dependence on k when clear. Note that any regular function family is also approximable-preimage-size.

2.3 Entropy Measures

Let U_n denote the uniform distribution over $\{0,1\}^n$. Given a function $f : \{0,1\}^n \to \{0,1\}^\ell$, we let $f(U_n)$ denote the distribution over $\{0,1\}^\ell$ induced by f operating on the uniform distribution. Given a distribution D over some set X, the *support* of D is defined to be the set $\{x \in X | D(x) > 0\}$. For D a distribution over some finite domain X, we use the following "measures" of entropy:

– The *min-entropy* of D is $H_\infty(D) \stackrel{\text{def}}{=} \min_{x \in X} \log(\frac{1}{D(x)})$.
– The *max-entropy* of D is $H_{max}(D) \stackrel{\text{def}}{=} \max_{x \in X} \log(\frac{1}{D(x)})$.

[3] Some previous definitions of regular functions do not require that r be poly-time computable. However, we do not know how to extend our results to this case.

[4] Our constructions generalize to the case where $r(k)$ (resp., $\tilde{D}_{\mathcal{F}}(y,k)$) are not computed precisely, but rather approximated to within an additive factor of $O(\log(k))$.

– The *Renyi entropy* of D is $H_2(D) \stackrel{\text{def}}{=} \log(\frac{1}{CP(D)})$, where $CP(D) \stackrel{\text{def}}{=} \sum_{x \in X} D(x)^2$ is the *collision probability* of D.

We will be interested in distributions of the form $D = f(U_n)$ for $f : \{0,1\}^n \to \{0,1\}^l$. Note that if f is r-regular, then D is uniform over some subset of $\{0,1\}^\ell$ and the above three measures coincide (and D has entropy $t = n - r$).

2.4 Universal Hashing and an Extended Chernoff Bound

Let $\mathcal{H} = \{H_k\}_{k \in \mathbb{N}}$ be a sequence of function families, where each H_k is a family of functions mapping strings of length $\ell(k)$ to strings of length $v(k)$. We say H_k is an $n(k)$-*universal hash family* (following [7]) if for any distinct $x_1, \ldots, x_{n(k)} \in \{0,1\}^{\ell(k)}$, and any $y_1, \ldots, y_{n(k)} \in \{0,1\}^{v(k)}$ we have:

$$\Pr_{h \leftarrow H_k}[h(x_1) = y_1 \wedge \cdots \wedge h(x_n) = y_n] = 2^{-v(k) \cdot n}.$$

In this paper, it is convenient to assume that for every k, the size of H_k is a power of two. This allows us to identify functions $h \in H_k$ with binary strings. We use $s(k)$ to denote the length of these strings.

We say that $\mathcal{H}$ is an $n(k)$-universal hash family if for every k, H_k is an $n(k)$-universal hash family and furthermore there is a polynomial time algorithm that given 1^k, $x \in \{0,1\}^{n(k)}$ and a string $h \in \{0,1\}^{s(k)}$ outputs $h(x)$ (where $h \in H_k$ is the function described by the string $h \in \{0,1\}^{s(k)}$). It is well-known that there is a family of functions with this property for every choice of ℓ and v with $s(k) = O(n(k) \cdot \max(\ell(k), v(k)))$.

The following Chernoff-like bound will be useful in our analysis:

Lemma 1. (Extended Chernoff Bound [32, Theorem 5]) *Let X be the sum of (any number of) n-wise independent random variables, each taking values in the interval $[0,1]$, such that $E[X] = \mu$. Then for any $\varepsilon \leq 1$ for which $n \geq \lfloor \varepsilon^2 \mu e^{-1/3} \rfloor$ we have $\Pr[|X - \mu| \geq \varepsilon\mu] \leq e^{-\lfloor \varepsilon^2 \mu/3 \rfloor}$.*

2.5 Interactive Hashing and the Construction of [25]

Interactive hashing was introduced by Ostrovsky, et al. [27, 28], and used by Naor, et al. [25] to construct a statistically-secure (actually, perfectly-secure) commitment scheme based on any one-way permutation family. We review interactive hashing, as well as the resulting commitment scheme, below. In what follows, we let $x \cdot y$ denote $\sum_{i=1}^m x_i y_i \bmod 2$ for $x, y \in \{0,1\}^m$.

Construction 4 (Interactive hashing) *The protocol is defined by algorithms $\mathcal{S}$ and $\mathcal{R}$, where $\mathcal{S}$ begins with an m-bit value y (with m known to $\mathcal{R}$), and proceeds as follows:*

1. *The parties interact in $m-1$ stages. In stage i (for $i = 1, \ldots, m-1$), $\mathcal{R}$ chooses $r_i \in \{0,1\}^{m-i}$ uniformly at random and sends the "query" $q_i = 0^{i-1}1r_i$ to $\mathcal{S}$ (in case $\mathcal{R}$ aborts, $\mathcal{S}$ simply takes q_i to be some default value); in response, $\mathcal{S}$ sends $c_i = q_i \cdot y$.*

2. *At the conclusion of the above, there are exactly two strings* $y_0, y_1 \in \{0,1\}^m$ *satisfying the system of equations* $\{q_i \cdot X = c_i\}_{1 \le i \le m-1}$*; let* y_0 *denote the lexicographically smaller of the two. Both parties compute* (y_0, y_1)*, and* $\mathcal{S}$ *chooses* v *such that* $y = y_v$.

We define the output of the protocol to be (y_0, y_1, v) *for* $\mathcal{S}$ *and* (y_0, y_1) *for* $\mathcal{R}$*. We denote by* $IH(y)$ *an execution of the interactive hashing protocol, where* $\mathcal{S}$ *begins with input* y.

The above was used in [25] to construct a perfectly-secure commitment scheme based on one-way permutations via the following approach:

Construction 5 *Let* $\mathcal{F} = \{f_k : \{0,1\}^{n(k)} \to \{0,1\}^{\ell(k)}\}$ *be a function family. Commitment scheme* $(\mathcal{S}, \mathcal{R}_1, \mathcal{R}_2)$ *is defined as follows:* $\mathcal{S}(1^k, b)$ *chooses* $x \in \{0,1\}^{n(k)}$ *uniformly at random, computes* $y = f_k(x)$*, and then executes* $IH(y)$ *with* $\mathcal{R}_1$*; this protocol results in output* (y_0, y_1, v) *for* $\mathcal{S}$ *and* (y_0, y_1) *for* $\mathcal{R}_1$*. The commitment phase concludes by having* $\mathcal{S}$ *send* $\hat{v} = v \oplus b$ *to* $\mathcal{R}_1$*. Finally,* $\mathcal{S}$ *outputs* $\mathsf{decom} = x$ *while* $\mathcal{R}_1$ *outputs state* $s = (y_0, y_1, \hat{v})$.

In the decommitment phase, $\mathcal{R}_2((y_0, y_1, \hat{v}), x)$ *proceeds as follows: if* $f_k(x) = y_0$*, output* $\hat{v}$*; if* $f_k(x) = y_1$*, output* $\hat{v} \oplus 1$*; otherwise, output* $\perp$.

It is relatively easy to observe that the above protocol is perfectly hiding if $\mathcal{F}$ is a permutation family (regardless of whether $\mathcal{F}$ is one-way). The main result of [25] was to prove that the above is *computationally binding* when $\mathcal{F}$ is a *one-way* permutation family. In fact, careful examination of their proof shows the above commitment scheme is computationally binding under a *weaker* condition on $\mathcal{F}$; it suffices for $\mathcal{F}$ to be what we call "one-way over its range", defined as follows:

Definition 6. Let $n, \ell = \mathsf{poly}(k)$ be poly-time computable functions and let $\mathcal{F} = \{f_k : \{0,1\}^{n(k)} \to \{0,1\}^{\ell(k)}\}_{k \in \mathbb{N}}$ be a function family. We say $\mathcal{F}$ is *one-way over its range* if the following hold:

– (**efficiently computable**) There exists a (deterministic) polynomial-time algorithm E such that, for all k and all $x \in \{0,1\}^{n(k)}$, $E(1^k, x) = f_k(x)$.
– (**one-way over range**) For all PPT A, the following is negligible (in k):

$$\Pr_{y \leftarrow \{0,1\}^{\ell(k)}}[f_k(A(1^k, y)) = y].$$

Theorem 1 (Implicit in [25]). *If* $\mathcal{F}$ *is one-way over its range, then Construction 5 is computationally binding.*

3 Statistical Hiding from Balanced Functions

In this section we define a notion of "balance" and show that if a function family $\mathcal{F}$ is "sufficiently balanced" then Construction 5 yields a protocol that is "somewhat hiding". Roughly speaking, a distribution D on $\{0,1\}^\ell$ is balanced if D is "close" to uniform "most" of the time. A function $f : \{0,1\}^n \to \{0,1\}^\ell$ is then defined to be balanced if the distribution $f(U_n)$ is balanced. Formally:

Definition 7. Distribution D on $\{0,1\}^\ell$ is (α,δ)-*balanced* if there is a set $\mathsf{Bad} \subset \{0,1\}^\ell$ such that:

1. $|\mathsf{Bad}| \leq \alpha \cdot 2^\ell$.
2. $\Pr_{y \leftarrow D}[y \in \mathsf{Bad}] \leq \alpha$.
3. For every $y_0 \notin \mathsf{Bad}$, $\left|\Pr_{y \leftarrow D}[y = y_0] - \frac{1}{2^\ell}\right| \leq \frac{\delta}{2^\ell}$ (we will always have $\delta < 1$).

Function $f : \{0,1\}^n \to \{0,1\}^\ell$ is (α,δ)-*balanced* if the distribution $f(U_n)$ is (α,δ)-balanced. Function family $\mathcal{F} = \{f_k : \{0,1\}^{n(k)} \to \{0,1\}^{\ell(k)}\}$ is (α,δ)-*balanced* if, for all k large enough, f_k is $(\alpha(k),\delta(k))$-balanced.

Our main result of this section is the following:

Theorem 2. *If $\mathcal{F} = \{f_k : \{0,1\}^{n(k)} \to \{0,1\}^{\ell(k)}\}$ is an (α,δ)-balanced function family, then Construction 5 is ρ-hiding for $\rho = 2\alpha + \delta + \alpha\delta$.*

Proof. Fix k large enough so that f_k is $(\alpha(k),\delta(k))$-balanced; from now on we simply write f,α,δ,ρ without explicitly indicating their dependence on k. For a given execution of the scheme, let τ denote the initial transcript resulting from the interactive hashing sub-protocol; thus, the view of $\mathcal{R}_1^*$ consists of τ and the bit $\hat{v}$ sent in the final round. Given a particular (deterministic) $\mathcal{R}_1^*$, we therefore write $\mathsf{Exp(b)} \stackrel{\text{def}}{=} (\tau,\hat{v}) \leftarrow \mathsf{view}_{\langle \mathcal{S}(b),\mathcal{R}_1^*\rangle}$ (cf. Definition 3) to denote the experiment in which $\mathcal{S}$ chooses a uniform random tape and then executes the protocol with $\mathcal{R}_1^*$ using this random tape and the bit b, resulting in view $(\tau,\hat{v})$ for $\mathcal{R}_1^*$. Below, we define a "good" set of initial transcripts Good, and show that:

Claim. With probability at least $1-\alpha(2+\delta)$, we have $\tau \in \mathsf{Good}$.

Claim. The following holds for all $\tau^* \in \mathsf{Good}$ and $\hat{v}^* \in \{0,1\}$:

$$\left|\Pr_{\mathsf{Exp(0)}}[\hat{v} = \hat{v}^* \mid \tau = \tau^*] - \Pr_{\mathsf{Exp(1)}}[\hat{v} = \hat{v}^* \mid \tau = \tau^*]\right| \leq \delta.$$

These claims suffice to prove the Theorem, since the statistical difference between the view of $\mathcal{R}_1^*$ when the sender commits to 0 (i.e., $b=0$) and the view of $\mathcal{R}_1^*$ when the sender commits to 1 (i.e., $b=1$) may be bounded as follows:

$$\begin{aligned}
&\frac{1}{2}\sum_{\tau^*,\hat{v}^*}\left|\Pr_{\mathsf{Exp(0)}}[(\tau,\hat{v}) = (\tau^*,\hat{v}^*)] \; - \Pr_{\mathsf{Exp(1)}}[(\tau,\hat{v}) = (\tau^*,\hat{v}^*)]\right| \\
&= \frac{1}{2}\sum_{\tau^*,\hat{v}^*}\left|\Pr_{\mathsf{Exp(0)}}[\tau = \tau^*]\Pr_{\mathsf{Exp(0)}}[\hat{v} = \hat{v}^* | \tau = \tau^*] \; - \Pr_{\mathsf{Exp(1)}}[\tau = \tau^*]\Pr_{\mathsf{Exp(1)}}[\hat{v} = \hat{v}^* | \tau = \tau^*]\right| \\
&\leq \Pr[\tau \notin \mathsf{Good}] + \frac{1}{2}\sum_{\tau^* \in \mathsf{Good},\hat{v}^*}\Pr[\tau = \tau^*]\left|\Pr_{\mathsf{Exp(0)}}[\hat{v} = \hat{v}^* | \tau = \tau^*] \; - \Pr_{\mathsf{Exp(1)}}[\hat{v} = \hat{v}^* | \tau = \tau^*]\right| \\
&\leq \alpha(2+\delta) + \frac{1}{2}\sum_{\tau^* \in \mathsf{Good};\hat{v}^*}\Pr[\tau = \tau^*] \cdot \delta \quad \leq \quad \alpha(2+\delta) + \delta,
\end{aligned}$$

where we use the fact that $\Pr_{\mathsf{Exp}(0)}[\tau = \tau^*] = \Pr_{\mathsf{Exp}(1)}[\tau = \tau^*]$ for any τ^*, since the initial transcript τ does not depend on b.

We proceed with the proof of the first claim by defining the set of good initial transcripts. Let $\mathsf{Bad} \subset \{0,1\}^\ell$ be the subset whose existence is guaranteed by Definition 7 (using the fact that f is balanced). Recall that the initial transcript τ defines two strings $y_0^\tau, y_1^\tau \in \{0,1\}^\ell$ (cf. Construction 4). We say $\tau \in \mathsf{Good}$ iff $y_0^\tau, y_1^\tau \notin \mathsf{Bad}$.

We first bound the probability that $y_v = y$ is in Bad (we are using here the notation from Construction 5). Since f is (α, δ)-balanced and since the value of y depends only on the choices of the sender (who is assumed honest here), it follows that this probability is at most α.

Next, we bound the probability that $y_v \notin \mathsf{Bad}$ but $y_{\bar{v}} \in \mathsf{Bad}$. Since f is balanced, we have $|\mathsf{Bad}| \leq \alpha 2^\ell$. Now, since $\mathcal{R}_1^*$ is deterministic, we have that $y_{\bar{v}}$ is uniquely determined by y_v. Let ϕ be the function mapping the sender's chosen value y_v to the second value $y_{\bar{v}}$ resulting from the interactive hashing protocol. Observe that if $\phi(y) = y'$ then $\phi(y') = y$; this is because, for either of these choices, the sender responds with the exact same answer to each of the receiver's queries during the interactive hashing sub-protocol. It follows that ϕ is a permutation. Letting $\mathsf{MapToBad} \stackrel{\text{def}}{=} \phi^{-1}(\mathsf{Bad})$, we get:

$$\begin{aligned}
\Pr\left[y_v \notin \mathsf{Bad} \bigwedge y_{\bar{v}} \in \mathsf{Bad}\right] &= \Pr\left[y_v \in \mathsf{MapToBad} \setminus \mathsf{Bad}\right] \\
&= \sum_{y^* \in \mathsf{MapToBad} \setminus \mathsf{Bad}} \Pr\left[y_v = y^*\right] \\
&\leq \sum_{y^* \in \mathsf{MapToBad} \setminus \mathsf{Bad}} (1+\delta)\frac{1}{2^\ell}
\end{aligned}$$

using the definition of Bad. Continuing:

$$\begin{aligned}
\sum_{y^* \in \mathsf{MapToBad} \setminus \mathsf{Bad}} (1+\delta)\frac{1}{2^\ell} &= |\mathsf{MapToBad} \setminus \mathsf{Bad}| \cdot (1+\delta)\frac{1}{2^\ell} \\
&\leq |\mathsf{MapToBad}| \cdot (1+\delta)\frac{1}{2^\ell} \\
&\leq (1+\delta) \cdot \alpha \qquad (1)
\end{aligned}$$

(using the fact that $|\mathsf{MapToBad}| = |\mathsf{Bad}|$). It follows that $\tau \notin \mathsf{Good}$ with probability at most $(2+\delta) \cdot \alpha$, completing the proof of the first claim.

We proceed to prove the second claim. Let $P(\tilde{y}) \stackrel{\text{def}}{=} \Pr_{x \in \{0,1\}^n}[f(x) = \tilde{y}]$. For any τ^* and any $\hat{v}^* \in \{0,1\}$ we have

$$\begin{aligned}
\Pr_{\mathsf{Exp(b)}}[\hat{v} = \hat{v}^* \mid \tau = \tau^*] &= \Pr_{\mathsf{Exp(b)}}[v = \hat{v}^* \oplus b \mid \tau = \tau^*] \\
&= \Pr_{\mathsf{Exp(b)}}[y = y^{\tau^*}_{\hat{v}^* \oplus b} \mid y \in \{y_0^{\tau^*}, y_1^{\tau^*}\}] \\
&= \frac{P(y^{\tau^*}_{\hat{v}^* \oplus b})}{P(y_0^{\tau^*}) + P(y_1^{\tau^*})}.
\end{aligned}$$

If $\tau^* \in \mathsf{Good}$, then $y_0^{\tau^*}, y_1^{\tau^*} \notin \mathsf{Bad}$ and so $P(y_0^{\tau^*}), P(y_1^{\tau^*})$ lie in the range $[(1-\delta)2^{-\ell}, (1+\delta)2^{-\ell}]$. It follows that when $\tau^* \in \mathsf{Good}$ the following holds for any $\hat{v}^* \in \{0,1\}$:

$$\left| \Pr_{\mathsf{Exp}(0)}[\hat{v} = \hat{v}^* \mid \tau = \tau^*] - \Pr_{\mathsf{Exp}(1)}[\hat{v} = \hat{v}^* \mid \tau = \tau^*] \right| = \frac{\left|P(y_0^{\tau^*}) - P(y_1^{\tau^*})\right|}{P(y_0^{\tau^*}) + P(y_1^{\tau^*})} \leq \delta,$$

which proves the claim. This completes the proof of the Theorem 2.

4 Achieving Our Main Result: a Roadmap

We now outline our approach to constructing statistically-secure commitment schemes based on assumptions weaker than one-way permutations. It follows from Theorems 1 and 2 that if we can construct an (α, δ)-balanced $\mathcal{F}$ that is also one-way over its range, then we can construct a ρ-secure commitment scheme for $\rho = O(\alpha + \delta)$. For α and δ sufficiently-small constants we thus obtain a ρ-secure commitment scheme for some constant $\rho < 1$. Using standard techniques, we can then "amplify" this scheme to obtain a statistically-secure commitment scheme. (Exact details of this amplification will appear in the full version.)

It remains to construct $\mathcal{F}$ with the desired properties. In Section 5 we show how to construct such an $\mathcal{F}$ based on any regular one-way function family, while in Section 6 we show how to base the construction on an approximable-preimage-size one-way function family. These, in turn, yield statistically-secure commitment schemes based on these assumptions. Altogether we conclude that:

Theorem 3 (Main Theorem). *If there exists an approximable-preimage-size one-way function family then there exists a statistically-secure commitment scheme.*

5 Starting from Regular One-Way Functions

In this section we show a construction of statistically-secure commitment based on any regular one-way function family. More concretely, given an $r(k)$-regular one-way function family $\mathcal{F}$, we show how to construct a balanced function $\mathcal{F}'$ which is also one-way over its range. Note that $n(k) - r(k)$ measures the entropy of the output distribution of f_k, and this holds for all the measures of entropy defined in this paper.

Construction 8 *Let $\mathcal{F} = \{f_k : \{0,1\}^{n(k)} \to \{0,1\}^{\ell(k)}\}_{k \in \mathbb{N}}$ be a family of functions, let $t = t(k)$ be a function, and let $c > 0$ be a constant. Let $\mathcal{H} = \{H_k\}$ be a $3k$-universal collection of hash families where each H_k is a family of functions mapping strings of length $\ell(k)$ to strings of length $t(k) - \log(ck)$, and furthermore $|H_k| = 2^{s(k)}$ where $s(k) = \mathsf{poly}(k)$. Define:*

$$\mathcal{F}' = \left\{ f'_k : H_k \times \{0,1\}^{n(k)} \to H_k \times \{0,1\}^{t(k) - \log(ck)} \right\}_{k \in \mathbb{N}}$$

such that $f'_k(h, x) = (h, h(f_k(x)))$.

The main result of this section is the following.

Theorem 4. *Let $0 < \delta < 1$ be an arbitrary constant. Let $\mathcal{F}$ be an $r(k)$-regular one-way function family. Set $t(k) = n(k) - r(k)$, $c = 6\ln 2/\delta^2$, and let $\mathcal{F}'$ be the function family defined in Construction 8. Then $\mathcal{F}'$ is a $(2^{-k}, \delta)$-balanced and one-way over its range.*

5.1 Showing that $\mathcal{F}'$ is Balanced

We begin by showing that $\mathcal{F}'$ is $(2^{-k}, \delta)$-balanced. Preparing for the case of approximable-preimage-size one-way function families, we prove a more general statement here.

Lemma 2. *Let $c > 6\ln 2$ be an arbitrary constant and $k \geq 2$ be an integer, and set $\delta = (6\ln 2/c)^{1/2}$ and $t > \log(ck)$. Let H be a $3k$-universal hash family mapping strings of length ℓ to strings of length $t - \log(ck)$, and let Z be a distribution on $\{0,1\}^\ell$ with $H_\infty(Z) \geq t$. Then the distribution $D = \{(h, h(z))\}_{h\leftarrow H, z\leftarrow Z}$ is $(2^{-k}, \delta)$-balanced.*

Note that it follows that $\mathcal{F}'$ is $(2^{-k}, \delta)$-balanced, as the output distribution of f_k has min-entropy at least $t(k)$ (in fact, exactly $t(k)$).

Proof. For any $z \in \{0,1\}^\ell$ and $y \in \{0,1\}^{t-\log(ck)}$, define the random variable $X_{z,y}$ (over choice of $h \in H$) to take the value $2^t \cdot \Pr_Z[z]$ if $h(z) = y$, and 0 otherwise. Note that $X_{z,y} \in [0,1]$ since Z has min-entropy at least t. Let $X_y \stackrel{\text{def}}{=} \sum_{z\in\{0,1\}^\ell} X_{z,y}$. For any z, y we have $E[X_{z,y}] = \Pr_{h\leftarrow H}[h(z) = y] \cdot 2^t \cdot \Pr_Z[z] = 2^{-(t-\log(ck))} \cdot 2^t \cdot \Pr_Z[z] = ck \cdot \Pr_Z[z]$. It follows that

$$\mu \stackrel{\text{def}}{=} E[X_y] = \sum_z E[X_{z,y}] = ck.$$

Furthermore, since H is a $3k$-universal hash family, the random variables $\{X_{z,y}\}$ are $3k$-wise independent. Thus, by Lemma 1, we have that (for any y)

$$\Pr_h\left[\left|X_y - ck\right| \geq \delta ck\right] \leq e^{-\lfloor \mu\delta^2/3 \rfloor} < 2^{-k} \tag{2}$$

Define $\phi(h, y) \stackrel{\text{def}}{=} 2^t \cdot \sum_{z:h(z)=y} \Pr_Z[z]$, and $\mathsf{Bad} = \{(h, y) : |\phi(h, y) - ck| > \delta ck\}$. We show that, setting $\alpha = 2^{-k}$, the set Bad satisfies the three requirements of Definition 7. (Note that the quantity 2^ℓ in the text of Definition 7 becomes $|H| \cdot 2^{t-\log(ck)}$ in the current context.) Noting that $\phi(h, y) = 2^t \Pr_{z\leftarrow Z}[h(z) = y]$, observe that

$$\begin{aligned}
|\mathsf{Bad}| &= \sum_y |H| \cdot \Pr_h[(h, y) \in \mathsf{Bad}] \\
&= \sum_y |H| \cdot \Pr_h\left[\left|2^t \cdot \Pr_{z\leftarrow Z}[h(z) = y] - ck\right| > \delta ck\right] \\
&\leq 2^{t-\log(ck)} \cdot |H| \cdot 2^{-k},
\end{aligned}$$

using Eq. (2) and the fact that, once h is chosen, $X_y = 2^t \cdot \Pr_{z \leftarrow Z}[h(z) = y]$. This proves property 1.

We move on to property 2. We proceed as above except that now, for each $\xi, z \in \{0,1\}^\ell$, we define the binary random variable $R_{z,\xi}$ to be $2^t \cdot \Pr_Z[z]$ if $h(z) = h(\xi)$, and 0 otherwise. Again, $R_{z,\xi} \in [0,1]$. Let $R_\xi \stackrel{\text{def}}{=} \sum_{z \in \{0,1\}^\ell} R_{z,\xi}$. For an arbitrary $z \in \{0,1\}^\ell \setminus \{\xi\}$ we have $E[R_{z,\xi}] = 2^{-(t-\log(ck))} \cdot 2^t \cdot \Pr_Z[z] = ck \Pr_Z[z]$; also $R_{\xi,\xi} = 2^t \Pr_Z[\xi]$ with probability 1. It follows that

$$\mu' \stackrel{\text{def}}{=} E[R_\xi] = \sum_z E[R_{z,\xi}] = ck + (2^t - ck) \Pr_Z[\xi]$$

for any ξ. Note that $ck \le \mu' \le ck + 1$. Furthermore, since H is a $3k$-universal hash family, the random variables $\{R_{z,\xi}\}$ are $(3k-1)$-wise independent. Thus, by Lemma 1 we have

$$\Pr\left[|R_\xi - \mu'| \ge \frac{3}{4}\delta\mu'\right] \le e^{-\lfloor 3\mu'\delta^2/16 \rfloor} \le 2^{-k}, \tag{3}$$

where we use the fact that $\mu' \frac{9}{16}\delta^2 e^{-1/3} \le (ck+1)\frac{9}{16}\delta^2 e^{-1/3} \le 3k - 1$ (recall $k \ge 2$). We then derive:

$$\begin{aligned}\Pr_{(h,y)\leftarrow D}[(h,y) \in \mathsf{Bad}] &= \sum_\xi \Pr_Z[\xi] \cdot \Pr_h\left[\left|\phi(h, h(\xi)) - ck\right| > \delta ck\right] \\ &\le \sum_\xi \Pr_Z[\xi] \cdot \Pr_h\left[\left|R_\xi - E[R_\xi]\right| \ge \frac{3}{4}\delta E[R_\xi]\right] \le 2^{-k},\end{aligned}$$

where the first inequality uses the stated bounds on μ' and the fact that, once h is chosen, $R_\xi = 2^t \cdot \Pr_{z \leftarrow Z}[h(z) = h(\xi)]$, while the second inequality uses Eq. (3). This gives property 2.

Property 3 holds, since for any (h_0, y_0) we have

$$\Pr_{(h,y)\leftarrow D}[(h,y) = (h_0,y_0)] = \Pr_{h \leftarrow H}[h = h_0] \cdot \sum_{z : h_0(z) = y_0} \Pr_Z[z] = \frac{\phi(h_0,y_0)}{|H|2^t}.$$

If $(h_0, y_0) \notin \mathsf{Bad}$, this probability is in the range $(1 \pm \delta)\frac{ck}{|H|2^t}$ as needed.

5.2 Showing that $\mathcal{F}'$ is One-Way Over Its Range

We now show that if the initial function family $\mathcal{F}$ is one-way, then the derived function family $\mathcal{F}'$ is one-way over its range. Preparing for the case of approximable-preimage-size one-way function families, we once more prove a more general statement here. For this purpose we define the following:

Definition 9. Distribution D has (t_{Renyi}, t_{max})-*entropy* if (1) $H_2(D) \ge t_{Renyi}$, and (2) $H_{max}(D) \le t_{max}$. Function $f : \{0,1\}^n \to \{0,1\}^\ell$ has (t_{Renyi}, t_{max})-*entropy* if the distribution $f(U_n)$ has (t_{Renyi}, t_{max})-entropy. A function family $\mathcal{F} = \{f_k : \{0,1\}^{n(k)} \to \{0,1\}^{\ell(k)}\}$ has (t_{Renyi}, t_{max})-*entropy* if, for all k large enough, f_k has $(t_{Renyi}(k), t_{max}(k))$-entropy.

Note that if f is a member of an r-regular function family then it has (t,t)-entropy for $t = n - r$. The following lemma shows that Construction 8, when given a (t_{Renyi}, t_{max})-entropy family of one-way functions, produces a function family which is one-way over its range.

Lemma 3. *Let $\mathcal{F} = \{f_k : \{0,1\}^{n(k)} \to \{0,1\}^{\ell(k)}\}$ be a (t_{Renyi}, t_{max})-entropy one-way function family and let $c > 0$ be a constant. Let $t(k)$ be a function and let $m \geq 0$ be a constant such that $t_{max}(k) - m\log(k) \leq t(k) \leq t_{Renyi}(k)$. Let $\mathcal{F}'$ be the result of applying Construction 8 with $\mathcal{F}$, $t(\cdot)$, and c. Then $\mathcal{F}'$ is one-way over its range.*

Note that it follows that $\mathcal{F}'$ is one-way over its range by using the aforementioned observation that the regular function family $\mathcal{F}$ has $(t(k), t(k))$-entropy. We remark that the proof uses only the fact that $\mathcal{H}$ is 2-universal.

Proof. Let $v(k) \stackrel{\text{def}}{=} t(k) - \log(ck)$. We start by proving that the Renyi-entropy of the output of $\mathcal{F}'$ is high. We then use this fact to show that $\mathcal{F}'$ is one-way (in the usual sense). Finally we derive that $\mathcal{F}'$ is one-way over its range.

Claim. $H_2(f'_k(U_{s(k)}, U_{n(k)})) \geq s(k) + v(k) - 1$.

Proof.

$$\begin{aligned}
&CP(f'_k(U_{s(k)}, U_{n(k)})) \\
&= \sum_{(h,y)\in \mathsf{image}(f_k)} \left(\Pr_{(h',x)\leftarrow(H_k \times \{0,1\}^{n(k)})} [f'_k(h',x) = (h,y)] \right)^2 \\
&= \sum_{y\in\{0,1\}^{v(k)}} \sum_{h\in H_k} \frac{1}{2^{2s(k)}} \left(\sum_{z\in h^{-1}(y)} \Pr_{x\leftarrow\{0,1\}^{n(k)}} [f_k(x) = z] \right)^2 .
\end{aligned}$$

Continuing, we have:

$$\begin{aligned}
&CP(f'_k(U_{s(k)}, U_{n(k)})) \\
&= \frac{1}{2^{2s(k)}} \sum_{y\in\{0,1\}^{v(k)}} \sum_{h\in H_k} \sum_{z\in h^{-1}(y)} \left(\Pr_{x\leftarrow\{0,1\}^{n(k)}} [z] \right)^2 \\
&+ \frac{1}{2^{2s(k)}} \sum_{y\in\{0,1\}^{v(k)}} \sum_{h\in H_k} \sum_{z_1\neq z_2\in h^{-1}(y)} \left(\Pr_{x\leftarrow\{0,1\}^{n(k)}} [f_k(x) = z_1] \cdot \Pr_{x\leftarrow\{0,1\}^{n(k)}} [f_k(x) = z_2] \right) \\
&= \frac{1}{2^{s(k)}} CP(f_k(U_{n(k)})) \\
&+ \frac{1}{2^{2s(k)}} \sum_{y\in\{0,1\}^{v(k)}} \frac{2^{s(k)}}{2^{2v(k)}} \sum_{z_1\neq z_2\in\{0,1\}^{l(k)}} \left(\Pr_{x\leftarrow\{0,1\}^{n(k)}} [f_k(x) = z_1] \cdot \Pr_{x\leftarrow\{0,1\}^{n(k)}} [f_k(x) = z_2] \right) \\
&\leq \frac{1}{2^{s(k)}} (CP(f_k(U_{n(k)})) + \frac{1}{2^{v(k)}}) \leq \frac{2}{2^{s(k)+v(k)}} .
\end{aligned}$$

Therefore $H_2(f'_k(U_{s(k)}, U_{n(k)})) = -\log(CP(f'_k(U_{s(k)}, U_{n(k)}))) \geq s(k) + v(k) - 1$.

We now use the above claim to prove the one-wayness of $\mathcal{F}'$.

Claim. $\mathcal{F}'$ is one-way (in the usual sense).

Proof. Let A' be a PPT adversary attempting to invert $\mathcal{F}'$ and let $\mathsf{Expt}_{A'}(k)$ denote the experiment "$h \leftarrow H_k; x \leftarrow \{0,1\}^{n(k)}; (h, y) = f'_k(h, x); (h', x') \leftarrow A'(1^k, h, y)$". Let

$$\mathsf{Adv}_{A',\mathcal{F}'}(k) \stackrel{\text{def}}{=} \Pr[\mathsf{Expt}_{A'}(k) : f'_k(h', x') = (h, y)]. \tag{4}$$

Now construct a PPT adversary A (attempting to invert $\mathcal{F}$) as follows:

$\underline{A(1^k, z)}$ $\quad // \ z = f_k(x)$ for some $x \in \{0,1\}^{n(k)}$ chosen at random.
 Choose $h \in H_k$ at random, and set $y = h(z)$;
 Run $A'(1^k, h, y)$ and obtain output h', x';
 Output x'.

Note that the distribution over the inputs of A' in the above experiment is identical to the distribution over the inputs of A' in Equation 4. For any $k \in \mathbb{N}$, $h \in H_k$ and $y \in \{0,1\}^{v(k)}$ such that $\Pr_{x \leftarrow \{0,1\}^{n(k)}}[f'_k(h, x) = (h, y)] > 0$ let:

$$\theta_h(y) \stackrel{\text{def}}{=} \frac{\min_{z \in \mathsf{image}(f_k) \bigwedge h(z)=y} \left\{\Pr_{x \leftarrow \{0,1\}^{n(k)}} [f_k(x) = z]\right\}}{\Pr_{x \leftarrow \{0,1\}^{n(k)}}[f'_k(x, h) = (h, y)]}.$$

Observe that:

$$\begin{aligned} \mathsf{Adv}_{A,\mathcal{F}}(k) &\stackrel{\text{def}}{=} \Pr_{x \leftarrow \{0,1\}^{n(k)}; z = f_k(x); x' \leftarrow A(1^k, z)}[f_k(x') = z] \\ &\geq \sum_{\hat{h}, \hat{y}} \Pr_{\mathsf{Expt}_{A'}(k)}[h(f_k(x')) = y \bigwedge (h, y) = (\hat{h}, \hat{y})] \cdot \theta_{\hat{h}}(\hat{y}). \end{aligned}$$

We will make use of the following standard fact (proof in full version).

Claim. Let D be a distribution over some finite domain X such that $H_2(D) \geq k$ and let ε be any positive constant, then there exists a set $B \subseteq X$ such that the following hold: (1) $\Pr_D[B] \leq 4\varepsilon$, and (2) $\forall y \notin B \ \Pr_D[y] \leq \frac{2^{1-k}}{\varepsilon}$.

Let $\varepsilon \stackrel{\text{def}}{=} \mathsf{Adv}_{A',\mathcal{F}'}(k)$. Using the previous claims we have that there exists a set $\mathsf{Bad} \subseteq (H_k \times \{0,1\}^{v(k)})$ such that:

1. $\Pr_{(h,x) \leftarrow (H_k \times \{0,1\}^n)}[f'_k(h, x) \in \mathsf{Bad}] \leq \dfrac{\varepsilon}{2}$
2. $\forall (h', y') \notin \mathsf{Bad} \quad \Pr_{(h,x) \leftarrow (H_k \times \{0,1\}^n)}[f'_k(h, x) = (h', y')] \leq \dfrac{32}{\varepsilon 2^{s(k)+v(k)}}$.

Moreover, by our choice of the probability of Bad the following holds,

$$\Pr_{\mathsf{Expt}_{A'}(k)}[f'_k(h', x') = (h, y) \bigwedge (h, y) \notin \mathsf{Bad}] \geq \frac{\varepsilon}{2}.$$

Finally, by the definition of $v(k)$ the following holds for any $(h,y) \notin \mathsf{Bad}$

$$\theta_h(y) \geq \frac{\varepsilon 2^{v(k)}}{32 \cdot 2^{t_{max}(k)}} \geq \frac{\varepsilon}{32 \cdot (ck) \cdot k^m} = \frac{\varepsilon}{32 \cdot c \cdot k^{m+1}}$$

Hence:

$$\begin{aligned}
\mathsf{Adv}_{A,\mathcal{F}}(k) &\geq \sum_{(\hat{h},\hat{y}) \notin \mathsf{Bad}} \Pr_{\mathsf{Expt}_{A'}(k)}[h(f_k(x')) = y \bigwedge (h,y) = (\hat{h},\hat{y})] \cdot \theta_{\hat{h}}(\hat{y}) \\
&\geq \frac{\varepsilon}{32 \cdot c \cdot k^{m+1}} \sum_{(\hat{h},\hat{y}) \notin \mathsf{Bad}} \Pr_{\mathsf{Expt}_{A'}(k)}[h(f_k(x')) = y \bigwedge (h,y) = (\hat{h},\hat{y})] \\
&= \frac{\varepsilon}{32 \cdot c \cdot k^{m+1}} \Pr_{\mathsf{Expt}_{A'}(k)}[f'_k(h',x') = (h,y) \bigwedge (h,y) \notin \mathsf{Bad}] \\
&\geq \frac{\varepsilon}{32 \cdot c \cdot k^{m+1}} \cdot \frac{\varepsilon}{2} = \frac{\varepsilon^2}{64 \cdot c \cdot k^{m+1}}.
\end{aligned}$$

Since $\mathsf{Adv}_{A,\mathcal{F}}(k)$ is negligible by assumption, it must be the case that $\mathsf{Adv}_{A',\mathcal{F}'}(k)$ is negligible as well and thus $\mathcal{F}'$ is one way.

To finish the proof we show that $\mathcal{F}'$ is one-way over its range.

Claim. $\mathcal{F}'$ is one-way over its range.

Proof. Consider any PPT algorithm A'' inverting $\mathcal{F}'$ "over its range". The advantage of A'' (in this sense) is given by:

$$\begin{aligned}
\mathsf{Adv}^*_{A'',\mathcal{F}'} &\stackrel{\text{def}}{=} \Pr_{h \leftarrow H_k; y \leftarrow \{0,1\}^{v(k)}; (h',x') \leftarrow A''(1^k,h,y)}[f'_k(h',x') = (h,y)] \\
&= \frac{1}{2^{s(k)+v(k)}} \cdot \sum_{h \in H_k} \sum_{y \in \{0,1\}^{t(k)}} \Pr[A'' \text{ inverts } (h,y)],
\end{aligned}$$

where "A'' inverts (h,y)" has the obvious meaning.

Consider now the advantage of A'' in inverting $\mathcal{F}'$ in the standard sense:

$$\begin{aligned}
\mathsf{Adv}_{A'',\mathcal{F}'} &\stackrel{\text{def}}{=} \Pr_{h \leftarrow H_k; x \leftarrow \{0,1\}^{n(k)}}[A'' \text{ inverts } (h, h(f_k(x)))] \\
&= \frac{1}{2^{s(k)+n(k)}} \sum_{h \in H_k} \sum_{x \in \{0,1\}^{n(k)}} \Pr[A'' \text{ inverts } (h, h(f_k(x)))] \\
&= \frac{1}{2^{s(k)+n(k)}} \sum_{h \in H_k} \sum_{z \in \mathsf{image}(f_k)} \Pr_{x \leftarrow \{0,1\}^{n(k)}}[f_k(x) = z] \cdot \Pr[A'' \text{ inverts } (h, h(z))] \\
&\geq \frac{1}{2^{s(k)+t_{max}(k)}} \sum_{h \in H_k} \sum_{y \in \mathsf{image}(h(f_k))} \sum_{z \in h^{-1}(y)} \Pr[A'' \text{ inverts } (h, h(z))] \\
&\geq \frac{1}{2^{s(k)+t_{max}(k)}} \sum_{h \in H_k} \sum_{y \in \{0,1\}^{v(k)}} \Pr[A'' \text{ inverts } (h,y)] \\
&= \frac{2^{s(k)+v(k)}}{2^{s(k)+t_{max}(k)}} \mathsf{Adv}^*_{A'',f'} \quad \geq \quad \frac{\mathsf{Adv}^*_{A'',\mathcal{F}'}}{c \cdot k^{m+1}}.
\end{aligned}$$

Since $\mathsf{Adv}_{A'',\mathcal{F}'}$ is negligible (by the one-wayness of $\mathcal{F}'$), $\mathsf{Adv}^*_{A'',\mathcal{F}'}$ is negligible as well. This completes the proof that $\mathcal{F}'$ is one-way over its range.

6 Starting from Approximable-Preimage-Size One-Way Functions

Given an approximable-preimage-size one-way function family we first use a result by Håstad et al. [20] to transform it into a one-way function family that is "closer" to regular. From there we use the same construction of the previous section with a more careful analysis. The main result of this section is the following:

Theorem 5. *If there exists an approximable-preimage-size one-way function family then for any $0 < \delta < 1$ there exists a (δ, δ)-balanced function family which is one-way over its range.*

6.1 From Approximable to Dense

The following construction appeared in [20]:

Construction 10 *Let $\mathcal{F} = \{f_k : \{0,1\}^{n(k)} \to \{0,1\}^{\ell(k)}\}_{k\in\mathbb{N}}$ be an approximable-preimage-size one-way function family and let $\mathcal{H} = \{H_k\}$ be a 2-universal collection of hash families where each H_k is a family of functions mapping strings of length $n(k)$ to strings of length $n(k)$, and furthermore $|H_k| = 2^{s(k)}$ where $s(k) = \mathsf{poly}(k)$. Define:*

$$\hat{\mathcal{F}} = \left\{\hat{f}_k : H_k \times \{0,1\}^{n(k)} \to H_k \times \{0,1\}^{l(k)+n(k)}\right\}_{k\in\mathbb{N}}$$

such that $\hat{f}_k(h,x) = (f_k(x), h(x)_{1\ldots(\tilde{D}_{\mathcal{F}}(f_k(x),k)+2)}, 0^{n-(\tilde{D}_{\mathcal{F}}(f_k(x),k)+2)}, h)$, where $h(x)_{1\ldots m}$ stands for the first m bits of $h(x)$.

The following lemma, proven in [20, Lemma 5.2], shows that $\hat{\mathcal{F}}$ is a family of $(s(k)+n(k)-1, s(k)+n(k))$-entropy one-way functions:

Lemma 4. *$\hat{\mathcal{F}}$ as defined in Construction 10 is one-way, and for all $k \in \mathbb{N}$, $H_2(\hat{f}_k(U_{s(k)}, U_{n(k)})) > s(k) + n(k) - 1$.*

6.2 Starting from a Dense One-Way Function

Given an approximable-preimage-size one-way function family, we can transform it using Lemma 4 into a one-way function family $\mathcal{F}$ that has $(n(k)-1, n(k))$-entropy. Intuitively, such a function is "close" to being 1-regular. The following lemma shows how to use this property to construct a balanced function family which is one-way over its range.

Lemma 5. *Let $\mathcal{F} = \{f_k : \{0,1\}^{n(k)} \to \{0,1\}^{\ell(k)}\}_{k\in\mathbb{N}}$ be an $(n(k)-1, n(k))$-entropy family of one-way functions, let $c > 24\ln 2$ be an arbitrary constant, let $\delta = (24\ln 2/c)^{1/2}$ and let $\mathcal{F}'$ be the result of applying Construction 8 with $\mathcal{F}$, $t(k) = n(k) - 1 - \log(n(k))$ and c. Then $\mathcal{F}'$ is $(2^{-k} + 12/\delta n(k), \delta)$-balanced as well as one-way over its range.*

Theorem 5 follows immediately.

Proof. (of Lemma 5) Note that since $n(k)$ is polynomial in k, there exists a constant $m \geq 0$ such that $t(k) \geq n(k) - m\log(k)$. Hence by applying Lemma 3 we have that $\mathcal{F}'$ is one-way on range. It is left to prove that $\mathcal{F}'$ is $(2^{-k}+12/\delta n(k), \delta)$-balanced. We use the following standard fact. The proof appears in the full version.

Claim. Let D be a distribution over some finite domain X such that $H_2(D) \geq k$ then for every $\varepsilon > 0$ there exists a distribution D' over X such that $H_\infty(D') \geq k - log(\frac{1}{\varepsilon})$ and $SD(D, D') \leq \varepsilon$.

Since the Renyi-entropy of $f_k(U_{n(k)})$ is at least $(n(k) - 1)$, we have that $f_k(U_{n(k)})$ is $1/n(k)$-close to having min-entropy $(n(k) - log(n(k)) - 1)$. We now apply Lemma 2 and deduce that the output distribution of f'_k, that is $(h, h(f_k(U_{n(k)})))$, is $1/n(k)$-close to a distribution that is $(2^{-k}, \delta/2)$-balanced. The proof concludes by the following claim.

Claim. Let P' be a distribution over $\{0,1\}^\ell$ that is ε-close to some distribution P that is (α, δ)-balanced. Then, P' is $((\alpha + 6\varepsilon/\delta), 2\delta)$-balanced.

Proof. Let Bad be the set of bad elements for P. Let A be the set of elements $y \notin \mathsf{Bad}$ such that $|\Pr_{P'}[y] - 1/2^\ell| > 2\delta/2^\ell$. Note that the set of bad elements Bad' of P' is a subset of $(\mathsf{Bad} \cup A)$ and therefore it is enough to bound the size and probability of this set. Note that since $A \bigcap \mathsf{Bad} = \emptyset$ we have that $\forall y \in A$ $|\Pr_P[y] - 1/2^\ell| \leq \delta/2^\ell$ and thus $|\Pr_{P'}[y] - \Pr_P[y]| > \delta/2^\ell$. Thus $SD(P', P) \geq \frac{1}{2}|A| \cdot \delta/2^\ell$. As the two distributions are ε-close, it follows that $\frac{1}{2}|A| \cdot \delta/2^\ell \leq \varepsilon$ or equivalently that $|A| \leq \frac{2\varepsilon \cdot 2^\ell}{\delta}$. Therefore we have that

$$\Pr_P[\mathsf{Bad}'] \leq \Pr_P[\mathsf{Bad} \bigcup A] \leq \Pr_P[\mathsf{Bad}] + \Pr_P[A].$$

Since for all $y \in A$ we have $\Pr_P[y] \leq (1+\delta)/2^\ell$, it follows that

$$\Pr_P[\mathsf{Bad}'] \leq \alpha + |A|(1+\delta)/2^\ell \leq \alpha + (1+\delta)\frac{2\varepsilon}{\delta} = \alpha + 2\varepsilon + \frac{2\varepsilon}{\delta}.$$

Hence:

$$\Pr_{P'}[\mathsf{Bad}'] \leq \alpha + 2\varepsilon + \frac{2\varepsilon}{\delta} + 2\varepsilon = \alpha + 4\varepsilon + \frac{2\varepsilon}{\delta} \leq \alpha + \frac{6\varepsilon}{\delta}.$$

To complete the proof we have to show that $|\mathsf{Bad}'| \leq (\alpha + \frac{6\varepsilon}{\delta})2^\ell$. But $|\mathsf{Bad}'| \leq |\mathsf{Bad}| + |A| \leq \alpha 2^\ell + \frac{2\varepsilon \cdot 2^\ell}{\delta} = (\alpha + \frac{2\varepsilon}{\delta})2^\ell$.

Acknowledgments

We are grateful to Virgil Gligor, Oded Goldreich, Danny Harnik, Omer Reingold, and Alon Rosen for helpful conversations. The third author thanks Yan Zong Ding for reading a preliminary version of this manuscript and for his encouragement. We thank the anonymous referees for comments that improved the presentation.

References

1. M. Bellare and S. Micali. How to sign given any trapdoor permutation. *J. ACM*, 39(1):214–233, 1992.
2. M. Blum. Coin flipping by phone. In *IEEE COMPCOM*, 1982.
3. M. Blum and S. Micali. How to generate cryptographically-strong sequences of pseudorandom bits. *SIAM J. Computing*, 13(4):850–864, 1984.
4. M. Blum, A. De Santis, S. Micali, and G. Persiano. Non-interactive zero-knowledge. *SIAM J. Computing*, 20(6):1084–1118, 1991.
5. J.F. Boyar, S.A. Kurtz, and M.W. Krentel. Discrete logarithm implementation of perfect zero-knowledge blobs. *Journal of Cryptology*, 2(2):63–76, 1990.
6. G. Brassard, D. Chaum, and C. Crépeau. Minimum disclosure proofs of knowledge. *J. Computer and System Sciences*, 37(2):156–189, 1988.
7. J.L. Carter and M.N. Wegman. Universal classes of hash functions. *J. Computer and System Sciences*, 18(2):143–154, 1979.
8. I. Damgård, T. Pedersen, and B. Pfitzmann. On the existence of statistically-hiding bit commitment and fail-stop signatures. In *Crypto*, 1993.
9. U. Feige, D. Lapidot, and A. Shamir. Multiple non-interactive zero-knowledge proofs under general assumptions. *SIAM J. Computing*, 29(1):1–28, 1999.
10. O. Goldreich. *Foundations of Cryptography, vol. 1: Basic Tools*. Cambridge University Press, 2001.
11. O. Goldreich. *Foundations of Cryptography, vol. 2: Basic Applications*. Cambridge University Press, 2004.
12. O. Goldreich, S. Goldwasser, and S. Micali. On the cryptographic applications of random functions. In *Crypto '84*.
13. O. Goldreich, S. Goldwasser, and S. Micali. How to construct random functions. *J. ACM*, 33(4):792–807, 1986.
14. O. Goldreich, R. Impagliazzo, L. Levin, R. Venkatesan, and D. Zuckerman. Security preserving amplification of hardness. In *FOCS*, 1990.
15. O. Goldreich and A. Kahan. How to construct constant-round zero-knowledge proof systems for NP. *Journal of Cryptology*, 9(3):167–190, 1996.
16. O. Goldreich, H. Krawczyk, and M. Luby. On the existence of pseudorandom generators. *SIAM J. Computing*, 22(6):1163–1175, 1993.
17. O. Goldreich and L.A. Levin. Hard-core predicates for any one-way function. In *STOC*, 1989.
18. S. Goldwasser, S. Micali, and R.L. Rivest. A digital signature scheme secure against adaptive chosen-message attacks. *SIAM J. on Computing*, 17(2):281–308, 1988.
19. S. Halevi and S. Micali. Practical and provably-secure commitment schemes from collision-free hashing. In *Crypto*, 1996.
20. J. Håstad, R. Impagliazzo, L.A. Levin, and M. Luby. A pseudorandom generator from any one-way function. *SIAM J. Comput.*, 28(4):1364–1396, 1999.

21. R. Impagliazzo and M. Luby. One-way functions are essential for complexity-based cryptography. In *FOCS*, 1989.
22. R. Impagliazzo and S. Rudich. Limits on the provable consequences of one-way permutations. In *STOC*, 1989.
23. Y. Lindell. Parallel coin-tossing and constant-round secure two-party computation. *Journal of Cryptology*, 16(3):143–184, 2003.
24. M. Naor. Bit commitment using pseudorandomness. *Journal of Cryptology*, 4(2):151–158, 1991.
25. M. Naor, R. Ostrovsky, R. Venkatesan, and M. Yung. Perfect zero-knowledge arguments for NP using any one-way permutation. *J. Crypto.*, 11(2):87–108, 1998.
26. M. Naor and M. Yung. Universal one-way hash functions and their cryptographic application. In *STOC*, 1989.
27. R. Ostrovsky, R. Venkatesan, and M. Yung. Secure commitment against a powerful adversary. In *STACS*, 1992.
28. R. Ostrovsky, R. Venkatesan, and M. Yung. Fair games against an all-powerful adversary. In *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, volume 13, 1993.
29. J. Rompel. One-way functions are necessary and sufficient for secure signatures. In *STOC*, 1990.
30. A. Russel. Necessary and sufficient conditions for collision-free hashing. *J. Cryptology*, 8(2):87–100, 1995.
31. A. De Santis and M. Yung. On the design of provably-secure cryptographic hash functions. In *Eurocrypt*, 1990.
32. J.P. Schmidt, A. Siegel, and A. Srinivasan. Chernoff-Hoeffding bounds for applications with limited independence. *SIAM J. Discrete Math.*, 8(2):223–250, 1995.