Locally Computable UOWHF with Linear Shrinkage

Benny Applebaum* and Yoni Moses

School of Electrical Engineering, Tel-Aviv University
bennyap@post.tau.ac.il, ymoses@gmail.com

Abstract. We study the problem of constructing locally computable Universal One-Way Hash Functions (UOWHFs) $\mathcal{H} : \{0,1\}^n \to \{0,1\}^m$. A construction with constant *output locality*, where every bit of the output depends only on a constant number of bits of the input, was established by [Applebaum, Ishai, and Kushilevitz, SICOMP 2006]. However, this construction suffers from two limitations: (1) It can only achieve a sub-linear shrinkage of $n - m = n^{1-\epsilon}$; and (2) It has a super-constant *input locality*, i.e., some inputs influence a large super-constant number of outputs. This leaves open the question of realizing UOWHFs with constant output locality and linear shrinkage of $n - m = \epsilon n$, or UOWHFs with constant input locality and minimal shrinkage of $n - m = 1$.

We settle both questions simultaneously by providing the first construction of UOWHFs with linear shrinkage, constant input locality, and constant output locality. Our construction is based on the one-wayness of "random" local functions – a variant of an assumption made by Goldreich (ECCC 2000). Using a transformation of [Ishai, Kushilevitz, Ostrovsky and Sahai, STOC 2008], our UOWHFs give rise to a digital signature scheme with a minimal *additive* complexity overhead: signing n-bit messages with security parameter κ takes only $O(n + \kappa)$ time instead of $O(n\kappa)$ as in typical constructions. Previously, such signatures were only known to exist under an *exponential* hardness assumption. As an additional contribution, we obtain new locally-computable hardness amplification procedures for UOWHFs that preserve linear shrinkage.

1 Introduction

The question of minimizing the parallel time complexity of cryptographic primitives has been the subject of an extensive body of research. At the extreme, one would aim for an ultimate level of efficiency at the form of *constant*-parallel time implementation. Namely, the goal is to have "local" cryptographic constructions in which each bit of the output depends only on a small constant number of input bits, and each bit of the input influences only a constant number of outputs. Achieving both constant *input locality* and constant *output locality* allows an implementation by constant-depth circuit of bounded fan-in and bounded

* Supported by Alon Fellowship, ISF grant 1155/11, Israel Ministry of Science and Technology (grant 3-9094), and GIF grant 1152/2011.

T. Johansson and P. Nguyen (Eds.): EUROCRYPT 2013, LNCS 7881, pp. 486–502, 2013.

fan-out [7]. Furthermore, such local constructions have turned to be surprisingly helpful in speeding-up the *sequential complexity* of cryptography [17]. At a more abstract level, the study of locally computable cryptography allows to understand whether extremely simple functions can generate cryptographic hardness.

Intuitively, one may suspect that functions with local input-output dependencies may be vulnerable to algorithmic attacks. Still, during the last decade it was shown that, under standard intractability assumptions, many cryptographic tasks can be implemented by local functions [6,5,7]. This includes basic primitives such as one-way functions and pseudorandom generators, as well as, more complicated primitives such as public-key encryption schemes. One notable exception, for which such a result is unknown, is hash functions with *linear* shrinkage.

A collection of hash functions $\mathcal{H} = \{h : \{0,1\}^n \to \{0,1\}^m\}$ shrinks a long n-bit string into a shorter string of length $m < n$ such that, given a random function $h \stackrel{R}{\leftarrow} \mathcal{H}$ and a target string x, it is hard to find a sibling $y \neq x$ that collide with x under h. The exact specification of the above game corresponds to different notions of hashing. We will mainly consider *universal one-way hash functions* (UOWHFs) [21], in which the adversary specifies the target string x without seeing the function h. (This property is also known as *target collision resistance* [8], TCR in short.) A central parameter of a hash function is the amount of shrinkage it provides. We measure this as the difference between the output length m and the input length n, namely the *additive shrinkage* $n - m$. We say that the shrinkage is linear if $n - m = \Omega(n)$, i.e., $m < (1 - \varepsilon)n$ for some constant ε. In this paper we ask:

> Are there UOWHFs with *linear shrinkage* and *constant* output and/or input locality ?

Previous Results. The results of [6] show that any log-space computable UOWHF can be converted into a UOWHF with constant output locality and sub-linear shrinkage of $n - m = n^{\varepsilon}$, for a constant $\varepsilon < 1$. (A similar result holds for collision-resistance hash functions.) This gives rise to UOWHFs with constant output locality based on standard cryptographic assumptions (e.g., factoring), or, more generally, on any log-space computable one-way function [21,24,15]. Although there are several ways to amplify the shrinkage of a UOWHF (cf. [21,8]), none of these transformations preserve low locality, and so the question of obtaining UOWHFs with linear shrinkage and constant output locality has remained wide open.

The situation is even worse for constant input locality. In [7] it was shown that tasks which involve secrecy (e.g., one-wayness, pseudorandomness, symmetric or public-key encryption) can be implemented with constant input locality (under plausible assumptions), while tasks which require some form of non-malleability (e.g., MACs, signatures, non-malleable encryption) cannot be implemented with constant input locality. Interestingly, hash functions escaped this characterization. Although it is easy to find near-collisions in a function with constant input locality (simply flip the first bit of the target x), it is unknown how to extend

this to a full collision. Overall, the question of computing UOWHFs with constant input locality has remained open, even for the case of a single-bit shrinkage $n - m = 1$.[1]

1.1 Main Result

We construct the first locally computable UOWHF with linear shrinkage. Our construction has both constant input locality and constant output locality, and is based on the one-wayness of random local functions (also known as Goldreich's one-way function [14]). The latter assumption asserts that a random local function $f : \{0,1\}^n \to \{0,1\}^m$ is one-way where f is chosen uniformly at random as follows. View the n inputs and m outputs as vertices in a bipartite graph G and connect each output node y_i to a random set of d distinct input nodes. To compute the i-th output apply some fixed d-local predicate $P : \{0,1\}^d \to \{0,1\}$ to the d inputs that are connected to y_i. This experiment defines a distribution $\mathcal{F}_{P,n,m}$ over functions with output locality of d. (See Section 3 for a formal definition.) We prove the following theorem.

Theorem 1. *There exists a constant d and a predicate $P : \{0,1\}^d \to \{0,1\}$ for which the following holds. If the collection $\mathcal{F}_{P,n,m=O(n^3)}$ is one-way then there exists a collection $\mathcal{H}$ of UOWHF with linear shrinkage, constant input locality, and constant output locality.*

The theorem is constructive, and can be applied to every predicate which satisfies a simple criteria. In particular, we show that the predicate $\mathsf{MST}_{d_1,d_2}(x,y) = (x_1 \wedge \ldots \wedge x_{d_1}) \oplus (y_1 \oplus \ldots \oplus y_{d_2})$ defined by [20] satisfies the theorem for every $d_1 \geq 2$ and every sufficiently large constant d_2. The hypothesis of the theorem (one-wayness of random local functions) was extensively studied in the last few years and it is supported both experimentally [22,13] and theoretically [14,2,13,10]. In fact, recent evidence suggest that, for a proper predicate, this collection may even be pseudorandom [4,3]. Interestingly, Theorem 1 can be proved under the (possibly weaker) assumption that $\mathcal{F}_{P,n,m=O(n)}$ is a weak pseudorandom generator (i.e., its output cannot be distinguished from truly random string with advantage better than, say, 0.1).

There are several interesting corollaries that follow from Theorem 1. First, it is possible to reduce the output locality to 4 (which is almost optimal) while preserving (tiny) linear shrinkage (i.e., $m = (1-\varepsilon)n$ for some small ε) via the compiler of [6].[2] Second, by self-composing $\mathcal{H}$ a constant number times, one can get arbitrary linear shrinkage (i.e., $m = \varepsilon n$ for arbitrary constant $\varepsilon > 0$) at the expense of increasing the locality to a larger constant. Furthermore, by iterating $\mathcal{H}$ a logarithmic number of times we get a linear-time computable hash function $\mathcal{H}'$ with polynomial shrinkage factor of $m = n^{\varepsilon}$ (the i-th level of the

[1] We note that standard transformations from one-way functions to UOWHFs [21,24,15] are inherently non-local as they employ primitives such as k-wise independent hash functions which cannot be computed locally.

[2] When applied to *local functions*, the AIK compiler preserves linear shrinkage.

circuit contains $O(n/2^i)$ gates). As observed by [17], one can then employ the Naor-Yung transform [21] and sign n-bit messages with linear time complexity and only *additive cryptographic overhead*, i.e., $O(n + \kappa)$. (See [17] for details.) This is contrasted with standard signature schemes whose complexity grows multiplicatively with the security parameter, i.e., $O(n\kappa)$. Previously, such linear-time computable UOWHFs and signatures were only known to exist assuming that Goldreich's collection is *exponentially*-hard to invert [17].[3]

1.2 Techniques

Hashing via Random Local Functions? As a starting point, we ask whether the collection $\mathcal{F}_{P,n,m=n(1-\varepsilon)}$ itself can be used, even heuristically, as a UOWHF. To make the question non-trivial, let us assume that the distribution of the input-output dependency graph is slightly modified such that the graph is (c, d)-regular, i.e., each input affects c outputs and each output depends on d inputs. (Otherwise, we are likely to have some inputs of degree 0, with no influence at all.) For concreteness let us think of P as the majority predicate. A moment of reflection suggests that collisions are easy to find even with respect to a random target string x. Indeed, suppose that there exists an input variable x_i that all of its neighboring inputs (i.e., the inputs that share an output with x_i) turn to be zero. In this case, we can flip the *insensitive* input x_i without affecting the output of the function, and this way obtain a trivial collision. Observe that each input variable has a constant probability of being insensitive as it has at most $cd = O(1)$ neighbors. Overall, one is likely to find $\Omega(n)$ insensitive inputs. Furthermore, by collecting an independent set I of insensitive inputs (that do not share any common output) one can simultaneously flip any subset of the inputs in I without changing the output. Hence, we find exponentially many collisions x' which form a "ball" around x of diameter $\Omega(n)$. It is not hard to show that a similar attack can be applied to $\mathcal{F}_{P,n,m}$ for every predicate P except for XOR or its negation. (Unfortunately, in the latter case collisions can be found via Gaussian elimination.)

Despite this failure, let us keep asking: Can $\mathcal{F}_{P,n,m}$ achieve some, possibly weak, form of collision resistance ? Specifically, one may hope to show that it is hard to find collisions which are β-far from the target x, for some (non-trivial) constant β. This assumption is intuitively supported by study of the *geometry of the solutions* of random Constraint Satisfaction Problems (e.g., Random SAT) [1]. Thinking of each output as inducing a local constraint on the inputs, it can be essentially showed that, for under-constraint problems where $m < n$, the space of solutions (siblings of x) is shattered into far-apart clusters of Hamming-close solutions. It is believed that efficient algorithms cannot move from one cluster to another as such a transition requires to pass through solutions x' which violate many constraints (i.e., $f(x')$ is far, in Hamming distance, from $f(x)$). Therefore, it seems plausible to conjecture that the collection $\mathcal{F}_{P,n,m}$ is secure with respect to β-far collisions.

[3] Exponential hardness assumptions do not seem to help in the context of *locally computable* UOWHFs.

As our main technical contribution, we prove that a weak form of this conjecture holds assuming the one-wayness of $\mathcal{F}_{P,n,m'}$ (where $m' > n > m$). Specifically, we prove that, for some constants $\varepsilon, \beta, \delta \in (0,1)$, it is hard to find β-far target collisions in $\mathcal{F}_{P,n,(1-\varepsilon)n}$ with probability better than δ. To prove Theorem 1, we show that (δ, β)-target collision resistance (TCR) can be *locally* amplified into standard TCR while preserving *linear* shrinkage. Let us sketch the main ideas behind each of these steps.

One-wayness $\Rightarrow$ (δ, β)*-TCR.* Assume that we have an algorithm $\mathcal{A}$ that, given a random function $h \stackrel{R}{\leftarrow} \mathcal{F}_{P,n,m=(1-\varepsilon)n}$ and a random target w, finds a β-far sibling with probability δ. We show how to use $\mathcal{A}$ to invert the collection $\mathcal{F}_{P,n,m'}$ with output length of $m' \approx 2m$. Given a random function $f_G \stackrel{R}{\leftarrow} \mathcal{F}_{P,n,m'}$ specified by a random input-output dependencies graph G, and an image $y = f_G(x)$ of a random point $x \stackrel{R}{\leftarrow} \{0,1\}^n$, we will recover the preimage x as follows. First, we choose a target w uniformly at random and partition the graph G into two subgraphs: G_0 which contains only the output nodes for which $f_G(w)$ agrees with y (and all input nodes), and G_1 which contains the remaining subgraph. Assuming that P is balanced, each subgraph contains roughly m' outputs. Next, we define $h = f_{G_0}$ to be the restriction of f_G to the output nodes for which $f_G(w)$ agrees with y, and ask $\mathcal{A}$ for a β-far sibling w' of w under h. Let us (optimistically) assume that w' is statistically independent of the sub-graph G_1 that was not used by h. That is, imagine that this part of the dependencies graph is chosen uniformly at random after w' is obtained. Since w is far from w', this pair is expected to disagree on a constant fraction γ of the remaining coordinate of f_{G_1}. Remembering that the pair (w, x) did not agree on any of these coordinates, we conclude that x and w agree on a fraction of $\frac{1}{2} + \gamma/2$ of the outputs of f_G (i.e., γ-fraction of the coordinates of f_{G_1} and all the coordinates of $h = f_{G_0}$). Assuming that P is *sensitive* enough, it follows that w' and x must be *correlated* – their Hamming distance is bounded by a constant which is strictly smaller than $\frac{1}{2}$. At this point we employ a result of [9] that allows to fully recover x given such a correlated string w' (and additional $O(n)$ outputs).

The above argument is over-optimistic, as there is no reason to assume that w' is statistically independent of the subgraph G_1. Fortunately, we can show that a failure of the above approach allows to distinguish the string $y = f(x)$ from a truly random string. At this point, we employ the result of [3] which shows that this string is somewhat pseudorandom assuming the one-wayness of $\mathcal{F}_{P,n,m''}$ for larger m''. Hence, we are in a win-win situation: we invert $\mathcal{F}$ either by finding a correlated string, or by distinguishing its output from a random string. (See Section 4 for details.)

(δ, β)*-TCR* $\Rightarrow$ δ*-TCR.* The above reduction leaves us with a δ-secure β-TCR $\mathcal{H}$ of linear shrinkage $n - m = \varepsilon n$, where $\delta, \beta, \varepsilon$ are constants. Our first goal is to get rid of β (i.e., obtain security with respect to standard, possibly close, collisions). A tempting approach would be to compose $\mathcal{H}$ with an error correcting code C, i.e., map an input x to a codeword $C(x)$ and hash the result via $h \in \mathcal{H}$.

A code of constant relative distance larger than β and constant rate smaller than ε will fully eliminate β-close collisions (in an information theoretic sense), while preserving linear shrinkage. Unfortunately, this transformation is inherently non-local, as local functions cannot compute codes with constant relative distance and constant rate.[4] We solve the problem via a dual approach: Instead of computing a codeword $C(x)$ and composing the result with h, we concatenate $h(x)$ with the syndrome Mx where M is a sparse parity-check matrix M whose dual relative distance is β. It is not hard to show that a pair of β-close strings x and x' will always be mapped by M to different outputs $y \neq y'$, and so the mapping $x \mapsto (h(x), Mx)$ is immunized against β-close collisions. Unlike the case of sparse generating matrices, whose distance is deemed to be non-constant, the dual distance of sparse parity-check matrices can be constant (aka LDPC) and so the transformation is locally computable. (See Section 5.2.)

δ-hard TCR $\Rightarrow$ TCR. We move on to amplify the error parameter δ from constant to negligible. Typically this is done via t-wise direct-product, i.e., $x \mapsto (h_1(x), \ldots, h_t(x))$ where the h_i's are chosen independently from $\mathcal{H}$. The error δ decreases exponentially fast and so any super-logarithmic t leads to a negligible error [11]. Unfortunately, in our case even a super-constant t will completely ruin the shrinkage and the input locality. An alternative, more economic, approach is to first stretch the input x into a longer string $C(x) = (c_1, \ldots, c_t) \in (\{0,1\}^n)^t$ via an error-correcting code C, and then apply t-wise direct product [18,11]. If the code has a constant relative distance, any collision (x', x) is translated into a pair $C(x), C(x')$ which collide under $\Omega(t)$ of the coordinates of $(h_1, \ldots, h_t) \stackrel{R}{\leftarrow} \mathcal{H}^t$. Hence, the error parameter decreases exponentially with t while keeping the shrinkage linear (for properly chosen parameters). Unfortunately, this optimization is inherently non-local as it requires a code with good distance. Nevertheless, we observe that even if C is replaced with a sparse generating matrix G, the resulting transformation is not completely useless. Although the distance of G is bad, it can be shown any pair of β-far inputs x, x' will be mapped by G to a pair (y, y') which is $\Omega(t)$ far apart. As a result, the modified construction amplifies hardness with respect to β-far collisions, but does not amplify hardness with respect to close collisions. Fortunately, such collisions can be again eliminated via LDPCs.[5] (See Section 5.3.)

We note that the above transformations can also be used to locally amplify *collision resistance*.

[4] In fact, such codes are as bad as possible as their relative distance is $O(1/n)$.

[5] One can change the order of the transformations, namely, transform (δ, β)-TCR to β-TCR and then to TCR. This allows to use LDPCs only once. Still we prefer the current order as once β is eliminated (in the first step), it is easy to amplify the shrinkage factor to a small constant via a constant number of self-compositions. Overall, this results in a more flexible reduction that works for a wider range of parameters.

2 Preliminaries

General. By default, logarithms are taken to base 2. For reals $p, q \in (0,1)$ we let $D_2(p\|q) := p\log(\frac{p}{q}) + (1-p)\log(\frac{1-p}{1-q})$ denote the relative entropy function. Observe that $1 - D_2(p\|\frac{1}{2})$ equals to the binary entropy function $H_2(p) := -p\log(p) - (1-p)\log(1-p)$. We will use the following additive form of Chernoff-Hoeffding bound. Let $X_1, \ldots, X_n$ be i.i.d. random variables where $X_i \in [0,1]$ and $\mathsf{E}[X_i] = p$. Then, for every $\varepsilon > 0$, the average $\bar{X} = n^{-1}\sum_i X_i$ satisfies

$$\Pr\left[\bar{X} \geq p + \varepsilon\right] < 2^{-D_2(p+\varepsilon\|p)n} \quad \text{and} \quad \Pr\left[\bar{X} \leq p - \varepsilon\right] < 2^{-D_2(p-\varepsilon\|p)n}.$$

A simpler form follows by noting that $D_2(p+\varepsilon\|p) > 2\varepsilon^2$.

Collection of Functions. We model cryptographic primitives as collections of functions $\mathcal{F} = \left\{f_k : \{0,1\}^n \to \{0,1\}^{m(n)}\right\}_{k\in\{0,1\}^{s(n)}}$ equipped with a pair of efficient algorithms: (1) an evaluation algorithm which given $(k \in \{0,1\}^s, x \in \{0,1\}^n)$ outputs $f_k(x)$; and (2) a key-sampling algorithm $\mathcal{K}$ which given 1^n samples a index $k \in \{0,1\}^{s(n)}$. We will sometimes keep the key-sampler implicit and write $f \xleftarrow{R} \mathcal{F}$ to denote the experiment where $k \xleftarrow{R} \mathcal{K}(1^n)$ and $f = f_k$. A collection of functions has constant *output locality* (resp., constant *input locality*) if there exists a constant d (which does not grow with n) such that for every fixed k each output of the function f_k depends on at most d inputs (resp., each input of f_k affects at most d outputs). The collection is locally computable if it has both constant input locality and constant output locality.

One-wayness and Pseudorandomness. A collection of functions $\mathcal{F}$ is δ-secure β *approximation-resilient one-way* (in short, (δ,β) one-way) if for every efficient adversary $\mathcal{A}$ the following event happens with probability at most δ: Given $f \xleftarrow{R} \mathcal{F}$ and $y = f(x)$ for random $x \xleftarrow{R} \{0,1\}^n$, the adversary $\mathcal{A}$ outputs a list of candidates X' which contains some string x' which is β-close to some preimage of y. The special case of $\beta = 0$ corresponds to the standard notion of δ-secure one-wayness, or simply one-wayness when $\delta = \mathrm{neg}(n)$. A collection of functions $\mathcal{F}$ is δ-pseudorandom if the distribution ensemble $(f, f(x))$ is δ computationally-indistinguishable from the distribution ensemble (f, y), where $f \xleftarrow{R} \mathcal{F}, x \xleftarrow{R} \{0,1\}^n$ and $y \xleftarrow{R} \{0,1\}^m$.

Hash Functions. Let $m = m(n) < n$ be an integer-valued function. A collection of functions $\mathcal{H} = \{h : \{0,1\}^n \to \{0,1\}^m\}$ is δ-secure β *target-collision resistance* ((δ,β)-TCR) if for every pair of efficient adversaries $\mathcal{A} = (\mathcal{A}_1, \mathcal{A}_2)$ it holds that

$$\Pr_{\substack{(x,r)\xleftarrow{R}\mathcal{A}_1(1^n)\\ h\xleftarrow{R}\mathcal{H}}}[\mathcal{A}_2(h,x,r) = x' \text{ s.t. } \Delta(x',x) > \beta \text{ and } h(x) = h(x')] \leq \delta,$$

where $\Delta(\cdot,\cdot)$ denotes relative Hamming distance. That is, first the adversary $\mathcal{A}_1$ specifies a target string x and a state information r, then a random hash

function h is selected, and then $\mathcal{A}_2$ tries to form a β-far collision x' with x under h. The collection is δ-secure β *random target-collision resistance* ((δ, β) RTCR) if the above holds in the special case where $\mathcal{A}_1$ outputs a uniformly chosen target string $x \xleftarrow{R} \{0,1\}^n$ and empty state information. (As we will see, there are standard local transformations from RTCR to TCR). The standard notions of δ-RTCR and δ-TCR correspond to the case where $\beta = 0$ (or just $\beta < 1/n$). If, in addition, δ is negligible we obtain standard RTCR and TCR. The *shrinking factor* of $\mathcal{H}$ is the ratio m/n. When $m/n < 1/(1+H_2(\beta))$ and $\delta = o(1)$ TCR and RTCR become non-trivial in the sense that their existence implies the existence of one-way functions. For an extensive study of hash functions see [8,23].

3 Random Local Functions and Sensitivity

Let $P : \{0,1\}^d \to \{0,1\}$ be a predicate, and let $G = (S_1, \ldots, S_m)$ where each $S_i \subseteq [n]$ is a subset of $[n]$ that contains d distinct ordered elements $S_{i,1}, \ldots, S_{i,d} \in [n]$. We will think of G as a bipartite graph with n input vertices and m output vertices where each output i is connected to the d inputs in S_i. We define the function $f_{G,P} : \{0,1\}^n \to \{0,1\}^m$ as follows: Given an n-bit input x, the i-th output bit y_i is computed by applying P to the restriction of x to the i-th set S_i, i.e., $y_i = P(x_{S_i})$. For $m = m(n)$ and some fixed predicate $P : \{0,1\}^d \to \{0,1\}$, we let $\mathcal{F}_{P,n,m}$ denote the collection $\left\{f_{G,P} : \{0,1\}^n \to \{0,1\}^{m(n)}\right\}$ where the key G is sampled by selecting $m(n)$ sets uniformly and independently at random from all the possible $n \cdot (n-1) \cdot \ldots \cdot (n-d+1)$ ordered sets. We refer to the latter distribution as the uniform distribution over (n, m, d) graphs and denote it by $\mathcal{G}_{n,m,d}$. When the predicate P is clear from the context, we omit it from the subscript and write f_G and $\mathcal{F}_{n,m}$.

By definition, the ensemble $\mathcal{F}_{P,n,m}$ has a constant output locality of d. However, some inputs will have large (super-constant) locality. Still, one can show, via simple probabilistic argument, that the locality of most inputs will be close to the expectation md/n which is constant when $m = O(n)$. We will later use this fact to reduce the input locality to constant.

3.1 Sensitivity

Let $P : \{0,1\}^d \to \{0,1\}$ be a d-local predicate. For a pair of strings $x, x' \in \{0,1\}^n$ let $s_P(x, x')$ be the expected relative Hamming distance between the images $f(x)$ and $f(x')$ where f is randomly chosen from $\mathcal{F}_{P,n,m}$. Equivalently, we may write $s_P(x, x')$ as

$$\Pr_S[P(x_S) \neq P(x'_S)], \tag{1}$$

where S is a random set of d distinct indices $i_1, \ldots, i_d$ which are chosen from $[n]$ uniformly at random without replacement. Imagine the following experiment: first x is chosen uniformly at random, and then an α-far string x' is chosen adversarially in order to minimize $s_P(x, x')$. We will be interested in predicates P for which, except with negligible probability, the value of $s_P(x, x')$ in the

above experiment will be relatively high (as a function of α). To analyze this property we make several simple observations. By symmetry, the strategy of the adversary boils down to selecting the fraction $\alpha_{0,1}$ of 0's which are flipped to 1, and the fraction $\alpha_{1,0}$ of 1's which are flipped to 0's (where $\alpha = \alpha_{0,1} + \alpha_{1,0}$). Furthermore, it suffices to analyze a simpler experiment in which x is a random string of Hamming weight $n/2$ and the set S (from Eq. 1) is chosen by selecting d indices uniformly at random from $[n]$ *with replacement* (i.e., the tuple may not be distinct). We will show (in Lemma 1) that, with all but negligible probability over x, these simplifications add a small $o(1)$ error to the value of the experiment.

The above observations motivate a new quantitative measure of sensitivity which refines the standard notion of *noise sensitivity*. For $\alpha_{0,1}, \alpha_{1,0} \in [0, \frac{1}{2}]$, let $\mathcal{D}(\alpha_{0,1}, \alpha_{1,0})$ be a distribution over pairs $w, w' \in \{0,1\}^d$ where w is chosen uniformly at random and the i-th bit of w' is obtained by flipping the i-th bit of w with probability $2\alpha_{0,1}$ if $w_i = 0$, and with probability $2\alpha_{1,0}$ if $w_i = 1$. (Hence, $\Pr[(w_i, w'_i) = (01)] = \alpha_{01}$, and $\Pr[(w_i, w'_i) = (00)] = \frac{1}{2} - \alpha_{01}$, etc.) For $\alpha \in [0,1]$ let $s_P(\alpha)$ denote the infimum of $\Pr_{(w,w') \xleftarrow{R} \mathcal{D}(\alpha_0, \alpha_1)}[P(w) \neq P(w')]$ taken over all $\alpha_{0,1}$ and $\alpha_{1,0}$ which sum-up to α. Call x *typical* if its Hamming weight is $n/2 \pm n^{2/3}$. By Chernoff bound, a random string is typical with all but negligible probability. The following lemma, whose proof is deferred to the full version, relates $s_P(x, x')$ to $s_P(\alpha)$.

Lemma 1. *For every predicate P, the function $s_P(\alpha)$ is well defined and continuous. Also, for every typical x and every string x' $s_P(x, x') \geq s_P(\Delta(x, x')) - o(1)$.*

Good Predicates. We say that P is (β, γ) *good* if: (1) The value of $s_P(\cdot)$ is lower-bounded by γ in the interval $[\beta, 1]$; and (2) P has a *sensitive coordinate* meaning that $P(w) = w_1 \oplus P'(w_2, \ldots, w_d)$ for some $(d-1)$-local predicate P'. Observe that the latter condition implies that P is balanced and that $s_P(\frac{1}{2}) = \frac{1}{2}$.

In the next section we will use (β, γ)-good predicate to construct β-RTCRs with shrinkage $1 - \varepsilon$ where $\varepsilon \in (0, \frac{1}{2})$ satisfies the inequality

$$\varepsilon < 1 - \frac{1}{2(1 - H_2(\frac{1}{2} - \gamma))}, \tag{2}$$

where H_2 is the binary entropy function. In general, we would like to have a small value of $\beta > 0$ and a large value of $\gamma \leq \frac{1}{2}$ (which leads to a larger ε and better shrinkage). It turns out that by increasing the locality, one can simultaneously push β arbitrarily close to 0 and γ arbitrarily close to $\frac{1}{2}$. This is illustrated by the following family of predicates which generalizes the predicate from [20]. Let MST_{d_1,d_2} be the $(d_1 + d_2)$ local predicate $(x_1 \wedge \ldots \wedge x_{d_1}) \oplus (y_1 \oplus \ldots \oplus y_{d_2})$. In the full version we will prove the following lemma.

Lemma 2. *For every constants $\gamma < \frac{1}{2}$, $\beta > 0$ and integer $d_1 \geq 2$ there exists a constant d_2 for which MST_{d_1,d_2} is (β, γ)-good.*

4 Random Local Functions Are (δ, β)-RTCR

Let P be (β, γ) good predicate. Assume that Eq 2 holds for some $\varepsilon > 0$ and let $m = (1-\varepsilon)n$. In Section 4.1 we prove the following.

Theorem 2. *For every $\delta_1, \delta_2 \in (0,1)$ there exists a constant $\mu > 0$ such that if $\mathcal{F}_{P,n,2m}$ is both δ_1-pseudorandom and $(\delta_2, \frac{1}{2}-\mu)$ one-way then $\mathcal{F}_{P,n,m}$ is δ'-secure β-RTCR where $\delta' = 2(\delta_1+\delta_2) + \mathrm{neg}(n)$.*

It was shown in [9, Thm. 1.3] and [3, Prop. 3.4] that if $\mathcal{F}_{n,m}$ is one-way for sufficiently long output length m, then it is also approximate one-way and pseudorandom for shorter output lengths. Together with Theorem 2, we get:

Corollary 1. *For every constant $\delta > 0$, there exists a constant c such that if $\mathcal{F}_{P,n,cn^3}$ is one-way then $\mathcal{F}_{P,n,(1-\varepsilon)n}$ is δ-secure β-RTCR. Furthermore, if $\mathcal{F}_{P,n,(1-\varepsilon)n}$ is δ-secure β-RTCR then for every constant $\eta > 0$ there exists a δ-secure $\frac{\beta}{1-\eta}$-RTCR $\mathcal{H}$ with constant input and constant output locality and shrinkage factor of $\frac{1-\varepsilon}{1-\eta}$.*

The "furthermore" part is obtained by randomly fixing a small fraction of the inputs of $\mathcal{F}_{P,n,m}$ (the ones with maximal influence). See full version for details.

4.1 Proof of Theorem 2

Assume, towards a contradiction, that $\mathcal{F}_{P,n,m}$ is not δ'-secure β-RTCR. Namely, there exists an efficient adversary $\mathcal{A}$ which, given a random target $w \xleftarrow{R} \{0,1\}^n$ and a random graph $G \xleftarrow{R} \mathcal{G}_{n,m,d}$, finds, with probability δ', a string z which is a β-far sibling of w under f_G. Assume that $\mathcal{F}_{n,2m}$ is δ_1-pseudorandom. We construct an attacker $\mathcal{B}$ who breaks the $(\delta_2, \frac{1}{2}-\mu)$ one-wayness of $\mathcal{F}_{n,2m}$ for some constant μ whose value will be determined later. Given a graph $G = (S_1, \ldots, S_{2m})$ and a string $y \in \{0,1\}^{2m}$, the algorithm $\mathcal{B}$ is defined as follows:

1. Randomly choose $w \xleftarrow{R} \{0,1\}^n$ and let $r = f_{G,P}(w) \oplus y$.
2. *Fail*, if the number of 0's in r is smaller than m or larger than $m + m^{2/3}$.
3. Let I_0 be the set of the first m indices i for which $r_i = 0$, and $I_1 = \{i : r_i = 1\}$. Let $G_0 = \{S_i : i \in I_0\}$ and $G_1 = \{S_i : i \in I_1\}$.
   (Note that $f_{G_0,P}(w) = y_{I_0}$ and that $f_{G_1,P}(w) = \mathbf{1} \oplus y_{I_1}$.)
4. Apply $\mathcal{A}$ to (G_0, w) and let $z \in \{0,1\}^n$ denote the resulting output.
5. If $P(z_{S_i}) = y_i$ for at least $m(1+\gamma) - 2m^{2/3}$ of indices $i \in [2m]$ output z; Otherwise, *Fail*.

We begin by bounding the failure probability of the algorithm. Intuitively, the algorithm does not fail due to the following reasoning. Assuming that z is a collision, we have that $P(z_{S_i}) = y_i$ for all the m indices $i \in I_0$. In addition, if z is β-far from w and statistically independent of G_1 then (since P is (β, γ) good), the outputs $f_{G_1,P}(w)$ and $f_{G_1,P}(z)$ are expected to disagree on a set of γm coordinates. Since $f_{G_1,P}(w) = \mathbf{1} \oplus y_{I_1}$, this translates to γm indices in I_1 for

which $P(z_{S_i}) = y_i$. The above analysis is inaccurate as the random variables z and G_1 are statistically dependent (via the random variable (w, G_0)). Still the above approach can be used when the input y (as well as the graph G) is truly random.

Claim 3. $\Pr_{G \stackrel{R}{\leftarrow} \mathcal{G}_{n,2m,d}, y \stackrel{R}{\leftarrow} \{0,1\}^{2m}}[\mathcal{B}(G,y) \textit{ does not fail}] > \delta'/2 - \mathrm{neg}(n)$.

Proof. When the pair (G, y) is uniformly chosen, the process $\mathcal{B}(G, y)$ can be equivalently described as follows. In the first step, we choose $S_1, \ldots, S_{2m}$ uniformly at random, choose a random string $w \stackrel{R}{\leftarrow} \{0,1\}^n$, and a random string $r \stackrel{R}{\leftarrow} \{0,1\}^{2m}$. We let $y = f_{G,P}(w) \oplus r$. Then steps 2–5 are performed exactly as before. This process is clearly equivalent to $\mathcal{B}(G, y)$, but easier to analyze. The main observation is that the string w is *statistically independent* of the graphs G_0 and G_1 which are just random graphs (whose size is determined by the random variable r).

Specifically, consider the following event: (1) The number of zeroes in r is larger than $m/2$; (2) The number of zeroes in r is smaller than $m/2 + m^{2/3}$; (3) $\mathcal{A}$ outputs β-far collision z with w under $f_{G_0,P}$; (4) The Hamming weight of w is $n/2 \pm n^{2/3}$; (5) $P(z_{S_i}) = y_i$ for at least $m(1+\gamma) - 2m^{2/3}$ of indices $i \in [2m]$.

Event (1) happens with probability $\frac{1}{2}$ (this follows from the "mean in the median" result for the binomial distribution, cf. [19]), and Event (2) happens with all but negligible probability due to a Chernoff bound. Hence, by a union bound (1) and (2) happen together with probability $\frac{1}{2} - \mathrm{neg}(n)$. Fix some r which satisfies both (1) and (2) and let $m_1 \geq m - m^{2/3}$ be the Hamming weight of r. Now, w is a random string and $G_0 \stackrel{R}{\leftarrow} \mathcal{G}_{n,m,d}$, hence, $\mathcal{A}$ is invoked on the "right" probability distribution and (3) happens with probability δ'. By a Chernoff bound, (4) happens with all but negligible probability. Therefore, by union bound, (3) and (4) happen simultaneously (conditioned on (1,2)) with probability $\delta' - \mathrm{neg}(n)$. Fix w and G_0 which satisfy (3) and (4), and let us move to (5).

Since w and z form a collision under $f_{G_0,P}$, we have that $f_{G_0,P}(z) = y_{I_0}$ and therefore $P(z_{S_i}) = y_i$ for all the m indices $i \in I_0$. Hence, it suffices to show that $P(z_{S_i}) = y_i$ for at least

$$(\gamma - m^{-1/3})m_1 \geq \gamma m - 2m^{2/3}$$

of the indices in I_1. (Recall that $m_1 > m - m^{2/3}$.) We claim that this happens with all but negligible probability (taken over the random choice of $G_1 \stackrel{R}{\leftarrow} \mathcal{G}_{n,m_1,d}$). To see this, define for every $i \in I_1$ a random variable ξ_i which equals to one if $P(z_{S_i}) = y_i$. Equivalently, $\xi_i = 1$ if $P(z_{S_i}) \neq P(w_{S_i})$. Furthermore, since the sets S_i are distributed uniformly and independently, each ξ_i takes the value 1 independently with probability at least

$$s_P(w, z) \geq s_P(\Delta(w, z)) - o(1) > \gamma$$

where the first inequality follows from Lemma 1 and the fact that w is "typical" (of Hamming weight $n/2 \pm n^{2/3}$); and the second inequality follows from the goodness of P and the fact that $\Delta(w, z) \geq \beta$. Therefore, by Chernoff's bound,

$$\Pr\left[\sum \xi_i < (\gamma - m^{-1/3})m_1\right] < 2^{-D_2(\gamma - m^{-1/3}\|\gamma)m_1} < e^{-\Omega(m^{1/3})},$$

which is negligible in n and so the claim follows. □

Moving back to the case where y is an image of a random string x, we show that when $\mathcal{B}$ does not fail its output is likely to be correlated with x.

Claim 4. *There exists a constant μ such that the following holds. With all but negligible probability over the choice of $x \xleftarrow{R} \{0,1\}^n$ and $G \xleftarrow{R} \mathcal{G}_{n,2m,d}$, there is no string z such that $f_{G,P}(x)$ and $f_{G,P}(z)$ agree on at least $m(1+\gamma) - 2m^{2/3}$ coordinates but $\Delta(x,z) \in (\frac{1}{2} \pm \mu)$.*

Proof. Let $\mu > 0$ be a small constant for which the value of $s_P(\cdot)$ in the interval $(\frac{1}{2} \pm \mu)$ is lower bounded by a constant η which satisfies $\eta > \frac{1}{2} - \gamma$ and

$$2(1-\varepsilon)D_2(\frac{1}{2} - \gamma\|\eta) > 1. \tag{3}$$

Observe that for $\mu = 0$ we can take $\eta = \frac{1}{2}$ (as $s_P(\frac{1}{2}) = \frac{1}{2}$) and so Eq 3 translates to $2(1-\varepsilon)H_2(\frac{1}{2} - \gamma) > 1$ which follows from Eq 2. Since s_P is a continuous function, and the LHS of Eq 3 is also continuous in η, we conclude that Eq 3 also holds for sufficiently small constant $\mu > 0$.

Let us condition on the event that x is typical (as in Lemma 1), which, by a Chernoff bound, happens with all but negligible probability. Fix some string z for which $\Delta(x,z) \in (\frac{1}{2} \pm \mu)$. For a random d size set S we have, by Lemma 1, that $\Pr[P(x_S) \neq P(z_S)] \geq s_P(\Delta(x,z)) > \eta - o(1) > \frac{1}{2} - \gamma$. Let $G = (S_1, \ldots, S_m) \xleftarrow{R} \mathcal{G}_{n,2m,d}$. Since each set S_i is chosen independently and uniformly at random, we can upper-bound (via Chernoff) the probability that $f_{G,P}(x)$ and $f_{G,P}(z)$ *disagree* on less than $2m - (m(1+\gamma) - 2m^{2/3}) = (1-\gamma)m + 2m^{2/3}$ of the coordinates by

$$p = 2^{-2mD_2(\frac{1}{2} - \gamma + o(1)\|s(x,z))} \leq 2^{-2(1-\varepsilon)D_2(\frac{1}{2} - \gamma + o(1)\|\eta - o(1))n}.$$

By a union bound over all z's, we get that the claim holds with probability $p \cdot 2^n$ which is negligible since Eq.3 holds. □

We can now complete the proof of the theorem. Let $G \xleftarrow{R} \mathcal{G}_{n,2m,d}$ and $y = f_{G,P}(x)$ where $x \xleftarrow{R} \{0,1\}^n$. Consider the event that: (1) G and x satisfy Claim 4; and (2) $\mathcal{B}(G,y)$ does not fail and outputs the string z. In this case, either the string z or its negation has a non-trivial agreement of $\frac{1}{2} + \mu$ with x, which may happen with probability at most δ_2 due to the approximate one-wayness of $\mathcal{F}_{n,2m}$. Hence, it suffices to show that the above event happens with probability at least $\delta'/2 - \delta_1 - \mathrm{neg}(n)$. Indeed, (1) happens with all but negligible probability (due to Claim 4), and (2) happens with probability $\delta'/2 - \delta_1 - \mathrm{neg}(n)$ due to Claim 3 and the fact that (G,y) is δ_1-indistinguishable from (G,y') for truly random $y' \xleftarrow{R} \{0,1\}^{2m}$. □

5 From (δ, β)-RTCR to TCR

In this section we will start with δ-secure β-RTCR with shrinkage factor of $1-\varepsilon$ and gradually amplify each of the parameters via locally computable transformations (described in Sections 5.1–5.3). Formally, we prove the following theorem.

Theorem 5. *For every $\varepsilon \in (0,1)$ there exist universal constants $\delta, \beta \in (0,1)$ such that for every desired constant shrinkage factor $\varepsilon' \in (0,1)$ the following holds. Any locally computable δ-secure β-RTCR with shrinkage factor of $1-\varepsilon$ can be transformed into a locally computable TCR with shrinkage factor of ε'.*

We note that the proof of the theorem can be adopted to the setting of collision resistance hash functions. Namely, it allows to locally transform a δ-secure β-collision resistance hash function with shrinkage factor $1-\varepsilon$ into a standard collision resistance hash function with arbitrary constant shrinkage.

Observe that our main theorem (Theorem 1) follows by combining Theorem 5 with Corollary 1 instantiated with (β,γ)-good predicate P where $\beta < \beta^*$ and $\gamma > \gamma^*$ for some universal constants $\beta^* > 0$ and $\gamma^* < \frac{1}{2}$. The exact values of β^* and γ^* are determined by the quality of LDPC codes. (See section 5.2.)

5.1 Standard Transformations

We begin with two standard transformations.

Claim 6 (RTCR to TCR). *Let $\mathcal{H} = \{h_k\}$ be δ-secure β-RTCR with shrinkage factor of $1-\varepsilon$. Then the collection $\mathcal{H}' = \left\{h'_{k,y}\right\}$ defined by $h'_{k,y}(x) = h_k(x \oplus y)$ is δ-secure β-TCR.*

Assume that we already have δ-secure standard-TCR ($\beta = 0$) with shrinkage factor of $1-\varepsilon$. A standard way to amplify the shrinkage factor from $1-\varepsilon$ to $(1-\varepsilon)^t$ is via iterated self-composition [21]. We note that when $t = O(1)$ the locality remains constant.

Claim 7 (Amplifying the Shrinkage Factor). *Let $\mathcal{H} = \{h_k\}$ be a δ-secure TCR with shrinkage factor of $1-\varepsilon$ and key sampler $\mathcal{K}$. For any constant integer $t \geq 1$, the collection $\mathcal{H}^t$ (defined below) is $t\delta$-secure TCR with shrinkage factor of $(1-\varepsilon)^t$. The collection $\mathcal{H}^t$ is defined recursively, via*

$$\mathcal{H}^t = \{h_{k_1,\ldots,k_t}\}, \quad h_{k_1,\ldots,k_t}(x) = h_{k_t}(h_{k_1,\ldots,k_{t-1}}(x)), \quad \text{where } k_i \stackrel{R}{\leftarrow} \mathcal{K}(1^{n(1-\varepsilon)^{i-1}}).$$

A proof for $t = 2$ follows from [8, Lemma 3.2]. The case of arbitrary constant t follows by induction (or can be proven directly via a similar argument).

5.2 Reducing the Distance Parameter β

In this section we transform β-TCR to standard TCR (with some loss in hardness and shrinkage). Such a transformation can be easily obtained (non-locally) by encoding the input x via an error-correcting code. Here we provide a local alternative which employs low-density parity-check matrices (LDPC). Such matrices will also be used to amplify the hardness parameter δ in the next section.

LDPC. In order to amplify the distance parameter β we will need *sparse parity check matrices* of a good code. Let $m < n$ be an integer. We say that a matrix $M \in \mathbb{Z}_2^{m \times n}$ has a *dual (relative) distance* of $\beta \in (0,1)$ if the Hamming weight of every non-zero codeword $x \in \ker(M) = \{x | Mx = 0\}$ is larger than βn. We say that a family $\mathcal{M}_{m(n) \times n}$ of efficiently samplable distributions over matrices in $\{0,1\}^{m(n) \times n}$ is a low-density parity check code with error δ and distance β (in short, (δ, β)-LDPC) if (1) with probability at least $1-\delta$ a matrix $M \stackrel{R}{\leftarrow} \mathcal{M}_{m(n) \times n}$ has dual distance of β and (2) all matrices M in the support of $\mathcal{M}$ are *sparse* in the sense that the number of ones in each row and each column is bounded by some absolute constant d which does not depend on n. We will make use of the following proposition due to [12, Thm. 7.1].

Proposition 1. *For every $\varepsilon \in (0,1)$ there exists an efficiently samplable distribution $\mathcal{M}_{\varepsilon n \times n}$ of $(0, \beta(\varepsilon))$-LDPC for some $\beta = \varepsilon / \mathrm{polylog}(1/\varepsilon)$.*

Lemma 3 (β-TCR to TCR). *Let $\varepsilon' < \varepsilon$ and let $\mathcal{M}_{\varepsilon' n \times n}$ be an (δ', β)-LDPC. Let $\mathcal{H} = \{h_k\}$ be δ-secure β-TCR with shrinkage factor of $1-\varepsilon$ and key sampler $\mathcal{K}$, and define*

$$\mathcal{H}' = \left\{h'_{k,M}\right\} \quad h'_{k,M} = (h_k(x), Mx), \qquad \text{where } (k, M) \stackrel{R}{\leftarrow} (\mathcal{K}(1^n), \mathcal{M}_{\varepsilon' n \times n})$$

Then, $\mathcal{H}'$ is $(\delta + \delta')$-secure TCR with shrinkage factor of $1 - \varepsilon + \varepsilon'$.

Proof. We need the following observation: when $M \stackrel{R}{\leftarrow} \mathcal{M}$ has a dual distance of β, any pair of distinct strings x and x' which collide under $h'_{k,M}$ must be β-far. Indeed, if this is not the case then, since $Mx = Mx'$, the vector $x \oplus x'$ is a non-zero vector in the kernel of M whose Hamming weight is smaller than βn, in contrast to our assumption. The lemma now follows easily.

Let $\mathcal{A}_2$ be an TCR adversary that, given $(x, r) \stackrel{R}{\leftarrow} \mathcal{A}_1(1^n)$ and $h_{k,M} \stackrel{R}{\leftarrow} \mathcal{H}'$, finds a collision x' with x under $h_{k,M}$ with probability $\delta_{\mathcal{A}}$. To prove the lemma we define an adversary $\mathcal{B}$ that finds a β-close collision x' with $x \stackrel{R}{\leftarrow} \mathcal{A}_1(1^n)$ under $h_k \stackrel{R}{\leftarrow} \mathcal{H}$ with probability $\delta_{\mathcal{B}} \geq \delta_{\mathcal{A}} - \delta'$. Given a key $k \stackrel{R}{\leftarrow} \mathcal{K}(1^n)$ and a target/state pair $(x, r) \stackrel{R}{\leftarrow} \mathcal{A}_1(1^n)$, the adversary $\mathcal{B}$ samples $M \stackrel{R}{\leftarrow} \mathcal{M}$ and call $\mathcal{A}_2$ with $h_{k,M}$. Let good be the set of matrices whose dual distance is β and let us say that $\mathcal{A}$ *wins* if it outputs a valid collision x' with x under $h_{k,M}$, i.e., $x' \neq x, h_k(x) = h_k(x')$ and $Mx = Mx'$. Then we can write

$$\begin{aligned}
\delta_{\mathcal{A}} &= \Pr_{k,M,x,r}[\mathcal{A}_1(k, M, x, r) \text{ wins } | M \in \text{good}] \cdot \Pr_M[M \in \text{good}] \\
&\quad + \Pr_{k,M,x,r}[\mathcal{A}_1(k, M, x, r) \text{ wins } | M \notin \text{good}] \cdot \Pr_M[M \notin \text{good}] \\
&\leq \Pr_{k,M,x,r}[\mathcal{A}_1(k, M, x, r) \text{ wins } | M \in \text{good}] \cdot (1 - \delta') + \delta' \\
&\leq \delta_{\mathcal{B}} + \delta',
\end{aligned}$$

where the last inequality follows from the observation. □

Observe that the above transformation is local since the family $\mathcal{M}$ is sparse.

5.3 Hardness Amplification

We move on to amplify the hardness parameter δ from constant to negligible. In addition to LDPCs (i.e., sparse shrinking linear transformations), we employ *Distance Amplifiers* (i.e., sparse linear transformations which expands the input) which has the property of mapping any pair (x, x') of far-apart inputs to a pair of far apart outputs (y, y'). This can be seen as a relaxation of standard error-correcting codes which amplify the distance between any pair of *distinct* inputs.

Distance Amplifiers. Let $m > n$ be an integer and $\beta, \gamma \in (0,1)$ be constants. We say that a matrix $T \in \mathbb{Z}_2^{m \times n}$ is $(\beta \to \gamma)$-*distance amplifying* if for every pair $x, x' \in \{0,1\}^n$ of β-far strings the m-bit strings Tx and Tx' are γ-far. (Jumping ahead, we note that γ is allowed to be smaller than β as long as it is larger than the hardness parameter δ.) We say that a family $\mathcal{T}_{m(n)\times n}$ of efficiently samplable distributions over matrices in $\{0,1\}^{m(n)\times n}$ is a $(\beta \to \gamma)$ sparse distance amplifier (in short, $(\beta \to \gamma)$-SDA) if (1) with all but negligible probability a matrix $T \stackrel{R}{\leftarrow} \mathcal{T}_{m(n)\times n}$ is $(\beta \to \gamma)$-distance amplifying and (2) all matrices T in the support of $\mathcal{T}$ are sparse, meaning that the number of ones in each row and each column is bounded by some absolute constant d which does not depend on n. In the full version we prove the following proposition.

Proposition 2. *For every constant $\beta \in (0,1)$ and constant $\gamma \in (0, \frac{1}{2})$ there exists an efficiently samplable $(\beta \to \gamma)$-SDA $\mathcal{T}_{cn\times n}$ where $c = c(\beta, \gamma)$ is a constant.*

Let $T \in \{0,1\}^{cn^2 \times n^2}$. In the following we think of the linear mapping $x \mapsto Tx$ as a mapping from n^2-bit strings to a tuple of cn strings of length n each. Accordingly, for $i \in [cn]$ we let $(Tx)_i \in \{0,1\}^n$ denote the i-th entry of Tx.

Lemma 4 (Hardness Amplification). *Let $\mathcal{H} = \{h_k : \{0,1\}^n \to \{0,1\}^{\varepsilon_1 n}\}$ be δ-secure β-TCR with key sampler $\mathcal{K}$, let $\mathcal{M}_{\varepsilon_0 n^2 \times n^2}$ be a β-LDPC, and $\mathcal{T}_{cn^2\times n^2}$ be a $(\beta \to \gamma)$-SPA, where the constants $\varepsilon_0, \varepsilon, \gamma, \delta \in (0,1)$ and $c > 1$ satisfy $\varepsilon_0 + \varepsilon c < 1$ and $\delta < \gamma$. Then, the following collection $\mathcal{H}'$ which shrinks n^2-bit strings by a factor of $\varepsilon_0 + \varepsilon c$ is a standard TCR:*

$$h'_{(k_1,\ldots,k_{cn}),M,T} : x \mapsto (Mx, h_{k_1}((Tx)_1), \ldots, h_{k_{cn}}((Tx)_{cn}))\,,$$

where $M \stackrel{R}{\leftarrow} \mathcal{M}_{\varepsilon_0 n^2 \times n^2}, T \stackrel{R}{\leftarrow} \mathcal{T}_{cn^2\times n^2}$ and $k_i \stackrel{R}{\leftarrow} \mathcal{K}(1^n)$ for $i \in [cn]$.

Proof. Let $\mathcal{A} = (\mathcal{A}_1, \mathcal{A}_2)$ be an adversary that breaks $\mathcal{H}'$ with probability $\delta_{\mathcal{A}}$. We construct an adversary $\mathcal{B} = (\mathcal{B}_1, \mathcal{B}_2)$ that given cn independent samples of $\mathcal{H}$ finds collisions on γ fraction of them with probability $\delta_{\mathcal{B}}$. Namely, let $\delta_{\mathcal{B}}$ be

$$\Pr_{\substack{\boldsymbol{k}\stackrel{R}{\leftarrow}\mathcal{K}^{cn}(1^n)\\(\boldsymbol{y},R)\stackrel{R}{\leftarrow}\mathcal{B}_1(1^n)}} [\mathcal{B}_2(\boldsymbol{k}, \boldsymbol{y}, R) = \boldsymbol{y'} \text{ s.t. } |\{i : (y_i \neq y'_i) \wedge (h_{k_i}(y_i) = h_{k_i}(y'_i))\}| \geq \gamma cn],$$

where $\boldsymbol{k} = (k_1, \ldots, k_{cn}), \boldsymbol{y} = (y_1, \ldots, y_{cn})$, and $\boldsymbol{y'} = (y'_1, \ldots, y'_{cn})$. A general *threshold direct product theorem* of Impagliazzo and Kabanets [16, Thm 5.2]

shows that the advantage $\delta_{\mathcal{B}}$ is upper-bounded by $2^{-cnD(\gamma\|\delta)} + \mathrm{neg}(n) = \mathrm{neg}(n)$. Hence, to prove the lemma it suffices to show that

$$\delta_{\mathcal{A}} - \mathrm{neg}(n) \leq \delta_{\mathcal{B}}.$$

Let us define $\mathcal{B}$. The target sampler $\mathcal{B}_1(1^n)$ samples $M \xleftarrow{R} \mathcal{M}_{\varepsilon_0 n^2 \times n^2}, T \xleftarrow{R} \mathcal{T}_{cn^2 \times n^2}$ and $(x, r) \xleftarrow{R} \mathcal{A}_1(1^n)$. It outputs the state $R = (M, T, x, r)$ and the target vector $\boldsymbol{y} = (y_1, \ldots, y_{cn})$ where $y_i = (Tx)_i$. Given $(\boldsymbol{k}, \boldsymbol{y}, R = (x, r, M, T))$, the collision-finder $\mathcal{B}_2$ passes to $\mathcal{A}_2$ the key $(\boldsymbol{k}, M, T)$, the target x, and the state r, and asks for a collision x' under $h'_{\boldsymbol{k},M,T}$. The output of $\mathcal{B}_2$ is $\boldsymbol{y}' = (y'_1, \ldots, y'_{cn})$ where $y'_i = (Tx')_i$. We say that $\mathcal{A}_1(\boldsymbol{k}, M, T, x, r)$ *wins* if its output x' collide with x under $h'_{\boldsymbol{k},M,T}$ and $x \neq x'$. A pair (M, T) is good if M has dual distance of β and T is $(\beta \to \gamma)$ distance amplifying. We claim that

$$\delta_{\mathcal{A}} - \mathrm{neg}(n) \leq \Pr_{\boldsymbol{k},M,T,x,r}[\mathcal{A}_1(\boldsymbol{k}, M, T, x, r) \text{ wins } | (M, T) \in \text{good}] \leq \delta_{\mathcal{B}}.$$

The first inequality follows from Bayes' law together with $\Pr[\text{good}] > 1 - \mathrm{neg}(n)$. As for the second inequality, observe that if $\mathcal{A}$ wins and (M, T) are good then the collision x and x' must be β-far (as $Mx = Mx'$) and therefore Tx and Tx' must disagree on at least γcn^2 coordinates. Hence, for at least γ fraction of $i \in [cn]$ we have that $(Tx)_i \neq (Tx')_i$. Furthermore, $h_{k_i}((Tx)_i) = h_{k_i}((Tx')_i)$ for all $i \in [cn]$ since $\mathcal{A}$ wins. Hence, in this case $\mathcal{B}$ wins as well and the claim follows. □

Theorem 5 follows by combining Claims 6, 7 and Lemmas 3, 4 with properly chosen parameters. See full version for details.

Acknowledgement. We thank Uri Feige and Danny Vilenchik for valuable discussions.

References

1. Achlioptas, Ricci-Tersenghi: On the solution-space geometry of random constraint satisfaction problems. In: STOC: ACM Symposium on Theory of Computing (STOC) (2006)
2. Alekhnovich, M., Hirsch, E.A., Itsykson, D.: Exponential lower bounds for the running time of DPLL algorithms on satis_able formulas. J. Autom. Reasoning 35(1-3), 51–72 (2005)
3. Applebaum, B.: Pseudorandom generators with long stretch and low locality from random local one-way functions. In: STOC, pp. 805–816 (2012)
4. Applebaum, B., Bogdanov, A., Rosen, A.: A dichotomy for local small-bias generators. In: Cramer, R. (ed.) TCC 2012. LNCS, vol. 7194, pp. 600–617. Springer, Heidelberg (2012)
5. Applebaum, B., Ishai, Y., Kushilevitz, E.: Computationally private randomizing polynomials and their applications. Journal of Computational Complexity 15(2), 115–162 (2006)
6. Applebaum, B., Ishai, Y., Kushilevitz, E.: Cryptography in NC^0. SIAM Journal on Computing 36(4), 845–888 (2006)

7. Applebaum, B., Ishai, Y., Kushilevitz, E.: Cryptography with constant input locality. Journal of Cryptology 22(4), 429–469 (2009)
8. Bellare, M., Rogaway, P.: Collision-resistant hashing: Towards making UOWHFs practical. In: Kaliski Jr., B.S. (ed.) CRYPTO 1997. LNCS, vol. 1294, pp. 470–484. Springer, Heidelberg (1997)
9. Bogdanov, A., Qiao, Y.: On the security of goldreich's one-way function. In: Dinur, I., Jansen, K., Naor, J., Rolim, J. (eds.) APPROX and RANDOM 2009. LNCS, vol. 5687, pp. 392–405. Springer, Heidelberg (2009)
10. Bogdanov, A., Rosen, A.: Input locality and hardness amplification. In: Ishai, Y. (ed.) TCC 2011. LNCS, vol. 6597, pp. 1–18. Springer, Heidelberg (2011)
11. Canetti, R., Rivest, R., Sudan, M., Trevisan, L., Vadhan, S.P., Wee, H.M.: Amplifying collision resistance: A complexity-theoretic treatment. In: Menezes, A. (ed.) CRYPTO 2007. LNCS, vol. 4622, pp. 264–283. Springer, Heidelberg (2007)
12. Capalbo, Reingold, Vadhan, Wigderson: Randomness conductors and constant-degree lossless expanders. In: STOC: ACM Symposium on Theory of Computing (STOC) (2002)
13. Cook, J., Etesami, O., Miller, R., Trevisan, L.: Goldreich's one-way function candidate and myopic backtracking algorithms. In: Reingold, O. (ed.) TCC 2009. LNCS, vol. 5444, pp. 521–538. Springer, Heidelberg (2009)
14. Goldreich, O.: Candidate one-way functions based on expander graphs. Electronic Colloquium on Computational Complexity (ECCC) 7(090), `citeseer.nj.nec.com/382413.html` (2000)
15. Haitner, I., Holenstein, T., Reingold, O., Vadhan, S.P., Wee, H.: Universal one-wayhash functions via inaccessible entropy. IACR Cryptology ePrint Archive 2010, 120 (2010)
16. Impagliazzo, R., Kabanets, V.: Constructive proofs of concentration bounds. In: Serna, M., Shaltiel, R., Jansen, K., Rolim, J. (eds.) APPROX and RANDOM 2010. LNCS, vol. 6302, Springer, Heidelberg (2010)
17. Ishai, Y., Kushilevitz, E., Ostrovsky, R., Sahai, A.: Cryptography with constant computational overhead. In: Proc. of 40th STOC, pp. 433–442 (2008)
18. Knudsen, Preneel: Construction of secure and fast hash functions using nonbinary error-correcting codes. IEEETIT: IEEE Transactions on Information Theory 48 (2002)
19. Lord, N.: Binomial averages when the mean is an integer. The Mathematical Gazette 94, 331–332 (2010)
20. Mossel, E., Shpilka, A., Trevisan, L.: On _-biased generators in NC^0. Proc. 44th FOCS, 136–145 (2003)
21. Naor, Yung: Universal one-way hash functions and their cryptographic applications. STOC: ACM Symposium on Theory of Computing (STOC) (1989)
22. Panjwani, S.K.: An experimental evaluation of goldreich's one-way function. Tech.rep., IIT, Bombay oded/PS/ow-report.ps (2001), `http://www.wisdom.weizmann.ac.il/`
23. Rogaway, P., Shrimpton, T.: Cryptographic hash-function basics: Definitions, implications, and separations for preimage resistance, second-preimage resistance, and collision resistance. In: Roy, B., Meier, W. (eds.) FSE 2004. LNCS, vol. 3017, pp. 371–388. Springer, Heidelberg (2004)
24. Rompel: One-way functions are necessary and su_cient for secure signatures. STOC: ACM Symposium on Theory of Computing (STOC) (1990)