Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 98 (2013)

Locally Computable UOWHF with Linear Shrinkage*

Benny Applebaum† Yoni Moses‡

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 fan-out [8]. Furthermore, such local constructions have turned to be surprisingly helpful in speeding-up the *sequential complexity* of cryptography [19]. At a more abstract level, the study of locally computable cryptography allows us to understand whether extremely simple functions can generate cryptographic hardness.

*An extended abstract of this work appears in Eurocrypt 2013. The full version appears in Journal of Cryptology, volume 30, pages 672-698 (2017).

†School of Electrical Engineering, Tel-Aviv University, bennyap@post.tau.ac.il. Supported by Alon Fellowship, ISF grant 1155/11, Israel Ministry of Science and Technology (grant 3-9094), and GIF grant 1152/2011.

‡School of Computer Science, Tel-Aviv University, ymoses@gmail.com.

ISSN 1433-8092

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 [7, 6, 8]. 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) [23], in which the adversary specifies the target string x without seeing the function h. (This property is also known as *target collision resistance* [9], 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. In [7] it is shown 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-resistant 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 [23, 26, 17]. Although there are several ways to amplify the shrinkage of a UOWHF (cf. [23, 9]), 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 [8] 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] Put differently, high input locality (as captured by the so-called Confusion/Diffusion or Avalanche principle) is typically viewed as a desired property for collision resistance – but is it really necessary?

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

[1] We note that standard transformations from one-way functions to UOWHFs [23, 26, 17] are inherently non-local as they employ primitives such as k-wise independent hash functions which cannot be computed locally.

random local functions (also known as Goldreich's one-way function [16]). 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 nodes 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-ary 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 2 for a formal definition.) We prove the following theorem.

Theorem 1.1 (Main theorem)**.** *There exists a constant d and a predicate $P : \{0,1\}^d \to \{0,1\}$ for which the following holds. If the collection $\mathcal{F}_{P,n,m=\Omega(n^3)}$ 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 condition. In particular, we show that the predicate $\mathsf{MST}_{d_1,d_2}(x,y) = (y_1 \oplus \ldots \oplus y_{d_1}) \oplus (x_1 \wedge \ldots \wedge x_{d_2})$, defined by [22], satisfies the condition for every $d_2 \geq 2$ and every sufficiently large odd constant d_1. 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 [24, 13] and theoretically [16, 2, 13, 21, 14, 20, 11]. In fact, recent evidence suggest that, for a proper predicate, this collection may even be pseudorandom [5, 4]. Interestingly, Theorem 1.1 can be proved under the (possibly weaker) assumption that $\mathcal{F}_{P,n,m=\Omega(n)}$ 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 1.1. First, it is possible to reduce the output locality to 3 (which is optimal) while preserving (tiny) linear shrinkage (i.e., $m = (1-\varepsilon)n$ for some small ε) via the compiler of [7].[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 circuit contains $O(n/2^i)$ gates). As observed by [19], one can then employ the Naor-Yung transform [23] and sign n-bit messages with linear time complexity and only *additive cryptographic overhead*, i.e., $O(n + \kappa)$. (See Section 6 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 [19].[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

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

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

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.

As our main technical contribution, we prove that a weak form of this conjecture holds assuming the pseudorandomness of $\mathcal{F}_{P,n,m'}$ (where $m' > n > m$). Specifically, we prove the following theorem. (See Section 4 for details).

Theorem 1.2. *There exists a predicate P, constants $\varepsilon, \beta \in (0, \frac{1}{2})$ and $c > 1$ such that for every $\delta > 0$, if $\mathcal{F}_{P,n,cn}$ is $(\delta/3)$-pseudorandom then it is hard to find β-far target collisions in $\mathcal{F}_{P,n,(1-\varepsilon)n}$ with probability better than δ.*

We mention that we can base the theorem on the one-wayness of random local functions using the reduction of [4].

Proof idea. Let $m = (1-\varepsilon)n$. Let P be a balanced predicate, which, in addition, enjoys the following *sensitivity* properties:[4]

$$\forall x, x' \in \{0,1\}^n, \Delta(x,x') > \beta \Rightarrow \mathop{\mathbb{E}}_{f \xleftarrow{R} \mathcal{F}_{P,n,m}} [\Delta(f(x), f(x'))] > \gamma \qquad \text{for some constants } \beta, \gamma > 0$$

$$\forall x, x' \in \{0,1\}^n, \Delta(x,x') = \frac{1}{2} \Rightarrow \mathop{\mathbb{E}}_{f \xleftarrow{R} \mathcal{F}_{P,n,m}} [\Delta(f(x), f(x'))] = \frac{1}{2},$$

where $\Delta(\cdot,\cdot)$ denotes the relative Hamming distance and $\mathbb{E}$ denotes expectation. An example of such a predicate is parity $\oplus_d$ with an odd arity d. A relaxation of the above properties (e.g., by considering only x of Hamming weight $\frac{1}{2}$) allows us to use richer families of predicates including MST_{d_1,d_2} for every $d_2 \geq 2$ and every odd constant d_1. (Larger d_1 pushes β towards zero and increases γ towards $\frac{1}{2}$.)

[4] It can be shown that the above properties are actually independent of the output length m, and corresponds to a generalized notion of noise sensitivity of P. See Section 3.

Assume that we have an algorithm $\mathcal{A}$ that, given a random function $h \xleftarrow{R} \mathcal{F}_{P,n,m}$ and a random target w, finds a β-far sibling with probability δ. Let us first try to use $\mathcal{A}$ to invert the collection $\mathcal{F}_{P,n,m'}$ with output length of $m' \approx 2m$. Given a random function $f_G \xleftarrow{R} \mathcal{F}_{P,n,m'}$ specified by a random input-output dependencies graph G, and an image $y = f_G(x)$ of a random point $x \xleftarrow{R} \{0,1\}^n$, we will recover the preimage x as follows.

First, we choose a target w uniformly at random and partition the graph G into two graphs: $G_=$ which contains only the output nodes for which $f_G(w)$ agrees with y, and $G_{\neq}$ which contains the remaining output nodes. Hence,

$$f_{G_=}(x) = f_{G_=}(w) \qquad \text{and} \qquad f_{G_{\neq}}(x) = \overline{f_{G_{\neq}}(w)},$$

where $\overline{z}$ denotes the bit-wise complement of the string z. Since P is balanced, each subgraph contains roughly m' outputs. Next, we ask $\mathcal{A}$ for a β-far sibling w' of w under the function $f_{G_=}$. As we will see next w' is likely to be *correlated* with the preimage x, in the sense that for some constant $\alpha > 0$, either w' or its complement $\overline{w'}$ agree with x on $(\frac{1}{2}+\alpha)$-fraction of their coordinates. At this point, we will employ a result of [10] that allows us to fully recover x given such a correlated string w' (and additional $O(n)$ outputs).

It remains to show that w' is likely to be correlated with the preimage x. Using the sensitivity properties of the predicate P, this boils down to proving that $f_G(w')$ and $f_G(x)$ agree on $\frac{1}{2} + \alpha'$ of their coordinates, for some constant $\alpha' > 0$. Let us first (optimistically) assume that w' is *statistically independent* of the subgraph $G_{\neq}$ that was not submitted to the adversary. That is, imagine that this part of the dependencies graph is chosen uniformly at random after w' is obtained. Since w is β-far from w', this pair is expected to disagree on a constant fraction γ of the remaining coordinates of $f_{G_{\neq}}$. Namely,

$$\Delta(f_{G_{\neq}}(w), f_{G_{\neq}}(w')) > \gamma.$$

Since $f_{G_{\neq}}(x) = \overline{f_{G_{\neq}}(w)}$ it follows that

$$\Delta(f_{G_{\neq}}(x), f_{G_{\neq}}(w')) < 1 - \gamma.$$

Furthermore, since w' collides with w under $f_{G_=}$ we have that

$$f_{G_=}(x) = f_{G_=}(w) = f_{G_=}(w').$$

We conclude that x and w agree on a fraction of $1 - \frac{1}{2}(1-\gamma) = \frac{1}{2} + \gamma/2$ of the outputs of f_G (γ-fraction of the coordinates of $f_{G_{\neq}}$ and all the coordinates of $f_{G_=}$).

The above argument is over-optimistic, since it is not clear that w' is statistically independent of the subgraph $G_{\neq}$. (Indeed, the adversary $\mathcal{A}$ chooses w' based on $(w, G_=)$ which contain some information on x and, therefore, also on $G_{\neq}$.) Fortunately, we can show that a failure of the above approach allows to distinguish the string $y = f_G(x)$ from a truly random string. Hence, we are in a win-win situation: we can either invert $\mathcal{F}$ by finding a correlated string, or we can distinguish its output from a random string. So the theorem can be based on the pseudorandomness of $\mathcal{F}_{P,n,n+\Omega(n)}$. □

The above reduction leaves us with a δ-secure β-TCR $\mathcal{H}$ of linear shrinkage $n - m = \varepsilon n$. To prove Theorem 1.1, we show that, for sufficiently small constants $\delta, \beta > 0$, any (δ, β)-target collision resistance (TCR) $\mathcal{H}$ can be *locally* amplified into standard TCR while preserving *linear* shrinkage. This is done via the following steps.

Amplifying hardness. First we reduce the security parameter δ to be negligible at the expense of slightly increasing the distance parameter β to, say, $\beta+2\delta$. This is done by taking t independent copies of $\mathcal{H}$ and applying them to t independent inputs, i.e., $h'(x_1,\ldots,x_t)=(h_1(x_1),\ldots,h_t(x_t))$. It is not hard to see that any $(\beta+2\delta)$-far collision $x=(x_1,\ldots,x_t)$ and $y=(y_1,\ldots,y_t)$ under h' induces β-far collisions (x_i,y_i) for at least 2δ-fraction of the copies of h. Standard Threshold Direct Product Theorems (e.g., [18, Theorem 5.2]) guarantee that the latter task cannot be achieved with more than negligible probability.

Eliminating close collisions. In the second step we eliminate β-close collisions by letting $h'(x)=(h(x),Mx)$ where M is a parity-check matrix 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 function h' is immunized against β-close collisions. Since there are sparse parity-check matrices with constant dual relative distance (aka LDPC), the transformation is locally computable.[5] Finally note that although the shrinkage factor is slightly degraded, we can locally amplify it to any constant via a constant number of self compositions.

Organization. Section 2 gives the necessary preliminaries. In Section 3, we present a new notion of sensitivity for predicates, study its properties and identify a class of "good" predicates for which our results apply. In Section 4 we reduce the one-wayness of random local functions to (δ,β) target-collision resistance. Later, in Section 5, we show how to transform (δ,β) TCR to standard TCR while preserving constant locality and linear shrinkage. Finally, in Section 6 we combine the results of the previous sections and derive the main theorem and its applications.

2 Preliminaries

General. We let $[n]$ denote the set $\{1,\ldots,n\}$. For a pair of strings $x,x'\in\{0,1\}^n$, we let $\Delta(x,x')$ denote the relative Hamming distance between x and x', i.e., $|\{i\in[n]:x_i\neq x'_i\}|/n$. A pair of strings is α-close if $\Delta(x,x')\leq\alpha$ and α-far if $\Delta(x,x')>\alpha$. By default, logarithms are taken to base 2. For reals $p,q\in(0,1)$ we let $H_2(p):=-p\log(p)-(1-p)\log(1-p)$ denote the binary entropy function, and $D_2(p\|q):=p\log(\frac{p}{q})+(1-p)\log(\frac{1-p}{1-q})$ denote the relative entropy function (also known as the binary Kullback-Leibler divergence). Observe that $D_2(p\|\frac{1}{2})=1-H_2(p)$. We will use the following form of Chernoff-Hoeffding:

Fact 2.1 (Additive Chernoff bound). *Let $X_1,\ldots,X_n$ be i.i.d. random variables where $X_i\in[0,1]$ and $\mathbb{E}[X_i]=p$. Then, for every $\varepsilon>0$,*

$$\Pr\left[n^{-1}\sum_i X_i\geq p+\varepsilon\right]\leq 2^{-D_2(p+\varepsilon\|p)n},\qquad \Pr\left[n^{-1}\sum_i X_i\leq p-\varepsilon\right]\leq 2^{-D_2(p-\varepsilon\|p)n}$$

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

[5] A dual approach would be to pre-code the input x via an error-correcting code with constant relative distance and constant rate. While this approach eliminates close collisions, it is inherently non-local. Indeed, it can be shown that local functions cannot compute good error correcting codes.

Locality and Degree. Let $f : \{0,1\}^n \to \{0,1\}^l$ be a function. We say that the i-th output variable y_i *depends* on the j-th input variable x_j (or equivalently, x_j *affects* the output y_i) if there exists a pair of input strings which differ only on the j-th location whose images differ on the i-th location. The locality of an output variable (resp., input variable) is the number of inputs on which it depends (resp., on which it affects). We say that an output has degree d if it can be expressed as a multivariate polynomial of degree d in the inputs over the binary field $\mathbb{F}_2$. The locality of an output variable trivially upper bounds its degree.

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 (resp., input) of the function f_k has locality of at most d. Similarly, the collection has constant *algebraic degree* of d if for every fixed k each output of the function f_k has degree of at most d. The collection is *locally computable* if it has both constant input locality and constant output locality. When $\mathcal{F}$ is used as a primitive we will always assume that the adversary that tries to break it gets the collection index as a public parameter. Moreover, our constructions are all in the "public-coin" setting, and so they remain secure even if the adversary gets the coins used to sample the index of the collection.

One-wayness and Pseudorandomness. Let $\delta(n) \in (0,1)$ and $\beta(n) \in (0, \frac{1}{2})$. We say that a collection of functions $\mathcal{F} = \left\{f_k : \{0,1\}^n \to \{0,1\}^{m(n)}\right\}$ 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 $\delta(n)$: Given $k \xleftarrow{R} \mathcal{K}(1^n)$ and $y = f_k(x)$ 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. Note the size of the list is bounded by the running-time of the adversary, which is polynomial in n. The special case of $\beta = 0$ corresponds to the standard notion of δ-one-wayness, or simply one-wayness when $\delta = \mathrm{neg}(n)$. This is consistent with standard one-wayness (cf. [15]) as when $\delta = 0$, the algorithm can efficiently check which of the candidates (if any) is a preimage and output only a single candidate z rather than a list. A collection of functions $\mathcal{F}$ is δ-pseudorandom if $|\Pr[\mathcal{A}(k, f_k(x)) = 1] - \Pr[\mathcal{A}(k, y) = 1]| \leq \delta(n)$, where $k \xleftarrow{R} \mathcal{K}(1^n)$, $x \xleftarrow{R} \{0,1\}^n$ 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

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

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 an 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 [9, 25].

Random Local Functions. Let $P : \{0,1\}^d \to \{0,1\}$ be a predicate, and let $G = (S_1, \ldots, S_m)$ where each S_i is a d-tuple $(S_{i,1}, \ldots, S_{i,d})$ whose entries are d distinct elements of $[n]$. We will think of G as a bipartite graph with n input nodes and m output nodes where each output i is connected to the d (ordered) 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 tuple S_i, i.e.,

$$y_i = P(x_{S_i}) = P(x_{S_{i,1}}, \ldots, x_{S_{i,d}}).$$

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 $m(n)$ tuples uniformly and independently at random from all the possible d-tuples with distinct elements. 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 Sensitivity

3.1 Overview

Let $P : \{0,1\}^d \to \{0,1\}$ be a d-ary 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 d-tuple with distinct elements $(i_1, \ldots, i_d)$ 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 over the choice of x, 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 tuple S (from

Eq. 1) is chosen by selecting d indices uniformly at random from $[n]$ *with replacement* (i.e., the entries may not be distinct). We will show (in Lemma 3.1) that, with all but negligible probability over x, these simplifications have only a minor effect on the value of the experiment (the error tends to zero with n). We will later show (Lemma 3.3) that for every constants $\beta > 0$ and $\gamma < \frac{1}{2}$ there are some concrete (non-linear) highly sensitive predicates for which a modification of more than β fraction of the inputs, flips the output with probability larger than γ.

3.2 Generalized Noise Sensitivity

The above discussion motivates 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, the pair (w_i, w'_i) takes the value 01 (respectively, 00,10, and 11) with probability α_{01} (respectively, $\frac{1}{2} - \alpha_{01}$, α_{10} and $\frac{1}{2} - \alpha_{10}$). For $\alpha \in [0,1]$ let $s_P(\alpha)$ denote the infimum of $\Pr_{(w,w') \stackrel{R}{\leftarrow} \mathcal{D}(\alpha_0,\alpha_1)}[P(w) \neq P(w')]$ taken over all $\alpha_{0,1}$ and $\alpha_{1,0}$ which sum-up to α. Call $x \in \{0,1\}^n$ *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 relates $s_P(x, x')$ to $s_P(\alpha)$.

Lemma 3.1. *For every predicate P, the function $s_P(\alpha)$ is well defined and continuous. Also, for every typical $x \in \{0,1\}^n$ and every string $x' \in \{0,1\}^n$*

$$s_P(x, x') \geq s_P(\Delta(x, x')) - \delta(n),$$

where the error term $\delta(n) = o(1)$.

Proof. Fix P and let $s(\cdot) = s_P(\cdot)$. Let $\mathcal{D}$ be an arbitrary probability distribution over pair of bits, which is described by the probability vector $z = (z_{00}, z_{01}, z_{10}, z_{11})$. Sample a pair of d-bit strings $w, w' \in \{0,1\}^d$ by collecting d independent samples of bit pairs (w_i, w'_i) from $\mathcal{D}$. Then the quantity $\Pr_{(w,w')}[P(w) \neq P(w')]$ can be written as a degree d multivariate polynomial in z:

$$Q(z) = \sum_{\substack{w,w' \in \{0,1\}^d \\ P(w) \neq P(w')}} \prod_{i=1}^{d} z_{w_i w'_i}.$$

Specifically, for every fixed α we can write $s(\alpha)$ as

$$\inf_{\alpha_{0,1} \in [\max(0, \alpha - \frac{1}{2}), \min(\alpha, \frac{1}{2})]} Q(z)$$

where

$$z_{01} = \alpha_{0,1}, z_{00} = \frac{1}{2} - \alpha_{0,1}, z_{10} = \alpha - \alpha_{0,1}, z_{11} = 1 - (\alpha - \alpha_{0,1}).$$

Hence, we are minimizing a degree d univariate polynomial over a closed interval, and so $s(\alpha)$ is well defined. We also conclude that $s(\alpha)$ is continuous since it is defined to be the minimum over the interval $[0, \alpha]$ of a continuous function (univariate polynomial).

We move on to the second part of the lemma. Fix some pair of n-bit strings x and x', and define the *frequency vector* $z = (z_{00}, z_{01}, z_{10}, z_{11})$ to be $z_{\sigma_1\sigma_2} = |\{i : (x_i x'_i) = \sigma_1\sigma_2\}|/n$. Imagine

that we were choosing the tuple S uniformly at random from $[n]^d$ allowing repetitions. Then, $s(x,x') = \Pr_S[P(x_S) \neq P(x'_S)] = Q(z)$. Since this is not the case and the elements of S are chosen without repetitions, the quantity $s(x,x')$ equals to

$$Q_n(z_{00},z_{01},z_{10},z_{11}) = \sum_{\substack{a,b\in\{0,1\}^d \\ P(a)\neq P(b)}} \prod_{i=1}^{d} (z_{a_ib_i} - \delta(a,b,i,z_{a_ib_i})), \tag{2}$$

where $\delta(a,b,i,z_{a_ib_i}) = \min(|\{j<i : (a_j,b_j)=(a_i,b_i)\}|/n, z_{a_ib_i})$. For every $z=(z_{00},z_{01},z_{10},z_{11})$ we have

$$Q(z) - 2^{2d}d^2/n \leq Q_n(z) \leq Q(z),$$

where the left inequality follows by noting that $\delta(a,b,i) \leq d/n$ and that for reals $p_i \geq \delta_i$ and integer t we have $\prod_{i=1}^{t}(p_i-\delta_i) \geq (\prod_i p_i) - \sum_i \delta_i$. Now assume that x is typical, and let z be the frequency vector of x and x'. By definition, $z_{00}+z_{01} \in [\frac{1}{2} \pm n^{-1/3}]$ and $z_{01}+z_{10} = \Delta(x,x')$. By adding/subtracting a small quantity of at most $n^{-1/3}$ to each coordinate of z, we can define a related *balanced* frequency vector z' for which $z'_{00}+z'_{01} = \frac{1}{2}$ and $z'_{01}+z'_{10} = \Delta(x,x')$. Observe that in this case $Q(z) \geq Q(z') - 2^{2d}d/n^{1/3}$, and overall it follows that

$$s(x,x') = Q_n(z) \geq Q(z) - 2^{2d}d^2/n \geq Q(z') - (2^{2d}d^2/n) - (2^{2d}d/n^{1/3}).$$

Since, by definition, $Q(z')$ is lower-bounded by $s(z'_{01}+z'_{10}) = s(\alpha)$ it follows that $s(x,x') \geq s(\alpha) - \delta(n)$ where $\delta(n) = 2^{2d}d^2/n + 2^{2d}d/n^{1/3} = o(1)$ and the lemma follows. □

3.3 Good Predicates

Definition 3.2 (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 $(d-1)$-ary predicate P'.*

Motivation. Recall that in Section 3.1, we described a game in which an adversary is given a random string x and outputs an α-far string x' with the hope of minimizing $s_P(x,x')$. (The latter quantity essentially approximates the distance between $f_{G,P}(x)$ and $f_{G,P}(x')$ for random G.) Property (1) above guarantees that as long as $\alpha > \beta$, the value of $s_P(x,x')$ will be at least γ (except for the negligible event where x is non-typical). The second property of Definition 3.2 is needed for two reasons. First, it allows us to use a theorem from [4] which reduces the pseudorandomness of the ensemble $\mathcal{F}_{P,n,m}$ to its one-wayness. In addition, it is not hard to verify that this condition implies that $s_P(\frac{1}{2}) = \frac{1}{2}$. The latter property implies that for proper output length ℓ, the ensemble $\mathcal{F}_{P,n,\ell}$ satisfies the following: If a pair of images $y=f(x)$ and $y'=f(x')$ is highly correlated, then the preimages x and x' must also have a non-trivial correlation. This property (to be formalized in Claim 4.4) will turn to be useful later.

Usage. In Section 4 we will use (β, γ)-good predicate P to construct β-RTCRs with shrinkage factor of $1 - \varepsilon$ for a constant $\varepsilon \in (0, \frac{1}{2})$ which satisfies the inequality

$$\varepsilon < 1 - \frac{1}{2(1 - H_2(\frac{1}{2} - \gamma))}, \tag{3}$$

where H_2 denotes the binary entropy function. As a result, we would like to have a small value of $\beta > 0$ and a large value of $\gamma < \frac{1}{2}$ (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.

Lemma 3.3. *Let Q be c-ary predicate for which $s_Q(1) \leq \frac{1}{2}$. For every constants $\gamma < \frac{1}{2}$, and $\beta > 0$ there exists a constant d for which the predicate*

$$P(x_1, \ldots, x_d, x_{d+1}, \ldots, x_{d+c}) = (x_1 \oplus \ldots \oplus x_d) \oplus Q(x_{d+1}, \ldots, x_{d+c})$$

is (β, γ)-good.

Proof. Fix some constants $\gamma \in (0, \frac{1}{2})$ and $\beta > 0$. We will show that for sufficiently large odd d (whose value will be determined later) the predicate P is (β, γ)-good. Clearly the predicate has a sensitive coordinate, and so it is left to show that for every $\alpha \in (\beta, 1)$ and every $\alpha_{01} \in [\max(0, \alpha - \frac{1}{2}), \min(\alpha, \frac{1}{2})]$

$$\Pr_{(w,w') \xleftarrow{R} \mathcal{D}(\alpha_{01}, \alpha - \alpha_{01})} [P(w) \neq P(w')] > \gamma, \tag{4}$$

Since P is computed by applying the predicate Q and the d-wise XOR predicate on *disjoint* inputs and XOR-ing the outcomes, we can write the LHS of Eq. 4 as

$$q(\alpha, \alpha_{01}) \cdot (1 - p_{\oplus}(\alpha, \alpha_{01})) + (1 - q(\alpha, \alpha_{01})) \cdot p_{\oplus}(\alpha, \alpha_{01}), \tag{5}$$

where

$$q(\alpha, \alpha_{01}) = \Pr_{(w,w') \xleftarrow{R} \mathcal{D}(\alpha_{01}, \alpha - \alpha_{01})} \left[Q(w) \neq Q(w') \right],$$

and

$$p_{\oplus}(\alpha, \alpha_{01}) = \Pr_{(w,w') \xleftarrow{R} \mathcal{D}(\alpha_{01}, \alpha - \alpha_{01})} \left[\bigoplus_{i=1}^{d} w_i \neq \bigoplus_{i=1}^{d} w'_i \right].$$

Letting χ_i denote $w_i \oplus w'_i$, we can rewrite $p_{\oplus}(\alpha, \alpha_{01})$ as $\Pr[\chi_1 \oplus \ldots \oplus \chi_d = 1]$. Observe that the χ_i's are independent Bernoulli variables with mean α. Therefore $p_{\oplus}(\alpha, \alpha_{01})$ is just the probability of seeing an odd number of successes when tossing d independent α-biased coins. It is not hard to verify (e.g., by induction on d) that for odd d we have

$$p_{\oplus}(\alpha, \alpha_{01}) = p_{\oplus}(\alpha) = \frac{1}{2} + \frac{1}{2}(2\alpha - 1)^d.$$

Overall, Eq. 5 simplifies to

$$q(\alpha, \alpha_{01})(1 - 2p_{\oplus}(\alpha)) + p_{\oplus}(\alpha). \tag{6}$$

Fix some positive $\varepsilon < \frac{1}{2} - \gamma$ and let $\delta \in (0, \frac{1}{2})$ be a small constant for which $q(\alpha, \alpha_{01}) \leq \frac{1}{2} + \epsilon$ for every $\alpha \in [1-\delta, 1]$ and $\alpha_{01} \in [\alpha - \frac{1}{2}, \frac{1}{2}]$. Such a δ is promised to exist since $q(1, 1/2) = s_Q(1) \leq \frac{1}{2}$ and since $q(\cdot, \cdot)$ is continuous (as shown in the proof of Lemma 3.1). Let d be an odd integer which is larger than $\max(\frac{\log(1-2\gamma)}{\log(1-2\beta)}, \frac{\log(1-2\gamma)}{\log(1-2\delta)})$. We prove that (6) is larger than γ via case analysis.

Case 1: $\beta \leq \alpha \leq \frac{1}{2}$**.** Observe that both $q(\alpha, \alpha_{01})$ and $1 - 2p_{\oplus}(\alpha) = -(2\alpha - 1)^d$ are non-negative (as $\alpha \leq \frac{1}{2}$ and d is odd). Therefore, (6) is lower-bounded by

$$p_{\oplus}(\alpha) \geq p_{\oplus}(\beta) = \frac{1}{2} + \frac{1}{2}(2\beta - 1)^d > \gamma,$$

where the first inequality follows from the fact that $p_{\oplus}(\cdot)$ is increasing in the interval $[\beta, \frac{1}{2}]$, and the last inequality holds as $d > \log(1-2\gamma)/\log(1-2\beta)$.

Case 2: $\frac{1}{2} \leq \alpha \leq 1 - \delta$**.** Since $1 - 2p_{\oplus}(\alpha) = -(2\alpha - 1)^d$ is negative and $q(\alpha, \alpha_{01}) \leq 1$, (6) is lower-bounded by

$$1 - 2p_{\oplus}(\alpha) + p_{\oplus}(\alpha) = \frac{1}{2} - \frac{1}{2}(2\alpha - 1)^d$$

which is monotonously decreasing in the interval $[\frac{1}{2}, 1 - \delta]$. It follows that the last term is lower-bounded by $\frac{1}{2} - \frac{1}{2}(1 - 2\delta)^d$ which is larger than γ since $d > \log(1-2\gamma)/\log(1-2\delta)$.

Case 3: $1 - \delta \leq \alpha \leq 1$**.** Since $\alpha > \frac{1}{2}$ the term $1 - 2p_{\oplus} = -(2\alpha - 1)^d$ is negative, and so (6) is minimized when $q(\alpha, \alpha_{01})$ is maximized. Recall that $q(\alpha, \alpha_{01}) \leq \frac{1}{2} + \epsilon$ for $\alpha > 1 - \delta$. Overall, (6) is lower-bounded by

$$\left(\frac{1}{2} + \varepsilon\right)(1 - 2p_{\oplus}(\alpha)) + p_{\oplus}(\alpha) = \frac{1}{2} + \epsilon - 2\epsilon p_{\oplus}(\alpha) \geq \frac{1}{2} - \epsilon > \gamma,$$

as required. □

Concrete instantiation. Observe that the condition $s_Q(1) \leq \frac{1}{2}$ simply means that $\Pr_w[Q(w) \neq Q(\bar{w}) \leq \frac{1}{2}]$, where w is a random c-bit string and $\bar{w}$ is the complement of w. Concretely, we suggest to let Q be the c-wise AND, for an arbitrary constant $c \geq 2$. (In this case, $\Pr_w[\bigwedge(w) \neq \bigwedge(\bar{w}) \leq 2 \cdot 2^{-c}] \leq \frac{1}{2}$.) This leads to the following family of *good* predicates

$$\mathsf{MST}_{d,c} = x_1 \oplus \ldots \oplus x_d \oplus (x_{d+1} \wedge \ldots \wedge x_{d+c}) \tag{7}$$

which generalizes the predicate from [22]. The previous lemma implies that for every constants $\gamma < \frac{1}{2}$, $\beta > 0$ and integer $c \geq 2$ there exists a constant d for which $\mathsf{MST}_{d,c}$ is (β, γ)-good.

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

In Section 4.1 we prove the following theorem.

Theorem 4.1. *Let P be a (β, γ)-good predicate. Assume there exists a constant $\varepsilon \in (0, \frac{1}{2})$ which satisfies Eq. 3, and let $m = (1 - \varepsilon)n$. Then, there exists a constant $\mu > 0$, such that for every $\delta_1(n)$ and $\delta_2(n)$ 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 $\delta' = \delta_1 + \delta_2 + \mathrm{neg}(n)$.*

In our proof the negligible overhead in the δ' expression is shown to be $2^{-\Omega(n^{1/3})}$. We did not attempt to optimize this term and it seems that a more careful analysis yields an exponential expression of 2^{-cn} where the constant c depends on the predicate P.

It turns out that, for random local functions, approximate one-wayness follows from one-wayness [10], which, in turn, (trivially) follows from pseudorandomness. Therefore, the implication of Theorem 4.1 can be based solely on pseudorandomness.[6] Formally, we derive the following corollary.

Corollary 4.2. *Let P be a (β, γ)-good d-ary predicate and assume that $\varepsilon > 0$ is a constant that satisfies Eq. 3. Then, there exists a constant $c = c(P, \varepsilon) > 0$ such that for every δ, if $\mathcal{F}_{P,n,cn}$ is δ-pseudorandom then $\mathcal{F}_{P,n,(1-\varepsilon)n}$ is 3δ-secure β-RTCR.*

Proof. Fix P, ε, and let $\mu = \mu(\varepsilon, P) > 0$ be the constant guaranteed by Theorem 4.1. Let δ be an arbitrary inverse polynomial. Assume that $\mathcal{F}_{P,n,cn}$ is δ-pseudorandom for some sufficiently large constant $c = c(P, \mu)$ whose value will be determined later. By employing Theorem 4.1 with $\delta_1 = \delta_2 = \delta$, it suffices to show that $\mathcal{F}_{P,n,2(1-\varepsilon)n}$ is both δ-pseudorandom and $(\delta, \frac{1}{2} - \mu)$ one-way. The pseudorandomness condition is trivially satisfied for $c > 2$. To establish approximation-resilient one-wayness, we first observe that since $\mathcal{F}_{P,n,cn}$ is δ-pseudorandom it must also be δ' one-way for $\delta' = \delta + 2^{(1-c)n} < \delta + o(1)$, assuming that $c > 1$. (To see this just use the hypothetical δ'-inverter as a distinguisher in the straightforward way, cf. [15, Section 3.3.6].) Next, we employ a theorem of Bogdanov and Qiao [10, Theorem 1.3] which asserts that for every constant $\mu > 0$ there exists a constant $k = k(\mu, d)$ such that if $\mathcal{F}_{P,n,kn}$ is δ' one-way then it is also $(\delta' + o(1), \frac{1}{2} - \mu)$ one-way (for every inverse polynomial δ'). Letting c be a sufficiently large constant (e.g., larger than $\max(k, 2)$), the corollary follows. □

We note that the corollary is valid even if δ decreases with n (as long as it is inverse polynomial), although we will employ it only with small constant values.

4.1 Proof of Theorem 4.1

Let μ be a constant which depends on P and ε whose value will be determined later. 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. It will convenient to further assume that $\mathcal{A}$ has a similar success probability when G is a random (n, m', d) graph for $m' < m$. This is without loss of generality, since such a graph G can be always padded into a random (n, m, d) graph H; clearly any β-far sibling of w under f_H is also a β-far sibling of w under f_G.

Assume that $\mathcal{F}_{P,n,2m}$ is δ_1-pseudorandom. We construct an attacker $\mathcal{B}$ who breaks the $(\delta_2, \frac{1}{2} - \mu)$ one-wayness of $\mathcal{F}_{P,n,2m}$. 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 $r = f_{G,P}(w) \oplus y$.
   (Think of r as representing the set of indices for which y and the image of w disagree.)
2. *Fail*, if m_0 the number of 0's in r is smaller than $m - m^{2/3}$ or larger than $m + m^{2/3}$.

[6] We will later reduce pseudorandomness to one-wayness as well.

3. Let I_0 be the set of the first $\min(m_0, m)$ 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 $m(1+\gamma) - 3m^{2/3}$ 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 4.3. $\Pr_{G \xleftarrow{R} \mathcal{G}_{n,2m,d}, y \xleftarrow{R} \{0,1\}^{2m}}[\mathcal{B}(G,y) \textit{ does not fail}] > \delta' - 2^{-\Omega(n^{1/3})}$.

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 $w \xleftarrow{R} \{0,1\}^n$, and a random string $r \xleftarrow{R} \{0,1\}^{2m}$. 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 Hamming weight of r is $m/2 \pm m^{2/3}$;

2. $\mathcal{A}$ outputs a β-far collision z;

3. The Hamming weight of w is $n/2 \pm n^{2/3}$;

4. $P(z_{S_i}) = y_i$ for at least $m(1+\gamma) - 3m^{2/3}$ of indices $i \in [2m]$.

By a Chernoff bound, Event (1) happens with probability $1 - 2^{-\Omega(n^{1/3})}$. Fix some r which satisfies (1) and let $m_1 \in m \pm m^{2/3}$ be the Hamming weight of r. Now, w is a random string and $G_0 \xleftarrow{R} \mathcal{G}_{n,m,d}$, hence, $\mathcal{A}$ is invoked on the "right" probability distribution and (2) happens with probability δ'. By a Chernoff bound, (3) happens with all but probability $2^{-\Omega(n^{1/3})}$. Therefore, by union bound, (1–3) and happen simultaneously with probability $\delta' - 2^{-\Omega(n^{1/3})}$. Fix some w and G_0 which satisfy (2) and (3), and let us move to (4).

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 indices $i \in I_0$. Recalling that $|I_0| \geq m - m^{2/3}$, 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 $G_1 \xleftarrow{R} \mathcal{G}_{n,m_1,d}$). 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 tuples 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 3.1 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,

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

By summing all the error terms, we derive the claim. □

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 4.4. *Assume that ε and γ satisfy Eq. 3, then there exists a constant $\mu = \mu(P, \varepsilon)$ such that the following holds. With probability $1 - 2^{-\Omega(n^{1/3})}$ 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 $m(1+\gamma) - 3m^{2/3}$ 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

$$2(1 - \varepsilon)D_2(\frac{1}{2} - \gamma \| \eta) > 1. \tag{8}$$

To see that such μ exists, we make the following observations. First, for $\mu = 0$ the conditions are satisfied. Indeed, η can be taken to be $\frac{1}{2}$ (recall that $s_P(\frac{1}{2}) = \frac{1}{2}$) and Eq. 8 simplifies to $2(1 - \varepsilon)H_2(\frac{1}{2} - \gamma) > 1$ which follows from Eq. 3. Now, since s_P is a continuous function, and the LHS of Eq. 8 is also continuous in η, we conclude that both conditions also hold for sufficiently small constant $\mu > 0$.

Let us condition on the event that x is typical (as in Lemma 3.1), which, by a Chernoff bound, happens with all but $2^{-\Omega(n^{1/3})}$ probability. Fix some string z for which $\Delta(x, z) \in (\frac{1}{2} \pm \mu)$. For a random d-size tuple S we have, by Lemma 3.1, 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 tuple 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 $2m - (m(1+\gamma) - 3m^{2/3}) = (1-\gamma)m + 3m^{2/3}$ 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, by Eq.8, is upper-bounded by $2^{-\Omega(n)}$. □

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 4.4; 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 $\delta' - \delta_1 - 2^{-\Omega(n^{1/3})}$. Indeed, (1) happens with all but probability $2^{-\Omega(n^{1/3})}$ (due to Claim 4.4), and (2) happens with probability $\delta' - 2^{-\Omega(n^{1/3})} - \delta_1$ due to Claim 4.3 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 transform δ-secure β-RTCR with shrinkage factor of $1-\varepsilon$ and constant output locality into a (standard) TCR with constant shrinkage factor ε', constant input locality, and constant output locality. Interestingly, we can do this without increasing the algebraic degree. Formally, we prove the following theorem.

Theorem 5.1. *For every constant $\varepsilon \in (0,1)$ there exist universal constants $\delta, \beta \in (0,1)$ for which any δ-secure β-RTCR $\mathcal{H}$ with shrinkage factor of $1-\varepsilon$ and constant output locality can be transformed into a TCR $\mathcal{H}'$ with shrinkage factor of $1-\varepsilon/4$, constant input locality, constant output locality and the same algebraic degree as $\mathcal{H}$. Furthermore, one can obtain an arbitrary constant shrinkage factor of ε' at the expense of further increasing the input and output localities to a larger constant (which grows exponentially in $\log(\varepsilon')/\log(1-\varepsilon)$).*

The proof relies on a sequence of transformations (described in Sections 5.1–5.2) in which we gradually amplify each of the parameters of the underlying collection while keeping the output locality constant.[7] We refer to such transformations as *local*. Finally, we observe that once constant output locality and constant shrinkage factor are achieved, constant input locality can be also guaranteed (with a minor loss in the shrinkage).

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

5.1 Standard Transformations

We begin with two standard transformations.

Claim 5.2 (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 with the same shrinkage, output locality, input locality and algebraic degree as $\mathcal{H}$.*

Proof. Let $(\mathcal{A}_1, \mathcal{A}_2)$ be an adversary that contradicts the claim. We construct an adversary $\mathcal{B}$ that contradicts the hypothesis. Given a random RTCR challenge $x \xleftarrow{R} \{0,1\}^n, h_k \xleftarrow{R} \mathcal{H}$, the adversary $\mathcal{B}$ computes $\mathcal{A}_1(1^n)$ and obtains a target y and a state r. Then, $\mathcal{B}$ invokes $\mathcal{A}_2$ with the function $h'_{k,x\oplus y}$, state information r, and target y. Finally, $\mathcal{B}$ outputs $x' = y' \oplus y \oplus x$ where y' is the output of $\mathcal{A}_2$. The claim follows by noting that if y and y' is a β-far collision under $h'_{k,x\oplus y}$ then x and x'

[7] In fact, these transformations also preserve constant input locality.

is a β-far collision under h_k. Clearly, $\mathcal{H}'$ has the same shrinkage, output locality, input locality and algebraic degree as $\mathcal{H}$. □

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 [23].

Claim 5.3 (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*

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

Furthermore, the construction is local: if $\mathcal{H}$ has an output (resp., input) locality of $d = O(1)$, then the new family $\mathcal{H}^t$ has an output (resp., input) locality of $d^t = O(1)$.

A proof can be found in [23] (see also [9]).

5.2 Hardness Amplification

We move on to amplify the hardness parameter δ from constant to negligible at the expense of slightly increasing the distance parameter. Our construction is based on a simple direct product.[8]

Lemma 5.4 (Hardness Amplification)**.** *Let $\mathcal{H} = \{h_k : \{0,1\}^n \to \{0,1\}^{\varepsilon n}\}$ be δ-secure β-TCR with key sampler $\mathcal{K}$. Then, for every polynomial $t = t(n)$ and every $\gamma > \delta$, the t-direct product collection $\mathcal{H}'$ defined via*

$$h'_{(k_1,\ldots,k_t)} : (x_1,\ldots,x_t) \mapsto (h_{k_1}(x_1),\ldots,h_{k_t}(x_t)), \qquad \text{where } x_i \in \{0,1\}^n, k_i \xleftarrow{R} \mathcal{K}(1^n),$$

is $(2^{-t(\gamma-\delta)^2} + \mathrm{neg}(n))$-secure $(\beta+\gamma)$-TCR with the same shrinkage factor, output locality, input locality and algebraic degree as $\mathcal{H}$.

By taking $t = n$ and letting γ be a constant which is strictly larger than δ, we reduce the security error to negligible.

Proof. We will need the following simple observation: Consider a pair of stings $\vec{x} = (x_1,\ldots,x_t) \in (\{0,1\}^n)^t$ and $\vec{y} = (y_1,\ldots,y_t) \in (\{0,1\}^n)^t$ which are $(\beta+\gamma)$ far in Hamming distance. Then, by an averaging argument, it holds that $\Delta(x_i, y_i) > \beta$ for at least γ-fraction of the i's.

We can now prove the lemma. Let $\mathcal{A} = (\mathcal{A}_1, \mathcal{A}_2)$ be an adversary that finds $(\beta+\gamma)$-far collisions under $\mathcal{H}'$ with probability δ'. Using the above observation it follows that

$$\delta' \leq \Pr_{\substack{\vec{k}\xleftarrow{R}\mathcal{K}^t(1^n) \\ (\vec{x},R)\xleftarrow{R}\mathcal{A}_1(1^n)}} [\mathcal{A}_2(\vec{k},\vec{x},R) = \vec{y} \text{ s.t. } |\{i : \Delta(x_i,y_i) > \beta \wedge h_{k_i}(x_i) = h_{k_i}(y_i)\}| \geq \gamma n],$$

where $\vec{k} = (k_1,\ldots,k_t)$, $\vec{x} = (x_1,\ldots,x_t)$ and $\vec{y} = (y_1,\ldots,y_t)$. That is, given $t(n)$ independent samples of $\mathcal{H}$, the adversary $\mathcal{A}$ finds β-far collisions on γ fraction of them with probability δ'. A general *threshold direct product theorem* of Impagliazzo and Kabanets [18, Theorem 5.2] shows that, in this case, the advantage δ' is upper-bounded by $2^{-tD_2(\gamma\|\delta)} + \mathrm{neg}(n) < 2^{-t(\gamma-\delta)^2} + \mathrm{neg}(n)$. The lemma follows. □

[8] In the conference version we used a more complicated solution based on sparse *Distance Amplifiers* (which have the property of mapping any pair (x,x') of far-apart inputs to a pair of far apart outputs (y,y')). We thank the anonymous referee for pointing out the current, simpler variant.

5.3 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 an infinite sequence of $m(n) \times n$ binary matrices $\mathcal{M}_{m(n)\times n} = \{M_n\}_{n\in\mathbb{N}}$ is a *low-density parity check code* with distance β if for every n the matrix M_n has dual distance of β and M_n is *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. To make our construction efficient we need an LDPC $\mathcal{M}_{m(n)\times n}$ whose n-th member can be computed in $\mathrm{poly}(n)$ time. Such a construction is given in [12].

Proposition 5.5. *For every $\varepsilon \in (0,1)$ there exists a sequence $\mathcal{M}_{\varepsilon n\times n} = \{M_n\}_{n\in\mathbb{N}}$ of $\beta(\varepsilon)$-LDPC for some $\beta = \varepsilon/\mathrm{polylog}(1/\varepsilon)$. Furthermore, there exists an efficient algorithm that given 1^n outputs the matrix M_n in $\mathrm{poly}(n)$-time.*

Proof. For every constant ε, Theorem 7.1 of Capalbo et al. [12] provides an explicit (efficiently computable) family of unbalanced bipartite graphs $\mathcal{G}$ with n "column" nodes and $m = \varepsilon n$ "row" nodes with constant degree on each side, such that each set of at most βn column nodes has almost full expansion of $0.99d\beta n$ row nodes where d is the degree of the column nodes. Sipser and Spielman [27] showed that the adjacency matrix of such a graph is β-LDPC. □

Lemma 5.6 (β-TCR to TCR)**.** *Let $\varepsilon' < \varepsilon$ and let $\mathcal{M}_{\varepsilon' n\times n} = \{M_n\}$ be a 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*

$$\mathcal{H}' = \left\{h'_k\right\} \quad h'_k = (h_k(x), M_n x), \qquad \text{where } k \xleftarrow{R} (\mathcal{K}(1^n).$$

Then, $\mathcal{H}'$ is δ-secure TCR with shrinkage factor of $1-\varepsilon+\varepsilon'$. Furthermore, the transformation is local and the algebraic degree of $\mathcal{H}'$ is the same as the degree of $\mathcal{H}$.

Proof. First, observe that the above transformation is local since $\mathcal{M}$ is d-sparse for $d = O(1)$. Specifically, both the input locality and the output locality grow by an additive factor of d. Moreover, since M is used as a linear operator the algebraic degree of $\mathcal{H}'$ is equal to the algebraic degree of $\mathcal{H}$. We move on to prove the security. Let $\mathcal{A}_2$ be a TCR adversary that, given $(x,r) \xleftarrow{R} \mathcal{A}_1(1^n)$ and $h'_k \xleftarrow{R} \mathcal{H}'$, finds a collision x' with x under h'_k with probability $\delta_{\mathcal{A}}$. We claim that such a collision must be β-far and so, by our assumption, $\delta_{\mathcal{A}} < \delta$. Indeed, by definition, $h_k(x) = h_k(x')$ and, in addition, $M_n x = M_n x'$. It follows that the difference vector $x \oplus x'$ is a non-zero vector in the kernel of M_n, and therefore $x \oplus x'$ has a relative weight of β. The lemma follows. □

Remark 5.7 (Using LDPC ensembles)**.** *Lemma 5.6 easily generalizes to the case where $\mathcal{M}_{\varepsilon' n\times n}$ forms an* ensemble *of β-LDPC's, i.e., there exists an efficient sampler that given 1^n samples a*

sparse $(\varepsilon' n \times n)$ *binary matrix that, with probability* $1-\delta'$*, has a dual distance of* β*. In this case, we modify the key sampler of* $\mathcal{H}'$ *to sample a matrix* M *from* $\mathcal{M}_{\varepsilon' n \times n}$ *together with a key* $k \xleftarrow{R} \mathcal{K}(1^n)$ *and let* $h'_{k,M}(x) = (h_k(x), Mx)$*. It is not hard to show that the resulting collection is* $(\delta+\delta')$*-secure TCR. While our (theoretical) results can be derived without this extension (based on the explicit LDPC's from Proposition 5.5), the use of ensembles may be beneficial in terms of concrete efficiency. Indeed, most practical LDPC codes (e.g., based on random sparse matrices) yield efficiently samplable ensembles of LDPC's.*

5.4 Reducing the Input Locality

We next show how to reduce the input locality of a TCR with constant output locality and constant shrinkage factor.

Lemma 5.8 (Reducing Input Locality)**.** *Assume that there exists a TCR* $\mathcal{H}$ *with output locality* d *and shrinkage factor* ε*. Then, for every* $\alpha \in (0,1)$ *there exists a TCR* $\mathcal{H}'$ *with output locality* d*, input locality* $d/(\varepsilon \cdot \alpha)$ *and shrinkage factor* $\varepsilon/(1-\alpha)$*.*

Proof. Let us assume, without loss of generality, that for every function $h_k : \{0,1\}^n \to \{0,1\}^{\varepsilon n}$ in the collection $\mathcal{H}$, the input variables $(x_1, \ldots, x_n)$ are ordered according to their input locality. Namely, if x_i affects t_i outputs, then $t_1 \leq t_2 \leq \cdots \leq t_n$.[9] We define $h'_k : \{0,1\}^{n'} \to \{0,1\}^{n' \cdot \varepsilon/(1-\alpha)}$ by mapping an $n' = n(1-\alpha)$-bit string x to the value $h_k(x, 0^{\alpha n}) \in \{0,1\}^{n' \cdot \varepsilon/(1-\alpha)}$. Let $\mathcal{H}'$ denote the collection $\{h'_k : h_k \in \mathcal{H}\}$ equipped with the key-sampler $\mathcal{K}'(1^{n'}) = \mathcal{K}(1^n)$ where $\mathcal{K}$ is the key-sampler of $\mathcal{H}$.

Observe that for every fixed index k, the average input locality of h_k is at most cd, and therefore, by Markov's inequality, the fraction of inputs whose locality is larger than cd/α is at most α. It follows that the input locality of h'_k is at most $cd/\alpha = O(1)$, as claimed.

In addition, it is not hard to prove that $\mathcal{H}'$ is a TCR. Specifically, given a TCR-adversary $\mathcal{A}' = (\mathcal{A}'_1, \mathcal{A}'_2)$ which breaks $\mathcal{H}'$ with success probability of $\delta(n')$, we define an adversary $\mathcal{A} = (\mathcal{A}_1, \mathcal{A}_2)$ which breaks $\mathcal{H}$ with similar success probability $\delta(n') = \delta(n(1-\alpha))$. The target specifier $\mathcal{A}_1(1^n)$ computes $(x', r) \xleftarrow{R} \mathcal{A}'_1(1^n)$ and outputs the target string $x = (x', 0^{\alpha n})$ and state information r. The collision finder $\mathcal{A}_2(k, x, r)$ computes $y' \xleftarrow{R} \mathcal{A}'_2(k, x', r)$ and outputs the string $y = (y', 0^{\alpha n})$. The claim follows by noting that for every index k if the pair (x', y') forms a collision under h'_k, then the padded pair (x, y) forms a collision under h_k. □

5.5 Proof of Theorem 5.1

Let $\varepsilon > 0$ be the given shrinkage parameter. Let $\varepsilon_{\text{LDPC}} = \varepsilon/2$ and let $\mathcal{M}_{\varepsilon_{\text{LDPC}} n \times n}$ be an efficient β_{LDPC}-LDPC for some constant $\beta_{\text{LDPC}} > 0$ whose existence is promised by Proposition 5.5. We will show how to obtain TCR with shrinkage $1 - \varepsilon/4$ from any δ-secure β-RTCR $\mathcal{H}$ with shrinkage factor $1-\varepsilon$, where $\beta + \delta < \beta_{\text{LDPC}}$.

Start by transforming $\mathcal{H}$ into δ-secure β-TCR with shrinkage factor of $1-\varepsilon$ via Claim 5.2. Then, amplify the security security error to negligible by employing Lemma 5.4 with $t = n$ and $\gamma = \beta_{\text{LDPC}} - \beta > \delta$. This yields a $\text{neg}(n)$-secure β_{LDPC}-TCR with shrinkage factor of $1-\varepsilon$.

[9] Since the input locality of every variable can be computed efficiently [3, Chp. 2] (regardless of the actual representation of the collection index k), one can always efficiently permute the order of inputs to guarantee this property. Furthermore, the permuted function is still TCR as the permutation is efficiently computable.

Next, apply distance amplification (Lemma 5.6) with $\mathcal{M}_{\varepsilon_{\mathrm{LDPC}}n\times n}$ and obtain a TCR with standard security, shrinkage factor of $1-\varepsilon+\varepsilon_{\mathrm{LDPC}} = 1-\varepsilon/2$ and constant output locality. Finally, reduce the input locality via Lemma 5.8 (instantiated with $\alpha = \frac{\varepsilon}{4}$) at the expense of increasing the shrinkage factor to $1-\varepsilon/4$. Observe that all these transformations preserve the algebraic degree and so we derive the main part of the theorem. The "furthermore" part now follows immediately from Claim 5.3. □

6 Putting It Together

In this section we combine the results of the previous sections and derive the main theorem and its applications.

6.1 Locally Computable UOWHFs

Theorem 6.1. *There exist some universal constants $\beta > 0$ and $0 < \gamma < \frac{1}{2}$ such that for every (β,γ)-good predicate P the following holds. Assuming that $\mathcal{F}_{P,n,cn}$ is $1/c$-pseudorandom or that $\mathcal{F}_{P,n,c'n^3}$ is one-way for some constants $c = c(P)$ and $c' = c'(P)$, there exists a locally computable UOWHF $\mathcal{H}$ with constant shrinkage factor. Moreover, the algebraic degree of $\mathcal{H}$ is equal to the degree of the predicate P.*

By Claim 5.3, one can further reduce the shrinkage factor to an arbitrary constant $\varepsilon' \in (0,1)$ at the expense of increasing the output/input locality and degree to a larger constant.

Proof. Fix some $\varepsilon > 0$. By Theorem 5.1 it suffices to prove the existence of δ-secure β-RTCR with shrinkage factor of $1-\varepsilon$ and constant output locality, for some universal constant β and δ. Let γ be a constant for which Eq. 3 is satisfied with ε. (E.g., for $\varepsilon = 0.3$ it suffices to let $\gamma = 0.46$.) Let P be a (β,γ)-good predicate. By Corollary 4.2, there exists a constant $k = k(P)$ for which $\mathcal{F}_{P,n,(1-\varepsilon)n}$ is δ-secure β-RTCR assuming that $\mathcal{F}_{P,n,kn}$ is $\delta/3$-pseudorandom. Taking $c = \max(k, 3/\delta)$ completes the proof of the first part of the theorem. To prove the second ("one-wayness") part, we employ Corollary 6.2 of [4] which asserts that for predicates with a sensitive coordinate (as in property 2 of Def. 3.2), $1/c$-pseudorandomness of $\mathcal{F}_{P,n,cn}$ is implied by the one-wayness of $\mathcal{F}_{P,n,c'n^3}$ for a constant c' which depends on the constant c and (the locality of) P. □

We suggest instantiating the theorem with the predicate MST_{d_1,d_2} defined in Eq. 7, which XORs together a d_1-ary XOR with a d_2-ary AND (over d_1+d_2 distinct inputs). Recall that Lemma 3.3 guarantees that for sufficiently large odd d_1 and every $d_2 \geq 2$ the predicate MST_{d_1,d_2} satisfies the goodness condition needed in Theorem 6.1. Concretely, the results of [11] support the following assumption:

Assumption 6.2. *For every $d \geq 3$ the collection $\mathcal{F}_{\mathsf{MST}_{d,2},n,cn}$ is $1/c$-pseudorandom for arbitrary large constant c.*

In fact, based on our existing knowledge, it seems that the above assumption holds even for $c = n^{\varepsilon}$ for some small constant ε. Alternatively, one can start with one-wayness as captured by the following assumption.

Assumption 6.3. *For all sufficiently large constants d_1 and d_2 the collection $\mathcal{F}_{\mathsf{MST}_{d_1,d_2},n,cn^3}$ is one-way for arbitrary large constant c.*

Again, based on known attacks, one may conjecture that a much stronger version of the assumption holds. Namely, that for every constant c and all sufficiently large constants $d_1, d_2 > d(c)$ the collection $\mathcal{F}_{\mathsf{MST}_{d_1,d_2},n,n^c}$ is one-way.[10] We further mention that the latter conjecture is supported by the results of [13].

Combined with Theorem 6.1, any of the above assumptions implies the existence of locally computable UOWHF with constant shrinkage factor, and so Theorem 1.1 follows.

6.2 Optimizing the Output Locality

One can further optimize the output locality (while preserving constant input locality and linear shrinkage) via the AIK-compiler [7].

Proposition 6.4. *If there exists a UOWHF $\mathcal{H}$ with constant shrinkage factor constant output locality and constant input locality, then there exists a UOWHF $\hat{\mathcal{H}}$ with constant shrinkage factor, constant input locality, and output locality of 4. Moreover, if the algebraic degree of $\mathcal{H}$ is 2 then the output locality of $\hat{\mathcal{H}}$ is 3.*

Proof. In [7] it is shown that, for some small (universal) constant c, any UOWHF family $\mathcal{H} : \{0,1\}^n \to \{0,1\}^{m(n)}$ that each of its output bits is computable by an $\mathbf{NC^1}$ circuit of size $l(n)$ can be transformed into a UOWHF $\hat{\mathcal{H}} : \{0,1\}^{n+m(n)\cdot l(n)^c} \to \{0,1\}^{m(n)+m(n)\cdot l(n)^c}$ with output locality 4. Moreover, in the special case where the degree of $\mathcal{H}$ is 2 the output locality of $\hat{\mathcal{H}}$ is 3. (Originally, these implications are proven for collision resistance hash function, though the proof easily generalizes to the case of UOWHF as well.)

Typically in [7], $l(n)$ is superconstant and so the shrinkage $n - m(n)$ of the resulting UOWHF $\hat{\mathcal{H}}$ is only *sublinear* in its input length $n + m(n) \cdot l(n)^c$. However, when $\mathcal{H}$ has a constant output locality each output bit is computable by a constant size circuit and so $l(n) = O(1)$. In this case, linear shrinkage is preserved, i.e., since $n - m(n) = \Theta(n)$ and $l(n) = O(1)$, the shrinkage of the resulting UOWHF which is $n - m(n) = \Theta(n)$ is still linear in its input length $n + O(n) + O(n) \cdot O(1)$. Finally, we note that, when $\mathcal{H}$ enjoys constant output locality, the above transformation preserves constant input locality as well. □

As observed by Goldreich [16] functions with output locality 2 are efficiently invertible (due to the easiness of 2-SAT). Therefore, one cannot hope for output locality smaller than 3. Indeed, such an optimal locality can be achieved based on Assumption 6.2.

Corollary 6.5. *Under Assumption 6.2 there exists a UOWHF with output locality 3, constant input locality and constant shrinkage factor.*

Proof. Since $\mathsf{MST}_{d,2}$ is a degree 2 predicate, Theorem 6.1 yields a degree-2 locally computable UOWHF with constant shrinkage factor. The corollary now follows from Proposition 6.4. □

6.3 Applications

As mentioned in the introduction locally computable UOWHFs with constant shrinkage factor also allows us to optimize the sequential complexity of cryptography. In the following we measure the time complexity $T(n)$ of a collection $\mathcal{H}$ of UOWHF, as the the sum of the sampling time and the

[10] Known attacks of [22] imply that $d(c) \geq c/2$.

evaluation time. Namely, $T(n)$ measures the time which takes to sample $h \xleftarrow{R} \mathcal{H}$ and evaluate it on a given n-bit input x.

Proposition 6.6 (Fast UOWHFs and Signatures)**.** *Assume the existence of a UOWHF $\mathcal{H}$ with constant shrinkage factor and constant output locality. Then:*

1. *For every constant $\varepsilon > 0$ there exists a UOWHF $\hat{\mathcal{H}} : \{0,1\}^n \to \{0,1\}^{n^\varepsilon}$ with polynomial shrinkage factor which is computable in linear-time in the RAM model. Furthermore, each function $h \in \hat{\mathcal{H}}$ is described by a string of length $O(n^\varepsilon)$.*

2. *There exists a digital signature scheme whose time complexity (both for signing and for verifying) is linear in the message length in the RAM model.*

Proof. The proof follows the outline of [19] and is given here for completeness. Fix $\varepsilon > 0$, and let $\mathcal{H} = \{h_k\}$ be the underlying locally computable UOWHF whose shrinkage factor is constant. Without loss of generality (by Claim 5.3), we may assume that the collection shrinks n-bit strings to $(n/2)$-bit strings. Let us denote the RAM complexity of the key-sampling algorithm by $O(n^c)$ for some constant $c > 0$ and assume, without loss of generality, that $c > 1/\varepsilon$. We further assume that the key of the collection k is simply a canonic description of the (linear-size) $\mathbf{NC^0}$ circuit which computes the function h_k. (This can be always guaranteed by learning a canonic description of h_z – see [3, Proposition 2.4.1].) Observe that, given k and x, the value of $h_k(x)$ can be computed in linear-time by a RAM machine.

To reduce the time complexity of the sampling procedure (as well as the length of the key) we use direct-product collection $\mathcal{H}' = \{h'_k\}$, defined by a single key $k \xleftarrow{R} \mathcal{K}(1^{n^{1/c}})$ and $h'_k(x) = (h_k(x^1), \ldots, h_k(x^t))$ where $x \in \{0,1\}^n$ is partitioned to $t = n^{1-1/c}$ blocks each of size $n^{1/c}$. It is not hard to verify that the resulting collection is still a UOWHF with shrinkage factor 2, and that the total RAM complexity of sampling a key and evaluating the function is $t(n) = O(n)$. Furthermore, the description length of the key is $n^{1/c}$.

As in Claim 5.3, we can amplify the shrinkage by composing $(1-\varepsilon)\log n$ functions from $\mathcal{H}$ where the i-th function h_{k_i} shrinks $n/2^i$ bits to $n/2^{i+1}$ bits and $k_i \xleftarrow{R} \mathcal{K}(1^{n/2^i})$. The resulting collection $\hat{\mathcal{H}}$ is a UOWHF (see [23]) which shrinks n bits to n^ε bits. The RAM-complexity of the i-th level is $t(n/2^i)$, and so the overall complexity is $T(n) = \sum_{i=0}^{(1-\varepsilon)\log n} t(n/2^i) = O(n)$, the description of the key is of length $O(n^\varepsilon)$. This completes the proof of the first item.

We move on to the proof of the second item. Let (G, S, V) be a standard signature scheme (whose existence follows from the the existence of UOWHF [23]). Assume that the complexity of verification and signature is $O(n^b)$ for some constant $b > 0$. We define a new signature scheme (G, S', V') by employing the Naor-Yung transformation instantiated with the aforementioned linear-time computable UOWHF $\mathcal{H} : \{0,1\}^n \to \{0,1\}^{n^\varepsilon}$ whose keys are of length $O(n^\varepsilon)$ where $\varepsilon = 1/b$. Namely, to sign an n-bit message m, the new signing algorithm $S'_{\mathsf{sk}}(m)$ samples a key k for $\mathcal{H}$ and outputs $(k, S_{\mathsf{sk}}(k, h_k(m)))$. To verify whether a tag (k, β) is a valid signature of a document $m \in \{0,1\}^n$ use V_{pk} to check whether β is a valid signature of $h_k(m)$ under the original scheme. The overall complexity in both cases is $O(n)$ (for the hashing) plus $O(n^{b\varepsilon}) = O(n)$ (for applying the original signing/verifying algorithm on an input of length $O(n^\varepsilon)$). This completes the proof of the second item. □

We mention that [19] construct signatures which are linear-time computable in the (stronger)

circuit model, at the expense of using a stronger assumption (namely, that random local functions are exponentially one-way.)

Acknowledgement. We thank Uri Feige and Danny Vilenchik for valuable discussions. We are also grateful to the anonymous reviewers whose suggestions helped to improve this writeup, and in particular for simplifying the hardness amplification step (Lemma 5.4).

References

[1] D. Achlioptas and F. Ricci-Tersenghi. On the solution-space geometry of random constraint satisfaction problems. In J. M. Kleinberg, editor, *38th ACM STOC*, pages 130–139. ACM Press, May 2006.

[2] M. Alekhnovich, E. A. Hirsch, and D. Itsykson. Exponential lower bounds for the running time of DPLL algorithms on satisfiable formulas. *J. Autom. Reasoning*, 35(1-3):51–72, 2005.

[3] B. Applebaum. *Cryptography in Constant Parallel Time*. Phd thesis, Technion, Israel Institute of Technology, August 2007.

[4] B. Applebaum. Pseudorandom generators with long stretch and low locality from random local one-way functions. *SIAM Journal on Computing*, 42(5), 2013.

[5] B. Applebaum, A. Bogdanov, and A. Rosen. A dichotomy for local small-bias generators. In R. Cramer, editor, *TCC 2012*, volume 7194 of *LNCS*, pages 600–617. Springer, Mar. 2012.

[6] B. Applebaum, Y. Ishai, and E. Kushilevitz. Computationally private randomizing polynomials and their applications. *Journal of Computational Complexity*, 15(2):115–162, 2006.

[7] B. Applebaum, Y. Ishai, and E. Kushilevitz. Cryptography in NC^0. *SIAM Journal on Computing*, 36(4):845–888, 2006.

[8] B. Applebaum, Y. Ishai, and E. Kushilevitz. Cryptography with constant input locality. *Journal of Cryptology*, 22(4):429–469, 2009.

[9] M. Bellare and P. Rogaway. Collision-resistant hashing: Towards making UOWHFs practical. In B. S. Kaliski Jr., editor, *CRYPTO'97*, volume 1294 of *LNCS*, pages 470–484. Springer, Aug. 1997.

[10] A. Bogdanov and Y. Qiao. On the security of goldreich's one-way function. In *Proc. of 13th RANDOM*, pages 392–405, 2009.

[11] A. Bogdanov and A. Rosen. Input locality and hardness amplification. In Y. Ishai, editor, *TCC 2011*, volume 6597 of *LNCS*, pages 1–18. Springer, Mar. 2011.

[12] M. R. Capalbo, O. Reingold, S. P. Vadhan, and A. Wigderson. Randomness conductors and constant-degree lossless expanders. In *34th ACM STOC*, pages 659–668. ACM Press, May 2002.

[13] J. Cook, O. Etesami, R. Miller, and L. Trevisan. Goldreich's one-way function candidate and myopic backtracking algorithms. In O. Reingold, editor, *TCC 2009*, volume 5444 of *LNCS*, pages 521–538. Springer, Mar. 2009.

[14] S. O. Etesami. Pseudorandomness against depth-2 circuits and analysis of goldreich's candidate one-way function. Technical Report EECS-2010-180, UC Berkeley, 2010.

[15] O. Goldreich. *Foundations of Cryptography: Basic Tools*. Cambridge University Press, 2001.

[16] O. Goldreich. Candidate one-way functions based on expander graphs. In *Studies in Complexity and Cryptography*, pages 76–87. 2011. Available as TR00-090 of ECCC.

[17] I. Haitner, T. Holenstein, O. Reingold, S. P. Vadhan, and H. Wee. Universal one-way hash functions via inaccessible entropy. In H. Gilbert, editor, *EUROCRYPT 2010*, volume 6110 of *LNCS*, pages 616–637. Springer, May 2010.

[18] R. Impagliazzo and V. Kabanets. Constructive proofs of concentration bounds. In *APPROX-RANDOM*, pages 617–631, 2010.

[19] Y. Ishai, E. Kushilevitz, R. Ostrovsky, and A. Sahai. Cryptography with constant computational overhead. In R. E. Ladner and C. Dwork, editors, *40th ACM STOC*, pages 433–442. ACM Press, May 2008.

[20] D. Itsykson. Lower bound on average-case complexity of inversion of goldreich's function by drunken backtracking algorithms. In *Computer Science - Theory and Applications, 5th International Computer Science Symposium in Russia*, pages 204–215, 2010.

[21] R. Miller. Goldreich's one-way function candidate and drunken backtracking algorithms. Distinguished major thesis, University of Virginia, 2009.

[22] E. Mossel, A. Shpilka, and L. Trevisan. On e-biased generators in NC0. In *44th FOCS*, pages 136–145. IEEE Computer Society Press, Oct. 2003.

[23] M. Naor and M. Yung. Universal one-way hash functions and their cryptographic applications. In *21st ACM STOC*, pages 33–43. ACM Press, May 1989.

[24] S. K. Panjwani. An experimental evaluation of goldreich's one-way function. Technical report, IIT, Bombay, 2001.

[25] P. Rogaway and T. Shrimpton. Cryptographic hash-function basics: Definitions, implications, and separations for preimage resistance, second-preimage resistance, and collision resistance. In B. K. Roy and W. Meier, editors, *FSE 2004*, volume 3017 of *LNCS*, pages 371–388. Springer, Feb. 2004.

[26] J. Rompel. One-way functions are necessary and sufficient for secure signatures. In *22nd ACM STOC*, pages 387–394. ACM Press, May 1990.

[27] M. Sipser and D. A. Spielman. Expander codes. *IEEETIT: IEEE Transactions on Information Theory*, 42, 1996.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il