Universal One-Way Hash Functions via Inaccessible Entropy

Iftach Haitner*and Thomas Holenstein† and Omer Reingold‡
and Salil Vadhan§ and Hoeteck Wee¶

March 5, 2010

Abstract

This paper revisits the construction of Universal One-Way Hash Functions (UOWHFs) from any one-way function due to Rompel (STOC 1990). We give a simpler construction of UOWHFs, which also obtains better efficiency and security. The construction exploits a strong connection to the recently introduced notion of *inaccessible entropy* (Haitner et al. STOC 2009). With this perspective, we observe that a small tweak of any one-way function f is already a weak form of a UOWHF: Consider $F(x, i)$ that outputs the i-bit long prefix of $f(x)$. If F were a UOWHF then given a random x and i it would be hard to come up with $x' \neq x$ such that $F(x, i) = F(x', i)$. While this may not be the case, we show (rather easily) that it is hard to sample x' with almost full entropy among all the possible such values of x'. The rest of our construction simply amplifies and exploits this basic property.

With this and other recent works, we have that the constructions of three fundamental cryptographic primitives (Pseudorandom Generators, Statistically Hiding Commitments and UOWHFs) out of one-way functions are to a large extent unified. In particular, all three constructions rely on and manipulate computational notions of entropy in similar ways. Pseudorandom Generators rely on the well-established notion of pseudoentropy, whereas Statistically Hiding Commitments and UOWHFs rely on the newer notion of inaccessible entropy.

Keywords: computational complexity, cryptography, hashing, target collision-resistance, one-way functions

*Microsoft Research, New England Campus. E-mail: iftach@microsoft.com.

†

‡Microsoft Research - Silicon Valley and the Weizmann Institute of Science. E-mail:omreing@microsoft.com. Supported by US-Israel BSF grant 2006060.

§School of Engineering and Applied Sciences and Center for Research on Computation and Society, Harvard University. E-mail: salil@seas.harvard.edu. Supported by NSF grant CNS-0831289 and US-Israel BSF grant 2006060.

¶Queens College, CUNY. E-mail: hoeteck@cs.qc.cuny.edu. Supported in part by PSC-CUNY Award #6014939 40.

1 Introduction

Universal one-way hash functions (UOWHFs), as introduced by Naor and Yung [NY], are a weaker form of collision-resistant hash functions. The standard notion of collision resistance requires that given a randomly chosen function $f \xleftarrow{R} \mathcal{F}$ from the hash family, it is infeasible to find any pair of distinct inputs x, x' such that $f(x) = f(x')$. UOWHFs only require *target collision resistance*, where the adversary must specify one of the inputs x before seeing the description of the function f. Formally:

Definition 1.1. *A family of functions* $\mathcal{F}_k = \{F_z : \{0,1\}^{n(k)} \to \{0,1\}^{m(k)}\}_{z \in \{0,1\}^k}$ *is a family of* universal one-way hash functions (UOWHFs) *if it satisfies:*

1. *Efficiency: Given* $z \in \{0,1\}^k$ *and* $x \in \{0,1\}^{n(k)}$, $F_z(x)$ *can be evaluated in time* $\mathrm{poly}(n(k), k)$.
2. *Shrinking:* $m(k) < n(k)$.
3. *Target Collision Resistance: For every probabilistic polynomial-time adversary* A*, the probability that* A *succeeds in the following game is negligible in* k*:*
   (a) *Let* $(x, \mathsf{state}) \leftarrow A(1^k) \in \{0,1\}^{n(k)} \times \{0,1\}^*$.
   (b) *Choose* $z \xleftarrow{R} \{0,1\}^k$.
   (c) *Let* $x' \xleftarrow{R} A(\mathsf{state}, z) \in \{0,1\}^{n(k)}$.
   (d) A succeeds *if* $x \neq x'$ *and* $F_z(x) = F_z(x')$.

It turns out that this weaker security property suffices for many applications. The most immediate application given in [NY] is *secure fingerprinting*, whereby the pair $(f, f(x))$ can taken as a compact "fingerprint" of a large file x, such that it is infeasible for an adversary, seeing the fingerprint, to change the file x to x' without being detected. More dramatically, Naor and Yung [NY] also showed that UOWHFs can be used to construct secure digital signature schemes, whereas all previous constructions (with proofs of security in the standard model) were based on trapdoor functions (as might have been expected to be necessary due to the public-key nature of signature schemes). More recently, UOWHFs have been used in the Cramer–Shoup encryption scheme [CS] and in the construction of statistically hiding commitment schemes from one-way functions [HNO$^+$, HRVW].

Naor and Yung [NY] gave a simple and elegant construction of UOWHFs from any one-way *permutation*. Subsequently, Rompel [Rom1] gave a much more involved construction to prove that UOWHFs can be constructed from an arbitrary one-way function, thereby resolving the complexity of UOWHFs (as one-way functions are the minimal complexity assumption for complexity-based cryptography, and are easily implied by UOWHFs).[1] While complications may be expected for constructions from arbitrary one-way functions (due to their lack of structure), Rompel's analysis also feels quite ad hoc. In contrast, the construction of pseudorandom generators from one-way functions of [HILL], while also somewhat complex, involves natural abstractions (e.g., pseudoentropy) that allow for modularity and measure for what is being achieved at each stage of the construction.

In this paper, we give simpler constructions of UOWHFs from one-way functions, based on (a variant of) the recently introduced notion of *inaccessible entropy* [HRVW]. In addition, one of the constructions obtains slightly better efficiency and security.

[1] More details of Rompel's proof are worked out, with some corrections, in [Rom2, KK].

1.1 Inaccessible Entropy

For describing our construction, it will be cleaner to work with a variant of UOWHFs where there is a *single* shrinking function $F : \{0,1\}^n \to \{0,1\}^m$ (for each setting of the security parameter k) such that it is infeasible to find collisions with *random inputs*. That is, an adversary A is given a uniformly random $x \xleftarrow{R} \{0,1\}^n$, outputs an x' such that $F(x') = F(x)$, and succeeds if $x' \neq x$.[2] Note that we can assume without loss of generality that $x' = A(x)$ is always a preimage of $F(x)$ (A has the option of outputting x in case it does not find a different preimage); we refer to an algorithm A with this property as an *F-collision finder*.

Our construction is based on an entropy-theoretic view of UOWHFs. The fact that F is shrinking implies that there are many preimages x' available to A. Indeed, if we consider an (inefficient) adversary $A(x)$ that outputs $x' \xleftarrow{R} F^{-1}(F(x))$ and let X be a random variable uniformly distributed on $\{0,1\}^n$, then

$$\mathrm{H}(A(X)|X) = \mathrm{H}(X|F(X)) \geq n - m,$$

where $\mathrm{H}(\cdot|\cdot)$ denotes conditional Shannon entropy. (See Section 2for more definitional details.) We refer to the quantity $\mathrm{H}(X|F(X))$ as the *real entropy of F^{-1}*.

On the other hand, the target collision resistance means that effectively only one of the preimages is accessible to A. That is for every probabilistic polynomial-time F-collision finder A, we have $\Pr[A(X) \neq X] = \mathrm{neg}(n)$, which is equivalent to requiring that:

$$\mathrm{H}(A(X)|X) = \mathrm{neg}(n)$$

for all probabilistic polynomial-time F-collision finders A. (If A can find a collision X' with non-negligible probability, then it can achieve nonnnegligible conditional entropy by outputting X' with probability $1/2$ and outputting X with probability $1/2$.) We refer to the maximum of $\mathrm{H}(A(X)|X)$ over all efficient F-collision finders as the *accessible entropy of F^{-1}*. We stress that accessible entropy refers to an *upper bound* on a form of computational entropy, in contrast to the Håstad et al.'s notion of *pseudoentropy* [HILL].

Thus, a natural weakening of the UOWHF property is to simply require a noticeable gap between the real and accessible entropies of F^{-1}. That is, for every probabilistic polynomial-time F-collision finder A, we have $\mathrm{H}(A(X)|X) < \mathrm{H}(X|F(X)) - \Delta$, for some noticeable Δ, which we refer to as the *inaccessible entropy of F*.

1.2 Our Constructions

Our constructions of UOWHFs have two parts. First, we show how to obtain a function with noticeable inaccessible entropy from any one-way function. Second, we show how to build a UOWHF from any function with inaccessible entropy.

OWFs ⇒ Inaccessible Entropy. Given a one-way function $f : \{0,1\}^n \to \{0,1\}^m$, we show that a random truncation of f has inaccessible entropy. Specifically, we define $F(x,i)$ to be the first i bits of $f(x)$.

To see that this works, suppose for contradiction that F does not have noticeable inaccessible entropy. That is, we have an efficient adversary A that on input (x,i) can sample from the set $S(x,i) = \{x' : f(x')_{1\ldots i} = f(x)_{1\ldots i}\}$ with almost-maximal entropy, which is equivalent to sampling

[2] It is easy to convert any such function F into a standard UOWHF family by defining $F_z(x) = F(z+x)$.

according to a distribution that is statistically close to the uniform distribution on $S(x,i)$. We can now use A to construct an inverter *Inv* for f that works as follows on input y: choose $x_0 \stackrel{R}{\leftarrow} \{0,1\}^n$, and then for $i = 1, \ldots, n$ generate a random $x_i \stackrel{R}{\leftarrow} A(x_{i-1}, i-1)$ subject to the constraint that $f(x_i)_{1,\cdots,i} = y_{1,\cdots,i}$. The latter step is feasible, since we are guaranteed that $f(x_i)_{1,\ldots,i-1} = y_{1,\cdots,i-1}$ by the fact that A is an F-collision finder, and the expected number of trials needed get agreement with y_i is at most 2 (since $y_i \in \{0,1\}$, and y and $f(x_i)$ are statistically close). It is not difficult to show that when run on a random output Y of f, *Inv* produces an almost-uniform preimage of Y. This contradicts the one-wayness of f. Indeed, we only need f to be a *distributional* one-way function [IL], whereby it is infeasible to generate almost-uniform preimages under f.

Inaccessible Entropy $\Rightarrow$ UOWHFs. Once we have a non-negligible amount of inaccessible entropy, we can construct a UOWHF via a series of standard transformations.

1. Repetition: By evaluating F on many inputs, we can increase the amount of inaccessible entropy from $1/\operatorname{poly}(n)$ to $\operatorname{poly}(n)$. Specifically, we take $F^t(x_1, \ldots, x_t) = (F(x_1), \ldots, F(x_t))$ where $t = \operatorname{poly}(n)$. This transformation also has the useful effect of converting the real entropy of F^{-1} to *min-entropy*.

2. Hashing Inputs: By hashing the input to F (namely taking $F'(x,g) = (F(x), g(x))$ for a universal hash function g), we can reduce both the real (min-)entropy and the accessible entropy so that $(F')^{-1}$ still has a significant amount of real entropy, but has (weak) target collision resistance (on random inputs).

3. Hashing Outputs: By hashing the output to F (namely taking $F'(x,g) = g(F(x))$), we can reduce the output length of F to obtain a shrinking function that still has (weak) target collision resistance.

There are two technicalities that occur in the above steps. First, hashing the inputs only yields *weak* target collision resistance; this is due to the fact that accessible Shannon entropy is an average-case measure and thus allows for the possibility that the adversary can achieve high accessible entropy most of the time. Fortunately, this weak form of target collision resistance can be amplified to full target collision resistance using another application of repetition and hashing (similar to [CRS$^+$]).

Second, the hashing steps require having a fairly accurate estimate of the real entropy. This can be handled similarly to [HILL, Rom1], by trying all (polynomially many) possibilities and concatenating the resulting UOWHFs, at least one of which will be target collision resistant.

A More Efficient Construction. We obtain a more efficient construction of UOWHFs by hashing the output of the one-way function f before truncating. That is, we define $F(x,g,i) = (g, g(f(x))_{1\cdots i})$. This function is in the spirit of the function that Rompel [Rom1] uses as a first step, but our function uses three-wise independent hash function instead of n-wise independent one, and enjoys a much simpler structure.[3] Our analysis of this function is significantly simpler than Rompel's and can be viewed as providing a clean abstraction of what it achieves (namely, inaccessible entropy) that makes the subsequent transformation to a UOWHF much easier.

[3]Rompel started with the function $f'(z, g_1, g_2) := (g_2(f_0(g_1(z))), g_1, g_2)$, where g_1 and g_2 are n-wise independent hash-functions, and f_0 is defined as $f_0(x,y,i) = (f(x), y^{n-i}, 0^i)$.

We obtain improved UOWHF parameters over our first construction for two reasons. First, we obtain a larger amount of inaccessible entropy: $(\log n)/n$ bits instead of roughly $1/n^4$ bits. Second, we obtain a bound on a stronger form of accessible entropy, which enables us to get full target collision resistance when we hash the inputs, avoiding the second amplification step.

This construction yields better parameters than Rompel's original construction. A one-way function of input length n yields a UOWHF with output length $\tilde{O}(n^7)$, improving Rompel's bound of $\tilde{O}(n^8)$. Additionally, we are able to reduce the key length needed: Rompel's original construction uses a key of length $\tilde{O}(n^{12})$, whereas our construction only needs a key of length $\tilde{O}(n^7)$. If we allow the construction to utilize some nonuniform information (namely an estimate of the real entropy of F^{-1}), then we obtain output length $\tilde{O}(n^5)$, improving Rompel's bound of $\tilde{O}(n^6)$. For the key length, the improvement in this case is from $\tilde{O}(n^7)$ to $\tilde{O}(n^5)$. Of course, these bounds are still far from practical, but they illustrate the utility of inaccessible entropy in reasoning about UOWHFs, which may prove useful in future constructions (whether based on one-way functions or other building blocks).

1.3 Perspective

The idea of inaccessible entropy was introduced in [HRVW] for the purpose of constructing statistically hiding commitment schemes from one-way functions and from zero-knowledge proofs. There, the nature of statistically hiding commitments necessitated more involved notions of inaccessible entropy than we present here — inaccessible entropy was defined in [HRVW] for interactive protocols and for "generators" that output many blocks, where one considers adversaries that try to generate next-messages or next-blocks of high entropy. In such a setting, it is necessary to have the adversary privately "justify" that it is behaving consistently with the honest party, and to appropriately discount the entropy in case the adversary outputs an invalid justification.

Here, we are able to work with a much simpler form of inaccessible entropy. The simplicity comes from the noninteractive nature of UOWHFs (so we only need to measure the entropy of a single string output by the adversary), and the fact that we can assume without loss of generality that the adversary behaves consistently with the honest party. Thus, the definitions here can serve as a gentler introduction to the concept of inaccessible entropy. On the other hand, the many-round notions from [HRVW] allow for a useful "entropy equalization" transformation that avoids the need to try all possible guesses for the entropy. We do not know an analogous transformation for constructing UOWHFs. We also note that our simple construction of a function with inaccessible entropy by randomly truncating a one-way function (and its analysis) is inspired by the the construction of an "inaccessible entropy generator" from a one-way function in [HRVW].

Finally, with our constructions, the proof that one-way functions imply UOWHFs now parallels those of pseudorandom generators [HILL, HRV] and statistically hiding commitments [HNO$^+$, HRVW], with UOWHFs and statistically hiding commitments using dual notions of entropy (high real entropy, low accessible entropy) to pseudorandom generators (low real entropy, high pseudoentropy).

2 Preliminaries

Most of the material in this section is taken almost verbatim from [HRVW], and missing proofs can be found in that paper.

2.1 Random Variables

Let X and Y be random variables taking values in a discrete universe $\mathcal{U}$. We adopt the convention that when the same random variable appears multiple times in an expression, all occurrences refer to the same instantiation. For example, $\Pr[X = X]$ is 1. For an event E, we write $X|_E$ to denote the random variable X conditioned on E. The *support* of a random variable X is $\mathrm{Supp}(X) := \{x : \Pr[X = x] > 0\}$. X is *flat* if it is uniform on its support. For an event E, we write $I(E)$ for the corresponding indicatory random variable, i.e. $I(E)$ is 1 when E occurs and is 0 otherwise.

We write $\|X - Y\|$ to denote the *statistical difference* (a.k.a. variation distance) between X and Y, i.e.

$$\|X - Y\| = \max_{T \subseteq \mathcal{U}} |\Pr[X \in T] - \Pr[Y \in T]| .$$

If $\|X - Y\| \leq \varepsilon$ (respectively, $\|X - Y\| > \varepsilon$), we say that X and Y are ε-*close* (resp., ε-far).

2.2 Entropy Measures

We will refer to several measures of entropy in this work. The relation and motivation of these measures is best understood by considering a notion that we will refer to as the *sample-entropy*: For a random variable X and $x \in \mathrm{Supp}(X)$, we define the sample-entropy of x with respect to X to be the quantity

$$\mathrm{H}_X(x) := \log(1/\Pr[X = x]).$$

The sample-entropy measures the amount of "randomness" or "surprise" in the specific sample x, assuming that x has been generated according to X. Using this notion, we can define the *Shannon entropy* $\mathrm{H}(X)$ and *min-entropy* $\mathrm{H}_\infty(X)$ as follows:

$$\begin{aligned} \mathrm{H}(X) &:= \mathop{\mathrm{E}}_{x \xleftarrow{R} X} [\mathrm{H}_X(x)] \\ \mathrm{H}_\infty(X) &:= \min_{x \in \mathrm{Supp}(X)} \mathrm{H}_X(x) \end{aligned}$$

We will also discuss the *max-entropy* $\mathrm{H}_0(X) := \log(1/|\mathrm{Supp}(X)|)$. The term "max-entropy" and its relation to the sample-entropy will be made apparent below.

It can be shown that $\mathrm{H}_\infty(X) \leq \mathrm{H}(X) \leq \mathrm{H}_0(X)$ with equality if and only if X is flat. Thus, saying $\mathrm{H}_\infty(X) \geq k$ is a strong way of saying that X has "high entropy" and $\mathrm{H}_0(X) \leq k$ a strong way of saying that X as "low entropy".

Smoothed Entropies. Shannon entropy is robust in that it is insensitive to small statistical differences. Specifically, if X and Y are ε-close then $|\mathrm{H}(X) - \mathrm{H}(Y)| \leq \varepsilon \cdot \log |\mathcal{U}|$. For example, if $\mathcal{U} = \{0,1\}^n$ and $\varepsilon = \varepsilon(n)$ is a negligible function of n (i.e., $\varepsilon = n^{-\omega(1)}$), then the difference in Shannon entropies is vanishingly small (indeed, negligible). In contrast, min-entropy and max-entropy are brittle and can change dramatically with a small statistical difference. Thus, it is common to work with "smoothed" versions of these measures, whereby we consider a random variable X to have high entropy (respectively, low entropy) if X is ε-close to some X' with $\mathrm{H}_\infty(X) \geq k$ (resp., $\mathrm{H}_0(X) \leq k$) for some parameter k and a negligible ε.[4]

[4]The term "smoothed entropy" was coined by Renner and Wolf [RW], but the notion of smoothed min-entropy has commonly been used (without a name) in the literature on randomness extractors [NZ].

These smoothed versions of min-entropy and max-entropy can be captured quite closely (and more concretely) by requiring that the sample-entropy is large or small with high probability:

Lemma 2.1. *1. Suppose that with probability at least $1-\varepsilon$ over $x \stackrel{R}{\leftarrow} X$, we have $\mathrm{H}_X(x) \geq k$. Then X is ε-close to a random variable X' such that $\mathrm{H}_\infty(X') \geq k$.*

2. Suppose that X is ε-close to a random variable X' such that $\mathrm{H}_\infty(X') \geq k$. Then with probability at least $1-2\varepsilon$ over $x \stackrel{R}{\leftarrow} X$, we have $\mathrm{H}_X(x) \geq k - \log(1/\varepsilon)$.

Lemma 2.2. *1. Suppose that with probability at least $1-\varepsilon$ over $x \stackrel{R}{\leftarrow} X$, we have $\mathrm{H}_X(x) \leq k$. Then X is ε-close to a random variable X' such that $\mathrm{H}_0(X') \leq k$.*

2. Suppose that X is ε-close to a random variable X' such that $\mathrm{H}_0(X') \leq k$. Then with probability at least $1-2\varepsilon$ over $x \stackrel{R}{\leftarrow} X$, we have $\mathrm{H}_X(x) \leq k + \log(1/\varepsilon)$.

Think of ε as inverse polynomial or a slightly negligible function in $n = \log(|\mathcal{U}|)$. The above lemmas show that up to negligible statistical difference and a slightly superlogarithmic number of entropy bits, the min-entropy (resp. max-entropy) is captured by lower (resp. upper) bound on sample-entropy.

Conditional Entropies. We will also be interested in conditional versions of entropy. For jointly distributed random variables (X,Y) and $(x,y) \in \mathrm{Supp}(X,Y)$, we define the *conditional sample-entropy* to be $\mathrm{H}_{X|Y}(x|y) = \log(1/\Pr[X=x|Y=y])$. Then the standard *conditional Shannon entropy* can be written as:

$$\mathrm{H}(X|Y) = \mathop{\mathrm{E}}_{(x,y)\stackrel{\mathrm{R}}{\leftarrow}(X,Y)}\left[\mathrm{H}_{X|Y}(x|y)\right] = \mathop{\mathrm{E}}_{y\stackrel{\mathrm{R}}{\leftarrow}Y}\left[\mathrm{H}(X|_{Y=y})\right] = \mathrm{H}(X,Y) - \mathrm{H}(Y).$$

There is no standard definition of conditional min-entropy and max-entropy, or even their smoothed versions. For us, it will be most convenient to generalize the sample-entropy characterizations of smoothed min-entropy and max-entropy given above. Specifically we will think of X as having "high min-entropy" (resp., "low max-entropy") given Y if with probability at least $1-\varepsilon$ over $(x,y) \stackrel{\mathrm{R}}{\leftarrow} (X,Y)$, we have $\mathrm{H}_{X|Y}(x|y) \geq k$ (resp., $\mathrm{H}_{X|Y}(x|y) \leq k$).

Flattening Shannon Entropy. It is well-known that the Shannon entropy of a random variable can be converted to min-entropy (up to small statistical distance) by taking independent copies of this variable.

Lemma 2.3. *1. Let X be a random variable taking values in a universe $\mathcal{U}$, let $t \in \mathbb{N}$, and let $\varepsilon > 0$. Then with probability at least $1-\varepsilon-2^{-\Omega(t)}$ over $x \stackrel{R}{\leftarrow} X^t$,*

$$|\mathrm{H}_{X^t}(x) - t\cdot\mathrm{H}(X)| \leq O(\sqrt{t\cdot\log(1/\varepsilon)}\cdot\log(|\mathcal{U}|\cdot t)).$$

2. Let X,Y be jointly distributed random variables where X takes values in a universe $\mathcal{U}$, let $t \in \mathbb{N}$, and let $\varepsilon > 0$. Then with probability at least $1-\varepsilon-2^{-\Omega(t)}$ over $(x,y) \stackrel{R}{\leftarrow} (X^t,Y^t) := (X,Y)^t$,

$$\left|\mathrm{H}_{X^t|Y^t}(x|y) - t\cdot\mathrm{H}(X|Y)\right| \leq O(\sqrt{t\cdot\log(1/\varepsilon)}\cdot\log(|\mathcal{U}|\cdot t)).$$

Proof. 1. For $x = (x_1, \ldots, x_t)$, we have $\mathrm{H}_{X^t}(x) = \sum_{i=1}^{t} \mathrm{H}_X(x_i)$. Thus, when $x \stackrel{R}{\leftarrow} X^t$, $\mathrm{H}_{X^t}(x)$ is the sum of t independent random variables $\mathrm{H}_X(x_i)$, and thus we can obtain concentration around the expectation (which is $t \cdot \mathrm{H}(X)$) via Chernoff-Hoeffding Bounds. These random variables $\mathrm{H}_X(x_i)$ are not bounded (as is required to apply the standard Chernoff–Hoeffding Bound), but they are unlikely to be much larger than $O(\log |\mathcal{U}|)$. Specifically, for every $\tau > 0$ we have

$$\begin{aligned} \Pr_{x_i \stackrel{R}{\leftarrow} X}\left[\mathrm{H}_X(x_i) \geq \log(|\mathcal{U}|/\tau)\right] &\leq \sum_{x_i \in \mathcal{U}: \mathrm{H}_X(x_i) \geq \log(|\mathcal{U}|/\tau)} \Pr[X = x_i] \\ &\leq |\mathcal{U}| \cdot 2^{-\log(|\mathcal{U}|/\tau)} \\ &= \tau. \end{aligned}$$

A Chernoff Bound for random variables with such exponentially vanishing tails follows from [Vad], and it says that the probability that the sum deviates from the expectation by at least $\Delta \cdot (\log(|\mathcal{U}|/\tau)) + 2\tau t$ is at most $\exp(-\Omega(\Delta^2/t)) + \exp(-\Omega(\tau t))$, provided $\tau \in [0, 1]$. An appropriate choice of $\Delta = O(\sqrt{t \log(1/\varepsilon)})$ and $\tau = \min\{1, O(\log(1/\varepsilon)/t)\}$ completes the proof.

2. Similar, noting that $\mathrm{H}_{X^t|Y^t}(x|y) = \sum_{i=1}^{t} \mathrm{H}_{X|Y}(x_i|y_i)$.

□

2.3 Hashing

A family of functions $F = \{f : \{0,1\}^n \to \{0,1\}^m\}$ is *2-universal* if for every $x \neq x' \in \{0,1\}^n$, when we choose $F \stackrel{R}{\leftarrow} F$, we have $\Pr[F(x) = F(x')] \leq 1/|\{0,1\}^m|$. F is *t-wise independent* if for all distinct $x_1, \ldots, x_t \in \{0,1\}^n$, when we choose $F \stackrel{R}{\leftarrow} F$, the random variables $F(x_1), \ldots, F(x_t)$ are independent and each uniformly distributed over $\{0,1\}^m$.

F is *explicit* if given the description of a function $f \in F$ and $x \in \{0,1\}^n$, $f(x)$ can be computed in time $\mathrm{poly}(n, m)$. F is *constructible* if it is explicit and there is a probabilistic polynomial-time algorithm that given $x \in \{0,1\}^n$, and $y \in \{0,1\}^m$, outputs a random $f \stackrel{R}{\leftarrow} F$ such that $f(x) = y$.

It is well-known that there are constructible families of 2-universal functions (resp. t-wise independent functions) in which choosing a function $f \stackrel{R}{\leftarrow} F$ uses only $\max\{n, m\} + m$ (resp., $t \cdot n$) random bits.

3 Inaccessible Entropy for Inversion Problems

As discussed in the introduction, for a function F, we define the *real entropy* of F^{-1} to be the amount of entropy left in the input after revealing the output.

Definition 3.1. *Let n be a security parameter, and $F : \{0,1\}^n \to \{0,1\}^m$ a function. We say that F^{-1} has* real Shannon entropy *k if*

$$\mathrm{H}(X|F(X)) = k,$$

where X is uniformly distributed on $\{0,1\}^n$. We say that F^{-1} has real min-entropy *at least k if there is a negligible function $\varepsilon = \varepsilon(n)$ such that*

$$\Pr_{x \stackrel{R}{\leftarrow} X}\left[\mathrm{H}_{X|F(X)}(x|F(x)) \geq k\right] \geq 1 - \varepsilon(n).$$

We say that F^{-1} *has* real max-entropy *at most* k *if there is a negligible function* $\varepsilon = \varepsilon(n)$ *such that*

$$\Pr_{x \xleftarrow{R} X}\left[\mathrm{H}_{X|F(X)}(x|F(x)) \leq k\right] \geq 1 - \varepsilon(n).$$

Note that more concrete formulas for the entropies above are:

$$\begin{aligned}\mathrm{H}_{X|F(X)}(x|F(x)) &= \log|F^{-1}(F(x))| \\ \mathrm{H}(X|F(X)) &= \mathrm{E}\left[\log|F^{-1}(F(X))|\right].\end{aligned}$$

As our goal is to construct UOWHFs that are shrinking, achieving high real entropy is a natural intermediate step. Indeed, the amount by which F shrinks is a lower bound on the real entropy of F^{-1}:

Proposition 3.2. *If* $F : \{0,1\}^n \to \{0,1\}^m$*, then the real Shannon entropy of* F^{-1} *is at least* $n-m$*, and the real min-entropy of* F^{-1} *is at least* $n - m - s$ *for any* $s = \omega(\log n)$*.*

Proof. For Shannon entropy, we have

$$\mathrm{H}(X|F(X)) \geq \mathrm{H}(X) - \mathrm{H}(F(X)) \geq n - m.$$

For min-entropy, let $S = \{y \in \{0,1\}^m : \Pr[f(X) = y] < 2^{-m-s}\}$. Then $\Pr[f(X) \in S] \leq 2^m \cdot 2^{-m-s} = \mathrm{neg}(n)$, and for every x such that $f(x) \notin S$, we have

$$\begin{aligned}\mathrm{H}_{X|F(X)}(x|F(x)) &= \log\frac{1}{\Pr[X = x|F(X) = f(x)]} \\ &= \log\frac{\Pr[f(X) = f(x)]}{\Pr[X = x]} \\ &\geq \log\frac{2^{-m-s}}{2^{-n}} \\ &= n - m - s.\end{aligned}$$

□

To motivate the definition of accessible entropy, we now present an alternative formulation of real entropy in terms of the entropy that computationally unbounded "collision-finding" adversaries can generate.

Definition 3.3. *For a function* $F : \{0,1\}^n \to \{0,1\}^m$*, an* F-collision-finder *is a randomized algorithm* A *such that for every* $x \in \{0,1\}^n$ *and coin tosses* r *for* A*, we have* $A(x;r) \in F^{-1}(F(x))$*.*

Note that A is required to *always* produce an input $x' \in \{0,1\}^n$ such that $F(x) = F(x')$. This is a reasonable constraint because A has the option of outputting $x' = x$ if it does not find a true collision. We consider A's goal to be maximizing the entropy of its output $x' = A(x)$, given a random input x. If we let A be computationally unbounded, then the optimum turns out to equal exactly the real entropy:

Proposition 3.4. *Let* $F : \{0,1\}^n \to \{0,1\}^m$*. Then the real Shannon entropy of* F^{-1} *equals the maximum of* $\mathrm{H}(A(X;R)|X)$ *over all (computationally unbounded)* F*-collision finders* A*, where the random variable* X *is uniformly distributed in* $\{0,1\}^n$ *and* R *is uniformly random coin tosses for* A*. That is,*

$$\mathrm{H}(X|F(X)) = \max_A \mathrm{H}(A(X;R)|X),$$

where the maximum is taken over all F*-collision finders* A*.*

Proof. The F-collision finder A that maximizes $\mathrm{H}(A(X)|X)$ is the algorithm A^* that, on input x, outputs a uniformly random element of $f^{-1}(f(x))$. Then

$$\mathrm{H}(A^*(X;R)|X) = \mathrm{E}[\log |f^{-1}(f(X))|] = \mathrm{H}(X|F(X)).$$

□

The notion of *accessible entropy* simply restricts the above to efficient adversaries, e.g. those that run in probabilistic polynomial time (PPT for short):

Definition 3.5. *Let n be a security parameter and $F : \{0,1\}^n \to \{0,1\}^m$ a function. We say that F^{-1} has* accessible Shannon entropy *at most k if for every* PPT *F-collision-finder A, we have*

$$\mathrm{H}(A(X;R)|X) \leq k$$

for all sufficiently large n, where the random variable X is uniformly distributed on $\{0,1\}^n$ and R is uniformly random coin tosses for A.

As usual, it is often useful to have an upper bound not only on Shannon entropy, but on the max-entropy (up to some negligible statistical distance). Recall that a random variable Z has max-entropy at most k iff the support of Z is contained in a set of size 2^k. Thus, we require that $A(X;R)$ is contained in a set $L(X)$ of size at most 2^k, except with negligible probability:

Definition 3.6. *Let n be a security parameter and $F : \{0,1\}^n \to \{0,1\}^m$ a function. For $p = p(n) \in [0,1]$, we say that F^{-1} has p-*accessible max-entropy *at most k if for every* PPT *F-collision-finder A, there exists a family of sets $\{L(x)\}_{x \in \mathrm{Supp}(X)}$ each of size at most 2^k such that $x \in L(x)$ for all $x \in \mathrm{Supp}(X)$ and*

$$\Pr\left[A(X;R) \in L(X)\right] \geq 1 - p$$

for all sufficiently large n, where random variable X is uniformly distributed on $\{0,1\}^n$ and R is uniformly random coin tosses for A. In addition, if $p = \varepsilon(n)$ for some negligible function $\varepsilon(\cdot)$, then we simply say that F^{-1} has accessible max-entropy *at most k.*

The reason that having an upper bound on accessible entropy is useful as an intermediate step towards constructing UOWHFs is that accessible max-entropy 0 is equivalent to target collision resistance (on random inputs):

Definition 3.7. *Let $F : \{0,1\}^n \to \{0,1\}^m$ be a function. For $q = q(n) \in [0,1]$, we say that F is* q-collision-resistant on random inputs *if for every* PPT *F-collision-finder A,*

$$\Pr[A(X;R) \ = X] \geq q,$$

for all sufficiently large n, where random variable X is uniformly distributed on $\{0,1\}^n$ and R is uniformly random coin tosses for A. In addition, if $q = 1 - \varepsilon(n)$ for some negligible function $\varepsilon(\cdot)$, we say that F is collision-resistant on random inputs.

Lemma 3.8. *Let n be a security parameter and $F : \{0,1\}^n \to \{0,1\}^m$ be a function. Then, for any $p = p(n) \in (0,1)$, the following statements are equivalent:*

(1) *F^{-1} has p-accessible max-entropy 0.*

(2) *F is $(1-p)$-collision-resistant on random inputs.*

In particular, F^{-1} has accessible max-entropy 0 iff F is collision-resistant on random inputs.

Proof. Note that (1) implies (2) follows readily from the definition. To see that (2) implies (1), simply take $L(x) = \{x\}$. □

While bounding p-accessible max-entropy with negligible p is our ultimate goal, one of our constructions will work by first giving a bound on accessible Shannon entropy, and then deducing a bound on p-accessible max-entropy for a value of $p < 1$ using the following lemma:

Lemma 3.9. *Let n be a security parameter and $F : \{0,1\}^n \to \{0,1\}^m$ be a function. If F^{-1} has accessible Shannon entropy at most k, then F^{-1} has p-accessible max-entropy at most $k/p + O(2^{-k/p})$ for any $p = p(n) \in (0,1)$.*

Proof. Fix any PPT F-collision-finder A. From the bound on accessible Shannon entropy, we have that $\mathrm{H}(A(X;R)|X) \leq k$. Applying Markov's inequality, we have

$$\Pr_{x \xleftarrow{R} X, r \xleftarrow{R} R}\left[\mathrm{H}_{A(X;R)|X}(A(x;r)|x) \leq k/p\right] \geq 1-p$$

Take $L(x)$ to be the set:

$$L(x) = \{x\} \cup \left\{x' : \mathrm{H}_{A(X;R)|X}(x'|x) \leq k/p\right\}$$

We may rewrite $L(x)$ as $\{x\} \cup \{x' : \Pr_r[A(x;r) = x'] \geq 2^{-k/p}\}$. It is easy to see that $|L(x)| \leq 2^{k/p}+1$ and thus F^{-1} has p-accessible max-entropy at most $k/p + O(2^{-k/p})$. □

Once we have a bound on p-accessible max-entropy for some $p < 1$, we need to apply several transformations to obtain a function with a good bound on $\mathrm{neg}(n)$-accessible max-entropy.

Our second construction (which achieves better parameters), starts with a bound on a different average-case form of accessible entropy, which is stronger than bounding the accessible Shannon entropy. The benefit of this notion it that it can be converted more efficiently to $\mathrm{neg}(n)$-accessible max-entropy, by simply taking repetitions.

To motivate the definition, recall that a bound on accessible Shannon entropy means that the sample entropy $\mathrm{H}_{A(X;R)|X}(x'|x)$ is small on average over $x \xleftarrow{R} X$ and $x' \xleftarrow{R} A(x;R)$. This sample entropy may depend on both the input x and the x' output by the adversary (which in turn may depend on its coin tosses). A stronger requirement is to say that we have upper bounds $k(x)$ on the sample entropy that depend *only on* x. The following definition captures this idea, thinking of $k(x) = \log|L(x)|$. (We work with sets rather than sample entropy to avoid paying the $\log(1/\varepsilon)$ loss in Lemma 2.2.)

Definition 3.10. *Let n be a security parameter and $F : \{0,1\}^n \to \{0,1\}^m$ a function. We say that F^{-1} has* accessible average max-entropy *at most k if for every* PPT *F-collision-finder A, there exists a family of sets $\{L(x)\}_{x \in \mathrm{Supp}(X)}$ and a negligible function $\varepsilon = \varepsilon(n)$ such that $x \in L(x)$ for all $x \in \mathrm{Supp}(X)$, $\mathrm{E}[\log|L(X)|] \leq k$ and*

$$\Pr\left[A(X;R) \in L(X)\right] \geq 1 - \varepsilon(n),$$

for all sufficiently large n, where random variable X is uniformly distributed on $\{0,1\}^n$ and R is uniformly random coin tosses for A.

We observe that bounding accessible average max-entropy is indeed stronger than bounding accessible Shannon entropy:

Proposition 3.11. *If F^{-1} has accessible average max-entropy at most k, then for every constant c, F^{-1} has accessible Shannon entropy at most $k + 1/n^c$.*

Proof. Given an F-collision-finder A, let $\{L(x)\}$ be the sets guaranteed in Definition 3.10. Let random variable X be uniformly distributed in $\{0,1\}^n$, let R be uniformly random coin tosses for A, and let I be the indicator random variable for $A(X;R) \in L(X)$. So $\Pr[I = 0] = \text{neg}(n)$. Then:

$$\begin{aligned}
\mathrm{H}(A(X;R)|X) &\leq \mathrm{H}(A(X;R)|X,I) + \mathrm{H}(I) \\
&\leq \Pr[I=1] \cdot \mathrm{H}(A(X;R)|X, I=1) + \Pr[I=0] \cdot \mathrm{H}(A(X;R)|X, I=1) + \text{neg}(n) \\
&\leq \Pr[I=1] \cdot \mathrm{E}[\log |L(X)||I=1] + \Pr[I=0] \cdot n + \text{neg}(n) \\
&\leq \mathrm{E}[\log |L(X)|] + \text{neg}(n) \\
&\leq k + \text{neg}(n).
\end{aligned}$$

□

4 Inaccessible Entropy from One-way Functions

In Section 4.1 we show that *any* one-way function can be very slightly altered into a function with inaccessible entropy. In Section 4.2 we show that an additional hashing step implies a stronger form of inaccessible entropy (which we can then use for a more efficient construction of UOWHF). Still, we find the more direct construction of Section 4.1 and its analysis to be striking in its simplicity.

4.1 A Direct Construction

Theorem 4.1 (Inaccessible Shannon entropy from one-way functions)**.** *Let $f\colon \{0,1\}^n \mapsto \{0,1\}^n$ be a one-way function and define F over $\{0,1\}^n \times [n]$ as $F(x,i) = f(x)_{1,\ldots,i}$. Then F^{-1} has accessible Shannon entropy at most $\mathrm{H}(Z|F(Z)) - 1/(2^9 \cdot n^4 \cdot \log^2 n)$, where $Z = (X,I)$ is uniformly distributed over $\{0,1\}^n \times [n]$.*[5]

Proof. Suppose on the contrary that there exists a PPT F-collision-finder A such that

$$\mathrm{H}(Z|F(Z)) - \mathrm{H}(A(Z;R)|Z) < \varepsilon = 1/(2^9 \cdot n^4 \cdot \log^2 n)$$

for infinitely many n's, and R is uniformly distributed over the random coins of A. Since I is determined by $F(Z)$, and since Z also determines the second part of A's output (since A is an F-collision-finder), it follows that

$$\mathrm{H}(X|F(Z)) - \mathrm{H}(A'(Z;R)|Z) < \varepsilon$$

where A' is the algorithm that on input $(z;r)$ outputs the first component of $A(z;r)$'s output. In the following use A' to construct an efficient algorithm that inverts f with constant probability. We do so in two steps: 1. Constructing such an inverter under the assumption that we have access to an (inefficient) oracle Sam_{Ideal} defined shortly, and 2. Showing how to efficiently approximate Sam_{Ideal} using A'.

[5] We believe that the actual gap between the real and accessible entropy of F^{-1} is $\Omega(1/n^2)$, or possibly even $\Omega(1/n)$, and not $\Omega(1/n^4)$ as stated. Since even the optimistic $\Omega(1/n)$ bound does not yield as efficient overall construction as the one resulting from Section 4.2, we defer a tighter analysis to the final version of the paper.

Algorithm 4.2 (Sam_{Ideal})**.**

Input: $x \in \{0,1\}^n$, $i \in [n]$ *and* $b \in \{0,1\}$.

Return a random $x' \in F^{-1}(F(x,i-1))_1$ *such that* $f(x')_i = b$ *(return an arbitrarily value if no such* x' *exists), where* $F^{-1}(F(x,j))_1 = \{x' \in \{0,1\}^n \colon F(x',j) = F(x,j)\}$.

That is, Sam_{Ideal} outputs uniformly at random x' such that $f(x')_{1,\ldots,i} = (f(x)_{1,\ldots,i-1}, b)$. We define an algorithm Inv with access to an oracle Sam. When $Sam = Sam_{\mathsf{Ideal}}$, it will be easy to argue that Inv inverts f with probability one.

Algorithm 4.3 (Inv^{Sam})**.**

Input: $y \in \{0,1\}^n$.

Oracle: Sam.

For $i = 1$ *to* n *do:*
  let $x^i = Sam(x^{i-1}, i, y_i)$ *(where* x^0 *is chosen arbitrarily)*

Output x^n.

It is immediate that $Inv^{Sam_{\mathsf{Ideal}}}$ inverts f with probability one. We now turn to showing that A' can be used to efficiently approximate Sam_{Ideal}. The resulting algorithm Sam_δ will be sufficiently similar to Sam_{Ideal} and as a result Inv^{Sam_δ} will still invert f with high probability. A property of Inv that will come handy is that, on a uniform value $y = f(x)$, the first coordinate of each individual query that the inverter $Inv^{Sam_{\mathsf{Ideal}}}$ makes (i.e., x^i) is uniform in $\{0,1\}^n$ (the queries are correlated of course).

Recall that the output of A' has high Shannon entropy - almost as high as the uniform distribution over its set of prescribed outputs. Claim 4.5 (which is rather standard), shows that this also implies that the distribution of A''s output is statistically close to this uniform distribution.

Definition 4.4. *For* $\delta \in [0,1]$ *let* $\mathcal{A}_\delta$ *be the family of efficient* F*-collision-finders with the following guarantee: for every* $A'' \in \mathcal{A}_\delta$ *there exist infinitely many* n*'s such that*
$\| (Z, A''(Z;R)) - (Z, \mathrm{F}^{-1}(F(Z))_1) \| \leq \delta$*, where* R *is uniformly distributed over the random-coins of* A'' *and* $\mathrm{F}^{-1}(F(x,i))_1$ *is uniformly distributed over* $F^{-1}(F(x,i))_1$.

Showing that the output of A' is statistically close to uniform can therefore be formalized by showing the following claim:

Claim 4.5. $A' \in \mathcal{A}_{\sqrt{\varepsilon}}$.

Proof.

$$\begin{aligned}
\| (Z, \mathrm{F}^{-1}(F(Z))_1) - (Z, A'(Z;R)) \| &= \mathop{\mathrm{E}}_{z \leftarrow Z}\left[\|\mathrm{F}^{-1}(F(z))_1 - A'(z;R))\|\right] \\
&\leq \mathop{\mathrm{E}}_{z \leftarrow Z}\left[\sqrt{\mathrm{H}(\mathrm{F}^{-1}(F(z))_1) - \mathrm{H}(A'(z;R))}\right] \\
&\leq \sqrt{\mathop{\mathrm{E}}_{z \leftarrow Z}\left[\mathrm{H}(\mathrm{F}^{-1}(F(z))_1) - \mathrm{H}(A'(z;R))\right]} \\
&= \sqrt{\mathrm{H}(X|F(Z)) - \mathrm{H}(A'(Z;R))} \\
&\leq \sqrt{\varepsilon},
\end{aligned}$$

where the first inequality uses the fact that if W is a random variable whose support is contained in a set S and U is the uniform distribution on S, then $\|U - W\| \leq \sqrt{\mathrm{H}(U) - \mathrm{H}(W)}$. (See [CT, Lemma 11.6.1].) □

As we just shown that $A' \in \mathcal{A}_{\sqrt{\varepsilon}}$ it is enough to show how to use an algorithm $A'' \in \mathcal{A}_\delta$ to approximate Sam_{Ideal} (with error which depends on δ). In order to keep notation simple, we abuse notation and denote by $\mathcal{A}_\delta$ some $A'' \in \mathcal{A}_\delta$. Fix $\delta \in [0,1]$ and consider the following efficient approximation of Sam_{Ideal}:

Algorithm 4.6 (Sam_δ)**.**

Input: $x \in \{0,1\}^n$, $i \in [n]$ *and* $b \in \{0,1\}$.

Oracle: $\mathcal{A}_\delta$.

Repeat $16n \cdot \log n$ *times:*

1. *Let* $x' = \mathcal{A}_\delta(x, i-1)$
2. *If* $f(x')_i = b$, *return* x'.

Abort.

Let Inv_δ denote Inv^{Sam_δ} and Inv_{Ideal} denote $Inv^{Sam_{\mathsf{Ideal}}}$. We will show that the output of Inv_δ (on a random value $f(x)$) is statistically close to that of Inv_{Ideal}. As Inv_{Ideal} inverts f with probability one, we will conclude that Inv_δ inverts f with high probability as well. To analyze the statistical distance between the outputs of the two inverters, we consider hybrid inverters that use the ideal Sam_{Ideal} in the first queries and use Sam_δ in the rest of the queries: For $i \in [n+1]$ let Inv^i_δ be the variant of Inv that uses Sam_{Ideal} in the first $i-1$ queries and Sam_δ for the rest of the queries. The next claim will allow us to easily bound the difference between the output distribution of any two neighboring hybrid inverters:

Claim 4.7. *Let* $i \in [n]$ *and let* $\delta_i = \left\|(X, \mathcal{A}_\delta(X, i; R)) - (X, \mathrm{F}^{-1}(F(X,i))_1)\right\|$, *then* $\|(X, Sam_{\mathsf{Ideal}}(X, i, f(X)_i)) - (X, Sam_\delta(X, i, f(X)_i))\| \leq 1/2n + 16 \cdot n \cdot \log n \cdot \delta_i$.

Proof. Sam_δ is imperfect for two reasons (which our analysis handles separately). The first reason is that Sam_δ relies on the output of $\mathcal{A}_\delta$ that returns an inverse that is only close to uniform (rather than fully uniform). The error accumulated in each query to $\mathcal{A}_\delta$ is δ_i and there are only $16 \cdot n \cdot \log n$ such queries, which altogether contributes $16 \cdot n \cdot \log n \cdot \delta_i$ to the statistical distance bounded by the claim. The second source of error is that after $16 \cdot n \cdot \log n$ unsuccessful repetitions, Sam_δ aborts without retrieving a correct inverse x'. As we now argue, such failure will only happen with small probability (contributing $\frac{1}{2n}$ to the bound in the claim).

To separate our analysis of the two sources of error, we start by considering the case that $\delta_i = 0$. Note that in this case $\mathcal{A}_\delta(x, i; R) = \mathcal{A}_0(x, i; R)$ is identical to $\mathrm{F}^{-1}(F(x,i))_1$. For $x \in \{0,1\}^n$, $i \in [m]$ and $b \in \{0,1\}$, let $\alpha(x,i,b) := \Pr_{y \xleftarrow{R} f(X)}[y_i = b \mid y_{1,\ldots,i-1} = f(x)_{1,\ldots,i-1}]$. Note that for every i, $\Pr[\alpha(X, i, f(X)_i)) < \beta] < \beta$ for every $\beta > 0$. We also note that $Sam_\delta(x, i, f(x)_i)$ aborts with probability at most $(1 - \frac{1}{4n})^{16n \cdot \log n} < \frac{1}{4n}$ in the case that $\alpha(x, i, f(x)_i) \geq \frac{1}{4n}$, and that in case it

does not abort, (since we assume that $\delta_i = 0$) it returns the same distribution as $Sam_{\mathsf{Ideal}}(x, i, b)$ does. Hence, for the case that $\delta_i = 0$ we have that

$$\begin{aligned}
&\|(X, Sam_{\mathsf{Ideal}}(X, i, f(X)_i)) - (X, Sam_\delta(X, i, f(X)_i))\| \\
&\leq \Pr[\alpha(X, i, f(X)_i) < \frac{1}{4n}] + \Pr[Sam_\delta(X, i, f(X)_i) \text{ aborts} \mid \alpha(X, i, f(X)_i) \geq \frac{1}{4n}] \\
&< \frac{1}{4n} + \frac{1}{4n} \\
&\leq \frac{1}{2n}.
\end{aligned}$$

We now want to analyze the general case where δ_i may be larger than zero. The statistical distance between the output distribution of $Sam_\delta(X, i, f(X)_i)$ in the case $\delta_i = 0$ and in the general case is at most the maximal number of calls to $\mathcal{A}_\delta$ made by Sam_δ times $\|(X, \mathcal{A}_\delta(X, i)) - (X, \mathcal{A}_0(X, i))\|$, we therefore have that

$$\begin{aligned}
&\|(X, Sam_{\mathsf{Ideal}}(X, i, f(X)_i) - (X, Sam_\delta(X, i, f(X)_i)))\| \\
&\leq \frac{1}{2n} + 16n \cdot \log n \cdot \|(X, \mathcal{A}_\delta(X, i) - (X, \mathcal{A}_0(X, i)\| \\
&= \frac{1}{2n} + 16n \cdot \log n \cdot \left\|(X, \mathcal{A}_\delta(X, i) - (X, \mathrm{F}^{-1}(F(X, i))_1))\right\| \\
&= \frac{1}{2n} + 16 \cdot n \cdot \log n \cdot \delta_i.
\end{aligned}$$

□

Now note that the i'th query of $Inv_\delta^i(f(X))$ (a query to Sam_δ) and the i'th query of $Inv_\delta^{i+1}(f(X))$ (a query to Sam_{Ideal}) are both distributed as $(X, i, f(X)_i)$. Therefore Claim 4.7 yields that for every $i \in [n]$,

$$\left\|Inv_\delta^{i+1}(f(X)) - Inv_\delta^i\right\| \leq \frac{1}{2n} + 16 \cdot n \cdot \log n \cdot \delta_i.$$

Hence,

$$\begin{aligned}
&\Pr[Inv_\delta(f(X)) \in f^{-1}(f(X))] \\
&\geq 1 - \sum_{i=1}^{n} \left\|Inv_\delta^{i+1}(f(X)) - Inv_\delta^i(f(X))\right\| \\
&\geq 1 - \sum_{i=1}^{n} \frac{1}{2n} + 16 \cdot n \cdot \log n \cdot \|(X, Sam_{\mathsf{Ideal}}(X, i, f(X)_i) - (X, Sam_\delta(X, i, f(X)_i)\| \\
&\geq \frac{1}{2} - 16 \cdot n^2 \cdot \log n \cdot \delta.
\end{aligned}$$

Let Inv be the instantiation of Inv_δ obtained when we implement Sam_δ using A'. Claim 4.5 yields that $\Pr[Inv(f(X)) \in f^{-1}(f(X))] \geq \Pr[Inv_{\sqrt{\varepsilon}}(f(X)) \in f^{-1}(f(X))] \geq 1/2 - 16 \cdot n^2 \cdot \log n \cdot \sqrt{\varepsilon} > 1/4$.
□

4.2 A More Efficient Construction

The following theorem shows that a simplified variant of the first step of [Rom1] (which is also the first step of [KK]) yields inaccessible entropy with much stronger guarantees than those obtained in Section 4.1. The function we construct is $F(x,g,i) = (g(f(x))_{1,\ldots,i}, g)$, where $g : \{0,1\}^n \to \{0,1\}^n$ is a three-wise independent function. Since the composition of g and f is still a one-way function then Theorem 4.1 already implies that F^{-1} has inaccessible entropy. The benefits of the additional hashing step are that 1. we get more inaccessible entropy ($\tilde{\Theta}(1/n)$ bits rather than $\tilde{\Theta}(1/n^4)$ bits), and 2. we get a bound on accessible average max-entropy rather than accessible Shannon entropy. These allow for a simpler and more efficient transformation of F into a UOWHF.

Theorem 4.8 (Inaccessible average max-entropy from one-way functions)**.** *Let $f\colon \{0,1\}^n \mapsto \{0,1\}^n$ be a one-way function and let $\mathcal{G} = \{g : \{0,1\}^n \to \{0,1\}^n\}$ be a family of constructible,three-wise independent hash functions. Define F with domain $\mathrm{Dom}(F) := \{0,1\}^n \times \mathcal{G} \times [n]$ by*

$$F(x,g,i) = (g(f(x))_{1,\ldots,i}, g).$$

Then, for every constant d, F^{-1} has accessible average max-entropy at most $\mathrm{H}(Z|F(Z)) - (d\log n)/n$ for every $d > 0$, where $Z = (X,G,I)$ is uniformly distributed over $\mathrm{Dom}(F)$.

Proof. Let c be a sufficiently large constant (whose value we determine later, depending on the constant d in the theorem statement) and define for every $y \in \{0,1\}^n$ and $i \in [n]$:

$$L(y,i) = \left\{y' \in \{0,1\}^n \colon \mathrm{H}_{f(X)}(y') \geq (i + c\cdot\log n) \vee y' = y\right\}.$$

(Recall that the sample entropy is defined as $\mathrm{H}_{f(X)}(y) = \log(1/\Pr[f(X)=y]) = n - \log\left|f^{-1}(y)\right|$, so the "heavy" images, where $f^{-1}(y)$ is large, have low sample entropy.) Namely, $L(y,i)$ consists, in addition to y itself, of "i-light" images with respect to f.

We later show that the sets $L'(x,g,i) = f^{-1}(L(f(x),i)) \times \{(g,i)\}$ satisfy the properties required to show that the accessible max-entropy of F^{-1} is as stated in the theorem.[6] Towards this goal, we first show that the only accessible inputs of F come from preimages of $L(y,i)$.

Claim 4.9. *For every* PPT *F-collision-finder A and every constant $c > 0$, it holds that*

$$\Pr[A_1(X,G,I;R) \notin f^{-1}(L(f(X),I))] \leq \mathrm{neg}(n),$$

where (X,G,I) is uniformly distributed over $\mathrm{Dom}(F)$, R is uniformly distributed over the random coins of A, and A_1 denotes the first component of A's output.

Note that the above immediately yields that $\Pr[A(X,G,I;R) \notin L'(X,G,I)] \leq \mathrm{neg}(n)$, since the other two output components of A are required to equal (g,i), due to the fact that $F(x,g,i)$ determines (g,i).

Proof. Suppose on the contrary that there exist an efficient F-collision-finder A, $c > 0$ and a non-negligible function $\varepsilon = \varepsilon(n)$ such that $\Pr[A_1(X,G,I;R) \notin f^{-1}(L(f(X),I))] \geq \varepsilon$. Fix a triple (x,i,r) and let

$$\varepsilon_{x,i,r} = \Pr[A_1(x,G;r) \notin f^{-1}(L(f(x),i))].$$

[6] We are working with the set L, and not with L', as it significantly simplifies notations. Note that the sets L' are independent of the adversary, even though the definition of accessible average max-entropy allows the sets to depend on the adversary. Further, note that the sets L are independent of $\mathcal{G}$.

Define $A'(g) = A_1(x, g; r)$. We will show how to use any such A' to invert f with probability at least $\varepsilon_{x,i,r}/n^c$. By picking (x, i, r) at random, we will invert f with probability at least $\mathrm{E}_{x,i,r}[\varepsilon_{x,i,r}/n^c] = \varepsilon/n^c$, which contradicts the one-wayness of f. Our inverter works as follows, on input $y \in \{0,1\}^n$.

Inv(y)**:** choose g uniformly at random from $\mathcal{G}$ subject to the constraint $g(y)_{1\cdots i} = g(f(x))_{1\cdots i}$,[7] and output $A'(g)$.

To analyze the success probability *Inv*, we first rewrite the success probability of A' as follows:

$$\begin{aligned}
\varepsilon_{x,i,r} &\leq \Pr[A'(G) \notin f^{-1}(L(f(x), i))] \\
&= \sum_{y \notin L(f(x),i)} \Pr[A'(G) \in f^{-1}(y)] \\
&= \sum_{y \notin L(f(x),i)} \Pr[G(y)_{1\cdots i} = G(f(x))_{1\cdots i}] \\
&\qquad \cdot \Pr[A'(G) \in f^{-1}(y) | G(y)_{1\cdots i} = G(f(x))_{1\cdots i}] \\
&= 2^{-i} \cdot \sum_{y \notin L(f(x),i)} \Pr[A'(G) \in f^{-1}(y) | G(y)_{1\cdots i} = G(f(x))_{1\cdots i}].
\end{aligned}$$

Above the second equality follows because A is an F-collision finder (so it is always the case that $x' = A'(g) = A(x, g, i)_1$ has the property that $g(f(x'))_{1\cdots i} = g(f(x))_{1\cdots i}$), and the third inequality follows by the two-wise independence of $\mathcal{G}$ ($y \notin L(f(x), i)$ implies that $y \neq f(x)$). Now, we can bound the success probability of *Inv* in finding a preimage of $Y = f(X)$ by:

$$\begin{aligned}
&\Pr[Inv(Y) \in f^{-1}(Y)] \\
&\quad = \sum_y \Pr[Y = y] \cdot \Pr[A'(G) \in f^{-1}(y) | G(y)_{1\cdots i} = f(x)_{1\cdots i}] \\
&\quad \geq \sum_{y \notin L(f(x),i)} \Pr[Y = y] \cdot \Pr[A'(G) \in f^{-1}(y) | G(y)_{1\cdots i} = f(x)_{1\cdots i}] \\
&\quad \geq \frac{1}{2^{i + c \log n}} \cdot \sum_{y \notin L(f(x),i)} \Pr[A'(G) \in f^{-1}(y) | G(y)_{1\cdots i} = f(x)_{1\cdots i}] \\
&\quad \geq \varepsilon_{x,i,r}/n^c,
\end{aligned}$$

where the penultimate inequality holds because every $y \notin L(f(x), i)$ satisfies $\mathrm{H}_{f(X)}(y) < (i + c \cdot \log n)$. □

We have seen that sets $f^{-1}(L(y, i))$ capture the accessible inputs of F; now it remains to show that the expected logarithm of their size is sufficiently smaller than the real entropy $\mathrm{H}(Z|F(Z)) = \mathrm{E}[\log |F^{-1}(F(Z))|]$ (again, this property immediately propagates to L').

Claim 4.10. *For every constant $c > 8$, it holds that*

$$\mathrm{E}\left[\log \left|f^{-1}(L(f(X), I))\right|\right] \leq \mathrm{E}\left[\log \left|F^{-1}(F(Z))\right|\right] - \Omega\left(\frac{c \log n}{n}\right),$$

where $Z = (X, G, I)$ is uniformly distributed in $\mathrm{Dom}(F)$.

[7] This can be done by first choosing $z \xleftarrow{\mathrm{R}} \{0,1\}^{n-i}$ and then using the constructibility of $\mathcal{G}$ to generate a random g such that $g(y) = (g(f(x))_{1\cdots i}, z)$.

Proof. We assume for simplicity that n is a power of 2 (otherwise, we "pad" f) and that c is a power of 2, and let $c' = c/2$. For ease of notation, we will work in entropy units of $c' \log n$. Namely, for $i \in \{0, \cdots, m = n/(c' \log n)\}$ and $y \in \{0,1\}^n$, let $y_{\{1\},\ldots,\{i\}}$ be the first $i \cdot c' \log n$ bits of y, define

$$\mathrm{H}_f(y) := \frac{\mathrm{H}_{f(X)}(y)}{c' \log n}.$$

and let

$$q_i = \Pr[\mathrm{H}_f(f(X)) \in [i, i+1)].$$

Recall that (X, G, I) is uniformly distributed in $\mathrm{Dom}(F)$. We define additional random variables that categorize the "non trivial collisions" induced by F into two separate categories:

$$\begin{aligned}\mathrm{Light} := |\{x' \in \{0,1\}^n \colon & f(x') \neq f(X) \wedge G(f(x'))_{\{1\},\ldots,\{I\}} = G(f(X))_{\{1\},\cdots,\{I\}} \\ & \wedge \mathrm{H}_f(f(x')) \geq I + 2\}|.\end{aligned}$$

Namely, Light consists of the preimages that collide with $f(X)$, different from $f(X)$, and "light" — have few preimages. Similarly, let

$$\begin{aligned}\mathrm{Heavy} := |\{x' \in \{0,1\}^n \colon & f(x') \neq f(X) \wedge G(f(x'))_{\{1\},\ldots,\{I\}} = G(f(X))_{\{1\},\cdots,\{I\}} \\ & \wedge \mathrm{H}_f(f(x')) < I + 2\}|.\end{aligned}$$

Namely, Heavy consists of the preimages that collide with $f(X)$, different from $f(X)$, and "heavy" — have many preimages. Note that

$$\left|F^{-1}(F(Z))\right| = \mathrm{Light} + \mathrm{Heavy} + |f^{-1}(f(X))|$$

(recall that the all elements $F^{-1}(F(x,g,i))$ are of the form (x',g,i)) and

$$\left|f^{-1}(L(f(X),I))\right| \leq \mathrm{Light} + |f^{-1}(f(X))|.$$

Thus, we have

$$\begin{aligned}&\mathrm{E}[\log\left|F^{-1}(F(Z))\right|] - \mathrm{E}[\log\left|f^{-1}(L(f(X),I))\right|] \\ &\quad \geq \mathrm{E}\left[\log \frac{\mathrm{Light} + \mathrm{Heavy} + |f^{-1}(f(X))|}{\mathrm{Light} + |f^{-1}(f(X))|}\right]\end{aligned} \tag{1}$$

We manipulate this as follows:

$$\begin{aligned}&\mathrm{E}\left[\log \frac{\mathrm{Light} + \mathrm{Heavy} + |f^{-1}(f(X))|}{|f^{-1}(f(X))| + \mathrm{Light}}\right] \\ &\geq \mathrm{E}\left[\log\left(1 + \frac{\mathrm{Heavy}}{|f^{-1}(f(X))| + \mathrm{Light} + \mathrm{Heavy}}\right)\right] \\ &\geq \mathrm{E}\left[\frac{\mathrm{Heavy}}{|f^{-1}(f(X))| + \mathrm{Light} + \mathrm{Heavy}}\right],\end{aligned} \tag{2}$$

where the last inequality uses the fact that $\log(1+\alpha) \geq \alpha$ for $\alpha \leq 1$. The proof of Claim 4.10 easily follows from the next claim, which yields that with constant probability, Heavy is a significant term in $(|f^{-1}(f(X))| + \mathrm{Light} + \mathrm{Heavy})$.

Claim 4.11. *Let $\alpha \geq 1$, $i \in \{0, \ldots, m-1\}$ and $x \in \{0,1\}^n$. Condition on $I = i$ and $X = x$, and define the following events (over the random variable G):*

$$\begin{aligned} E_i^1 : &\quad (\text{Light} + \text{Heavy}) \leq 3 \cdot 2^{n - i \cdot (c' \log n)} \\ E_i^2 : &\quad \text{Heavy} \geq (q_{i+1} - \alpha \cdot \sqrt{1/n^{c'}}) \cdot 2^{n - i \cdot c' \log n - 1} \end{aligned}$$

Then $\Pr[E_i^1] \geq 2/3$, *and* $\Pr[E_i^2] \geq 1 - 4/\alpha^2$.

Proof. For E_i^1, we note that $\mathrm{E}[\text{Light} + \text{Heavy}] \leq 2^{n - i \cdot (c' \log n)}$ by two-universality of $\mathcal{G}$, and apply Markov's Inequality.

For E_i^2, let

$$S := \left\{ x' \in \{0,1\}^n \colon f(x') \neq f(x) \wedge \mathrm{H}_f(f(x')) \in [i+1, i+2) \right\}.$$

Note that $|S| \geq (q_{i+1} - \mathrm{neg}(n)) \cdot 2^n$, where we subtract $\mathrm{neg}(n)$ for not taking into account the preimages of $f(x)$. For $g \in \mathcal{G}$, let

$$\begin{aligned} S_g := \{ x' \in \{0,1\}^n \colon & f(x') \neq f(x) \wedge \mathrm{H}_f(f(x')) \in [i+1, i+2) \\ & \wedge\ g(f(x'))_{\{1\},\ldots,\{i\}} = g(f(x))_{\{1\},\cdots,\{i\}} \}, \end{aligned}$$

note that (conditioned on $I = i$ and $X = x$) $\text{Heavy} \geq |S_G|$. We write $|S_g| = \sum_{y \in f(S)} 1_{g,y} \cdot |f^{-1}(y)|$, where $1_{g,y}$ is the indicator for $g(y)_{\{1\},\ldots,\{i\}} = g(f(x))_{\{1\},\cdots,\{i\}}$. By the three-wise independence of $\mathcal{G}$, the $1_{G,y}$'s are pairwise independent Bernoulli random variables, each with expectation $2^{-i \cdot c' \log n}$. Thus, $\mathrm{E}[|S_G|] \geq (q_{i+1} - \mathrm{neg}(n)) \cdot 2^{n - i \cdot c' \log n}$. Assuming that $q_{i+1} > \alpha \cdot \sqrt{1/n^{c'}} \geq \sqrt{1/n^{c'}}$ (as otherwise the claim about E_i^2 holds trivially), it follows that

$$\mathrm{E}[|S_G|] > q_{i+1} \cdot 2^{n - i \cdot c' \log n - 1}$$

By the pairwise independence of $1_{G,y}$'s, we also have

$$\begin{aligned} \mathrm{Var}[|S_G|] &= \sum_{y \in f(S)} \mathrm{Var}[1_{G,y} \cdot |f^{-1}(y)|] \\ &\leq 2^{-i \cdot c' \log n} \cdot \sum_{y \in f(S)} \left| f^{-1}(y) \right|^2 \\ &\leq 2^{-i \cdot c' \log n} \cdot |S| \cdot \max_{y \in f(S)} |f^{-1}(y)| \leq 2^{-i \cdot c' \log n} \cdot 2^n \cdot 2^{n - (i+1) \cdot c' \log n} \\ &= \left(\frac{2}{\sqrt{n^{c'}}} \cdot 2^{n - i \cdot c' \log n - 1} \right)^2, \end{aligned}$$

and thus by Chebyshev inequality

$$\begin{aligned} \Pr[E_i^2] &\geq \Pr\left[|S_G| \geq (q_{i+1} - \alpha \cdot \sqrt{1/n^{c'}}) \cdot 2^{n - i \cdot c' \log n - 1} \right] \\ &\geq 1 - \Pr\left[\left| |S_G| - \mathrm{E}[|S_G|] \right| \geq \frac{\alpha}{2} \cdot \sqrt{\mathrm{Var}[|S_G|]} \right] \geq 1 - \frac{4}{\alpha^2}. \end{aligned}$$

□

Noting that $\mathrm{H}_f(f(X)) \geq i$ means $|f^{-1}(f(X))| \leq 2^{n-i\cdot(c'\log n)}$, and applying Claim 4.11 with $\alpha = 4$, we have

$$\begin{aligned}
&\mathrm{E}[\log|F^{-1}(F(Z))|] - \mathrm{E}[\log\left|f^{-1}(L(Y,I))\right|] \\
&\geq \mathrm{E}\left[\frac{\text{Heavy}}{|f^{-1}(f(X))| + \text{Light} + \text{Heavy}}\right] \\
&\geq \frac{1}{m}\cdot\sum_{i=0}^{m-1}\Pr[\mathrm{H}_f(f(X)) \geq i]\cdot\Pr[E_i^1 \wedge E_i^2 \mid \mathrm{H}_f(f(X)) \geq i] \\
&\quad\cdot\ \frac{\left(q_{i+1} - \frac{4}{n^{c'/2}}\right)\cdot 2^{n-i\cdot c'\log n-1}}{2^{n+2-i\cdot(c'\log n)}} \\
&\geq \frac{1}{m}\cdot\sum_{i=0}^{m-1}(q_{i+1}+\cdots+q_m)\cdot\left(1-\frac{1}{3}-\frac{1}{4}\right)\cdot\left(\frac{q_{i+1}-4/n^{c'/2}}{8}\right) \\
&\geq \frac{1}{48m}\cdot\left(\sum_{j,i\in\{0,\ldots,m\}} q_i\cdot q_j\right) - O\left(\frac{m}{n^{c'/2}}\right) \\
&\geq \frac{1}{48m} - O\left(\frac{m}{n^{c'/2}}\right),
\end{aligned}$$

where the first inequality is by Equation 2, and the third inequality holds since $q_0 = 0$ for every one-way function, which implies that $\sum_{1\leq i\leq j\leq m} q_i\cdot q_j = \sum_{0\leq i\leq j\leq m} q_i\cdot q_j \geq \frac{1}{2}\cdot\sum_{j,i\in\{0,\ldots,m\}} q_i\cdot q_j$. Thus, Claim 4.10 holds with respect to any $c = 2c' \geq 8$. □

By Claims 4.9 and 4.10 and the fact that $F(x,g,i)$ determines g and i, the sets $L'(x,g,i) = f^{-1}(L(f(x),i))\times\{(g,i)\}$ satisfy the properties required to show that the accessible max-entropy of F^{-1} is at most $\mathrm{H}(Z|F(Z)) - \Omega(c(\log n)/n)$. Taking c to be a sufficiently large constant times d, completes the proof. □

5 UOWHFs from Inaccessible Entropy

In this section we show how to construct a UOWHF from any efficiently computable function with a noticeable gap between real Shannon entropy and either accessible average max-entropy or accesssible Shannon entropy. Recall that the more efficient construction from Section 4.2 satisfies the former, and the more direct construction from Section 4.1 satisfies the latter. Combined with these constructions, we obtain two new constructions of UOWHFs from any one-way function.

In both cases, we first transform the entropy gap into a noticeable gap between real Shannon entropy and accessible *max*-entropy. We begin with the construction that starts from a gap between real Shannon entropy and accessible average max-entropy because the transformation involves fewer steps (and is also more efficient).

5.1 The More Efficient UOWHF

Theorem 5.1. *Suppose there exists a polynomial-time computable function $F : \{0,1\}^\lambda \to \{0,1\}^m$ such that F^{-1} has a noticeable gap Δ between real Shannon entropy and accessible average max-*

entropy. Then, there exists a family of universal one-way hash functions with output length $O(\lambda^4 s/\Delta^3)$ *and key length* $O(\lambda^4 s/\Delta^3 \cdot \log n)$ *for any* $s = \omega(\log n)$.

Combining this with Theorem 4.8, from any one way function $f : \{0,1\}^n \to \{0,1\}^n$, we get a UOWHF with output length $O(n^7)$ and key length $\tilde{O}(n^7)$ (where we instantiate the preceding theorem with $\lambda = O(n)$ and $\Delta = \log n/n$).
The construction proceeds via a series of transformations as outlined in Section 1.2; we begin by describing these transformations and establishing the properties they achieve.

Gap Amplification. Here, we show that a direct product construction increases the gap between real entropy and accessible entropy. Another useful effect of direct product (for certain settings of parameters) is turning real Shannon entropy into real min-entropy, and turning accessible average max-entropy into accessible max-entropy.

Lemma 5.2 (Gap amplification)**.** *Let* n *be a security parameter and* $F : \{0,1\}^n \to \{0,1\}^m$ *be a function. For* $t \in \text{poly}(n)$*, let* F^t *be the* t*-fold direct product of* F*. Then,* F^t *satisfies the following properties:*

i. If F^{-1} *has real Shannon entropy at least* k*, then* $(F^t)^{-1}$ *has real min-entropy at least* $t \cdot k - n \cdot \sqrt{st}$ *for any* $s = \omega(\log n)$.

ii. If F^{-1} *has accessible average max-entropy at most* k*, then* $(F^t)^{-1}$ *has accessible max-entropy at most* $t \cdot k + n \cdot \sqrt{st}$ *for any* $s = \omega(\log n)$.

Proof. In the following X and $X^{(t)} = (X_1, \ldots, X_t)$ are uniformly distributed over $\{0,1\}^n$ and $(\{0,1\}^n)^t$ respectively.

i. Follows readily from Lemma 2.3.

ii. Given any PPT F^t-collision-finder A', we construct a PPT F-collision-finder A that:

> On input x, picks a random i in $[t]$ along with random $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_t$, computes $A'(x_1, \ldots, x_t) \to (x'_1, \ldots, x'_t)$, and outputs x'_i.

By the bound on the accessible average max-entropy of F^{-1}, we know that there exists a family of sets $\{L(x)\}$ such that $\mathrm{E}\big[\log |L(X)|\big] \leq k$, $x \in L(x)$, and $\Pr[A(X) \notin L(X)] \leq \text{neg}(n)$. Consider the family of sets $\big\{L'(x^{(t)}) : x^{(t)} \in (\{0,1\}^n)^t\big\}$ given by:

$$L'(x^{(t)}) = L(x_1^{(t)}) \times L(x_2^{(t)}) \times \cdots \times L(x_t^{(t)}).$$

By linearity of expectations, we have $\mathrm{E}\big[\log |L'(X_1, \ldots, X_t)|\big] \leq t \cdot k$. Moreover, by the Chernoff-Hoeffding bound and using the fact that $\log |L(X)|$ assumes values in $[0, n]$, we have

$$\begin{aligned} &\Pr\big[\log |L'(X^{(t)})| \geq t \cdot k + n\sqrt{st}\big] \\ &= \Pr\big[\log |L(X_1^{(t)})| + \cdots + \log |L(X_t^{(t)})| \geq t \cdot k + n\sqrt{st}\big] \leq e^{-2s}. \end{aligned} \tag{3}$$

We claim that this implies that A' has accessible max-entropy at most $t \cdot k + n\sqrt{st}$. Suppose otherwise, then there exists a non-negligible function ϵ such that

$$\Pr[A'(F^t(X^{(t)})) \notin L'(X^{(t)})] \geq \epsilon - e^{-2s} \geq \epsilon/2$$

Therefore,

$$\Pr[A(F(X)) \notin L(X)] = \Pr[A'(F^t(X^{(t)})) \notin L'(X^{(t)})]/t \geq \epsilon/2t$$

which contradicts our assumption on A.

□

Entropy reduction. Next, we describe a construction that given F and any parameter ℓ, reduces the accessible max-entropy of F^{-1} by roughly ℓ bits, while approximately preserving the gap between real min-entropy and accessible max-entropy.

Lemma 5.3 (Reducing entropy)**.** *Let n be a security parameter and $F : \{0,1\}^n \to \{0,1\}^m$ be a function. Fix a family of 2-universal hash functions $\mathcal{G} = \{g : \{0,1\}^n \to \{0,1\}^\ell\}$. Then, $F' : \{0,1\}^n \times \mathcal{G} \to \{0,1\}^m \times \mathcal{G} \times \{0,1\}^\ell$ as given by $F'(x,g) = (F(x), g, g(x))$ satisfies the following properties:*

i. If F^{-1} has real min-entropy at least k, then $(F')^{-1}$ has real min-entropy at least $k - \ell - s$ for any $s = \omega(\log n)$.

ii. If F^{-1} has accessible max-entropy at most k, then $(F')^{-1}$ has accessible max-entropy at most $\max\{k - \ell + s, 0\}$ for any $s = \omega(\log n)$.

Proof. In the following X and G are uniformly distributed over $\{0,1\}^n$ and $\mathcal{G}$ respectively.

i. Fix $g \in G$ and let $S_g = \{z | \Pr[g(X) = z] \leq 2^{-\ell-s}\}$. Observe that:

- $\Pr[g(X) \in S_g] \leq 2^{-s}$ by a union bound over $z \in S_g$;
- In addition, whenever $z \notin S_g$, we have that:

$$\begin{aligned} \Pr[X = x \mid F(X) = F(x), g(x) = z] &\leq \frac{\Pr[X = x \mid F(X) = F(x)]}{\Pr[g(x) = z]} \\ &\leq \frac{\Pr[X = x \mid F(X) = F(x)]}{2^{-\ell-s}}. \end{aligned}$$

Combining the two observations, we have that for all $g \in G$, with probability $1 - 2^{-s}$ over $x \xleftarrow{R} X$, we have $\mathrm{H}_{X|F(X),g(X)}(x|F(x), g(x)) \geq \mathrm{H}_{X|F(X)}(x|F(x)) - (\ell + s)$. The bound on the real min-entropy of F' then follows from the bound on the real min-entropy of F.

ii. Given any PPT F'-collision-finder A', we construct a PPT F-collision-finder A as follows:

On input x, picks a pair (g, r) uniformly at random and output $A'(x, g; r)$.

By the bound on the accessible max-entropy of F^{-1}, we know that there exists a family of sets $\{L(x) \subseteq \{0,1\}^n : x \in \{0,1\}^n\}$ such that $|L(x)| \leq 2^k$, $x \in L(x)$, and

$$\Pr\big[A(X, G; R) \in L(X)\big] \geq 1 - \mathrm{neg}(n), \tag{4}$$

where R is uniformly distributed over the random coins of A. Let $L'(x, g) := \{(x', g) : x' \in L(x) \wedge g(x') = g(x)\}$. Equation 4 yields that

$$\Pr\big[A'(X, G; R) \in L'(X, G)\big] \geq 1 - \mathrm{neg}(n).$$

Next, we bound the size of the set $L'(x,g)$ via 2-universal hashing. Specifically, for all $x \in \{0,1\}^n$ it holds that

$$\Pr\big[|L'(x,G)| \leq 2^{k-\ell+s-1}+1\}\big] \geq 1-2^{-(s-1)},$$

where we are taking into account the possibility that $x \in L(x)$. Combining the last two inequalities, we obtain

$$\Pr\big[A'(X,G;R) \in L'(X,G) \wedge \big|L'(X,G)\big| \leq \max\Big\{2^{k-\ell+s},1\Big\}\big] \geq 1-\mathrm{neg}(n)-2^{-(s-1)}.$$

The above yields an upper bound of $\max\{k-\ell+s\}$, on the accessible max-entropy of $(F')^{-1}$. □

Reducing Output Length. The next transformation gives us a way to derive a function that is both length-decreasing and collision-resistant on random inputs.

Lemma 5.4 (Reducing output length)**.** *Let n be a security parameter and $F:\{0,1\}^n \to \{0,1\}^m$ be a function. Fix a family of pairwise independent hash functions $\mathcal{G} = \{g:\{0,1\}^m \to \{0,1\}^{n-\log n}\}$ and let $F':\{0,1\}^n \times \mathcal{G} \to \{0,1\}^{n-\log n} \times \mathcal{G}$ be defined by $F'(x,g) = (g, g(F(x)))$. The following holds: if F^{-1} has real min-entropy at least $\omega(\log n)$ and F is collision-resistant on random inputs, then F' is collision-resistant on random inputs.*

Proof. The bound on real min-entropy implies that there exists a subset $S \subseteq \{0,1\}^n$ of density at most $\mathrm{neg}(n)$, such that for all $x \notin S$ it holds that $|F^{-1}(F(x))| = n^{\omega(1)}$. Hence, it follows that $|\operatorname{Im} F| \leq \mathrm{neg}(n)\cdot 2^n$. Therefore, $|\operatorname{Im} F| \leq |S| + (|\bar{S}|/n^{\omega(1)}) \leq \mathrm{neg}(n)\cdot 2^n$. Next, observe that by 2-universal hashing,

$$\Pr\big[\exists y' \in \operatorname{Im} F\colon y' \neq F(X) \wedge G(y') = G(F(X))\big] \leq \frac{|\operatorname{Im} F|}{2^{n-\log n}} \leq \mathrm{neg}(n) \tag{5}$$

Namely, $g(F(x))$ uniquely determines $F(x)$ with high probability. In particular, a collision for $g \circ F$ is also a collision for F. Given any PPT F'-collision-finder A', we construct a PPT F-collision-finder A as follows:

> On input x, pick g and r at random and compute $x' = A'(x,g;r)$. If $F(x') = F(x)$, output x', else output x.

Equation 5 implies that $\Pr[A'(X,G;R) \neq (A(X;G,R),G)] \leq \mathrm{neg}(n)$. Therefore, $\Pr[A'(X,G;R) = (X,G)] \geq 1-\mathrm{neg}(n)$. Namely, F' is also collision-resistant on random inputs. □

Additional Transformations. We present two more standard transformations from folklore and previous work that are needed to complete the construction.

Lemma 5.5 (From random inputs to targets, folklore)**.** *Let n be a security parameter and $F:\{0,1\}^n \to \{0,1\}^m$ be a length-decreasing function. Suppose F is collision-resistant on random inputs. Then, $\{F'_y:\{0,1\}^n \to \{0,1\}^m\}_{y\in\{0,1\}^n}$ $F'_y(x) = F(y+x)$ as defined by $F'_y(x) = F(y+x)$ is a family of target collision-resistant hash functions.*

Proof. Given a PPT adversary A' that breaks target collision-resistance of F'_y, we can construct a PPT adversary A that breaks F as follows:

On input x, run $A'(1^n)$ to compute (x_0, state), and then run $A'(\mathsf{state}, x \oplus x_0)$ to compute x_1. Output $x \oplus x_o \oplus x_1$.

Note that (x_0, x_1) is a collision for $F'_{x \oplus x_0}$ iff $(x, x \oplus x_0 \oplus x_1)$ is a collision for F. It then follows quite readily that A breaks F with the same probability that A' breaks F'_y. □

The following result of Shoup [Sho] (improving on [NY, BR]) shows that we can construct target collision-resistant hash functions for arbitrarily long inputs starting from one for a fixed input length.

Lemma 5.6 (Increasing the input length [Sho])**.** *Let n be a security parameter, $t = \mathrm{poly}(n)$ be a parameter and let $\{F_y : \{0,1\}^{n+\log n} \to \{0,1\}^n\}$ be a family of target collision-resistant hash functions. Then, there exists a family of target collision-resistant hash functions $\{F'_{y'} : \{0,1\}^{n+t\log n} \to \{0,1\}^n\}$ where $|y'| = O(|y| \log t)$.*

Putting It Together. Recall that we started out with a function F with a gap Δ between real Shannon entropy and accessible average max-entropy. Let k_{REAL} denote the real Shannon entropy of F^{-1}. Throughout, let $s \in \omega(\log n)$ denote any super-logarithmic function.

STEP 1 **(gap amplification):** Let F_1 be the t-fold direct product of F. That is, $F_1(x_1, \ldots, x_t) = (F(x_1), \ldots, F(x_t))$ where $t \in O(\lambda^2 s/\Delta^2)$. Specifically, we require that

$$t \cdot k_{\mathrm{REAL}} - \lambda \cdot \sqrt{st} \geq t \cdot (k_{\mathrm{REAL}} - \Delta/2) + \lambda \cdot \sqrt{st} + 3s.$$

Lemma 5.2 yields that this repetition increases both the real and accessible entropies of F_1 by a factor of t (comparing to F_0). In addition, this repetition converts real Shannon entropy to real min-entropy and accessible average max-entropy to accessible max-entropy (up to additive terms that are sub-linear in t). More precisely:

- $F_1 : \{0,1\}^{\ell_1^{\mathrm{IN}}} \to \{0,1\}^{\ell_1^{\mathrm{OUT}}}$ where $\ell_1^{\mathrm{IN}}(n) = t \cdot \lambda$ and $\ell_1^{\mathrm{OUT}}(n) = t \cdot m$.
- F_1^{-1} has real min-entropy at least $t \cdot k_{\mathrm{REAL}} - m \cdot \sqrt{st}$, which by our choice of t is at least $t \cdot (k_{\mathrm{REAL}} - \Delta/2) + \lambda \cdot \sqrt{st} + 3s$.
- F_1^{-1} has accessible max-entropy at most $t \cdot (k_{\mathrm{REAL}} - \Delta) + \lambda \cdot \sqrt{st}$.

From the next step on, the construction is given an additional parameter k (a "good" estimate of k_{REAL}) such that $k \in [k_{\mathrm{REAL}}, k_{\mathrm{REAL}} + \Delta/2]$. This means that:

- F_1^{-1} has real min-entropy at least $t \cdot (k - \Delta) + m \cdot \sqrt{st} + 3s$.
- F_1^{-1} has accessible max-entropy at most $t \cdot (k - \Delta) + m \cdot \sqrt{st}$.

That is, there is a gap of $3s$ between real min-entropy and accessible max-entropy, and moreover, we "know" where the gap is (given k).

STEP 2 **(entropy reduction):** Apply entropy reduction to F_1 to obtain F_2. That is, $F_2(x, g) = (F_1(x), g, g(x))$, where $g : \{0,1\}^{\ell_1^{\mathrm{IN}}} \to \{0,1\}^{\ell}$ is selected from a family of 2-universal hash functions, where $\ell = \ell(n,k) = t \cdot (k - \Delta) + \lambda \cdot \sqrt{st} + s = O(t\lambda)$. Lemma 5.3 yields that this additional hashing reduces the real min-entropy and accessible max-entropy by ℓ (up to an additive term of s). More precisely,

- $F_2 : \{0,1\}^{\ell_2^{\text{IN}}} \to \{0,1\}^{\ell_2^{\text{OUT}}}$ where $\ell_2^{\text{IN}}(n,k) = O(t\lambda)$ and $\ell_2^{\text{OUT}}(n,k) = O(t\lambda)$.
- F_2^{-1} has real min-entropy at least s.
- F_2^{-1} has accessible max-entropy at most 0. Hence, F_2 is collision-resistant on random inputs (by Lemma 3.8).

STEP 3 **(reducing the output length):** First reduce the output length of F_2 by hashing the output to $\ell_2^{\text{IN}} - \log n$ bits. That is, $F_3(x,g) = (g, g(F_3(x)))$ where $g : \{0,1\}^{\ell_2^{\text{OUT}}} \to \{0,1\}^{\ell_2^{\text{IN}} - \log n}$ is selected from a family of pairwise-independent hash functions.

- $F_3 : \{0,1\}^{\ell_3^{\text{IN}}} \to \{0,1\}^{\ell_3^{\text{OUT}}}$ where $\ell_3^{\text{IN}}(n,k) = O(t\lambda)$ and $\ell_3^{\text{OUT}}(n,k) = \ell_3^{\text{IN}} - \log n$.
- By Lemma 5.4, F_3 remains collision-resistant on random inputs.

Next, transform F_3 into a family $\{F_y\}$ of target collision-resistant hash functions via a random shift, following Lemma 5.5. That is, $F_y(x) = F_3(y + x)$.

- This yields a non-uniform construction $\{F_y\}$ with input length and key length $\ell_3^{\text{IN}}(n,k) = O(t\lambda) = O(\lambda^3 s/\Delta^2)$, where the non-uniformity corresponds to choice of the parameter $k \in [k_{\text{REAL}}, k_{\text{REAL}} + \Delta/2]$.

STEP 4 **(removing non-uniformity):** To remove the non-uniform advice k, we "try all possibilities" from 0 to λ in steps of size $\Delta/2$, similar to the approach used in [Rom1] (see also [KK, Section 3.6])

i. First, we construct $\kappa = \lambda \cdot 2/\Delta$ families of functions $\{F_y^i\}$ for $i = 1, 2, \ldots, \kappa$, where $\{F_y^i\}$ is the family of functions obtained by instantiating Steps 1 through 3 with the parameter k set to the value $i\Delta/2$. This κ families of functions satisfy the following properties:
   - Each of $F_y^1, \ldots, F_y^\kappa$ is length-decreasing; in particular, F_y^i has input length $\ell_3^{\text{IN}}(n, i\Delta/2)$ and output length $\ell_3^{\text{IN}}(n, i\Delta/2) - \log n$. Note that $\ell_3^{\text{IN}}(n, i\Delta/2) \le \ell_3^{\text{IN}}(n,\lambda)$ for all i because $\ell(n,k)$ increases as a function of k. We may then assume that all κ functions $F_y^1, \ldots, F_y^\kappa$ have the same input length $\ell_3^{\text{IN}}(n,\lambda)$ and the same output length $\ell_3^{\text{IN}}(n,\lambda) - \log n$ by padding "extra part" of the input to the output.
   - At least one of $\{F_y^1\}, \ldots, \{F_y^\kappa\}$ is target collision-resistant; this is because $k_{\text{REAL}} \in [0,\lambda]$ so there exists some i for which $i\Delta/2$ lies between k_{REAL} and $k_{\text{REAL}} + \Delta/2$.

ii. Next, for each $i = 1, 2, \ldots, \kappa$, we construct a family of functions $\{\tilde{F}_y^i\}$ from $\{F_y^i\}$ with input length $\kappa \cdot \ell_3^{\text{IN}}(n,\lambda)$, key length $O(\ell_3^{\text{IN}}(n,\lambda) \cdot \log n)$ and output length $\ell_3^{\text{IN}}(n,\lambda) - \log n$, by following the construction given by Lemma 5.6. Again, at least one of $\{\tilde{F}_{\tilde{y}_1}^1\}, \ldots, \{\tilde{F}_{\tilde{y}_\kappa}^\kappa\}$ is target collision-resistant.

iii. Finally, we define a family of functions $\{F_{\tilde{y}_1,\ldots,\tilde{y}_\kappa}\}$ to be the concatenation of $\tilde{F}_{\tilde{y}_1}^1, \ldots, \tilde{F}_{\tilde{y}_\kappa}^\kappa$ on the same input. That is, $F_{\tilde{y}_1,\ldots,\tilde{y}_\kappa}(x) = \tilde{F}_{\tilde{y}_1}^1(x) \circ \cdots \circ \tilde{F}_{\tilde{y}_\kappa}^\kappa(x)$.
   - Note that F has input length $\kappa \cdot \ell_3^{\text{IN}}(n,\lambda)$ and output length $\kappa \cdot (\ell_3^{\text{IN}}(n,\lambda) - \log n)$, so F is length-decreasing.
   - Moreover, since at least one of $\{\tilde{F}_{\tilde{y}_1}^1\}, \ldots, \{\tilde{F}_{\tilde{y}_\kappa}^\kappa\}$ is target collision-resistant, $\{F_{\tilde{y}_1,\ldots,\tilde{y}_\kappa}\}$ must also be target collision-resistant. This is because a collision for $F_{\tilde{y}_1,\ldots,\tilde{y}_\kappa}$ is a collision for each of $\tilde{F}_{\tilde{y}_1}^1, \ldots, \tilde{F}_{\tilde{y}_\kappa}^\kappa$.

This yields a uniform construction of a UOWHF with output length length $O(\lambda/\Delta \cdot \lambda \cdot \lambda^2 s/\Delta^2) = O(\lambda^4 s/\Delta^3)$. and key length $\tilde{O}(\lambda/\Delta \cdot \lambda \cdot \lambda^2 s/\Delta^2 \cdot \log n) = O(\lambda^4 s/\Delta^3 \cdot \log n)$.

5.2 UOWHF via a Direct Construction

Theorem 5.7. *Suppose there exists a polynomial-time computable function $F : \{0,1\}^\lambda \to \{0,1\}^m$ such that F^{-1} has a noticeable gap Δ between real Shannon entropy and accessible Shannon entropy. Then, there exists a family of universal one-way hash functions with output length $O(\lambda^8 s^2/\Delta^7)$ and key length $O(\lambda^8 s^2/\Delta^7 \cdot \log n)$ for any $s = \omega(\log n)$.*

Combining this with Theorem 4.1, from any one way function $f : \{0,1\}^n \to \{0,1\}^n$, we get a UOWHF with output and key length $\tilde{O}(n^{36})$ (where we instantiate the preceding theorem with $\lambda = O(n)$ and $\Delta = \Theta(1/n^4 \log n)$).

As alluded to earlier, we show how to transform a noticeable gap between real Shannon entropy and accessible Shannon entropy to one between real Shannon entropy and accessible max-entropy, and then follow the construction from the previous section. To achieve this, we first need to establish some additional properties achieved by gap amplification and entropy reduction.

Lemma 5.8 (Gap amplification, continued)**.** *Let n be a security parameter and $F : \{0,1\}^n \to \{0,1\}^m$ be a function. For $t \in \text{poly}(n)$, let F^t be the t-fold direct product of F. Then, F^t also satisfies the following properties:*

i. If F^{-1} has real Shannon entropy at most k, then $(F^t)^{-1}$ has real max-entropy at most $t \cdot k + n \cdot \sqrt{st}$ for any $s = \omega(\log n)$.

ii. If F^{-1} has real min-entropy at least k, then $(F^t)^{-1}$ has real min-entropy at least $t \cdot k$.

iii. If F^{-1} has real max-entropy at most k, then $(F^t)^{-1}$ has real max-entropy at most $t \cdot k$.

iv. If F^{-1} has accessible Shannon entropy at most k, then $(F^t)^{-1}$ has accessible Shannon entropy at most $t \cdot k$.

v. If F^{-1} has accessible max-entropy at most k, then $(F^t)^{-1}$ has accessible max-entropy at most $t \cdot k$.

vi. If F is q-collision-resistant on random inputs and F^{-1} has real max-entropy at most k, then F^{-1} has accessible max-entropy at most $(1 - q/8) \cdot tk + t$, provided that $t = \omega((1/q) \cdot \log n)$.

Proof. Again, X and $X^{(t)} = (X_1, \ldots, X_t)$ are uniformly distributed over $\{0,1\}^n$ and $(\{0,1\}^n)^t$ respectively.

i. Follows readily from Lemma 2.3.

ii. This follows from a union bound and that fact that for all $x_1, \ldots, x_t$:

$$\mathrm{H}_{X^{(t)}}(x_1, \ldots, x_t | F^t(x_1, \ldots, x_t)) \quad = \quad \sum_{i=1}^{t} \mathrm{H}_{X|F(X)}(x_i | F(x_i))$$

iii. Same as previous part.

iv. Given any PPT F^t-collision-finder A', we construct the following PPT F-collision-finder A:

> On input x, pick a random i in $[t]$ along with random $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_t$, compute $A'(x_1, \ldots, x_t) \to (x'_1, \ldots, x'_t)$, and output x'_i.

Define the random variables $(X'_1, \ldots, X'_t) = A'(X_1, \ldots, X_t)$. Then,

$$\begin{aligned}
&\mathrm{H}(X'_1, \ldots, X'_t | X_1, \ldots, X_t) \\
&\quad \leq \mathrm{H}(X'_1|X_1) + \cdots + \mathrm{H}(X'_t|X_t) && \text{subadditivity of conditional Shannon entropy} \\
&\quad = t \cdot \mathrm{H}(X'_I|X_I) && \text{where } I \text{ has the uniform distribution over } [t] \\
&\quad = t \cdot \mathrm{H}(A(X)|X) && \text{by definition of } A \\
&\quad \leq t \cdot k && \text{by the bound on accessible Shannon entropy of } F^{-1}
\end{aligned}$$

v. Analogous to Lemma 5.2 part ii, but simpler, since we do not have to use the Chernoff-Hoeffding bound.

vi. Suppose on the contrary that there exists a PPT F^t-collision-finder A' that violates the guarantee on accessible max-entropy. For $x^{(t)} \in (\{0,1\}^n)^t$, let $B(x^{(t)}) := \left\{x'^{(t)} \in (\{0,1\}^n)^t \colon F^t(x^{(t)}) = F^t(x'^{(t)}) \wedge \left|\left\{i \in [t] \colon x'^{(t)}_i = x^{(t)}_i\right\}\right| \geq qt/8\right\}$. By the bound on real max-entropy, we have that $\Pr[\exists i \in [t] \colon |F^{-1}(F(X^{(t)}_i))| > 2^k] \leq t \cdot \mathrm{neg}(n) = \mathrm{neg}(n)$. Hence,

$$\Pr\left[|B(X^{(t)})| > \binom{t}{qt} 2^{(1-q/8)tk}\right] \leq \mathrm{neg}(n) \tag{6}$$

Since A' achieves accessible max-entropy greater than $(1 - q/8)tk + t$, there must exists a non-negligible function ϵ such that $\Pr[A'(X^{(t)}; R') \notin B(X^{(t)})] \geq \epsilon - t \cdot \mathrm{neg}(n) \geq \epsilon/2$, where R' is uniformly distributed over the random coins of A'. Namely, A' finds collisions on at least a $1 - q/8$ fraction of the coordinates with non-negligible probability.

Since F is q-collision resistant, this violates a standard Chernoff-type direct product theorem. We now give a self-contained proof, following a similar analysis done for standard collision resistance in [CRS$^+$]. Consider the following PPT F-collision-finder A:

> On input $x \in \{0,1\}^n$, pick a random $i \in [t]$ along with random $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_t$, compute $A'(x_1, \ldots, x_t) \to (x'_1, \ldots, x'_t)$, and output x'_i.

To analyze the success probability of A', fix any subset S of $\{0,1\}^n$ of density $q/2$. If $t = \omega(\log n/q)$, then a Chernoff bound yields that

$$\Pr[A'(X^{(t)}) \notin B(X^{(t)}) \wedge |\{i \in [t] \colon X^{(t)}_i \in S\}| \geq q/4] \geq \epsilon/4.$$

This means that

$$\Pr_{i \xleftarrow{R} [t]}[A'(X^{(t)}) \to (X'_1, \ldots, X'_t) \wedge X_i \in S \wedge X'_i \neq X_i] \geq \epsilon/4 \cdot q/8.$$

We may then deduce (following the same calculations in [CRS$^+$, Prop 2]) that

$$\Pr_{x \xleftarrow{R} X}\left[\Pr[A(x; R) \neq x] \geq \epsilon/4 \cdot q/8 \cdot 2/q\right] \geq 1 - q/2.$$

where R is uniformly distributed over the random coins of A. By repeating A a sufficient number of times, we may find collisions on random inputs of F with probability $1 - q$, contradicting our assumption that F is q-collision-resistant on random inputs.

□

Lemma 5.9 (Reducing entropy, continued)**.** *Let n be a security parameter and $F : \{0,1\}^n \to \{0,1\}^m$ be a function. Fix a family of 2-universal hash functions $\mathcal{G} = \{g : \{0,1\}^n \to \{0,1\}^\ell\}$. Then, $F' : \{0,1\}^n \times \mathcal{G} \to \{0,1\}^m \times \mathcal{G} \times \{0,1\}^\ell$ as given by $F'(x,g) = (F(x), g, g(x))$ satisfies the following properties:*

i. If F^{-1} has real max-entropy at most k, then $(F')^{-1}$ has real max-entropy at most $\max\{k - \ell + s, 0\}$ for any $s = \omega(\log n)$.

ii. If F^{-1} has p-accessible max-entropy at most k, then $(F')^{-1}$ has $p + 2^{-\Omega(s)}$-accessible max-entropy at most $\max\{k - \ell + s, 0\}$ for any s.

Proof. In the following X and G are uniformly distributed over $\{0,1\}^n$ and $\mathcal{G}$ respectively.

i. Fix an x such that $|F^{-1}(F(x))| \leq 2^k$. By 2-universal hashing,

$$\mathrm{E}\big[|G^{-1}(G(x)) \cap (F^{-1}(F(x)) \setminus \{x\})|\big] \leq (2^k - 1) \cdot 2^{-\ell} \leq 2^{k-\ell}.$$

The bound on the real max-entropy of F^{-1} and the Markov's inequality, yield that

$$\Pr\big[|G^{-1}(G(X)) \cap (F^{-1}(F(X)) \setminus \{x\})| \geq 2^{(s-1)} \cdot 2^{k-\ell}\big] \leq 2^{-(s-1)} + \mathrm{neg}(n).$$

The bound on the real max-entropy of $(F')^{-1}$ follows.

ii. Readily follows from the proof of Lemma 5.3 part ii.

□

Putting everything together. Recall that we started out with a function F with a gap Δ between real Shannon entropy and accessible Shannon entropy. Let k_{REAL} denote the real Shannon entropy of F^{-1}.

STEP 1 **(gap amplification):** Let F_1 be the t-fold direct product of F for a sufficiently large t to be determined later. That is, $F_1(x_1, \ldots, x_t) = (F(x_1), \ldots, F(x_t))$.

Lemma 5.2 yields that this repetition increases both the real and accessible entropies of F_1 by a factor of t. In addition, the repetition converts real Shannon entropy to real min-entropy and real max-entropy (up to an additive $o(t)$ term). More precisely:

- $F_1 : \{0,1\}^{\ell_1^{\text{IN}}} \to \{0,1\}^{\ell_1^{\text{OUT}}}$ where $\ell_1^{\text{IN}}(n) = t \cdot \lambda$ and $\ell_1^{\text{OUT}}(n) = t \cdot m$.
- F_1^{-1} has real min-entropy at least $t \cdot k_{\text{REAL}} - \lambda\sqrt{st}$ and real max-entropy at most $t \cdot k_{\text{REAL}} + \lambda\sqrt{st}$.
- F_1^{-1} has accessible Shannon entropy at most $t \cdot k_{\text{REAL}} - t\Delta$.

From the next step on, the construction is given an additional parameter k (a "good" estimate of k_{REAL}) such that $k \in [k_{\text{REAL}}, k_{\text{REAL}} + \Delta^2/128\lambda]$. This means that:

- F_1^{-1} has accessible Shannon entropy at most $tk - t\Delta$. Lemma 3.9 yields that F_1^{-1} has $(1 - \Delta/4k)$-accessible max-entropy at most $tk - t\Delta/2$.

STEP 2 **(entropy reduction):** Apply entropy reduction to F_1 with $\ell = \ell(n,k) = tk - t\Delta/2 + s$ to obtain F_2. That is, $F_2(x,g) = (F_1(x), g, g(x))$, where $g : \{0,1\}^{\ell_1^{\text{IN}}} \to \{0,1\}^{\ell}$ is selected from a family of 2-universal hash functions.

By Lemma 5.3 and Lemma 5.9, this reduces the accessible max-entropy to 0, which allows us to deduce that F_2 is weakly collision-resistant on random inputs.

- $F_2 : \{0,1\}^{\ell_2^{\text{IN}}} \to \{0,1\}^{\ell_2^{\text{OUT}}}$ where $\ell_2^{\text{IN}}(n,k) = O(t\lambda + \ell(n,k)) = O(t\lambda)$ and $\ell_2^{\text{OUT}}(n,k) = O(tm + \ell(n,k)) = O(t\lambda)$.
- F_2^{-1} has real min-entropy at least $t \cdot (k_{\text{REAL}} - k + \Delta/2) - \lambda\sqrt{st} - 2s$, which is at least
$$t \cdot (\Delta/2 - \Delta^2/128\lambda) - \lambda\sqrt{st} - 2s$$
and real max-entropy at most $t \cdot (k_{\text{REAL}} - k + \Delta/2) + \lambda\sqrt{st} \le t \cdot \Delta/2 + \lambda\sqrt{st}$.
- F_2^{-1} has $(1 - \Delta/4k + 2^{-\Omega(s)})$-accessible max-entropy at most 0. Thus, F_2 is q-collision-resistant on random inputs(by Lemma 3.8), for $q = \Delta/4k - 2^{-\Omega(s)}$.

STEP 3 **(gap amplification):** F_3 is t'-fold direct product of F_2, where $t' = s/q = O(ks/\Delta)$. That is, $F_3(x_1, \ldots, x_{t'}) = (F_2(x_1), \ldots, F_2(x_{t'}))$.

By Lemma 5.8, this allows us to amplify the weak collision-resistance property of F_2 to obtain a gap between real min-entropy and accessible max-entropy in F_3.

- F_3^{-1} has real min-entropy at least
$$t' \cdot \big(t \cdot (\Delta/2 - \Delta^2/128\lambda) - \lambda\sqrt{st} - 2s\big).$$
- F_3^{-1} has accessible max-entropy at most $t' \cdot \big((1 - q/8) \cdot (t\Delta/2 + \lambda\sqrt{st}) + 1\big)$, which is at most:
$$t' \cdot \big(t \cdot (\Delta/2 - \Delta q/16) + \lambda\sqrt{st}) + 1\big).$$
Now, $k \le \lambda$, so $q = \Delta/4k - 2^{-\Omega(s)} \ge \Delta/4\lambda - 2^{-\Omega(s)}$. This means F_3^{-1} has accessible max-entropy at most:
$$t' \cdot \big(t \cdot (\Delta/2 - \Delta^2/64\lambda + 2^{-\Omega(s)}) + \lambda\sqrt{st}) + 1\big).$$

Note that the gap is at least $t' \cdot \big(t \cdot \Delta^2/128\lambda - 2^{-\Omega(s)} - (2\lambda\sqrt{st} + 2s + 1)\big)$, which is at least $3s$ as long as:
$$t \cdot \Delta^2/128\lambda \ge 2^{-\Omega(s)} + 2\lambda\sqrt{st} + 2s + 1 + 3s/t'$$
Since $3s/t' = 3q \le 3\Delta$, we can set $t = O(\lambda/\Delta + \lambda s/\Delta^2 + \lambda^4 s/\Delta^4) = O(\lambda^4 s/\Delta^4)$ so that F_3^{-1} has a gap of $3s$ between real min-entropy and accessible max-entropy, and moreover, we know where this gap is (given k).

STEPS 4/5/6: We follow steps 2, 3, and 4 in the previous construction, with the following modifications in the parameters:

- We apply entropy reduction first, with
$$\ell = t' \cdot \big(t \cdot (\Delta/2 - \Delta q/16) + \lambda\sqrt{st}) + 1\big) + s.$$

- To remove the non-uniform advice k, we "try all possibilities" from 0 to λ in steps of size $\Delta^2/128\lambda$.

We then obtain a non-uniform construction of UOWHFs with output and key length $O(\lambda \cdot t \cdot t') = O(\lambda^6 s^2/\Delta^5)$, since $t = O(\lambda^4 s/\Delta^4)$ and $t' = O(\lambda s/\Delta)$. We also obtain a uniform construction with output length $O(\lambda/(\Delta^2/\lambda) \cdot \lambda \cdot t \cdot t' \cdot \log n) = O(\lambda^8 s^2/\Delta^7)$ and key length $O(\lambda^8 s^2/\Delta^7 \cdot \log n)$.

Acknowledgements

We are thankful to Ran Raz and Chiu-Yuen Koo for useful conversations.

References

[BR] M. Bellare and P. Rogaway. Collision-Resistant Hashing: Towards Making UOWHFs Practical. In *CRYPTO*, pages 470–484, 1997.

[CRS+] R. Canetti, R. L. Rivest, M. Sudan, L. Trevisan, S. P. Vadhan, and H. Wee. Amplifying Collision Resistance: A Complexity-Theoretic Treatment. In *CRYPTO*, pages 264–283, 2007.

[CT] T. M. Cover and J. A. Thomas. *Elements of information theory*. Wiley-Interscience, New York, NY, USA, second edition, 2006.

[CS] R. Cramer and V. Shoup. Design and analysis of practical public-key encryption schemes secure against adaptive chosen ciphertext attack. *SIAM Journal on Computing*, 33(1):167–226 (electronic), 2003.

[HNO+] I. Haitner, M. Nguyen, S. J. Ong, O. Reingold, and S. Vadhan. Statistically Hiding Commitments and Statistical Zero-Knowledge Arguments from Any One-Way Function. *SIAM Journal on Computing*, 39(3):1153–1218, 2009.

[HRV] I. Haitner, O. Reingold, and S. Vadhan. Efficiency Improvements in Constructions of Pseudorandom Generators. In *Proceedings of the 42th Annual ACM Symposium on Theory of Computing (STOC)*. ACM Press, 2010.

[HRVW] I. Haitner, O. Reingold, S. Vadhan, and H. Wee. Inaccessible Entropy. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing (STOC)*. ACM Press, 2009.

[HILL] J. Håstad, R. Impagliazzo, L. A. Levin, and M. Luby. A pseudorandom generator from any one-way function. *SIAM Journal on Computing*, 28(4):1364–1396, 1999. Preliminary versions in *STOC'89* and *STOC'90*.

[Hol] T. Holenstein. Rompels Construction in Average-Case Complexity. Unpublished manuscript, February 2009.

[IL] R. Impagliazzo and M. Luby. One-way Functions are Essential for Complexity Based Cryptography. In *Proceedings of the 30th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 230–235, 1989.

[KK] J. Katz and C. Koo. On Constructing Universal One-Way Hash Functions from Arbitrary One-Way Functions. Technical Report 2005/328, Cryptology ePrint Archive, 2005.

[NY] M. Naor and M. Yung. Universal One-Way Hash Functions and their Cryptographic Applications. In *Proceedings of the 21st Annual ACM Symposium on Theory of Computing (STOC)*, pages 33–43. ACM Press, 1989.

[NZ] N. Nisan and D. Zuckerman. Randomness is Linear in Space. *Journal of Computer and System Sciences*, 52(1):43–52, 1996.

[RW] R. Renner and S. Wolf. Smooth Renyi Entropy and Applications. In *IEEE International Symposium on Information Theory — ISIT 2004*, page 233. IEEE, June 2004.

[Rom1] J. Rompel. One-Way Functions are Necessary and Sufficient for Secure Signatures. In *Proceedings of the 22nd Annual ACM Symposium on Theory of Computing (STOC)*, pages 387–394, 1990.

[Rom2] J. Rompel. *Techniques for computing with low-independence randomness*. PhD thesis, Massachusetts Institute of Technology, 1990. `http://dspace.mit.edu/handle/1721.1/7582`.

[Sho] V. Shoup. A Composition Theorem for Universal One-Way Hash Functions. In *EUROCRYPT*, pages 445–452, 2000.

[Vad] S. P. Vadhan. Constructing Locally Computable Extractors and Cryptosystems in the Bounded-Storage Model. *Journal of Cryptology*, 17(1):43–77, January 2004. Extended abstract in *CRYPTO '03*.