UOWHFs from OWFs: Trading Regularity for Efficiency

Kfir Barhum and Ueli Maurer

Department of Computer Science, ETH Zurich, 8092 Zurich, Switzerland
{barhumk,maurer}@inf.ethz.ch

Abstract. A universal one-way hash function (UOWHF) is a shrinking function for which finding a second preimage is infeasible. A UOWHF, a fundamental cryptographic primitive from which digital signature can be obtained, can be constructed from any one-way function (OWF). The best known construction from any OWF $f : \{0,1\}^n \to \{0,1\}^n$, due to Haitner et. al. [2], has output length $\tilde{O}(n^7)$ and $\tilde{O}(n^5)$ for the uniform and non-uniform models, respectively. On the other hand, if the OWF is known to be injective, i.e., maximally regular, the Naor-Yung construction is simple and practical with output length linear in that of the OWF, and making only one query to the underlying OWF.

In this paper, we establish a trade-off between the efficiency of the construction and the assumption about the regularity of the OWF f. Our first result is a construction comparably efficient to the Naor-Yung construction but applicable to any close-to-regular function. A second result shows that if $|f^{-1}f(x)|$ is concentrated on an interval of size $2^{s(n)}$, the construction obtained has output length $\tilde{O}(n \cdot s(n)^6)$ and $\tilde{O}(n \cdot s(n)^4)$ for the uniform and non-uniform models, respectively.

Keywords: Complexity-Based Cryptography, One-Way Functions, Universal One-Way Hash Functions, Computational Entropy.

1 Introduction

1.1 Constructions of Cryptographic Primitives

A main task in cryptographic research is to construct a (strong) cryptographic primitive P from a (weaker) cryptographic primitive Q, for example to construct a pseudo-random generator from a one-way function (OWF). This paper is concerned with constructing a universal one-way hash function (UOWHF), a fundamental cryptographic primitive, from a OWF.

The term "construct" means that one gives an efficient reduction of the problem of breaking the underlying primitive Q to the problem of breaking the constructed primitive P. For two primitives P and Q, the most basic question is whether P can be constructed *in principle* from Q, meaning that the construction and the reduction must be efficient (i.e., polynomial-time) and that the reduction translates a non-negligible probability of breaking P into a non-negligible probability of breaking Q.

A. Hevia and G. Neven (Eds.): LATINCRYPT 2012, LNCS 7533, pp. 234–253, 2012.

The principle possibility of constructing a UOWHF from a OWF was proved by Rompel [7], using a highly inefficient construction and reduction. When trying to improve the construction, one can choose two orthogonal routes. Either one improves the construction for a general OWF, or one makes specific assumptions about the OWF allowing for special-purpose constructions that do not necessarily work in general, and can hence be more efficient. Of course, a key issue is how restrictive or how plausible the assumption one has to make is.

The best known *general* construction of a universal one-way hash function from any one-way function $f : \{0,1\}^n \to \{0,1\}^m$, due to Haitner et. al. [2], has output length $\tilde{O}(n^7)$ and $\tilde{O}(n^5)$ for the uniform and non-uniform cases, respectively. The best known special-purpose construction is due to Naor and Yung [6] and makes a single call to f (per argument to the constructed UOWHF), and the output length is linear in n, but the assumption one needs to make is that f is *injective*, which is a very strong assumption.

In this paper we investigate the middle grounds between completely general constructions and those requiring such a very specific assumption. Concretely, we investigate the trade-off between the regularity assumption for f and the efficiency of the construction. The regularity is characterized by how concentrated the *preimage size spectrum*, the random variable $|f^{-1}(f(X))|$ corresponding to the preimage size of the function value $f(X)$ for a uniformly random argument X, is. For injective functions, the preimage size spectrum is constant 1. Prior to our work, we do not know of any specific construction for a function which is anywhere between regular and arbitrary.

In this work we relate the assumptions made about the spectrum of f to the efficiency of the overall construction. Qualitatively speaking, the more is assumed about the regularity of f, the more efficient is the resulting construction.

1.2 Contributions of This Paper

A first result on the way to fully utilizing an assumption about the regularity of a function is an almost optimal construction of a universal one-way hash function from a regular (or almost regular) one-way function. Recall that a function is 2^r-regular if for every image there are 2^r preimages.

Following previous work, for simplicity of presentation, we assume that for a one-way function f the input length n is the security parameter. For this case, we get a construction with output length and key length $O(n \cdot \alpha(n) \cdot \log(n))$, where the construction makes $O(\alpha(n) \cdot \log(n))$ invocations to f for any super-constant function $\alpha(n)$. This improves on [8] by a factor of $\log(n)$ (see Section 1.3 for comparison with previous work).

We introduce a natural relaxation of the notion of regularity:

Definition 1 (roughly-regular function). *A function $f : \{0,1\}^n \to \{0,1\}^m$ is called (r,s)-*roughly-regular*, if for every x in $\{0,1\}^n$ it holds that $|f^{-1}(f(x))|$ lies in the interval $[r, rs]$. A family of functions $f = \{f_n\}_{n>0}$ is called (r,s)-*roughly-regular*, where $(r,s) = (r(n), s(n))$, if for every n it holds that f_n is $(r(n), s(n))$-roughly regular. Whenever $s(n) = n^c$ for some constant c the family is called r-*polynomially-roughly-regular*.*

We call r and s the *regularity* and the *roughness* parameters of f, respectively. Indeed, whenever the roughness parameter is trivial, that is, $s(n) = 1$ for all n, this definition coincides with the standard definition of an r-regular function. This definition, we argue, is both intuitive and quantifies the irregularity of a function.

For the case where f is a 2^r-polynomially-roughly-regular OWF, we observe that the construction for the 2^r-regular case with minor changes works (we omit the details in this extended abstract). For a pseudo-random generator based on a regular OWF an analog relaxation was already observed by [1].

Finally, in Section 4 we utilize the ideas developed in Section 3 and improve on [2] with the most general version (Theorem 3). We establish a trade-off between the regularity assumption made about the underlying one-way function and the overall efficiency of the construction. When f is a $(2^{r(n)}, 2^{s(n)})$-roughly regular one-way function, we show a construction with output length and key length of $\tilde{O}(n \cdot s^4)$ for the non-uniform model and of $\tilde{O}(n \cdot s^6)$ for the uniform model. Indeed, our construction ties up both ends of the existing constructions: When s is constant, we get an almost linear construction, and when $s = O(n)$ our construction matches that of [2].

The analysis of the construction presented in Section 3 improves by a factor of $O(\log^2(n))$ on the construction presented in Section 4 when instantiated with a 2^r-regular function.

1.3 Related Work

Inaccessible Entropy. Our work uses the framework of [2] for constructing UOWHFs from OWFs using the notion of inaccessible entropy. Inaccessible entropy was first introduced in [5] and along with work done in [3] and [4], it completes the construction of the fundamental cryptographic primitives: universal one-way hash functions, pseudo-random generators and commitment schemes using this notion.

A Regularity-Efficiency Trade-Off for the Construction of a UOWHF. In [8] it was first shown how to construct a UOWHF for the almost-regular case. Our construction achieves the same query complexity to the underlying one-way function ($O(\alpha(n) \cdot \log(n))$ calls), but is superior in two aspects: It makes its queries to the underlying one-way function in a non-adaptive manner, and our resulting primitive has an output (and seed) length of $n \cdot log(n) \cdot \alpha(n)$ whereas the construction from [8] has an additional $\log(n)$ factor.

While for the almost-regular case the improvement is not dramatic, we believe that our analysis, which extends the approach suggested in [2], sheds more light on what is achieved at each step. The way the almost-regularity property of the underlying one-way function is utilized later allows to generalize it to any level of regularity. This is in contrast to the construction in [8] which is more ad-hoc.

2 Preliminaries

2.1 Notations and Basics

Throughout the paper we use capital letters to denote random variables and small letters for specific values they assume. We denote by $\mathcal{N}$ the set of natural numbers. For an integer n we denote by $[n]$ the set $\{1, \ldots, n\}$. For two bit-strings x and y we denote their concatenation by $x \| y$. For a random variable X we denote by $\mathbf{E}[X]$ and $\mathbf{V}[X]$ its expectation and variance, accordingly. For an event A we denote its indicator random variable (which assumes the value 1 whenever A happens and 0 otherwise) by $\mathbb{1}_A$, and its complement event by $\overline{A}$. We implicitly make use of the fact that $\mathbf{E}[\mathbb{1}_A] = \Pr[A]$. The support of a random variable X is defined as $\mathrm{Supp}(X) = \{x : \Pr[X = x] > 0\}$. For a function $f : X \to Y$, we define the preimage spectrum function $\pi_f : X \to \mathcal{N}$, where $\pi_f(x) = |f^{-1}(f(x))|$.

For understood $Y_1 \times \cdots \times Y_n$ we denote by $\phi_i : Y_1 \times \cdots \times Y_n \to Y_i$ the projection onto the i'th component. We extend this to a set $S \subseteq Y_1 \times \cdots \times Y_n$ with $\phi_i(S) \stackrel{\text{def}}{=} \{\phi_i(s) : s \in S\}$. A non-decreasing function $f : \mathcal{N} \to \mathcal{N}$ is called super-constant if for all $c \in \mathcal{N}$ there exists an $n \in \mathcal{N}$, such that $f(n) > c$. All log functions are to the base 2.

We cite the Hoeffding bound and bring the definition of a t-wise independent hash family in Appendix A.

2.2 OWF and UOWHF

Definition 2 (OWF). *A family of functions $\{f : \{0,1\}^{n(\rho)} \to \{0,1\}^{m(\rho)}\}_{\rho \in \mathcal{N}}$, where ρ is a security parameter, is a* one-way function *if:*

1. *There exists an efficient algorithm that, given x, evaluates $f(x)$.*
2. *For any efficient randomized algorithm A:*

$$\Pr_{x \xleftarrow{r} \{0,1\}^{n(\rho)}} \left[A(1^\rho, f(x)) \in f^{-1}(f(x))\right] \leq \mathsf{negl}(\rho) \ .$$

Definition 3 (UOWHF). *A family of keyed functions $\{\{f_k : \{0,1\}^{n(\rho)} \to \{0,1\}^{m(\rho)}\}_{k \in K(\rho)}\}_{\rho \in \mathcal{N}}$, where ρ is a security parameter, is a* universal one-way hash function *if:*

1. *There exists an efficient algorithm that, given x and k, evaluates $f_k(x)$.*
2. *$m(\rho) < n(\rho)$.*
3. *For any pair of efficient randomized algorithms (A_1, A_2):*

$$\Pr_{k \xleftarrow{r} K(\rho), (x,\sigma) \xleftarrow{r} A_1(1^\rho)} \left[A_2(x, k, \sigma) \in {f_k}^{-1}(f_k(x)) \setminus \{x\}\right] \leq \mathsf{negl}(\rho) \ .$$

As noted, we focus on families of functions where the input domain parameter n equals to the security parameter, i.e., $n(\rho) = \rho$, in which case we parameterize the family by n. Additionally, we slightly abuse notation when referring to a function $f : \{0,1\}^n \to \{0,1\}^{m(n)}$, where formally $f = \{f_n : \{0,1\}^n \to \{0,1\}^{m(n)}\}_{n \in \mathcal{N}}$ is a parametrized family of functions, and often we omit the security parameter when referring to f_n or other parametrized values.

2.3 Entropy Measures

For a random variable X and $x \in \mathrm{Supp}(X)$ the point-wise entropy of X is $H_X(x) \stackrel{\text{def}}{=} -\log(\Pr[X = x])$. The Shannon entropy $H(X)$ and min-entropy $H_\infty(X)$ of X are defined as:

$$H(X) \stackrel{\text{def}}{=} \mathbf{E}\left[H_X(X)\right] \ , \qquad H_\infty(X) \stackrel{\text{def}}{=} -\log\left(\max_{x \in \mathrm{Supp}(X)} \Pr[X = x]\right) \ .$$

These measures extend naturally to the case of a joint distribution of two random variables X, Y. Namely, the conditional point-wise entropy for $(x, y) \in \mathrm{Supp}(X, Y)$ is $H_{X|Y}(x, y) \stackrel{\text{def}}{=} -\log(\Pr[X = x|Y = y])$ and the conditional Shannon entropy is

$$H(X|Y) = \mathop{\mathbf{E}}_{(x,y) \stackrel{r}{\leftarrow} (X,Y)} [H_{X|Y}(x, y)] = \mathop{\mathbf{E}}_{y \stackrel{r}{\leftarrow} Y} [H(X|Y = y)] = H(X, Y) - H(Y).$$

The next definition measures the average and absolute guarantees as for the preimage-size of f in terms of entropy bits.

Definition 4 (preimage entropy measures). *For a function $f : \{0,1\}^n \to \{0,1\}^{m(n)}$, define its* real preimage-entropy *as $H_p(f) \stackrel{\text{def}}{=} H(X|f(X))$, where X is uniformly distributed on $\{0,1\}^n$. f has* min-preimage-entropy *at least $k = k(n)$ (and denote this by $H_{p,\min}(f) \geq k$), if there is a negligible function $\epsilon = \epsilon(n)$ such that*

$$\Pr_{x \stackrel{r}{\leftarrow} \{0,1\}^n} [H_{X|f(X)}(x, f(x)) \geq k] \geq 1 - \epsilon.$$

As the argument X in the definition is uniform, we have that for all x it holds that $H_{X|f(X)}(x, f(x)) = \log(\pi_f(x))$.

2.4 Collision Finders and Accessible Entropy

Definition 4 captures the average and absolute preimage set size guarantees for f. Clearly, when f is shrinking it has high preimage-entropy. Recall that our goal is to build a universal one-way hash function, namely, a shrinking function for which there exist many preimages, but at the same time any efficient algorithm, when given an x, cannot compute a different preimage from $f^{-1}(f(x))$.

Definition 5 (f-collision-finder). *Let $f : \{0,1\}^n \to \{0,1\}^{m(n)}$ be a function. An f-collision-finder is a randomized algorithm A such that $A(x) \in f^{-1}(f(x))$ for every $x \in \{0,1\}^n$.*

The requirement that $A(x)$ outputs a preimage of $f(x)$ can be made without loss of generality, as every algorithm A can be changed to one that outputs x whenever $A(x) \notin f^{-1}(f(x))$.

Using the notion of an f-collision-finder, one can define a computational analog of the definitions of real- and min-preimage-entropy of f. The analogous definitions capture the maximal, average, and absolute size of the preimage sets that are accessible to any efficient algorithm.

Definition 6. *A function* $f : \{0,1\}^n \to \{0,1\}^{m(n)}$ *has* accessible max-preimage-entropy *at most* $k = k(n)$ *if there exists a family of sets* $\{S_x\}_{x\in\{0,1\}^n}$ *such that for any efficient randomized* f*-collision-finder* A*, there exists a negligible function* $\epsilon = \epsilon(n)$ *such that for all sufficiently large* n*:*

1. $\Pr_{x\xleftarrow{r}\{0,1\}^n}[A(x) \in S_x] \geq 1 - \epsilon$.
2. $\log(|S_x|) \leq k$ *for all* x.

Definition 7. *A function* $f : \{0,1\}^n \to \{0,1\}^{m(n)}$ *has* accessible average max-preimage-entropy *at most* $k = k(n)$ *if it satisfies Definition 6 where instead of (2.) we have:*

2. $\mathbf{E}_{x\xleftarrow{r}\{0,1\}^n}[\log(|S_x|)] \leq k$.

We stress that these two definitions[1] are different from the classical definitions of Shannon entropy. As they capture the inputs accessible only to *efficient* algorithms, both definitions only bound from above the performance of such algorithms. Specifically, for an arbitrary function, we do not know how to compute exactly (as in the standard definition of entropy) these bounds. Nevertheless, as we see next, these bounds are a useful tool (see also [5]). We use the notation $H^{\text{eff}}_{p,\max}(f) \leq k$ and $H^{\text{eff}}_{p,\text{avg}-\max}(f) \leq k$ to denote that the corresponding bound holds.

The next two definitions are used to distinguish between two types of 'entropy gaps':

Definition 8. *A function* $f : \{0,1\}^n \to \{0,1\}^{m(n)}$ *has an* average inaccessible preimage-entropy gap $\Delta = \Delta(n)$*, if there exists some* $k = k(n)$ *such that:*

$$H^{\text{eff}}_{p,\text{avg}-\max}(f) \leq k \leq k + \Delta \leq H_p(f) \ . \tag{1}$$

That is, there is a gap of Δ between its average accessible max-preimage-entropy and its preimage-entropy. At times we will refer to this gap as an average entropy gap or a weak type of gap.

Definition 9. *A function* $f : \{0,1\}^n \to \{0,1\}^{m(n)}$ *has an* absolute inaccessible preimage-entropy gap $\Delta = \Delta(n)$*, if there exists some* $k = k(n)$ *such that:*

$$H^{\text{eff}}_{p,\max}(f) \leq k \leq k + \Delta \leq H_{p,\min}(f) \ . \tag{2}$$

[1] In fact, one may consider a weaker notion of algorithm-dependent accessible max-preimage-entropy and algorithm-dependent accessible average max-preimage-entropy where the sets $\{S_x\}$ may also depend on the algorithm. Such a definition would only require that for every algorithm there exist sets $\{S_{A,x}\}$. This weaker variant of Definitions 6 and 7 is enough for the purpose of constructing a universal one-way hash function and potentially may be easier to satisfy. In this work we do not make use of the weaker definition.

At times we will refer to this gap as an absolute or strong gap.

An important observation is that UOWHFs are just length-decreasing functions with accessible max-preimage-entropy 0, and an appropriate absolute entropy gap. Haitner et. al. observed that it is possible to achieve a noticeable gap of inaccessible entropy as an intermediate step, and then amplify it and transform it into a UOWHF.

2.5 Entropy Measures for t-fold Parallel Repetitions

For a function $f : X \to Y$ we define its t-fold parallel repetition $f^t : X^t \to Y^t$ as $f^t(x_1, \ldots, x_t) = (f(x_1), \ldots, f(x_t))$. It is well-known that using the definition of conditional entropy, properties of $\log(\dot{})$ and noting that choosing a random $x^t \in X^t$ can be done by t independent choices of x,

$$H_p(f^t) = H(X_1, \ldots, X_t | f(X_1), \ldots, f(X_t)) = H(X_1|f(X_1)) + \ldots \\ +H(X_t|f(X_t)) = t \cdot H_p(f) \ . \qquad (3)$$

The corresponding computational bound is given by the following claim and its corollary. Namely, the accessible preimages of the t-fold repetition of f come from the product set of the accessible preimages set of f:

Lemma 1. *Let $f : X \to X$ with accessible max-preimage-entropy at most $k(n)$, with sets S_x (as in Definition 6). Then for $t = \mathrm{poly}(n)$ any efficient f^t-collision-finder A' outputs a collision (except with negligible probability) from the set $S_{x^t} \stackrel{\mathrm{def}}{=} S_{x_1} \times \cdots \times S_{x_t}$.*

Proof. Let A' be an f^t-collision-finder algorithm with probability ϵ to output a collision $x'_1, \ldots, x'_t$ outside of S_{x^t}. Observe that this implies that for a randomly chosen coordinate $i \stackrel{r}{\leftarrow} [t]$ it holds that $\Pr[\phi_I(f^t(X^t)) \notin S_{\phi_I(X^t)}] \geq \epsilon/t$. This calls for the following f-collision-finder A: on input x choose uniformly at random a location i from $[t]$ and uniformly at random inputs $x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_t$ from X. Set $x_i = x$ and return $\phi_i(A'(x_1, \ldots, x_t))$. It follows that A outputs a collision for f outside of S_x with probability greater than ϵ/t. The lemma follows. □

Using linearity of expectation, the union bound, Definitions 6 and 7, and the fact that $\log(|S_{x^t}|) = \Sigma_{i=1}^t \log(|S_{x_i}|)$, we get:

Corollary 1.

1. *If $H^{\mathrm{eff}}_{p,\max}(f) \leq k$ then $H^{\mathrm{eff}}_{p,\max}(f^t) \leq t \cdot k$.*
2. *If $H^{\mathrm{eff}}_{p,\mathrm{avg}-\max}(f) \leq k$ then $H^{\mathrm{eff}}_{p,\mathrm{avg}-\max}(f^t) \leq t \cdot k$.*

2.6 An Overview of the Construction of Haitner et. al.

The construction consists of two independent parts. First they show how to get a function with a noticeable gap of average inaccessible entropy from any one-way

function. Specifically, they show that a prefix of a random length of a three-wise independent hashing of the output already has some weak form of an average entropy gap. Namely, on average over the inputs to the new construction, there is a noticeable gap of $\Delta = \Omega(\log n/n)$ between the real preimage-entropy and the average accessible max-preimage-entropy.

The second part of the construction starts with any function with some noticeable gap Δ and shows how to obtain a UOWHF. This is achieved using the following steps:

1. Gap amplification and transformation of an average type gap into an absolute type of gap.
2. Entropy reduction.
3. Output length reduction.
4. Random inputs collision-resistance to a UOWHF.
5. Removing the non-uniformity.

The composition of Steps 2 through 5 of their construction[2] is summarized in the following theorem, which we later use in a black-box manner:

Theorem 1.

1. *There exists an explicit black-box construction taking parameters a function* $\psi = \{\psi_n\}_{n\in\mathcal{N}}$, *where* $\psi_n : \{0,1\}^{\lambda(n)} \to \{0,1\}^{m(n)}$, *and a number* $\tau = \tau(n)$ *such that if* $H^{\text{eff}}_{p,\text{avg}-\max}(\psi_n) + \omega(\log(n)) \leq \tau(n) \leq H_p(\psi_n)$ *holds, the construction implements a UOWHF with output length and key length* $O(\lambda(n))$.
2. *Moreover, for all efficiently computable* $l = l(n)$ *there exists an explicit black-box construction taking parameters* ψ *(as before) and sets of numbers* $\tau = \tau(n) = \{\tau_{n,i}\}_{i=1}^{l(n)}$, *such that if one of* $\{(\psi_n, \tau_{n,i})\}_{i=1}^{l(n)}$ *satisfies the condition of part (1.), the construction implements a UOWHF with output length* $O(\lambda(n) \cdot l(n))$ *and key length of* $O(\lambda(n) \cdot l(n) \cdot \log(l(n)))$.

3 UOWHF from a 2^r-Regular OWF

Let $f : \{0,1\}^n \to \{0,1\}^m$ be a 2^r-regular one-way function. Our construction also works in two steps: First we obtain an entropy gap of $O(\log(n))$ applying f only once and use a variant of the Naor-Yung construction. Next, we show that the type of gap that we get by our first step is almost of the required absolute type. Namely, the average entropy gap is essentially concentrated on a smaller interval of size almost $O(\log(n))$, and in this case the structured gap can be transformed to an absolute type of gap via taking only a super-logarithmic number of independent samples. The main result of this section is:

Theorem 2. *Let* $f : \{0,1\}^n \to \{0,1\}^{m(n)}$ *be a* 2^r*-regular one-way function, where* $r = r(n)$ *is efficiently computable. Then there exists an explicit black-box construction of a universal one-way hash function based on* f *with output length and key length of* $O(n \cdot \log(n) \cdot \alpha(n))$ *for any super-constant function* $\alpha(n)$.

[2] Lemmas $5.3 - 5.4, 5.6$ in [2].

3.1 Inaccessible Entropy from 2^r-Regular One-Way Functions

In this case f has exactly 2^{n-r} different images. If we randomly distribute the images among b buckets, we expect to have roughly $\frac{2^{n-r}}{b}$ images in each bucket. Consider the composed function, $F(x,g) = (g(f(x)), g)$ where g is the description of a three-wise independent hash-function from some family $\mathcal{G}$. We show that an appropriate choice of the family $\mathcal{G}$ allows us to reduce the preimage inaccessibility of F to the hardness of the underlying function f.

For injective one-way functions this was already observed in [6]. The difference is that for an injective f, the resulting F is already a universal one-way hash function, whereas in the case where f is regular we get:

Lemma 2. *Let $f : \{0,1\}^n \to \{0,1\}^{m(n)}$ be a 2^r-regular one-way function, where $r = r(n)$ is efficiently computable. Let $d > 0$ and let $\mathcal{G} = \mathcal{G}(n) \stackrel{\text{def}}{=} \mathcal{G}_n^{(n-r)-4d\log(n)}$ be a family of constructible three-wise independent hash functions. Then the function $F : \{0,1\}^n \times \mathcal{G} \to \{0,1\}^{(n-r)-4d\log(n)} \times \mathcal{G}$ given by: $F(x,g) = (g(f(x)), g)$ satisfies the following properties:*

1. $H_p(F) \geq r + 3d\log(n)$.
2. $H^{\text{eff}}_{p,\max}(F) \leq r$.

Proof.

1. Recall that when the input is chosen uniformly at random from the input space the preimage-entropy of F is just the expected log-value of the size of the preimage set. We first compute the expected number of preimages for a fixed x with some random g from $\mathcal{G}$. Since $F(x,g)$ already determines g it follows that any potential preimage must have the same g component.
   For all fixed x, g we have the set equality:

$$F^{-1}(F(x,g)) = \bigcup_{y' : g(f(x)) = g(y')} \left(f^{-1}(y') \times \{g\}\right) \tag{4}$$

   and the union is over disjoint sets. We get that:

$$\begin{aligned} \pi_F(x,g) &= \pi_f(x) + \sum_{y' \neq f(x)} \mathbf{1}_{g(f(x))=g(y')} \cdot |f^{-1}(y')| \\ &= 2^r \cdot \left(1 + \sum_{y' \neq f(x)} \mathbf{1}_{g(f(x))=g(y')}\right) \ , \end{aligned} \tag{5}$$

   where due to the regularity it holds that $\pi_f(x) = |f^{-1}(y)| = 2^r$.
   Now we observe that for every fixed $x \in \{0,1\}^n$ and $y' \neq f(x)$

$$\mathop{\mathbf{E}}_{g \stackrel{r}{\leftarrow} \mathcal{G}} \left[\mathbf{1}_{g(f(x))=g(y')}\right] = 2^{-(n-r-4d\log(n))} \ , \tag{6}$$

   where the equality is due to due to the pair-wise independence of $\mathcal{G}$.

Using (5), (6) and linearity of expectation we have that

$$\mathop{\mathbf{E}}_{g\xleftarrow{r}\mathcal{G}}[\pi_F(x,g)] = 2^r \cdot \left(1 + \sum_{y'\neq f(x)} \mathop{\mathbf{E}}_{g\xleftarrow{r}\mathcal{G}}[\mathbb{1}_{g(f(x))=g(y')}]\right) > 2^{r+4d\cdot\log(n)} \quad , \quad (7)$$

where again due to the regulariry the summation is over $2^{n-r}-1$ indicators. Furthermore, as the family is three-wise independent, we also have that for different $f(x), y', y''$ it holds that the random variables $\mathbb{1}_{g(f(x))=g(y')}$ and $\mathbb{1}_{g(f(x))=g(y'')}$ are independent (and in particular, uncorrelated) and therefore

$$\mathop{\mathbf{V}}_{g\xleftarrow{r}\mathcal{G}}[\pi_F(x,g)] = (2^r)^2 \cdot \sum_{y'\neq f(x)} \mathop{\mathbf{V}}_{g\xleftarrow{r}\mathcal{G}}\left[\mathbb{1}_{g(f(x))=g(y')}\right] \leq \left(2^{r+2\cdot d\cdot\log(n)}\right)^2 \quad . \quad (8)$$

where the equality holds for the sum of uncorrelated random variables and the the inequality holds as for all indicator random variables $\mathbf{V}[\mathbb{1}_A] \leq \mathbf{E}[\mathbb{1}_A]$ and using (6). Now, the Chebysev Inequality establishes that for all $\alpha > 0$:

$$\Pr_{g\xleftarrow{r}\mathcal{G}}\left[\left|\pi_F(x,g) - \mathop{\mathbf{E}}_{g\xleftarrow{r}\mathcal{G}}[\pi_F(x,g)]\right| > \alpha \cdot 2^{r+2\cdot d\cdot\log(n)}\right] < \frac{1}{\alpha^2} \quad . \quad (9)$$

Whenever the event in (9) does not happen plugging (7) we obtain

$$\pi_F(x,G) \geq 2^{r+4\cdot d\cdot\log(n)} - \alpha\cdot 2^{r+2\cdot d\cdot\log(n)} \geq 2^{r+3.5\cdot d\cdot\log(n)} \quad ,$$

for all fixed α and sufficiently large n.
Finally, recall that due to the regularity we always have $\pi_F(x,g) \geq 2^r$ and so using conditional expectation on the event from (9) with $\alpha = 5$ and plugging (7) we obtain:

$$\mathop{\mathbf{E}}_{g\xleftarrow{r}\mathcal{G}}[\log(\pi_F(x,g))] > r + \frac{24}{25}\cdot(3.5\cdot d\cdot\log(n)) \geq r + 3\cdot d\cdot\log(n) \quad . \quad (10)$$

As this holds for every fixed x, it also holds for a random one, and we are done.

2. We show that any efficient algorithm that finds a collision for a random input (X,G) outside of $f^{-1}(X)\times\{G\}$ leads to one that inverts f. Let A_F be an F-collision-finder. We denote the randomness taken by A_F explicitly (as an additional argument) by r, and for some fixed randomness r and a fixed input x, denote by $\epsilon_{x,r} \stackrel{\text{def}}{=} \Pr_{g\xleftarrow{r}\mathcal{G}}[A_F(x,g,r) \notin f^{-1}(f(x))\times\{g\}]$.
We show[3] that the algorithm B inverts $y = f(x')$, where $x' \xleftarrow{r} \{0,1\}^n$ uniformly at random, with probability at least ϵ/n^{-d}.

[3] In similar manner to [6] and [2].

Algorithm $B(y)$ // On input $y \in \{0,1\}^n$

(a) Choose x uniformly at random from $\{0,1\}^n$.
(b) Choose randomness r for A_F uniformly at random.
(c) Choose g uniformly at random subject to $g(f(x)) = g(y)$.
(d) Output $\phi_1(A_F(x,g,r))$.

It holds that

$$\epsilon_{x,r} = \sum_{y \neq f(x)} \Pr_{g \xleftarrow{r} \mathcal{G}} \left[\phi_1(A_F(x,g,r)) \in f^{-1}(y)\right] \tag{11}$$

$$= \sum_{y \neq f(x)} \Pr_{g \xleftarrow{r} \mathcal{G}} \left[\phi_1(A_F(x,g,r)) \in f^{-1}(y) \mid g(f(x)) = g(y)\right] \tag{12}$$

$$\cdot \Pr_{g \xleftarrow{r} \mathcal{G}} [g(f(x)) = g(y)]$$

$$= \sum_{y \neq f(x)} \Pr_{g \xleftarrow{r} \mathcal{G}} \left[\phi_1(A_F(x,g,r)) \in f^{-1}(y) \mid g(f(x)) = g(y)\right] \tag{13}$$

$$\cdot\, 2^{-(n-r-d\log(n))} \ .$$

It follows that conditioned on the random choices $X = x$ and $R = r$ of the algorithm B, we obtain that

$$\begin{aligned}
&\Pr_{x' \xleftarrow{r} \{0,1\}^n} [B(f(x')) \in f^{-1}(f(x')) \mid X = x \wedge R = r] \\
&= \Pr_{y \xleftarrow{r} f(\{0,1\}^n)} [B(y) \in f^{-1}(y) \mid X = x \wedge R = r] \\
&\geq \sum_{y \neq f(x)} \Pr[Y = y] \cdot \Pr_{g \xleftarrow{r} \mathcal{G}} \left[\phi_1(A_F(x,g,r)) \in f^{-1}(y) \mid g(f(x)) = g(y)\right] \\
&\geq 2^{-d\log(n)} \cdot \epsilon_{x,r} \ . \qquad (14)
\end{aligned}$$

As X and R are chosen uniformly at random, and by the fact that $\epsilon = \mathbf{E}[\epsilon_{X,R}]$ we get that the algorithm inverts f with probability at least $n^{-d} \cdot \epsilon$. Thus we have shown that A_F's output on input (x,g) is limited to $f^{-1}(f(x)) \times \{g\}$. By the regularity of f we have that its size is exactly 2^r. Thus F has accessible max-preimage-entropy at most r.

□

We next show that by a more careful analysis of the amplification results in an almost-linear construction.

3.2 Amplifying the Entropy Gap and Converting Average to Absolute Entropy Gaps

Lemma 3 (Fast gap amplification and real- to min- preimage-entropy conversion). *Let f and F be as in Lemma 2, F^t be the t-fold application of F and $\alpha(n)$ be any*

super-constant function. Then for $t = \alpha(n) \cdot \log(n)$, F^t *has a strong inaccessible entropy gap of* $\alpha(n) \cdot \log^2(n)$. *Moreover, the following entropy-gap holds:*

1. $H_{p,\min}(F^t) \geq t(r + 2d\log(n))$.
2. $H^{\text{eff}}_{p,\max}(F^t) \leq t \cdot r$.

Proof.

1. We first show that by the Markov inequality, the probability that the point-wise entropy exceeds its expected value by more than $\alpha(n) \cdot \log(n)$ bits is negligible. Specifically, from the first part of Lemma 2, we know that the expected value of the preimage size of inputs to F is $2^{r+4d\log(n)}$. The Markov inequality asserts that the probability we get an input with more than $2^{\alpha(n)\cdot\log(n)} \cdot 2^{r+4d\log(n)}$ preimages is at most $1/n^{\alpha(n)}$. Let us denote this 'bad' event as A. By Lemma 5 [4] we get that $\mathbf{E}[\log(\pi_F(X,G)) \mid \overline{A}] \geq r + 4d\log(n) - n^{1-\alpha(n)}$.

   From now on we assume that this unlikely event does not happen. We get that the value of the real entropy is limited to an interval of size $O(\alpha(n) \cdot \log(n))$, as we always have a at least r bits of point-wise entropy due to the regularity of f.

   Now we can apply the Hoeffding bound which asserts that for this case a super-logarithmic number of repetitions suffice to bound from below the min-preimage-entropy of the t-fold application of F. Specifically, we get that

$$\Pr_{(x,g)^t \xleftarrow{r} (\{0,1\}^n \times \mathcal{G})^t} \left[\log\left(\pi_{F^t}((x,g)^t)\right) \geq t(r + 2d\log(n))\right] \leq \exp\left(\frac{-2t}{\alpha^2(n)}\right).$$

   The choice of $t = O(\alpha^3(n) \cdot \log(n))$ ensures that this happens with probability at most $1/n^{\alpha(n)}$.
2. This is just the t-fold accessible max-preimage-entropy we get by the second part of Lemma 2 and Corollary 1.

□

Proof (Theorem 2). By Lemma 3, F^t already has the required strong type of entropy gap between its accessible max-preimage-entropy and its real min-preimage-entropy. Moreover, it tells us exactly where this gap is (there are at most $t \cdot r$ bits of accessible max-preimage-entropy). Now, note that if $\alpha(n)$ is a super-constant function, then so is $\alpha'(n) = \alpha^{1/3}(n)$. Finally, utilizing Theorem 1 with parameter $(F^t, t \cdot (r + 2d\log(n)))$ completes the construction and yields a UOWHF with output length and key length $O(n \cdot \log(n) \cdot \alpha(n))$. □

[4] We use Lemma 5 in the uniform setting, where $l = l(n) = n$ (as we consider the point-wise Shannon entropy) and A_n is an event that happens with some negligible probability, that is, $\Pr[A_n] < n^{-\alpha(n)}$ for some super-constant function $\alpha(n)$.

4 UOWHF from a $(2^{r(n)}, 2^{s(n)})$-Roughly-Regular OWF

The main theorem proved in this section is:

Theorem 3. *Let $f : \{0,1\}^n \to \{0,1\}^{m(n)}$ be a $(2^r, 2^s)$-roughly-regular one-way function, where $r = r(n)$ and $s = s(n)$ are efficiently computable. Then there exists an explicit construction of a UOWHF with output length and key length of $\tilde{O}(n \cdot s^6(n))$ (resp., $\tilde{O}(n \cdot s^4(n))$) in the uniform (resp., non-uniform) model.*

4.1 $\log(n)/s(n)$ Bits of Average Inaccessible Entropy

Haitner et. al. showed that for a general one-way function f, a random truncation of a hashing of $f(x)$ using a three-wise independent family of hash functions yields an average entropy gap of $\Omega(\log(n)/n)$ entropy bits. We observe that a modification of their first step achieves an average inaccessible entropy gap of $\log(n)/s(n)$ bits from any $(2^{r(n)}, 2^{s(n)})$-roughly-regular one-way function.

The idea is to divide the images $f(x)$ (and respectively, the inputs x) into buckets, such that every bucket contains images with roughly the same number of preimages. We set $m \stackrel{\text{def}}{=} s(n)/d\log(n)$ and $J \stackrel{\text{def}}{=} \{j_0, \ldots, j_{m-1}\}$, where $j_i \stackrel{\text{def}}{=} n - r(n) - s(n) + (i-1)d\log(n)$, and show that truncating the output of the application of a three-wise independent hashing of $f(x)$ to a random length from J yields a function with the required gap. Recall that $H_{f(X)}(f(x)) \in (j_i, j_{i+1}]$ if and only if $\pi_f(x) \in [2^{r+s-i(d\log(n))}, 2^{r+s-(i-1)(d\log(n))})$. Let us denote $q_i \stackrel{\text{def}}{=} \Pr[H_{f(X)}(f(x)) \in (j_i, j_{i+1}]]$. By the roughly-regularity assumption on f, it holds that $\sum_{i=1}^{m} q_j = 1$.Now we set $\mathcal{G} \stackrel{\text{def}}{=} \mathcal{G}_n^n$ a family of three-wise independent hash functions, $X \stackrel{\text{def}}{=} \{0,1\}^n$ and define $F : X \times \mathcal{G} \times J \to X \times \mathcal{G} \times J$ as $F(x,g,j) = (g(f(x))_{1,\ldots,j} \| 0^{n-j}, g, j)$, where we denote the domain and range of F by $\mathcal{Z} \stackrel{\text{def}}{=} X \times \mathcal{G} \times J$.

Lemma 4. *The function F as defined above has an average preimage-entropy gap of $s(n)/log(n)$ bits.*

Proof. Recall that our goal in this step is to achieve an average inaccessible entropy gap of $\Omega(\log(n)/s(n))$ bits. That is, we need to show that for each $z = (x, g, j_i)$ there exists a set S_z, such that: (1) any efficient collision-finder outputs an element of S_z (except for an event that happens with negligible probability) and (2) $\mathbf{E}_{z \xleftarrow{r} \mathcal{Z}}[\log(\pi_F(z)) - \log(|S_z|)] > \Omega(\log(n)/s(n))$.

In a similar manner to the regular case, the set of inputs accessible by an efficient algorithm is limited only to those with relatively few images, where "few" corresponds to the length of the random truncation. Essentially, we show that when we hash to length j_i, any preimages an efficient algorithm finds are either already preimages of $f(x)$ (we refer to these as 'trivial' collisions) [5] or stem from some non-trivial collision, that is $F(x', g, j_i) = F(x, g, j_i)$ but $f(x) \neq f(x')$.

[5] Note that the definition of a one-way function does not rule out the possibility that given a preimage it is difficult to compute other preimages from $f^{-1}(f(x))$.

For the latter, we further distinguish between those x that have significantly fewer preimages than expected for a random function with output length j_i, and the rest. More precisely, we consider those preimages $z' = (x', g, j_i)$ for which $\pi_f(x') \leq 2^{j_{i+2}}$ and call these 'j_{i+2}-light' preimages of $f(x)$. The remaining 'heavy' collisions stem from inputs z' for which $\pi_f(x') > 2^{j_{i+2}}$.

We define:

$$T_z \stackrel{\text{def}}{=} T_{(x,g,j_i)} = f^{-1}(f(x)) \times \{g\} \times \{j\},$$

$$L_z \stackrel{\text{def}}{=} L_{(x,g,j_i)} = \{x' \in \{0,1\}^n \mid g(f(x))_{1,\ldots,j_i} = g(f(x'))_{1,\ldots,j_i} \\ \wedge H_{f(X)}(f(x)) \geq j_{i+2} \wedge x' \notin f^{-1}(f(x))\} \times \{g\} \times \{j\}$$

and

$$H_z \stackrel{\text{def}}{=} H_{(x,g,j_i)} = \{x' \in \{0,1\}^n | g(f(x))_{1,\ldots,j_i} = g(f(x'))_{1,\ldots,j_i} \\ \wedge H_{f(X)}(f(x)) < j_{i+2} \wedge x' \notin f^{-1}(f(x))\} \times \{g\} \times \{j\} \ ,$$

where T, L and H stand for 'trivial', 'light' and 'heavy', respectively. It follows that for every z,

$$F^{-1}(F(z)) = T_z \cup L_z \cup H_z \ , \tag{15}$$

where the union is over disjoint sets.

The rest of the proof is involved with proving that indeed the only accessible sets to any efficient algorithm are $T_z \cup L_z$, and that they constitute a large fraction of the preimage set $F^{-1}(F(z))$. The analysis follows the construction from [3] and is brought for completeness in Appendix B. □

4.2 Faster Amplification of the Inaccessible Entropy Gap of F

Our goal in this section is to amplify the entropy gap of F from the previous section. We show how to construct a function F' with $\omega(\log(n))$ bits of inaccessible entropy with an absolute type of gap.

Haitner et. al. [2] assert that independent repetitions of F achieve both these goals. They show that $\tilde{O}(n^4)$ repetitions are enough for getting this gap from an arbitrary one-way function. We are able to utilize the information about the underlying f (and in turn, that of F) and get a faster convergence, using the roughly-regularity assumption.

Set $\Delta \stackrel{\text{def}}{=} (c \cdot \log(n)/s(n))$ as the entropy gap of F, where c is the constant corresponding to the Ω notation, and fix k, such that F has preimage-entropy $H_p(F) = k + \Delta$. Lemma 4 asserts that $H^{\text{eff}}_{p,\text{avg}-\max}(F) \leq k$. Using (3) and Corollary 1 we know that for the t-fold parallel repetition of F it holds that

$$H_p(F^t) = t \cdot (k + \Delta) \ , \tag{16}$$

$$H^{\text{eff}}_{p,\text{avg}-\max}(F^t) \leq t \cdot k \ . \tag{17}$$

Thus for F^t we obtain an average entropy gap of $t \cdot \Delta$ bits.

Using the analysis of Lemma 4 and Lemma 1 we get that for an input $z^t = (z_1, \ldots, z_t)$ to F^t, the only accessible inputs to F^t are those that are contained in $S_z = (T_{z_1} \cup L_{z_1}) \times \cdots \times (T_{z_t} \cup L_{z_t})$, and that the set of preimages of z^t is just $F^{t^{-1}}(F^t(z^t)) = (T_{z_1} \cup L_{z_1} \cup H_{z_1}) \times \cdots \times (T_{z_t} \cup L_{z_t} \cup H_{z_t})$, except for an event B_1 that occurs with negligible probability. Next, we would like to apply the Hoeffding bound to get the required gap. Similarly to Lemma 3 we show that although for some inputs the preimage size of F may be very large (a priori there may be inputs with up to 2^n preimages, but not more, since $F(x, g, j_i)$ determines (g, j_i) uniquely as part of its output), this is not likely. First observe that $\log(\pi_F(z)) \in [r, n]$ for all z. This is due to the fact that every image of $f(x)$ has at least $2^{r(n)}$ preimages. We show that we can bound this also from above: except with negligible probability we have that for any super-constant function $\alpha(n)$: $\log(|T_z \cup L_z|) \leq \log(\pi_F(z)) < r(n) + s(n) + d\log(n) + \alpha(n)\log(n)$. Consider $\pi_F(Z)$ for a uniformly chosen random input $Z = (X, G, J)$. This value is maximized for $J = j_0$ because of the inclusion $\phi_1(F^{-1}(F(x, g, j_i'))) \subset \phi_1(F^{-1}(F(x, g, j_i)))$ for $j_i \leq j_i'$. It follows that in order to bound $\mathbf{E}[\pi_F(X, G, J)]$ it is sufficient to bound $\mathbf{E}[\pi_F(X, G, j_0)]$.

As in Lemma 2, using the three-wise independence of $\mathcal{G}$, and the roughly-regularity of f we have that for fixed x it holds that:

$$\mathop{\mathbf{E}}_{g \xleftarrow{r} \mathcal{G}} [\pi_F(x, g, j_0)] \leq 2^{r(n)+s(n)+d\log(n)+2} .$$

Next, fix any super-constant function $\alpha(n)$. Markov's inequality asserts that

$$\Pr_{g \xleftarrow{r} \mathcal{G}} \left[\pi_F(x, g, j_0) \geq 2^{r(n)+s(n)+d\log(n)} \cdot 2^{\alpha(n)\log(n)}\right] \leq n^{-\alpha(n)} .$$

Denote the event that this happens in any of the repetitions by B_2 and note that it happens only with negligible probability (as t is polynomial in n and using the union bound). We summarize this as follows: whenever B_2 does not occur, we get that

$$r(n) \leq \log(|T_Z \cup L_Z|) \leq \log(\pi_F(Z)) \leq r(n) + s(n) + (d + \alpha(n))\log(n) .$$

When this is the case, both quantities are within an interval of size $s' \stackrel{\text{def}}{=} 3 \cdot \max\{s(n), \alpha(n) \cdot \log(n)\}$.

By Lemma 5 we know that the preimage-entropy and the average accessible max-preimage-entropy values change by at most a negligible quantity when ignoring an event of negligible probability. Specifically, we get that whenever $\overline{B_1} \wedge \overline{B_2}$ happen we have:

$$k' \stackrel{\text{def}}{=} \mathop{\mathbf{E}}_{z \xleftarrow{r} \mathcal{Z}} \left[\log(|S_z|) \,|\, \overline{B_1} \wedge \overline{B_2}\right] \leq \mathop{\mathbf{E}}_{z \xleftarrow{r} \mathcal{Z}} [\log(|S_z|)] + \mathsf{negl}(n) \tag{18}$$

and

$$k'' \stackrel{\text{def}}{=} \mathop{\mathbf{E}}_{z \xleftarrow{r} \mathcal{Z}} \left[\log(\pi_F(z)) \,|\, \overline{B_1} \wedge \overline{B_2}\right] \geq \mathop{\mathbf{E}}_{z \xleftarrow{r} \mathcal{Z}} [\log(\pi_F(z))] - \mathsf{negl}(n) \tag{19}$$

with a gap of $\Delta' \stackrel{\text{def}}{=} k'' - k' \geq \Delta - \mathsf{negl}(n)$.

The Hoeffding bound yields that setting $t \stackrel{\text{def}}{=} O(\frac{s'^2(n)\cdot s^2(n)}{\log(n)})$ assures that the inaccuracies due to the sampling of the independent inputs to F are already smaller than the accumulated gap. Specifically:

$$\Pr_{z^t \xleftarrow{r} \mathcal{Z}^t}\left[\log\left(|S_{z^t}|\right) > t\cdot k' + \frac{c}{6}\cdot s'(n)\cdot\sqrt{t\cdot\alpha(n)\cdot\log(n)}\right] \leq n^{-\alpha(n)} \ , \tag{20}$$

$$\Pr_{z^t \xleftarrow{r} \mathcal{Z}^t}\left[\log\left(\pi_{F^t}(z^t)\right) < t\cdot(k'+\Delta')) - \frac{c}{6}\cdot s'(n)\cdot\sqrt{t\cdot\alpha(n)\cdot\log(n)}\right] \leq n^{-\alpha(n)}. \tag{21}$$

Plugging (18) and (19) we get that except with negligible probability there is an absolute entropy gap of at least

$$t\cdot\Delta - t\cdot\mathsf{negl}(n) - \frac{c}{3}\cdot s'(n)\cdot\sqrt{t\cdot\alpha(n)\cdot\log(n)} \in \omega(\log(n)) \ . \tag{22}$$

4.3 A UOWHF in the Non-uniform Model

To finish the construction we would like to apply the first part of Theorem 1. We use the preimage-entropy of F from Section 4.1 (in the form of a non-uniform advice), which equals $k+\Delta$. By what we have shown in Section 4.2 it holds that F^t has real min-preimage-entropy of at least $\tau \stackrel{\text{def}}{=} t\cdot(k+\Delta) - \frac{c}{4}\cdot s'(n)\cdot\sqrt{t\cdot\alpha(n)\cdot\log(n)}$ bits. Additionally, it enjoys the required absolute entropy gap. The first part of Theorem 1 with parameters (F^t,τ) yields a UOWHF with output length and key length $O(n\cdot s'^2(n)\cdot s^2(n)/\log(n))$.

4.4 An Efficient Non-uniform to Uniform Reduction

As explained, the construction obtained requires a non-uniform advice (i.e., the Shannon preimage-entropy of f). We remove the non-uniformity by 'trying all possibilities'. However, as opposed to the case of a general one-way function, where we need to try $O(n^2)$ different values, we show that using the roughness regularity assumption we only need $O(s^2(n))$ tries.

Recall that by the roughly-regularity assumption on f, it follows that the preimage-entropy of F lies in the interval $[r+d\log(n), s+d\log(n)]$.

For $i \in [\lceil 4\cdot s(n)/c\log(n)\rceil]$ set $k_i \stackrel{\text{def}}{=} r+d\log(n)+i\cdot\frac{1}{4\cdot s(n)}$. It holds that one of the k_i is within an additive distance of $\frac{\Delta}{4}$ from the real value $k+\Delta$. Accordingly, set $\tau_i \stackrel{\text{def}}{=} t\cdot(k_i+\Delta) - \frac{c}{4}\cdot s'(n)\cdot\sqrt{t\cdot\alpha(n)\cdot\log(n)}$. It follows that for the same i, (F_t,τ_i) satisfies the premise of the first part of Theorem 1, and thus the second part of the theorem yields a construction of a UOWHF with output length of $O(n\cdot s'^2(n)\cdot s^4(n)/\log^3(n))$ and key length of $O(n\cdot s'^2(n)\cdot s^4(n)/\log^2(n))$.

5 Conclusions

We demonstrated how to obtain more efficient constructions of a UOWHF from different assumptions on the structure of the underlying OWF. For the case of known regularity the resulting construction is very efficient and makes an almost logarithmic number of calls to the underlying OWF. In the case when the underlying OWF is known to be either 2^{r_1}-regular or 2^{r_2}-regular (i.e., the construction is given r_1 and r_2 and should be secure when instantiated with any function of the corresponding regularity), we observe that one obtains a construction that makes $\tilde{O}(n)$ calls (combining our construction for the regular case with the second part of Theorem 1). Of course, the main open problem remains to further improve the construction of Haitner et. al. for a general OWF.

Acknowledgments. We would like to thank the anonymous reviewers for their helpful comments. The first author would like to thank David Adjiashvili and Sandro Coretti for their comments on an earlier version of this work.

References

1. Goldreich, O., Krawczyk, H., Luby, M.: On the existence of pseudorandom generators. SIAM J. Comput. 22(6), 1163–1175 (1993)
2. Haitner, I., Holenstein, T., Reingold, O., Vadhan, S., Wee, H.: Universal One-Way Hash Functions via Inaccessible Entropy. In: Gilbert, H. (ed.) EUROCRYPT 2010. LNCS, vol. 6110, pp. 616–637. Springer, Heidelberg (2010)
3. Haitner, I., Nguyen, M.-H., Ong, S.J., Reingold, O., Vadhan, S.P.: Statistically hiding commitments and statistical zero-knowledge arguments from any one-way function. SIAM J. Comput. 39(3), 1153–1218 (2009)
4. Haitner, I., Reingold, O., Vadhan, S.P.: Efficiency improvements in constructing pseudorandom generators from one-way functions. In: Schulman, L.J. (ed.) STOC, pp. 437–446. ACM (2010)
5. Haitner, I., Reingold, O., Vadhan, S.P., Wee, H.: Inaccessible entropy. In: Mitzenmacher, M. (ed.) STOC, pp. 611–620. ACM (2009)
6. Naor, M., Yung, M.: Universal one-way hash functions and their cryptographic applications. In: STOC, pp. 33–43. ACM (1989)
7. Rompel, J.: One-way functions are necessary and sufficient for secure signatures. In: STOC, pp. 387–394. ACM (1990)
8. De Santis, A., Yung, M.: On the Design of Provably-Secure Cryptographic Hash Functions. In: Damgård, I.B. (ed.) EUROCRYPT 1990. LNCS, vol. 473, pp. 412–431. Springer, Heidelberg (1991)

A Further Preliminaries

A.1 The Hoeffding Bound

For independent bounded random variables $X_1, \ldots, X_t$, where $X_i \in [a_i, b_i]$, set $S_t = \Sigma_{i=1}^{t} X_i$, then:

$$\Pr\left[|S_t - \mathbf{E}[S_t]| \geq k\right] \leq 2 \cdot \exp\left(\frac{-2k^2}{\Sigma_{i=1}^{t}(b_i - a_i)^2}\right) .$$

A.2 *t*-wise Independent Hashing

Let $\mathcal{G}_n^m \stackrel{\text{def}}{=} \{g_k : \{0,1\}^n \to \{0,1\}^m\}_{k\in K}$ be a family of keyed functions. $\mathcal{G}_n^m$ is t-wise independent if for all $y_1, \ldots, y_t \in \{0,1\}^m$ and all distinct $x_1, \ldots, x_t \in \{0,1\}^n$ it holds that

$$\Pr_{k\stackrel{r}{\leftarrow}K}\left[g_k(x_1) = y_1 \wedge \cdots \wedge g_k(x_t) = y_t\right] = 2^{-tm} \ .$$

The family is called constructible if, for all $y_1, \ldots, y_s$ and distinct $x_1, \ldots, x_s$ where $s \leq t$, it is possible to sample a function uniformly subject to $g_K(x_1) = y_1, \ldots, g_K(x_s) = y_s$.

It is well-known that if $\mathcal{G}_n^n$ is a t-wise independent family, then by truncating the last $n - l$ bits of $\mathcal{G}_n^n$ one gets a t-wise independent family $\mathcal{G}_n^l$. Moreover, if $\mathcal{G}_n^n$ is constructible, then so is $\mathcal{G}_n^l$.

The next lemma shows that ignoring an unlikely event of a random variable that takes a value in some limited range, does not change much its expected value. The standard proof is omitted in this extended abstract.

Lemma 5. *Let X be a random variable with* $\mathrm{Supp}(X) \subset [0, l]$ *and A an event that happens with probability at most ϵ. Then:*

$$\left|\mathbf{E}\left[X\right] - \mathbf{E}\left[X|\overline{A}\right]\right| \leq 2 \cdot l \cdot \epsilon \ . \tag{23}$$

B Proof of Lemma 4, continued.

Our next goal is to show that the sets $\{T_z \cup L_z\}_{z\in Z}$ satisfy the needed requirements. Claim 4.9 in [2] shows that any efficient collision-finder cannot (except with negligible probability) output a preimage of $F(z)$ in H_z, as such an algorithm can be used to invert f. Specifically, they show (again, using the constructibility of the three-wise independent hash family as in the second part of Lemma 2) how to efficiently convert any F-collision-finder that outputs a preimage from H_z with probability ϵ to one that inverts a random input of f with probability ϵ/n^d.

As the preimage sets $\{H_z\}_{z\in\mathcal{Z}}$ are inaccessible, it remains to show that they constitute a noticeable part of the preimage sets. In order to complete the proof, we need to bound:

$$\begin{aligned}
\mathop{\mathbf{E}}_{z\stackrel{r}{\leftarrow}\mathcal{Z}}[\pi_F(z)] - \mathop{\mathbf{E}}_{z\stackrel{r}{\leftarrow}\mathcal{Z}}[\log(|T_z \cup L_z|)] &= \mathop{\mathbf{E}}_{z\stackrel{r}{\leftarrow}\mathcal{Z}}\left[\log\left(\frac{\pi_F(z)}{|T_z \cup L_z|}\right)\right] && (24)\\
&= \mathop{\mathbf{E}}_{z\stackrel{r}{\leftarrow}\mathcal{Z}}\left[\log\left(\frac{|T_z| + |L_z| + |H_z|}{|T_z| + |L_z|}\right)\right] && (25)\\
&= \mathop{\mathbf{E}}_{z\stackrel{r}{\leftarrow}\mathcal{Z}}\left[\log\left(1 + \frac{|H_z|}{|T_z| + |L_z|}\right)\right] && (26)\\
&\geq \frac{1}{2}\mathop{\mathbf{E}}_{z\stackrel{r}{\leftarrow}\mathcal{Z}}\left[\frac{|H_z|}{|T_z| + |L_z| + |H_z|}\right] && (27)
\end{aligned}$$

where the second equality is due to the partition in (15) and the inequality uses the fact that $\log(1+x) \geq x/2$ for $x \in [0,1)$. Thus, it is left to show that indeed $|H_z|$ constitutes a noticeable part of $\pi_F(z)$.

Proposition 1. *Conditioned on $X = x$ and $J = j_i$, define the events:*

$$E^1_{j_i} \stackrel{\text{def}}{=} \left\{|H_z| + |L_z| \leq 3 \cdot 2^{n-j_i}\right\}$$
$$E^2_{j_i} \stackrel{\text{def}}{=} \left\{|H_z| \geq \left(q_i - 4 \cdot \sqrt{1/n^d}\right) \cdot 2^{n-j_i-1}\right\} \ .$$

Then $\Pr\limits_{g \xleftarrow{r} \mathcal{G}}[E^1_{j_i}] > 2/3$ *and* $\Pr\limits_{g \xleftarrow{r} \mathcal{G}}[E^2_{j_i}] > 3/4$ *hold.*

This is just Claim 4.11 from [2] [6]. It follows that:

$$= \mathop{\mathbf{E}}_{z \xleftarrow{r} \mathcal{Z}}[\log(\pi_F(z))] - \mathop{\mathbf{E}}_{z \xleftarrow{r} \mathcal{Z}}[\log(|T_z \cup L_z|)] \tag{28}$$

$$\geq \frac{1}{2} \mathop{\mathbf{E}}_{z \xleftarrow{r} \mathcal{Z}}\left[\frac{|H_z|}{|T_z| + |L_z| + |H_z|}\right] \tag{29}$$

$$= \frac{1}{2} \sum_{i=0}^{m-1} \Pr_{z \xleftarrow{r} \mathcal{Z}}[J = j_i] \cdot \mathop{\mathbf{E}}_{z \xleftarrow{r} \mathcal{Z}}\left[\frac{|H_z|}{|T_z| + |L_z| + |H_z|} \;\middle|\; J = j_i\right] \tag{30}$$

$$\geq \frac{1}{2m} \sum_{i=0}^{m-1} \Pr_{z \xleftarrow{r} \mathcal{Z}}[H_{f(X)}(f(X)) > j_i] \cdot \mathop{\mathbf{E}}_{z \xleftarrow{r} \mathcal{Z}}\left[\frac{|H_z|}{|T_z| + |L_z| + |H_z|} \;\middle|\; H_{f(X)}(f(X)) > j_i, J = j_i\right] \tag{31}$$

$$\geq \frac{1}{2m} \sum_{i=0}^{m-1} (q_i + \cdots + q_m) \cdot (1 - \frac{1}{3} - \frac{1}{4}) \cdot \mathop{\mathbf{E}}_{z \xleftarrow{r} \mathcal{Z}}\left[\frac{|H_z|}{|T_z| + |L_z| + |H_z|} \;\middle|\; E^1_{j_i}, E^2_{j_i}, H_{f(X)}(f(X)) > j_i, J = j_i\right] \tag{32}$$

$$\geq \frac{1}{2m} \sum_{i=0}^{m-1} (q_i + \cdots + q_m) \cdot (1 - \frac{1}{3} - \frac{1}{4}) \cdot \frac{(q_{i+1} - 4/(n^{d/2})) \cdot 2^{n-j_i-1}}{2^{n-j_i+2}} \tag{33}$$

$$\geq \frac{1}{96m} \sum_{0 \leq i \leq k \leq m-1} (q_i \cdot q_k) - O\left(\frac{1}{n^{d/2}}\right) \tag{34}$$

$$\geq \frac{1}{200m} - \mathsf{negl}(n) - O(n^{-d/2}) \ , \tag{35}$$

where we used conditional expectations, the union bound, the fact that $H_{f(X)}(f(x)) \geq j_i$ is equivalent to $|T_z| \leq 2^{n-j_i}$ and the roughness-regularity assumption that $\sum_{i=1}^{m} q_i = 1 - \mathsf{negl}(n)$.

[6] We use it with $\alpha = 4$ and note that $\texttt{Heavy} = |H_z|$ and $\texttt{Light} = |L_z|$.

We conclude that the log-size of the set of the accessible inputs to an efficient collision-finder is, on average, bounded away from the point-wise entropy. Put differently, we get a noticeable fraction of $\Omega(1/m) = \Omega(\log(n)/s(n))$ average inaccessible entropy bits.