On Constructing Universal One-Way Hash Functions from Arbitrary One-Way Functions

Jonathan Katz[*][†] Chiu-Yuen Koo[*]

Abstract

A fundamental result in cryptography is that a digital signature scheme can be constructed from an arbitrary one-way function. A proof of this somewhat surprising statement follows from two results: first, Naor and Yung defined the notion of *universal one-way hash functions* and showed that the existence of such hash functions implies the existence of secure digital signature schemes. Subsequently, Rompel showed that universal one-way hash functions could be constructed from arbitrary one-way functions. Unfortunately, despite the importance of the result, a complete proof of the latter claim has never been published. In fact, a careful reading of Rompel's original conference publication reveals a number of errors in many of his arguments which have (seemingly) never been addressed.

We provide here what is — as far as we know — the first complete write-up of Rompel's proof that universal one-way hash functions can be constructed from arbitrary one-way functions.

[*] Dept. of Computer Science, University of Maryland. {jkatz,cykoo}@cs.umd.edu

[†] This research was supported in part by NSF CAREER award #0447075.

1 Introduction and Motivation

A key focus of modern cryptography is the construction of cryptographic tools (encryption schemes, digital signatures, etc.) from ever-more-basic cryptographic primitives (trapdoor permutations, one-way functions, etc.). A central question in this area is to determine the *minimal* possible assumptions under which various cryptographic tools can be constructed. The case of digital signature schemes [5] is representative in this regard. Due to their wide-ranging importance, constructions of signature schemes have been the subject of much investigation. The first provably-secure constructions were based on specific, number-theoretic assumptions [6, 5] (more generally, claw-free trapdoor permutations), and this was subsequently improved to show a construction based on arbitrary trapdoor permutations [1]. Somewhat surprisingly, Naor and Yung [8] showed that the presence of a "trapdoor" is not necessary: they introduced the notion of universal one-way hash functions (UOWHFs), showed that UOWHFs suffice to construct signature schemes, and finally demonstrated that UOWHFs could be constructed from one-way permutations. De Santis and Yung [2] improved upon this result by generalizing the class of one-way functions under which UOWHFs could be constructed (roughly speaking, they show how to construct UOWHFs from *regular* one-way functions). Settling the question — since it is relatively easy to see that one-way functions are *necessary* for UOWHFs or secure signature schemes — Rompel [9] proved that one-way functions *suffice* to construct UOWHFs, and hence signature schemes.

Unfortunately, despite the fundamental importance of the above result no complete version of Rompel's proof seems to exist. Making matters worse, the conference version of Rompel's paper [9] (the only version of which we are aware) contains a number of errors and/or omissions[1] which, at best, means that no complete proof is available, and, at worst, calls into question the correctness of various details of Rompel's construction. The lack of a clear and rigorous proof of this result is regrettable, and we hope this paper adequately corrects the situation. We stress that the construction shown here essentially follows that of [9], except for some minor changes made for clarity. Our contribution is thus the *proof*, not the *construction.*

It is our hope also that making this result accessible to a wider audience will lead to improvements and/or simplifications of the construction, as well as further applications of the techniques.

1.1 A High-Level Overview of Rompel's Construction

The construction and its proof are rather technical, and we therefore begin by briefly outlining the main steps in the construction at a relatively high level. It is stressed that the following is necessarily informal, but all formal details are available in the relevant sections of this paper.

Let us begin by recalling the notion of universal one-way hash families [8, 4]. Throughout this work, we will let n denote our security parameter.

Definition 1 A collection of function families $\mathcal{H} = \{H_n\}_{n\in\mathbb{N}}$, where each H_n is a function family $H_n = \{h_s : \{0,1\}^{\ell_1(n)} \to \{0,1\}^{\ell_2(n)}\}_{s\in\{0,1\}^{k(n)}}$ is a *universal one-way hash family* if:

Efficient The functions $\ell_1(\cdot)$ and $k(\cdot)$ are polynomially-bounded; furthermore, given n and strings $s \in \{0,1\}^{k(n)}$ and $x \in \{0,1\}^{\ell_1(n)}$, the value $h_s(x)$ can be computed in poly(n) time. We make the simplifying assumption that $k(\cdot)$ is monotonically increasing so that the value $k(n)$ uniquely determines n (this allows us to simplify our notation, but is otherwise inessential).

[1] Throughout our write-up, we will sometimes point out these errors and omissions. We stress that, ultimately, Rompel's construction is correct. Nevertheless, we still believe that having a complete and correct proof of security for this construction is important.

Compressing For all n we have $\ell_2(n) < \ell_1(n)$.

"Universal one-way" For all PPT algorithms A, the following is negligible (in n):

$$\Pr[x \leftarrow A(1^n); s \leftarrow \{0,1\}^{k(n)}; x' \leftarrow A(1^n, s, x) : x, x' \in \{0,1\}^{\ell_1(n)} \wedge x \neq x' \wedge h_s(x) = h_s(x')].$$

$\diamondsuit$

For the remainder of this informal overview, we fix the security parameter n and so will not explicitly write it in what follows.

Given an arbitrary one-way function $f^0 : \{0,1\}^{\ell'} \to \{0,1\}^{\ell'}$, the construction proceeds in the following stages:

1.1.1 Constructing a one-way function with partially-known structure

A difficulty in dealing directly with f^0 is that we know nothing about its structure. Specifically, for $x \in \{0,1\}^{\ell'}$ define the *sibling set of* x (under f^0) as

$$\mathsf{siblings}_{f^0}(x) \stackrel{\text{def}}{=} \{x' : f^0(x') = f^0(x)\}$$

and, for $0 \leq i \leq \ell'$, let

$$\mathsf{size}_i(f^0) \stackrel{\text{def}}{=} \{x : 2^i \leq |\mathsf{siblings}_{f^0}(x)| < 2^{i+1}\};$$

i.e., $\mathsf{size}_i(f^0)$ consists of those $x \in \{0,1\}^{\ell'}$ which have at least 2^i siblings but fewer than 2^{i+1} siblings (under f^0). We know nothing about how big $\mathsf{size}_i(f^0)$ is for various i. To remedy this, we construct a function $f : \{0,1\}^{\ell} \to \{0,1\}^{\ell}$ such that (1) f is one-way if f^0 is, and (2) $|\mathsf{size}_i(f)|$ is identical for all i in the range (roughly) $[\frac{\ell}{5}, \frac{4\ell}{5}]$. This function f is used in everything that follows.

1.1.2 A Hash Family with Some Hard Siblings

Paraphrasing Definition 1 informally, our goal is to construct a family $\{h_s\}$ such that, for any x and randomly-chosen s, it is infeasible for a computationally-bounded adversary to output *any* sibling of x with respect to h_s (of course, the sibling should be different from x). Toward this, we first show how to construct a family $\{h_s\}$ with the properties that: (1) for any x and randomly-chosen s, it is infeasible for a computationally-bounded adversary to output any so-called "hard" sibling of x with respect to h_s (the precise definition of "hard" is unimportant for what follows, but we do stress that x is never a hard sibling of itself); and (2) for any x and with "high" probability over choice of s, the fraction of siblings of x that are hard is "large". Putting these properties together implies (very roughly speaking) that, on average, there is some noticeable fraction of the siblings of x which are difficult for a PPT adversary to find. We remark that, in some sense, this is the crux of the entire construction, and requires the most involved proof.

Of course, the above says nothing about how easy it might be for an adversary to find other (non-hard) siblings of x with respect to h_s. In what follows, however, we will gradually eliminate such "easy" siblings.

1.1.3 Making Most Siblings Hard

We define a new hash family $\{h'_{\vec{s}}\}$ by running multiple copies of $\{h_s\}$ (as in the previous section) in parallel. Specifically, set

$$h'_{s_1 \cdots s_I}(x_1 \cdots x_I) \stackrel{\text{def}}{=} h_{s_1}(x_1) \cdot h_{s_2}(x_2) \cdots h_{s_I}(x_I),$$

where I is a parameter whose value is unimportant right now. We then say that $x'_1 \cdots x'_I$ is a hard sibling of $x_1 \cdots x_I$ (with respect to $h'_{\vec{s}}$) if for *any* i it holds that x'_i is a hard sibling of x_i (with respect to h_{s_i}). This definition is justified by a simple hybrid argument which shows that it remains computationally difficult for a PPT adversary to find any hard sibling of a fixed $x_1 \cdots x_I$ with respect to $h'_{\vec{s}}$ (since, informally, any such adversary could be used to find a hard sibling of a fixed x_i with respect to h_{s_i}, for some i).

More interestingly, it is not too difficult to see — and not much more difficult to prove — that the above construction increases (on average, over random choice of $\vec{s}$) the fraction of siblings that are hard for any given string $x_1 \cdots x_I$. We will call those siblings of $x_1 \cdots x_I$ that are *not* hard (with respect to a given $h'_{\vec{s}}$) "easy".

1.1.4 Making All Siblings Hard

We now show how to make *all* siblings of a given initial string hard to find. Roughly speaking, we do this by ensuring that no easy siblings remain. The intuition behind the step is actually rather straightforward (although the formal details, which we do not discuss here, are trickier): Say family $\{h'_{\vec{s}}\}$ from the previous section maps ℓ_{in}-bit strings to ℓ_{out}-bit strings. From the previous section, we know that for any fixed string $y \in \{0,1\}^{\ell_{in}}$ and random choice of $\vec{s}$, the fraction of easy siblings of y with respect to $h'_{\vec{s}}$ is "small" on average. For the sake of argument, assume we knew that the number of easy siblings of y was at most 2^E with all but negligible probability over choice of $\vec{s}$.

Consider the family $\{h_{\mu,\vec{s}} : \{0,1\}^{\ell'_{in}} \to \{0,1\}^{\ell_{out}}\}$, where $\mu : \{0,1\}^{\ell'_{in}} \to \{0,1\}^{\ell_{in}}$ is an pairwise[2] independent function (cf. Definition 3, below), and function evaluation is defined via

$$h_{\mu,\vec{s}}(x) \stackrel{\text{def}}{=} h'_{\vec{s}}(\mu(x)).$$

For any string $x \in \{0,1\}^{\ell'_{in}(n)}$, say x' is a hard sibling of x (with respect to $h_{\mu,\vec{s}}$) if $\mu(x')$ is a hard sibling of $\mu(x)$ (with respect to $h'_{\vec{s}}$). It is straightforward to see that it remains computationally difficult for a PPT adversary to find any hard sibling of a fixed x with respect to $h_{\mu,\vec{s}}$ when $\mu, \vec{s}$ are chosen at random (since any such adversary could be used to find a hard sibling of the fixed string $\mu(x)$ with respect to $h'_{\vec{s}}$ for randomly-chosen $\vec{s}$). We now see under what conditions we can argue that *all* siblings of x are hard (with all but negligible probability).

Fix $y = \mu(x)$. We know that with all but negligible probability y has at most 2^E easy siblings. Assuming this to be the case, the probability that there *exists* an easy sibling of x with respect to $h_{\mu,\vec{s}}$ (which is the probability that there exists an $x' \neq x$ such that $\mu(x')$ is an easy sibling of y with respect to $h'_{\vec{s}}$) is at most

$$2^{\ell'_{in}} \cdot \frac{2^E}{2^{\ell_{in}}},$$

using pairwise independence of μ and a union bound. Setting ℓ'_{in} so that $\ell_{in} - E - \ell'_{in} = \Omega(n)$, we see that in this case there are *no* easy siblings of x except with negligible probability. (Of course, we also do not want ℓ'_{in} to be *too* small or else we will have a hard time achieving compression; see the discussion in the next section.)

1.1.5 Completing the Construction

It seems that we are done. That is not quite true, however, as there are two problems left to resolve. The first problem is that the functions $\{h_{\mu,\vec{s}}\}$ may not be length-decreasing! This is relatively easy

[2] In the actual construction, μ will actually be n-wise independent.

to resolve, though, by hashing the output of $h_{\mu,\vec{s}}$ using another pairwise independent function. The second problem is that we assumed knowledge of E in the construction of the previous section. This problem, too, is relatively easy to resolve (though some additional subtleties crop up) by simply enumerating all possible values for E, of which there are only polynomially-many.

2 Preliminaries

We let $x \cdot y$ denote the concatenation of strings x and y. A function $\varepsilon : \mathbb{N} \to [0,1]$ is *negligible* if for any $c > 0$ there exists an n_c such that $\varepsilon(n) < n^{-c}$ for all $n > n_c$. We use "non-negligible" and "not negligible" interchangeably. We say an event occurs with *all but negligible probability* if it occurs with probability $1 - \varepsilon(n)$ for some negligible function ε. A function $\rho : \mathbb{N} \to \mathbb{R}$ is *noticeable* if there exists a $c > 0$ and an n_c such that $\rho(n) > n^{-c}$ for $n > n_c$. Note that a function may be neither negligible nor noticeable.

2.1 One-Way Function Families

We first recall the standard definition of one-way function families:

Definition 2 A *(uniformly) one-way function family* $\mathcal{F} = \{f_n : \{0,1\}^{\ell(n)} \to \{0,1\}^{\ell(n)}\}_{n\in N}$ is a family of functions for which:

Efficient $f_n(x)$ can be computed in time poly(n) (note in particular that this implies that $\ell(\cdot)$ is a polynomially-bounded function).

Hard to invert For all probabilistic, polynomial time (PPT) algorithms A, the following is negligible (in n):

$$\Pr[x \leftarrow \{0,1\}^{\ell(n)}; x' \leftarrow A(1^n, f_n(x)) : f_n(x') = f_n(x)].$$

Using padding techniques (cf. [3, Sect. 2.2.3.2]), the assumption that f_n is length-preserving is without loss of generality. We will also assume that $\ell(n) \geq n^3$ and that $\ell(n)$ is strictly increasing. By padding appropriately, this is again without loss of generality. $\diamondsuit$

When dealing with a function family $\{f_n\}$, we will often simply let f denote f_n when n is understood. For a function f, let $\mathsf{domain}(f)$ denote the domain of f. For $S \subseteq \mathsf{domain}(f)$, we let $f(S)$ denote $\{f(x)\}_{x\in S}$. Using this notation, we let $\mathsf{image}(f) \stackrel{\text{def}}{=} f(\mathsf{domain}(f))$. We also define:

1. $\mathsf{siblings}_f(x) \stackrel{\text{def}}{=} \{x' : f(x') = f(x)\}$; i.e., all values mapped by f to $f(x)$.
2. $\mathsf{size}_i(f) \stackrel{\text{def}}{=} \{x : 2^i \leq |\mathsf{siblings}_f(x)| < 2^{i+1}\}$; i.e., the set of x's for which $\lfloor \log |\mathsf{siblings}_f(x)| \rfloor = i$.
3. $\mathsf{image}_i(f) \stackrel{\text{def}}{=} f(\mathsf{size}_i(f))$; i.e., the set of $y \in \mathsf{image}(f)$ whose inverses all lie in $\mathsf{size}_i(f)$.

For a function whose domain is $\{0,1\}^{\ell(n)}$, the last two notations are meaningful for i ranging from 0 to $\ell(n)$. For completeness, we define $\mathsf{size}_i(f) = \emptyset$ and $\mathsf{image}_i(f) = \emptyset$ when i is outside this range. We state the following facts for convenience, and will use them in the rest of the paper without further comment:

Fact 1 *If* $x' \in \mathsf{siblings}_f(x)$ *and* $x \in \mathsf{size}_i(f)$, *then* $\mathsf{siblings}_f(x') = \mathsf{siblings}_f(x)$ *and* $x' \in \mathsf{size}_i(f)$.

Fact 2 *For all* i *such that* $\mathsf{size}_i(f) \neq \emptyset$ *we have* $2^i \leq \frac{|\mathsf{size}_i(f)|}{|\mathsf{image}_i(f)|} < 2^{i+1}$.

2.2 n-Wise Independent Function Families

We use the following slight generalization of the standard notion of n-wise independent function families:

Definition 3 A collection of function families $\mathcal{U} = \{U_n\}_{n\in\mathbb{N}}$, where each U_n is a function family $U_n = \{\mu_s : \{0,1\}^{\ell_1(n)} \to \{0,1\}^{\ell_2(n)}\}_{s\in\{0,1\}^{k(n)}}$ is *n-wise independent* if:

Efficient As in Definition 1. In particular, we continue to make the simplifying assumption that $k(n)$ uniquely defines n.

n-wise independence For all n, any distinct values $x_1, \dots, x_n \in \{0,1\}^{\ell_1(n)}$, and any (arbitrary) values $y_1, \dots, y_n \in \{0,1\}^{\ell_2(n)}$ we have:

$$\Pr[\mu_s(x_1) = y_1 \wedge \cdots \wedge \mu_s(x_n) = y_n] = 2^{-n\cdot\ell_2(n)},$$

where the probability is over random selection of $s \in \{0,1\}^{k(n)}$.

Efficient sampling For all j, n with $1 \le j \le n$, any distinct $x_1, \dots, x_j \in \{0,1\}^{\ell_1(n)}$, and any $y_1, \dots, y_j \in \{0,1\}^{\ell_2(n)}$, one can sample uniformly in poly(n) time from the set

$$\left\{s \in \{0,1\}^{k(n)} : \mu_s(x_1) = y_1 \wedge \cdots \wedge \mu_s(x_j) = y_j\right\}.$$

$\diamondsuit$

We write $\mu \in U_n$ to mean that there exists an $s \in \{0,1\}^{k(n)}$ such that the functions μ and μ_s are identical. Similarly, the notation "$\mu \leftarrow U_n$" simply means that we choose s uniformly at random from $\{0,1\}^{k(n)}$ and set μ equal to the funtion μ_s.

2.3 Probabilistic Lemmas

In our analysis, we will rely on a number of "Chernoff-Hoeffding"-type bounds. Let us first state a version of the standard Chernoff-Hoeffding bound for reference:

Lemma 1 *Let X be the sum of independent random variables X_i, each in the interval* $[0,1]$, *and such that* $\mathbf{Exp}[X] = \mu$. *Then for any $a > 0$ we have* $\Pr\left[|X-\mu| \ge a\right] \le \max\left\{2\cdot e^{-\frac{a^2}{4\mu}},\ 2^{-a}\right\}$. *Furthermore, for $\mu > a > 0$ we have* $\Pr\left[|X-\mu| \ge a\right] \le 2\cdot e^{-a^2/3\mu}$.

Proof For the first inequality, let $\delta \stackrel{\text{def}}{=} a/\mu$ and distinguish two cases. When $\delta > 2e - 1 > 1$ (here, e is the base of natural logarithms), we have

$$\begin{aligned}\Pr\left[|X-\mu| \ge a\right] &= \Pr\left[X \ge (1+\delta)\mu\right]\\ &\le 2^{-(1+\delta)\mu} \quad \text{(see [7, Ex. 4.1])}\\ &\le 2^{-a}.\end{aligned}$$

When $\delta \le 2e - 1$, we have

$$\begin{aligned}\Pr\left[|X-\mu| \ge a\right] &= \Pr\left[X \ge (1+\delta)\mu\right] + \Pr\left[X \le (1-\delta)\mu\right]\\ &\le e^{-\mu\delta^2/4} + e^{-\mu\delta^2/2} \quad \text{(see [7, Thms. 4.2, 4.3])}\\ &< 2\cdot e^{-a^2/4\mu}.\end{aligned}$$

The second inequality is standard. ■

We now prove a variant of the Chernoff-Hoeffding bound for sums of independent random variables that do not necessarily lie in $[0, 1]$:

Lemma 2 *Let X be the sum of independent random variables X_i, each in the interval $[0, L]$, and such that $\mathbf{Exp}[X] = \mu$. Then for any $a > 0$ we have $\Pr[|X - \mu| \geq a] \leq \max\left\{2 \cdot e^{-\frac{a^2}{4L\mu}}, 2^{-\frac{a}{L}}\right\}$. Furthermore, for $0 < a < \mu$ we have $\Pr[|X - \mu| \geq a] \leq 2 \cdot e^{-a^2/3L\mu}$.*

Proof Define $Y_i \overset{\text{def}}{=} X_i/L$ and $Y \overset{\text{def}}{=} \sum_i Y_i$; note that each Y_i lies in the interval $[0, 1]$, $Y = X/L$, and $\nu \overset{\text{def}}{=} \mathbf{Exp}[Y] = \mu/L$. Now $\Pr[|X - \mu| \geq a] = \Pr[|Y - \nu| \geq a/L]$. Applying Lemma 1 gives the claimed result. ■

The following extension of the Chernoff-Hoeffding bound to the case of n-wise independent random variables (rather than completely independent random variables) will be used extensively in our analysis:

Lemma 3 ([10, Theorem 5]) *Let $n \geq 2$ and let X be the sum of (any number of) n-wise independent random variables, each in $[0, 1]$, such that $\mathbf{Exp}[X] = \mu$. Then:*

1. *If $\delta \in (0, 1]$:*
   - *(a) If $n \leq \lfloor \delta^2 \mu e^{-1/3} \rfloor$, then $\Pr[|X - \mu| \geq \delta\mu] \leq e^{-\lfloor n/2 \rfloor}$;*
   - *(b) If $n \geq \lfloor \delta^2 \mu e^{-1/3} \rfloor$, then $\Pr[|X - \mu| \geq \delta\mu] \leq e^{-\lfloor \delta^2 \mu/3 \rfloor}$.*
2. *If $\delta > 1$:*
   - *(a) If $n \leq \lfloor \delta \mu e^{-1/3} \rfloor$, then $\Pr[|X - \mu| \geq \delta\mu] \leq e^{-\lfloor n/2 \rfloor}$;*
   - *(b) If $n \geq \lfloor \delta \mu e^{-1/3} \rfloor$, then $\Pr[|X - \mu| \geq \delta\mu] \leq e^{-\lfloor \delta \mu/3 \rfloor}$.*

We also state for future reference the following immediate corollary:

Corollary 4 *Let $n \geq 2$ and let X be the sum of (any number of) n-wise independent random variables, each in $[0, 1]$, such that $\mathbf{Exp}[X] \leq \mu_{\max}$ and $\mu_{\max} \geq 2n$. Then for any $\delta \in (0, 1]$:*

$$\Pr[X \geq (1 + \delta) \cdot \mu_{\max}] \leq e^{-\lfloor \delta^2 n/3 \rfloor}.$$

Proof We simply perform a case-by-case analysis using Lemma 3. Let $\mu \overset{\text{def}}{=} \mathbf{Exp}[X]$. There are two cases: If $\mu \geq n$ then

$$\begin{aligned}
\Pr[X \geq (1 + \delta) \cdot \mu_{\max}] &\leq \Pr[X \geq (1 + \delta) \cdot \mu] \\
&\leq \Pr[|X - \mu| \geq \delta\mu] \\
&\leq \max\left\{e^{-\lfloor n/2 \rfloor}, e^{-\lfloor \delta^2 \mu/3 \rfloor}\right\} \\
&\leq e^{-\lfloor \delta^2 n/3 \rfloor},
\end{aligned}$$

where the third inequality uses Case 1 of Lemma 3.

If, on the other hand, $\mu < n$, then

$$\begin{aligned}\Pr[X \geq (1+\delta)\cdot\mu_{\max}] &\leq \Pr[X-\mu \geq \mu_{\max}-\mu] \\ &\leq \Pr[|X-\mu| \geq n] \\ &\leq \max\left\{e^{-\lfloor n/2 \rfloor}, e^{-\lfloor n/3 \rfloor}\right\} \\ &\leq e^{-\lfloor \delta^2 n/3 \rfloor},\end{aligned}$$

using Case 2 of Lemma 3 for the third inequality, and $\delta \leq 1$ for the last inequality. ■

We will also use a counterpart of Lemma 3 which applies to weighted sums of random variables. First, we recall the following result from [10]:

Lemma 5 ([10, Theorem 4(III)]) *Let $n \geq 2$ and let X be the sum of n-wise independent random variables, each in the interval $[0,1]$, such that $\mathbf{Exp}[X] = \mu$. Then, for any $\delta > 0$:*

$$\Pr[|X-\mu| \geq \delta\mu] \leq \left(\frac{nC}{e^{2/3}\delta^2\mu^2}\right)^{\lfloor n/2 \rfloor},$$

where $C \geq \max\{n, \sigma^2[X]\}$.

We then easily obtain the following:

Corollary 6 *Let $n \geq 2$, and let $\{X_i\}$ be n-wise independent random variables in $\{0,1\}$ such that $\Pr[X_i = 1] = p$ for all i. Let $X = \sum_i \lambda_i \cdot X_i$ for some constants $\lambda_i \geq 1$. Set $\lambda = \sum_i \lambda_i$ and $\lambda_{\max} = \max_i\{\lambda_i\}$, and let $\mu = \mathbf{Exp}[X] = p\lambda$. Then for any $\delta > 0$:*

$$\Pr\left[|X-\mu| \geq \delta\mu\right] \leq \left(\frac{nC\lambda_{\max}^2}{e^{2/3}\delta^2\mu^2}\right)^{\lfloor n/2 \rfloor},$$

where $C = \max\{n, \mu/\lambda_{\max}\}$.

Proof Define the random variables $Y_i = \lambda_i X_i/\lambda_{\max}$ and note that $Y_i \in [0,1]$ for all i. Set $Y = \sum_i Y_i = X/\lambda_{\max}$, and let $\nu = \mathbf{Exp}[Y] = \mu/\lambda_{\max}$. Applying Lemma 5 gives:

$$\begin{aligned}\Pr\left[|X-\mu| \geq \delta\mu\right] &= \Pr\left[|Y-\nu| \geq \delta\nu\right] \\ &\leq \left(\frac{nC}{e^{2/3}\delta^2\nu^2}\right)^{\lfloor n/2 \rfloor} \\ &= \left(\frac{nC\lambda_{\max}^2}{e^{2/3}\delta^2\mu^2}\right)^{\lfloor n/2 \rfloor}, \qquad (1)\end{aligned}$$

for $C \geq \max\{n, \sigma^2[Y]\}$. Now,

$$\begin{aligned}\mathbf{Exp}[Y] &= \sum_i \mathbf{Exp}[Y_i] \\ &\geq \sum_i \sigma^2[Y_i] \quad \text{(since } Y_i \in [0,1]\text{)} \\ &= \sigma^2[Y] \quad \text{(using pairwise independence of the } \{Y_i\}\text{)},\end{aligned}$$

so Eq. (1) certainly holds for $C \geq \max\{n, \mathbf{Exp}[Y]\}$. The corollary follows. ■

3 Constructing a Universal One-Way Hash Family

We now give the formal details of the steps outlined in Section 1.1. Our starting point is a one-way function family $\mathcal{F}^0 = \{f_n^0 : \{0,1\}^{\ell'(n)} \to \{0,1\}^{\ell'(n)}\}_{n\in\mathbb{N}}$. For simplicity in the proofs that follow, we assume that certain quantities are powers of two when convenient (this means we can avoid using floors and ceilings, and using appropriate padding this is anyway without loss of generality).

3.1 Constructing a One-Way Function with Partially-Known Structure

We first construct a one-way function family $\mathcal{F}$ whose structure can be better characterized.

Construction 1 Let $\ell(n) \stackrel{\text{def}}{=} 5\ell'(n) + \log \ell'(n) + 2$. Define function family $\mathcal{F} = \{f_n : \{0,1\}^{\ell(n)} \to \{0,1\}^{\ell(n)}\}_{n\in\mathbb{N}}$ as follows:

Let $x \in \{0,1\}^{\ell'(n)}$, $y \in \{0,1\}^{4\ell'(n)}$, and $z \in \{0,1\}^{\log 4\ell'(n)} = \{0,1\}^{\log \ell'(n)+2}$. Then:

$$f_n(x \cdot y \cdot z) \stackrel{\text{def}}{=} f_n^0(x) \cdot \left(y \wedge (0^z \cdot 1^{4\ell'(n)-z})\right) \cdot z,$$

where $y \wedge y'$ represents the bit-wise AND of y and y', and $z \in \{0,1\}^{\log 4\ell'(n)}$ is identified with an integer in the range $\{0, \ldots, 4\ell'(n) - 1\}$. ♣

It is trivial to see that $\mathcal{F}$ is a one-way function family if $\mathcal{F}^0$ is, and so we omit the proof. More interestingly:

Lemma 7 *For all n, i with $\ell'(n) \le i < 4\ell'(n)$, we have*

$$|\mathsf{size}_i(f_n)| = \frac{2^{\ell(n)}}{4\ell'(n)} \quad \textit{and} \quad \frac{2^{\ell(n)-i}}{8\ell'(n)} < |\mathsf{image}_i(f_n)| \le \frac{2^{\ell(n)-i}}{4\ell'(n)}.$$

Furthermore, for any $i \in \{0, \ldots, \ell(n)\}$ we have $|\mathsf{size}_i(f_n)| \le \frac{2^{\ell(n)}}{4\ell'(n)}$.

Proof Fix n and i as in the statement of the lemma and let f denote f_n and f^0 denote f_n^0. First, note that $f(x \cdot y \cdot z) = f(\bar{x} \cdot \bar{y} \cdot \bar{z})$ if and only if $z = \bar{z}$, $f^0(x) = f^0(\bar{x})$, and the final $(4\ell'(n) - z)$ bits of y and $\bar{y}$ are equal; in particular, then, the first z bits of y and $\bar{y}$ can be arbitrary. It follows that $x \cdot y \cdot z \in \mathsf{size}_i(f)$ if and only if $x \in \mathsf{size}_{i-z}(f^0)$. This, in turn, means that for arbitrary $\hat{z}$ we have

$$\begin{aligned} |\{x \cdot y \cdot \hat{z} \in \mathsf{size}_i(f)\}| &= 2^{|y|} \cdot |\mathsf{size}_{i-\hat{z}}(f^0)| \\ &= 2^{4\ell'(n)} \cdot |\mathsf{size}_{i-\hat{z}}(f^0)|. \end{aligned}$$

Summing over all $\hat{z}$, we obtain:

$$\begin{aligned} |\mathsf{size}_i(f)| &= \sum_{\hat{z}=0}^{4\ell'(n)-1} |\{x \cdot y \cdot \hat{z} \in \mathsf{size}_i(f)\}| = \sum_{\hat{z}=0}^{4\ell'(n)-1} 2^{4\ell'(n)} \cdot |\mathsf{size}_{i-\hat{z}}(f^0)| \\ &= 2^{4\ell'(n)} \cdot \sum_{j=i-4\ell'(n)+1}^{i} |\mathsf{size}_j(f^0)|. \end{aligned} \tag{2}$$

Recall that $\mathsf{size}_j(f^0) = \emptyset$ when $j \notin \{0, \ldots, \ell'(n)\}$. When $\ell'(n) \le i < 4\ell'(n)$, we thus have:

$$\begin{aligned} |\mathsf{size}_i(f)| &= 2^{4\ell'(n)} \cdot \sum_{j=0}^{\ell'(n)} |\mathsf{size}_j(f^0)| \\ &= 2^{4\ell'(n)} \cdot |\mathsf{domain}(f^0)| = 2^{5\ell'(n)} = \frac{2^{\ell(n)}}{4\ell'(n)}, \end{aligned}$$

as desired. The bound on $|\mathsf{image}_i(f_n)|$ follows by Fact 2. The final statement of the lemma follows using Eq. (2) and the observation that $\sum_j |\mathsf{size}_j(f^0)| \leq |\mathsf{domain}(f^0)|$. ■

3.2 A Hash Family with Some Hard Siblings

We now take the one-way function family $\mathcal{F}$ constructed in the previous section and construct a hash family for which it is computationally hard to find some noticeable fraction of siblings for any fixed x and randomly-chosen hash function from the family.

In the rest of the paper, we will omit the explicit dependence of certain values on n unless we want to explicitly highlight this dependency. Thus, for example, we will let $\ell = \ell(n)$ and $\ell' = \ell'(n)$ (as in Construction 1) in all that follows. We also sometimes write f instead of f_n for convenience.

Construction 2 Let $\mathcal{F} = \{f_n\}$ be as in Construction 1 and let $\mathcal{U}_1 = \{U_n^1\}_{n\in\mathbb{N}}$ and $\mathcal{U}_2 = \{U_n^2\}_{n\in\mathbb{N}}$ be n-wise independent function families such that $U_n^1 = \{\mu_{1,s} : \{0,1\}^{\ell/2} \to \{0,1\}^{\ell}\}_{s\in k_1(n)}$ and $U_n^2 = \{\mu_{2,s} : \{0,1\}^{\ell} \to \{0,1\}^{\ell/2-2\log\ell}\}_{s\in k_2(n)}$. (From now on, we drop explicit mention of the key s and simply speak of functions $\mu_1 \in U_n^1$ and $\mu_2 \in U_n^2$.) Construct $\mathcal{H} = \{H_n\}_{n\in\mathbb{N}}$ where $H_n = \{h_{\mu_1,\mu_2} : \{0,1\}^{\ell/2} \to \{0,1\}^{\ell/2-2\log\ell}\}_{\mu_1\in U_n^1;\, \mu_2\in U_n^2}$, and h_{μ_1,μ_2} is defined as follows:

$$h_{\mu_1,\mu_2}(x) \overset{\text{def}}{=} \mu_2(f_n(\mu_1(x))).$$

♣

To analyze the above construction, we first define a notion of "hard" siblings:

Definition 4 Given μ_1, μ_2, and $x \in \{0,1\}^{\ell/2}$, define the *hard sibling set* $\mathsf{hard}_{h_{\mu_1,\mu_2}}(x)$ to be the set of $x' \in \mathsf{siblings}_{h_{\mu_1,\mu_2}}(x)$ for which $f(\mu_1(x')) \neq f(\mu_1(x))$ and $\mu_1(x') \in \mathsf{size}_{\frac{\ell}{2}}(f)$. Note that the latter condition is equivalent to requiring that $f(\mu_1(x')) \in \mathsf{image}_{\frac{\ell}{2}}(f)$. Also, note that $x \notin \mathsf{hard}_{h_{\mu_1,\mu_2}}(x)$. ♢

We show that over random choice of μ_1, μ_2, it is computationally infeasible to find a hard sibling of any fixed x with respect to h_{μ_1,μ_2}.

Theorem 8 *Assuming $\mathcal{F}$ is a one-way function family, the following is negligible for all* PPT *A:*

$$\mathsf{Succ}_A^{\mathsf{hard}}(n) \overset{\text{def}}{=} \Pr[x \leftarrow A(1^n); \mu_1 \leftarrow U_n^1; \mu_2 \leftarrow U_n^2; \bar{x} \leftarrow A(1^n, \mu_1, \mu_2, x) : \bar{x} \in \mathsf{hard}_{h_{\mu_1,\mu_2}}(x)].$$

Proof Assume toward a contradiction that there exists a PPT A for which $\mathsf{Succ}_A^{\mathsf{hard}}(n)$ is not negligible. We construct an algorithm B which inverts $f = f_n$ with non-negligible probability. This gives the desired result.

B takes as input $\bar{z} = f(\bar{y})$ for some $\bar{y}$, and is defined as follows:

```
$\underline{B(1^n, \bar{z})}$
$x \leftarrow A(1^n)$
$\mu_1 \leftarrow U_n^1$
Pick $\mu_2$ uniformly at random from the set $\{\mu_2 \in U_n^2 : \mu_2(f(\mu_1(x))) = \mu_2(\bar{z})\}$
$\bar{x} \leftarrow \mathcal{A}(1^n, \mu_1, \mu_2, x)$
If $f(\mu_1(x)) = \bar{z}$, output $\mu_1(x)$
If $f(\mu_1(\bar{x})) = \bar{z}$, output $\mu_1(\bar{x})$
Otherwise, output $\perp$.
```

Let U_ℓ denote the uniform distribution over strings of length ℓ, and let size_i and image_i denote $\mathsf{size}_i(f)$ and $\mathsf{image}_i(f)$, respectively. Say that "B inverts $\bar{z}$" if $f(B(1^n, \bar{z})) = \bar{z}$. Our goal is to show that $\Pr_{\bar{z} \leftarrow f(U_\ell)}[B \text{ inverts } \bar{z}]$ is not negligible. The following claim indicates that if we can show that B succeeds in inverting $\bar{z}$ with non-negligible probability *when $\bar{z}$ is is uniformly distributed in* $\mathsf{image}_{\frac{\ell}{2}}$, the proof is complete. Informally, this is because: (1) $\mathsf{size}_{\frac{\ell}{2}}$, which is the pre-image of $\mathsf{image}_{\frac{\ell}{2}}$, is a noticeable fraction of $\mathsf{domain}(f)$ (and so a $\bar{y}$ chosen uniformly in $\mathsf{domain}(f)$ has noticeable probability of being in $\mathsf{size}_{\frac{\ell}{2}}$); and (2) for any two elements $\bar{z}, \bar{z}' \in \mathsf{image}_{\frac{\ell}{2}}$, the number of pre-images of $\bar{z}$ is within a factor of two of the number of pre-images of $\bar{z}'$ (and so choosing $\bar{z}$ uniformly in $\mathsf{image}_{\frac{\ell}{2}}$ is "close enough" to choosing $\bar{y}$ uniformly in $\mathsf{size}_{\frac{\ell}{2}}$ and setting $\bar{z} = f(\bar{y})$). Formally:

Claim 9 $\Pr_{\bar{z} \leftarrow f(U_\ell)}[B \textit{ inverts } \bar{z}] \geq \frac{1}{8\ell'} \cdot \Pr_{\bar{z} \leftarrow \mathsf{image}_{\frac{\ell}{2}}}[B \textit{ inverts } \bar{z}]$.

Proof (of claim) We have:

$$\begin{aligned}
\Pr_{\bar{z} \leftarrow f(U_\ell)}[B \text{ inverts } \bar{z}] &\geq \Pr_{\bar{z} \leftarrow f(U_\ell)}[B \text{ inverts } \bar{z} \bigwedge \bar{z} \in \mathsf{image}_{\frac{\ell}{2}}] \\
&= \sum_{z \in \mathsf{image}_{\frac{\ell}{2}}} \Pr[B \text{ inverts } z] \cdot \Pr_{\bar{z} \leftarrow f(U_\ell)}[\bar{z} = z]. \qquad (3)
\end{aligned}$$

Furthermore, for any $z \in \mathsf{image}_{\frac{\ell}{2}}$ we have

$$\begin{aligned}
\Pr_{\bar{z} \leftarrow f(U_\ell)}[\bar{z} = z] &= \Pr_{\bar{z} \leftarrow f(U_\ell)}[\bar{z} = z \mid \bar{z} \in \mathsf{image}_{\frac{\ell}{2}}] \cdot \Pr_{\bar{z} \leftarrow f(U_\ell)}[\bar{z} \in \mathsf{image}_{\frac{\ell}{2}}] \\
&= \Pr_{\bar{z} \leftarrow f(U_\ell)}[\bar{z} = z \mid \bar{z} \in \mathsf{image}_{\frac{\ell}{2}}] \cdot \Pr_{\bar{y} \leftarrow U_\ell}[\bar{y} \in \mathsf{size}_{\frac{\ell}{2}}] \\
&= \Pr_{\bar{z} \leftarrow f(U_\ell)}[\bar{z} = z \mid \bar{z} \in \mathsf{image}_{\frac{\ell}{2}}] \cdot \frac{|\mathsf{size}_{\frac{\ell}{2}}|}{2^\ell} \\
&= \frac{1}{4\ell'} \cdot \Pr_{\bar{z} \leftarrow f(U_\ell)}[\bar{z} = z \mid \bar{z} \in \mathsf{image}_{\frac{\ell}{2}}], \qquad (4)
\end{aligned}$$

using Lemma 7. Now, for any $z \in \mathsf{image}_{\frac{\ell}{2}}$

$$\begin{aligned}
\Pr_{\bar{z} \leftarrow f(U_\ell)}[\bar{z} = z \mid \bar{z} \in \mathsf{image}_{\frac{\ell}{2}}] = \Pr_{\bar{y} \leftarrow U_\ell}[f(\bar{y}) = z \mid \bar{y} \in \mathsf{size}_{\frac{\ell}{2}}] &= \frac{|\{\bar{y} : f(\bar{y}) = z\}|}{|\mathsf{size}_{\frac{\ell}{2}}|} \\
&> \frac{2^{\ell/2}}{2^{\ell/2+1} \cdot |\mathsf{image}_{\frac{\ell}{2}}|} \\
&= \frac{1}{2 \cdot |\mathsf{image}_{\frac{\ell}{2}}|}. \qquad (5)
\end{aligned}$$

Combining Eqs. (3)–(5), we obtain:

$$\begin{aligned}
\Pr_{\bar{z} \leftarrow f(U_\ell)}[B \text{ inverts } \bar{z}] &\geq \frac{1}{8\ell'} \cdot \sum_{z \in \mathsf{image}_{\frac{\ell}{2}}} \Pr[B \text{ inverts } z] \cdot \frac{1}{|\mathsf{image}_{\frac{\ell}{2}}|} \\
&= \frac{1}{8\ell'} \cdot \Pr_{z \leftarrow \mathsf{image}_{\frac{\ell}{2}}}[B \text{ inverts } z],
\end{aligned}$$

as desired. □

To complete the proof of the theorem, we proceed in two stages. First, we show that

$$\mathsf{Succ}'_A(n) \stackrel{\rm def}{=} \Pr\left[\begin{array}{c} x \leftarrow A(1^n); \bar{z} \leftarrow \mathsf{image}_{\frac{\ell}{2}}; \mu_1 \leftarrow U_n^1; \\ \mu_2 \leftarrow \{\mu_2 : \mu_2(f(\mu_1(x))) = \mu_2(\bar{z})\}; \bar{x} \leftarrow A(1^n, \mu_1, \mu_2, x) \end{array} : \bar{x} \in \mathsf{hard}_{h_{\mu_1,\mu_2}}(x)\right]$$

is within a constant multiplicative factor of $\mathsf{Succ}^{\mathsf{hard}}_A(n)$. (Note that $\mathsf{Succ}'_A(n)$ is the probability that A outputs a hard sibling of x when A is invoked by B, assuming the input to B is uniformly distributed in $\mathsf{image}_{\frac{\ell}{2}}$.) Next, we show that whenever A outputs a hard sibling of x (in the experiment described by Succ'_A), then B outputs an inverse of $\bar{z}$ with noticeable probability. Under the assumption that $\mathsf{Succ}^{\mathsf{hard}}_A$ is not negligible, these imply that $\Pr_{\bar{z}\leftarrow\mathsf{image}_{\frac{\ell}{2}}}[B \text{ inverts } \bar{z}]$ is not negligible; applying Claim 9 then completes the proof of the theorem.

Claim 10 *For n large enough,* $\mathsf{Succ}'_A(n) \geq \frac{1}{3} \cdot \mathsf{Succ}^{\mathsf{hard}}_A(n)$.

Proof (of claim) We may write

$$\mathsf{Succ}^{\mathsf{hard}}_A(n) = \sum_{\hat{\mu}_1,\hat{\mu}_2} \Pr[x \leftarrow A(1^n); \bar{x} \leftarrow A(1^n, \hat{\mu}_1, \hat{\mu}_2, x) : \bar{x} \in \mathsf{hard}_{h_{\hat{\mu}_1,\hat{\mu}_2}}(x)] \cdot \frac{1}{|U_n^1|} \cdot \frac{1}{|U_n^2|}$$

and

$$\begin{aligned} \mathsf{Succ}'_A(n) &= \sum_{\hat{\mu}_1,\hat{\mu}_2} \Pr\left[x \leftarrow A(1^n); \bar{x} \leftarrow A(1^n, \hat{\mu}_1, \hat{\mu}_2, x) : \bar{x} \in \mathsf{hard}_{h_{\hat{\mu}_1,\hat{\mu}_2}}(x)\right] \\ &\qquad \cdot \frac{1}{|U_n^1|} \cdot \Pr\left[\bar{z} \leftarrow \mathsf{image}_{\frac{\ell}{2}}; \mu_2 \leftarrow \{\mu_2 : \mu_2(f(\hat{\mu}_1(x))) = \mu_2(\bar{z})\} : \mu_2 = \hat{\mu}_2\right], \end{aligned}$$

where the above sums are taken over $\hat{\mu}_1 \in U_n^1$ and $\hat{\mu}_2 \in U_n^2$. Let $z \stackrel{\rm def}{=} f(\hat{\mu}_1(x))$. To show that $\mathsf{Succ}'_A(n) \geq \frac{1}{3} \cdot \mathsf{Succ}^{\mathsf{hard}}_A(n)$ (for n large enough), it suffices to show that for *any* $z \in \{0,1\}^\ell$, the value of

$$\Pr[B \text{ picks } \hat{\mu}_2 \mid z] \stackrel{\rm def}{=} \Pr\left[\bar{z} \leftarrow \mathsf{image}_{\frac{\ell}{2}}; \mu_2 \leftarrow \{\mu_2 : \mu_2(z) = \mu_2(\bar{z})\} : \mu_2 = \hat{\mu}_2\right]$$

is at least $\frac{1}{2\cdot|U_n^2|}$ except for a negligible fraction of the $\hat{\mu}_2 \in U_n^2$ (note that any individual term in either of the above sums is negligible, since $1/|U_n^1|$ is negligible).

Fix any $z \in \{0,1\}^\ell$, and let

$$\Pr[B \text{ picks } \hat{\mu}_2 \mid z, \bar{z}] \stackrel{\rm def}{=} \Pr\left[\mu_2 \leftarrow \left\{\mu_2 : \mu_2(z) = \mu_2(\bar{z})\right\} : \mu_2 = \hat{\mu}_2\right].$$

Define[3] $G_z(\hat{\mu}_2) \stackrel{\rm def}{=} \left\{\bar{z} : \bar{z} \in \mathsf{image}_{\frac{\ell}{2}} \bigwedge \hat{\mu}_2(\bar{z}) = \hat{\mu}_2(z)\right\}$. Then:

$$\begin{aligned} \Pr[B \text{ picks } \hat{\mu}_2 \mid z] &= \sum_{\bar{z}\in G_z(\hat{\mu}_2)} \Pr[B \text{ picks } \hat{\mu}_2 \mid z, \bar{z}] \cdot \Pr_{\bar{z}'\leftarrow\mathsf{image}_{\frac{\ell}{2}}}[\bar{z}' = \bar{z}] \\ &= \sum_{\bar{z}\in G_z(\hat{\mu}_2)} \frac{\Pr[B \text{ picks } \hat{\mu}_2 \mid z, \bar{z}]}{|\mathsf{image}_{\frac{\ell}{2}}|} \\ &= \sum_{\bar{z}\in G_z(\hat{\mu}_2)} \frac{1}{|\{\mu_2 : \mu_2(\bar{z}) = \mu_2(z)\}| \cdot |\mathsf{image}_{\frac{\ell}{2}}|}. \qquad (6) \end{aligned}$$

[3] We define $G_z(\hat{\mu}_2)$ differently from [9]. The reason is that, in contrast to what is claimed in [9], the extended Chernoff bound (Lemma 3) does not seem to apply to $G(\hat{\mu}_2)$ as defined there.

Consider first the case that $z \notin \mathsf{image}_{\frac{\ell}{2}}$ (and hence $z \notin G_z(\hat{\mu}_2)$). Since U_n^2 is an n-wise independent function family, we have $|\{\mu_2 : \mu_2(\bar{z}) = \mu_2(z)\}| = |U_n^2| \cdot 2^{-\frac{\ell}{2}+2\log\ell}$ for any $\bar{z} \in G_z(\hat{\mu}_2)$ (recall that $\ell/2 - 2\log\ell$ is the output-length of functions in U_n^2). Eq. (6) then gives:

$$\Pr[B \text{ picks } \hat{\mu}_2 \mid z] = |G_z(\hat{\mu}_2)| \cdot \frac{2^{\frac{\ell}{2}-2\log\ell}}{|U_n^2|} \cdot \frac{1}{|\mathsf{image}_{\frac{\ell}{2}}|}\,. \tag{7}$$

In the expression above, only $|G_z(\hat{\mu}_2)|$ depends on $\hat{\mu}_2$. Viewing $|G_z(\hat{\mu}_2)|$ as a random variable (over random choice of $\hat{\mu}_2$, with z fixed), we have:

$$\begin{aligned}\mathbf{Exp}_{\hat{\mu}_2 \leftarrow U_n^2}\left[\,|G_z(\hat{\mu}_2)|\,\right] &= \sum_{\bar{z} \in \mathsf{image}_{\frac{\ell}{2}}} \Pr_{\hat{\mu}_2 \leftarrow U_n^2}[\hat{\mu}_2(\bar{z}) = \hat{\mu}_2(z)] \\ &= |\mathsf{image}_{\frac{\ell}{2}}| \cdot 2^{-\frac{\ell}{2}+2\log\ell}.\end{aligned}$$

From Lemma 7, we have $|\mathsf{image}_{\frac{\ell}{2}}| = \Theta\left(\frac{2^{\ell/2}}{\ell'}\right)$. Using the fact that $\ell' = \Theta(\ell)$, we see that $\mathbf{Exp}_{\hat{\mu}_2 \leftarrow U_n^2}\left[\,|G_z(\hat{\mu}_2)|\,\right] = \Theta(\ell)$. Now, note that $|G_z(\hat{\mu}_2)|$ is a sum (over $\bar{z} \in \mathsf{image}_{\frac{\ell}{2}}$) of the indicator random variables $\delta_{\bar{z}}$ such that $\delta_{\bar{z}} = 1$ iff $\hat{\mu}_2(\bar{z}) = \hat{\mu}_2(z)$. Since U_n^2 is an n-wise independent function family, the $\{\delta_{\bar{z}}\}$ are n-wise independent and hence Lemma 3 applies. Setting $\delta = \frac{1}{2}$ and applying the lemma shows that for all but a negligible fraction of the $\hat{\mu}_2 \in U_n^2$, the value of $|G_z(\hat{\mu}_2)|$ is within a factor of two of its expectation; i.e., for all but a negligible fraction of $\hat{\mu}_2 \in U_n^2$ we have

$$|G_z(\hat{\mu}_2)| \geq \frac{1}{2} \cdot |\mathsf{image}_{\frac{\ell}{2}}| \cdot 2^{-\frac{\ell}{2}+2\log\ell}.$$

Plugging this into Eq. (7) gives the desired result that $\Pr[B \text{ picks } \hat{\mu}_2 \mid z] \geq \frac{1}{2\cdot|U_n^2|}$ for all but a negligible fraction of the $\hat{\mu}_2 \in U_n^2$.

For the case when $z \in \mathsf{image}_{\frac{\ell}{2}}$, the analysis proceeds as above except that we need to deal separately with the special case $z = \bar{z}$ (in which case $\{\mu_2 : \mu_2(\bar{z}) = \mu_2(z)\} = U_n^2$, which only helps). We omit the details. This concludes the proof of the claim. □

To conclude the proof, we show that whenever A outputs a hard sibling of x (in the experiment defining Succ'_A), B outputs an inverse of $\bar{z}$ with noticeable probability $\Omega(1/\ell)$. As shown in the proof of the preceding claim, for any $z = f(\mu_1(x))$ and all but a negligible fraction of $\hat{\mu}_2 \in U_n^2$ we have $|G_z(\hat{\mu}_2)| = \Theta(\ell)$. Given the entire view of A (and assuming $z \neq \bar{z}$, since B will anyway output the desired inverse in this case), $\bar{z}$ is uniformly distributed in $G_z(\hat{\mu}_2) \setminus \{z\}$. If A outputs a hard sibling $\bar{x}$ of x, then by definition $f(\mu_1(\bar{x})) \in G_z(\hat{\mu}_2) \setminus \{z\}$. Thus, conditioned on A's outputting a hard sibling, the probability that $f(\mu_1(\bar{x})) = \bar{z}$ and B outputs the correct inverse is at least $1/|G_z(\hat{\mu}_2) \setminus \{z\}| = \Omega(1/\ell)$. ■

The previous theorem shows that it is computationally infeasible to find "hard" siblings of any fixed x. We now show[4] that for any fixed x and with constant probability over choice of $\mu_1 \in U_n^1, \mu_2 \in U_n^2$, the hard siblings of x are a noticeable fraction of all siblings of x.

Theorem 11 *Let $x \in \{0,1\}^{\ell/2}$ be arbitrary. Then for ℓ large enough it holds that with probability at least $1/3$ (over random choice of $\mu_1 \in U_n^1$ and $\mu_2 \in U_n^2$) we have:*

$$\frac{|\mathsf{hard}_{h_{\mu_1,\mu_2}}(x)|}{|\mathsf{siblings}_{h_{\mu_1,\mu_2}}(x)|} \geq \frac{1}{\ell}\,.$$

[4] Theorem 11 corresponds to [9, Lemma 5], but the lemma as stated there is actually incorrect.

I.e., with probability at least $1/3$ *the hard siblings of* x *are a noticeable fraction of all siblings of* x*.*

Proof We show that $|\mathsf{siblings}_{h_{\mu_1,\mu_2}}(x)| \le \frac{4\ell^2}{5}$ with probability at least 34/100 and furthermore that $|\mathsf{hard}_{h_{\mu_1,\mu_2}}(x)| \ge \ell$ with all but negligible probability. Applying a union bound, we see that with probability at least 1/3 both bounds hold. The theorem follows.

We write size_i and image_i for $\mathsf{size}_i(f)$ and $\mathsf{image}_i(f)$, respectively, and also let $\mathsf{size}_{i,j} \overset{\text{def}}{=} \bigcup_{k=i}^{j} \mathsf{size}_k$ and $\mathsf{image}_{i,j} \overset{\text{def}}{=} \bigcup_{k=i}^{j} \mathsf{image}_k$. It will be useful to recall that $|\mathsf{size}_i| = \frac{2^\ell}{4\ell'}$ for $\ell' \le i < 4\ell'$ (cf. Lemma 7) and the fact that $\ell > 5\ell'$. We stress also that x is fixed throughout what follows.

Claim 12 *For* ℓ *large enough, with probability at least* $34/100$ *over choice of* μ_1, μ_2 *we have* $|\mathsf{siblings}_{h_{\mu_1,\mu_2}}(x)| \le \frac{4\ell^2}{5}$*.*

Proof (of claim) If $x' \in \mathsf{siblings}_{h_{\mu_1,\mu_2}}(x)$, then exactly one of the following is true:

1. $x' = x$;
2. $x' \ne x$ but $f(\mu_1(x')) = f(\mu_1(x))$;
3. $f(\mu_1(x')) \ne f(\mu_1(x))$ but $\mu_2(f(\mu_1(x'))) = \mu_2(f(\mu_1(x)))$.

The following two sub-claims bound the number of x' falling into the second and third categories, respectively.

Sub-claim *Let* $S_{\mu_1} \overset{\text{def}}{=} \{x' : x' \ne x \wedge f(\mu_1(x')) = f(\mu_1(x))\}$*; i.e.,* S_{μ_1} *contains the* x' *falling into the second category above. Then for* ℓ *large enough, with probability at least* $37/100$ *(over choice of* μ_1*) we have* $|S_{\mu_1}| < 2\ell$*.*

Proof (of sub-claim) We first show that the event "$\mu_1(x) \in \mathsf{size}_{0,\frac{\ell}{2}+\log\ell-1}$" occurs with probability at least 3/8 over choice of μ_1; we then show that the desired bound on $|S_{\mu_1}|$ holds with all but negligible probability conditioned on this event.

An straightforward calculation gives:

$$\begin{aligned}\Pr_{\mu_1 \leftarrow U_n^1}\left[\mu_1(x) \in \mathsf{size}_{0,\frac{\ell}{2}+\log\ell-1}\right] &= \sum_{w \in \mathsf{size}_{0,\frac{\ell}{2}+\log\ell-1}} \Pr_{\mu_1 \leftarrow U_n^1}[\mu_1(x) = w] \\ &\ge \sum_{w \in \mathsf{size}_{\ell',5\ell'/2}} 2^{-\ell} \\ &\ge \left(\frac{5\ell'}{2} - \ell'\right)\left(\frac{2^\ell}{4\ell'}\right)\frac{1}{2^\ell} \quad = \quad \frac{3}{8}, \end{aligned} \tag{8}$$

and so with probability at least 3/8 we have $\mu_1(x) \in \mathsf{size}_{0,\frac{\ell}{2}+\log\ell-1}$. Let Good denote this event, and let Good^* denote the event that $\mu_1(x) \in \mathsf{size}_{\frac{\ell}{2}+\log\ell-1}$. It is evident that

$$\Pr\left[|S_{\mu_1}| \ge 2\ell \,\middle|\, \mathsf{Good}\right] \quad \le \quad \Pr\left[|S_{\mu_1}| \ge 2\ell \,\middle|\, \mathsf{Good}^*\right], \tag{9}$$

since $\mu_1(x)$ has the most siblings when Good^* occurs. (An argument as above shows that the probability of Good^* is non-zero, so conditioning on this event is ok.) Now,

$$\begin{aligned}
&\mathbf{Exp}_{\mu_1 \leftarrow U_n^1}\left[\,|S_{\mu_1}|\;\middle|\;\mathsf{Good}^*\right] \\
&\quad= \sum_{x' \neq x} \Pr_{\mu_1 \leftarrow U_n^1}\left[f(\mu_1(x')) = f(\mu_1(x))\;\middle|\;\mathsf{Good}^*\right] \\
&\quad= \sum_{x' \neq x}\ \sum_{y \in \mathsf{size}_{\frac{\ell}{2}+\log\ell-1}} \Pr_{\mu_1 \leftarrow U_n^1}\left[\mu_1(x) = y \bigwedge \mu_1(x') \in \mathsf{siblings}_f(y)\;\middle|\;\mathsf{Good}^*\right] \\
&\quad= \sum_{x' \neq x}\ \sum_{y \in \mathsf{size}_{\frac{\ell}{2}+\log\ell-1}} \Pr_{\mu_1 \leftarrow U_n^1}\left[\mu_1(x) = y\;\middle|\;\mathsf{Good}^*\right] \cdot \Pr_{\mu_1 \leftarrow U_n^1}\left[\mu_1(x') \in \mathsf{siblings}_f(y)\right],
\end{aligned}$$

where we omit the (implicit) conditioning on the event "$\mu_1(x) = y$" in the second probability since U_n^1 is n-wise independent. Continuing:

$$\begin{aligned}
&\mathbf{Exp}_{\mu_1 \leftarrow U_n^1}\left[\,|S_{\mu_1}|\;\middle|\;\mathsf{Good}^*\right] \\
&\quad= \sum_{x' \neq x}\ \sum_{y \in \mathsf{size}_{\frac{\ell}{2}+\log\ell-1}} \frac{1}{|\mathsf{size}_{\frac{\ell}{2}+\log\ell-1}|} \cdot \left(2^{-\ell} \cdot |\mathsf{siblings}_f(y)|\right) \\
&\quad= \sum_{y \in \mathsf{size}_{\frac{\ell}{2}+\log\ell-1}} \frac{1}{|\mathsf{size}_{\frac{\ell}{2}+\log\ell-1}|} \cdot \left(2^{-\ell/2} - 2^{-\ell}\right) \cdot |\mathsf{siblings}_f(y)|.
\end{aligned}$$

For $y \in \mathsf{size}_{\frac{\ell}{2}+\log\ell-1}$, we have $\frac{\ell \cdot 2^{\ell/2}}{2} \leq \left|\mathsf{siblings}_f(y)\right| < \ell \cdot 2^{\ell/2}$, and therefore

$$\frac{\ell}{2} \cdot \left(1 - 2^{-\ell/2}\right) \;\leq\; \mathbf{Exp}_{\mu_1 \leftarrow U_n^1}\left[\,|S_{\mu_1}|\;\middle|\;\mathsf{Good}^*\right] \;<\; \ell \cdot \left(1 - 2^{-\ell/2}\right).$$

Letting $\delta_{x'}$ be an indicator random variable which is 1 if and only if $f(\mu_1(x')) = f(\mu_1(x))$, we see that $|S_{\mu_1}| = \sum_{x' \neq x} \delta_{x'}$. Relying on the fact that the $\delta_{x'}$ are $(n-1)$-wise independent[5] and applying Lemma 3, we see that $|S_{\mu_1}| \geq 2\ell$ with only negligible probability conditioned on occurrence of Good^*, and hence (by Eq. (9)) $|S_{\mu_1}| \geq 2\ell$ with only negligible probability conditioned on occurrence of Good. Since we have already argued that Good occurs with probability at least 3/8, we conclude that, for ℓ large enough, with probability at least 37/100 over choice of μ_1 we have $|S_{\mu_1}| < 2\ell$. □

Sub-claim *Let*

$$\begin{aligned}
S'_{\mu_1,\mu_2} &\stackrel{\text{def}}{=} \left\{x' : f(\mu_1(x')) \neq f(\mu_1(x)) \bigwedge \mu_2(f(\mu_1(x'))) = \mu_2(f(\mu_1(x)))\right\} \\
&= \mathsf{siblings}_{h_{\mu_1,\mu_2}}(x) \setminus (S_{\mu_1} \cup \{x\})\,;
\end{aligned}$$

i.e., S'_{μ_1,μ_2} contains the x' falling into the third category from above. Fix arbitrary constant $\delta \in (0,1]$. Then with probability at least $1 - \ell^{-1/2} - \mathsf{negl}(n)$ we have

$$|S'_{\mu_1,\mu_2}| \leq (1+\delta)^2 \frac{\ell^2}{4\ell'} \left(\frac{\ell}{2} + \frac{3}{2}\log\ell + 1\right).$$

[5] Since we are conditioning on $\mu_1(x) \in \mathsf{size}_{\frac{\ell}{2}+\log\ell-1}$, we are left with only $n-1$ degrees of freedom.

Proof (of sub-claim) First, we show that with probability at least $1 - \ell^{-1/2}$ over choice of μ_1, μ_2, the set $\mu_1(S'_{\mu_1,\mu_2})$ lies entirely within $\mathsf{size}_{0,\frac{\ell}{2}+\frac{3}{2}\log\ell}$. We then show that $|S'_{\mu_1,\mu_2} \cap \mu_1^{-1}(\mathsf{size}_{0,\frac{\ell}{2}+\frac{3}{2}\log\ell})|$ is at most the claimed quantity with all but negligible probability. Applying a union bound proves the claim.

Observe that

$$\begin{aligned}\left|\mathsf{image}_{\frac{\ell}{2}+\frac{3}{2}\log\ell+1,\ell}\right| &= \sum_{i=\frac{\ell}{2}+\frac{3}{2}\log\ell+1}^{\ell} |\mathsf{image}_i(f)| \\ &\leq \sum_{i=\frac{\ell}{2}+\frac{3}{2}\log\ell+1}^{\ell} \left(\frac{2^\ell}{4\ell'}\right)\cdot 2^{-i} \leq \frac{2^{\frac{\ell}{2}}}{16\ell'\ell^{\frac{3}{2}}}, \qquad (10)\end{aligned}$$

using Lemma 7. We now bound the probability that $\mu_1(S'_{\mu_1,\mu_2}) \subseteq \mathsf{size}_{0,\frac{\ell}{2}+\frac{3}{2}\log\ell}$ by:

$$\begin{aligned}&\Pr_{\mu_1,\mu_2}\left[\forall x' \in S'_{\mu_1,\mu_2} : \mu_1(x') \in \mathsf{size}_{0,\frac{\ell}{2}+\frac{3}{2}\log\ell}\right] \\ &= 1 - \Pr_{\mu_1,\mu_2}\Big[\exists x' : f(\mu_1(x')) \neq f(\mu_1(x)) \bigwedge \mu_2(f(\mu_1(x'))) = \mu_2(f(\mu_1(x))) \\ &\qquad\qquad \bigwedge \mu_1(x') \in \mathsf{size}_{\frac{\ell}{2}+\frac{3}{2}\log\ell+1,\ell}\Big] \\ &\geq 1 - \Pr_{\mu_1,\mu_2}\left[\exists z' \in \mathsf{image}_{\frac{\ell}{2}+\frac{3}{2}\log\ell+1,\ell} : z' \neq f(\mu_1(x)) \bigwedge \mu_2(z') = \mu_2(f(\mu_1(x)))\right]\end{aligned}$$

(setting $z' = f(\mu_1(x'))$ to obtain the final inequality). Continuing, and using Eq. (10), we have:

$$\begin{aligned}&\Pr_{\mu_1,\mu_2}\left[\forall x' \in S'_{\mu_1,\mu_2} : \mu_1(x') \in \mathsf{size}_{0,\frac{\ell}{2}+\frac{3}{2}\log\ell}\right] \\ &\geq 1 - \Pr_{\mu_1,\mu_2}\left[\exists z' \in \mathsf{image}_{\frac{\ell}{2}+\frac{3}{2}\log\ell+1,\ell} : z' \neq f(\mu_1(x)) \bigwedge \mu_2(z') = \mu_2(f(\mu_1(x)))\right] \\ &\geq 1 - \sum_{z\in\mathsf{image}_{\frac{\ell}{2}+\frac{3}{2}\log\ell+1,\ell}} \left(\frac{1}{2^{\frac{\ell}{2}-2\log\ell}}\right) \\ &\geq 1 - \left(\frac{2^{\frac{\ell}{2}}}{16\ell'\ell^{\frac{3}{2}}}\right)\left(\frac{\ell^2}{2^{\frac{\ell}{2}}}\right) \geq 1 - \ell^{-1/2},\end{aligned}$$

as desired.

We next work toward bounding the expected size of

$$\begin{aligned}S''_{\mu_1,\mu_2} &\stackrel{\text{def}}{=} \left\{x' : x' \in S'_{\mu_1,\mu_2} \bigwedge \mu_1(x') \in \mathsf{size}_{0,\frac{\ell}{2}+\frac{3}{2}\log\ell}\right\} \\ &= S'_{\mu_1,\mu_2} \bigcap \mu_1^{-1}\left(\mathsf{size}_{0,\frac{\ell}{2}+\frac{3}{2}\log\ell}\right)\end{aligned}$$

over choice of μ_1, μ_2. Toward this end, let $y = \mu_1(x)$ be arbitrary and define

$$W_{\mu_2} \stackrel{\text{def}}{=} \left\{y' : f(y') \neq f(y) \bigwedge \mu_2(f(y')) = \mu_2(f(y)) \bigwedge y' \in \mathsf{size}_{0,\frac{\ell}{2}+\frac{3}{2}\log\ell}\right\}.$$

We will first bound the expected size of W_{μ_2} (over choice of μ_2) and then use this and the fact that $\mu_1(S''_{\mu_1,\mu_2}) \subseteq W_{\mu_2}$ to bound the size of S''_{μ_1,μ_2} (over choice of μ_1, μ_2).

We can express $|W_{\mu_2}|$ as

$$|W_{\mu_2}| = \sum_{y' \in \mathsf{size}_{0,\frac{\ell}{2}+\frac{3}{2}\log\ell} \setminus \mathsf{siblings}_f(y)} \delta_{y'},$$

where the $\delta_{y'}$ are indicator random variables equal to 1 iff $\mu_2(f(y')) = \mu_2(f(y))$. Since for y' involved in the sum we have $f(y') \neq f(y)$, it holds that:

$$\left(|\mathsf{size}_{0,\frac{\ell}{2}+\frac{3}{2}\log\ell}| - 2^{\frac{\ell}{2}+\frac{3}{2}\log\ell+1}\right) \cdot \frac{\ell^2}{2^{\ell/2}} \leq \mathbf{Exp}_{\mu_2 \leftarrow U_n^2}[|W_{\mu_2}|] \leq |\mathsf{size}_{0,\frac{\ell}{2}+\frac{3}{2}\log\ell}| \cdot \frac{\ell^2}{2^{\ell/2}}$$

(depending on whether or not $y \in \mathsf{size}_{0,\frac{\ell}{2}+\frac{3}{2}\log\ell}$). Using Lemma 7, we see that

$$\left(\frac{\ell}{2} + \frac{3}{2}\log\ell - \ell'\right) \cdot \frac{2^\ell}{4\ell'} \leq |\mathsf{size}_{0,\frac{\ell}{2}+\frac{3}{2}\log\ell}| \leq \left(\frac{\ell}{2} + \frac{3}{2}\log\ell + 1\right) \cdot \frac{2^\ell}{4\ell'},$$

and so $|\mathsf{size}_{0,\frac{\ell}{2}+\frac{3}{2}\log\ell}| = \Theta(2^\ell)$ and $\mu \stackrel{\mathrm{def}}{=} \mathbf{Exp}_{\mu_2 \leftarrow U_n^2}[|W_{\mu_2}|]$ satisfies

$$\mu = \Theta(2^{\ell/2}\ell^2) \quad \text{and} \quad \mu \leq \left(\frac{\ell}{2} + \frac{3}{2}\log\ell + 1\right) \cdot \frac{2^{\ell/2}\ell^2}{4\ell'}.$$

Unfortunately, we *cannot* apply Lemma 3 to bound the deviation of $|W_{\mu_2}|$ from its expectation since the indicator random variables $\{\delta_{y'}\}$ are *not* independent (in particular, if $f(y'_1) = f(y'_2)$ then $\delta_{y'_1} = \delta_{y'_2}$). Instead, we express $|W_{\mu_2}|$ as a *weighted* sum of random variables as follows:

$$|W_{\mu_2}| = \sum_{z' \in \mathsf{image}_{0,\frac{\ell}{2}+\frac{3}{2}\log\ell} \setminus \{f(y)\}} \lambda_{z'} \cdot \delta'_{z'},$$

where the $\delta'_{z'}$ are indicator random variables equal to 1 iff $\mu_2(z') = \mu_2(f(y))$, and $\lambda_{z'}$ is the number of pre-images of z' under f. The $\{\delta'_{z'}\}$ *are* n-wise independent, and so we may apply Corollary 6. Note that $\lambda_{\max} = \max_{z'}\{\lambda_{z'}\}$ satisfies $\lambda_{\max} = \Theta(2^{\ell/2}\ell^{3/2})$ and so $\mu/\lambda_{\max} = \Theta(\ell^{1/2})$, implying that $\mu/\lambda_{\max} > n$ for n large enough (recall that $\ell = \Omega(n^3)$). So

$$\begin{aligned}
\Pr_{\mu_2 \leftarrow U_n^2}\left[|W_{\mu_2}| \geq (1+\delta) \cdot \left(\frac{\ell}{2} + \frac{3}{2}\log\ell + 1\right) \cdot \frac{2^{\ell/2}\ell^2}{4\ell'}\right] &\leq \Pr_{\mu_2 \leftarrow U_n^2}[|W_{\mu_2}| \geq (1+\delta) \cdot \mu] \\
&\leq \left(\frac{n\lambda_{\max}}{e^{2/3}\delta^2\mu}\right)^{\lfloor n/2 \rfloor} \\
&= \Theta\left(\left(\frac{n}{\ell^{1/2}}\right)^{\lfloor n/2 \rfloor}\right),
\end{aligned}$$

where $\delta \in (0,1]$ is an arbitrary constant to be fixed later. Using again the fact that $\ell = \Omega(n^3)$, the above is negligible.

Let SmallW denote the event that the bound

$$|W_{\mu_2}| < (1+\delta) \cdot \left(\frac{\ell}{2} + \frac{3}{2}\log\ell + 1\right) \cdot \frac{2^{\ell/2}\ell^2}{4\ell'}$$

holds. Then:

$$\begin{aligned}
\mathbf{Exp}_{\mu_1,\mu_2}\left[\left|S''_{\mu_1,\mu_2}\right| \mid \mathsf{SmallW}\right] &\leq \mathbf{Exp}_{\mu_1,\mu_2}\left[\left|\{x' : \mu_1(x') \in W_{\mu_2}\}\right| \mid \mathsf{SmallW}\right] \\
&= \sum_{x' \in \{0,1\}^{\ell/2}} \Pr_{\mu_1,\mu_2}\left[\mu_1(x') \in W_{\mu_2} \,\middle|\, \mathsf{SmallW}\right] \\
&\leq 2^{\ell/2} \cdot \left\{ 2^{-\ell} \cdot (1+\delta) \cdot \left(\frac{\ell}{2} + \frac{3}{2}\log \ell + 1\right) \cdot \frac{2^{\ell/2}\ell^2}{4\ell'} \right\} \\
&= (1+\delta) \cdot \frac{\ell^2}{4\ell'}\left(\frac{\ell}{2} + \frac{3}{2}\log \ell + 1\right).
\end{aligned}$$

(Recall that the set W_{μ_2} and the event SmallW depend only on μ_2 and the value of $y = \mu_1(x)$; the fact that μ_1 is chosen from an n-wise independent function family thus justifies the above calculation.) We may view $|S''_{\mu_1,\mu_2}|$ as a sum of the indicator random variables $\delta''_{x'}$ which take on the value 1 iff $\mu_1(x') \in W_{\mu_2}$. Since the $\{\delta''_{x'}\}$ are $(n-1)$-independent we may apply Corollary 4, and so with all but negligible probability (conditioned on occurrence of SmallW):

$$\left|S''_{\mu_1\mu_2}\right| \leq (1+\delta)^2 \frac{\ell^2}{4\ell'}\left(\frac{\ell}{2} + \frac{3}{2}\log \ell + 1\right). \tag{11}$$

Since SmallW occurs with all but negligible probability, it follows that the above bound on $|S''_{\mu_1,\mu_2}|$ holds with all but negligible probability over choice of μ_1, μ_2 (i.e., even without conditioning on occurrence of SmallW).

Putting everything together, we have shown that with probability at least $1 - \ell^{-1/2}$ it holds that $\mu_1(S'_{\mu_1,\mu_2}) \subseteq \mathsf{size}_{0,\frac{\ell}{2}+\frac{3}{2}\log \ell}$. So, with at least this probability we have $|S'_{\mu_1,\mu_2}| = |S''_{\mu_1,\mu_2}|$. We have also shown that with all but negligible probability Eq. (11) holds. Applying a union bound, we see that with probability at least $1 - \ell^{-1/2} - \mathsf{negl}(n)$ we have

$$\left|S'_{\mu_1\mu_2}\right| \leq (1+\delta)^2 \frac{\ell^2}{4\ell'}\left(\frac{\ell}{2} + \frac{3}{2}\log \ell + 1\right),$$

as desired. □

Combining the previous two sub-claims, we see that with probability at least 34/100 (for ℓ large enough):

$$\begin{aligned}
\left|\mathsf{siblings}_{h_{\mu_1,\mu_2}}(x)\right| = 1 + \left|S_{\mu_1} \cup S'_{\mu_1,\mu_2}\right| &\leq 1 + 2\ell + (1+\delta)^2 \frac{\ell^2}{4\ell'}\left(\frac{\ell}{2} + \frac{3}{2}\log \ell + 1\right) \\
&\leq 2\ell + (1+\delta)^2 \frac{\ell^2}{4\ell'}(3\ell') \quad \leq \quad \frac{4}{5}\ell^2
\end{aligned}$$

for ℓ large enough and by choosing (constant) δ small enough. □

Claim 13 *With all but negligible probability over choice of μ_1, μ_2 we have $|\mathsf{hard}_{h_{\mu_1,\mu_2}}(x)| \geq \ell$.*

Proof (of claim) Recall that

$$\mathsf{hard}_{h_{\mu_1,\mu_2}}(x) \overset{\text{def}}{=} \left\{ x' : f(\mu_1(x')) \neq f(\mu_1(x)) \bigwedge \mu_2(f(\mu_1(x'))) = \mu_2(f(\mu_1(x))) \bigwedge \mu_1(x') \in \mathsf{size}_{\frac{\ell}{2}} \right\}.$$

The analysis here is similar to the above, except that we now want to prove a lower bound. Let $y = \mu_1(x)$ be arbitrary, and define

$$W_{\mu_2} \stackrel{\text{def}}{=} \left\{ y' : f(y') \neq f(y) \bigwedge \mu_2(f(y')) = \mu_2(f(y)) \bigwedge y' \in \mathsf{size}_{\frac{\ell}{2}} \right\}.$$

We will bound the expected size of W_{μ_2} (over choice of μ_2) and then bound the expected number of x' (over choice of μ_1) such that $\mu_1(x') \in W_{\mu_2}$. As in the proof of the previous sub-claim, we may write $|W_{\mu_2}|$ as

$$|W_{\mu_2}| = \sum_{z' \in \mathsf{image}_{\frac{\ell}{2}} \setminus \{f(y)\}} \lambda_{z'} \cdot \delta_{z'} \,,$$

where the $\delta_{z'}$ are indicator random variables equal to 1 iff $\mu_2(z') = \mu_2(f(y))$, and $\lambda_{z'}$ is the number of pre-images of z' under f. The value of $\mu \stackrel{\text{def}}{=} \mathbf{Exp}_{\mu_2 \leftarrow U_n^2}[|W_{\mu_2}|]$ is exactly $\frac{\ell^2}{2^{\ell/2}} \cdot \left| \mathsf{size}_{\frac{\ell}{2}} \setminus \mathsf{siblings}(\mu_1(x)) \right|$, and so

$$\frac{\ell^2}{2^{\ell/2}} \cdot \left(\frac{2^\ell}{4\ell'} - 2^{\frac{\ell}{2}+1} \right) \;\leq\; \mu \;\leq\; \frac{\ell^2}{2^{\ell/2}} \cdot \frac{2^\ell}{4\ell'}$$

(using Lemma 7). In particular, $\mu = \Theta(\ell \cdot 2^{\ell/2})$. Furthermore, $\lambda_{\max} \stackrel{\text{def}}{=} \max\{\lambda_z\} = \Theta(2^{\ell/2})$. We now apply Corollary 6. Since $\mu/\lambda_{\max} = \Theta(\ell)$, we have $\mu/\lambda_{\max} > n$ for large enough n (recall that $\ell = \Omega(n^3)$). Let δ be a constant to be fixed later. Then, as in the previous sub-claim, with all but negligible probability we have

$$|W_{\mu_2}| \geq (1-\delta) \cdot \frac{\ell^2}{2^{\ell/2}} \cdot \left(\frac{2^\ell}{4\ell'} - 2^{\frac{\ell}{2}+1} \right).$$

Let LargeW be the event that the above bound holds. Conditioned on the occurrence of LargeW, the expected number of x' that μ_1 maps to W_{μ_2} (which is the expected value of $|\mathsf{hard}_{h_{\mu_1,\mu_2}}|$) is:

$$\begin{aligned}
\sum_{x' \in \{0,1\}^{\ell/2} \setminus \{x\}} 2^{-\ell} \cdot |W_{\mu_2}| \;&\geq\; \left(2^{\ell/2} - 1 \right) \cdot 2^{-\ell} \cdot (1-\delta) \cdot \frac{2^\ell/4\ell' - 2^{\ell/2+1}}{2^{\frac{\ell}{2}-r}} \\
&=\; \left(1 - \frac{1}{2^{\frac{\ell}{2}}} \right) (1-\delta) \left(\frac{\ell^2}{4\ell'} - \frac{2}{2^{\frac{\ell}{2}-2\log\ell}} \right) \;=\; \Theta(\ell).
\end{aligned}$$

(Here again, since the set W_{μ_2} and the event LargeW depend only on μ_2 and the value of $y = \mu_1(x)$; the fact that μ_1 is chosen from an n-wise independent function family justifies the above calculation.) Applying Lemma 3, we see that with all but negligible probability (conditioned on occurrence of LargeW):

$$\begin{aligned}
|\mathsf{hard}_{\mu_1,\mu_2}(x)| \;&\geq\; \left(1 - \frac{1}{2^{\frac{\ell}{2}}} \right) (1-\delta)^2 \left(\frac{\ell^2}{4\ell'} - \frac{2}{2^{\frac{\ell}{2}-2\log\ell}} \right) \\
&\geq\; \ell \left(1 - \frac{1}{2^{\frac{\ell}{2}}} \right) (1-\delta)^2 \left(\frac{5\ell'}{4\ell'} - \frac{2}{\ell 2^{\frac{\ell}{2}-2\log\ell}} \right) \;\;\geq\;\; \ell,
\end{aligned}$$

for ℓ large enough and by taking δ small enough. Since we have already shown that LargeW occurs with all but negligible probability, this completes the proof of the claim. □

Combining Claims 12 and 13 as discussed earlier completes the proof of Theorem 11. ■

3.3 Making Most Siblings Hard

The construction of the previous section has the property that for any x, the fraction of "hard" siblings of x is noticeable with *constant* probability (cf. Theorem 11). Here, we show how to "amplify" the construction so that the fraction of hard siblings is much larger; this is done by simply running many copies of the previous construction in parallel.

Construction 3 Let $\ell = \ell(n)$ and let $H_n = \{h_s : \{0,1\}^{\ell/2} \to \{0,1\}^{\ell/2 - 2\log\ell}\}_{s\in S}$ be as in Construction 2. Set $I = I(n) = 2\ell^5$. Construct $\mathcal{H}' = \{H'_n\}_{n\in\mathbb{N}}$ where $H'_n = \{h'_{\vec{s}} : \{0,1\}^{\ell^6} \to \{0,1\}^{\ell^6 - 4\ell^5\log\ell}\}_{\vec{s}\in S^I}$, and $h'_{\vec{s}}(\vec{x})$ is defined via:

$$h'_{\vec{s}}(x_1\cdots x_I) = h_{s_1}(x_1)\cdot h_{s_2}(x_2)\cdots h_{s_I}(x_I),$$

for $\vec{s} = s_1\cdot s_2\cdots s_I$ (with $s_i \in S$) and $\vec{x} = x_1\cdot x_2\cdots x_I$ (with $x_i \in \{0,1\}^{\ell/2}$). ♣

Now, a hard sibling with respect to $h'_{s_1,\ldots,s_I}$ is simply a sibling whose i^{th} component is a hard sibling of h_{s_i} for *some* i. Formally:

Definition 5 Given $\vec{s} = s_1,\ldots,s_I$ and $x = x_1\cdots x_I \in \{0,1\}^{\ell^6}$, define $\mathsf{hard}_{h'_{\vec{s}}}(x)$ to be the set of $\bar{x} = \bar{x}_1\cdots\bar{x}_I$ such that $\bar{x}_i \in \mathsf{hard}_{h_{s_i}}(x_i)$ for some i. We also define the *easy sibling set* $\mathsf{easy}_{h'_{\vec{s}}}(x) \stackrel{\text{def}}{=} \mathsf{siblings}_{h'_{\vec{s}}}(x) \setminus \mathsf{hard}_{h'_{\vec{s}}}(x)$. ♢

A straightforward hybrid argument in conjunction with Theorem 8 shows:

Lemma 14 *Assuming $\mathcal{F}$ is a one-way function family, the following is negligible for all* PPT *A:*

$$\Pr[x \leftarrow A(1^n); \vec{s} \leftarrow S^I; \bar{x} \leftarrow A(1^n, \vec{s}, x) : \bar{x} \in \mathsf{hard}_{h'_{\vec{s}}}(x)].$$

More interesting is the following, which shows that the fraction of hard siblings is now much larger.

Lemma 15 *Let $x \in \{0,1\}^{\ell^6}$ be arbitrary and define $\mathsf{Ratio}_{\vec{s}}(x) \stackrel{\text{def}}{=} \log\frac{|\mathsf{siblings}_{h'_{\vec{s}}}(x)|}{|\mathsf{easy}_{h'_{\vec{s}}}(x)|}$. Then $\overline{\mathsf{Ratio}} \stackrel{\text{def}}{=} \mathbf{Exp}_{\vec{s}\leftarrow S^I}[\mathsf{Ratio}_{\vec{s}}(x)]$ does not depend on x, and $\overline{\mathsf{Ratio}} \geq \ell^4/2$ for ℓ large enough.*

Proof The proof is straightforward. Let $x = x_1\cdots x_I$, and let ψ_i denote an indicator random variable which is equal to 1 if and only if

$$\frac{\left|\mathsf{easy}_{h_{s_i}}(x_i)\right|}{\left|\mathsf{siblings}_{h_{s_i}}(x_i)\right|} \leq 1 - \frac{1}{\ell}$$

(where $\mathsf{easy}_{h_{s_i}}(x_i)$ is defined in the natural way as $\mathsf{siblings}_{h_{s_i}}(x_i) \setminus \mathsf{hard}_{h_{s_i}}(x_i)$). Note that the $\{\psi_i\}$ are independent. By Theorem 11, each ψ_i takes on the value 1 with probability at least 1/3. Therefore, $\mathbf{Exp}[\sum_{i\in I}\psi_i] \geq 2\ell^5/3$. Using a standard Chernoff bound, we see that with all but negligible probability we have $\sum_{i\in I}\psi_i \geq 7\ell^5/12$. When this is the case, we have

$$\frac{\left|\mathsf{easy}_{h'_{\vec{s}}}(x)\right|}{\left|\mathsf{siblings}_{h'_{\vec{s}}}(x)\right|} = \prod_{1\leq i\leq I}\frac{\left|\mathsf{easy}_{h_{s_i}}(x_i)\right|}{\left|\mathsf{siblings}_{h_{s_i}}(x_i)\right|} \leq \left(1 - \frac{1}{\ell}\right)^{7\ell^5/12} \leq e^{-7\ell^4/12}.$$

Hence with all but negligible probability we have $\mathsf{Ratio}_{\vec{s}}(x) \geq 7\ell^4/12$. The lemma follows. ■

3.4 Making All Siblings Hard

We now give a construction in which *all* siblings are hard with high probability. The construction is parameterized by a value $B = B(n)$; the role of B will become clear from Theorem 16.

Construction 4 Let $\mathcal{U}_3 = \{U_n^3\}_{n\in\mathbb{N}}$ be an n-wise independent function family such that $U_n^3 = \{\mu_{3,s} : \{0,1\}^{\ell^6-B} \to \{0,1\}^{\ell^6}\}$. (From now on, we drop explicit mention of the key s and simply speak of functions $\mu_3 \in U_n^3$.) Let $H'_n = \{h'_{\vec{s}}\}_{\vec{s}\in S^I}$ be as in Construction 3. Construct $\mathcal{H}^B = \{H_n^B\}$ where $H_n^B = \{h^B_{\mu_3,\vec{s}} : \{0,1\}^{\ell^6-B} \to \{0,1\}^{\ell^6-4\ell^5\log\ell}\}_{\mu_3\in U_n^3;\, \vec{s}\in S^I}$ and $h^B_{\mu_3,\vec{s}}$ is defined as follows:

$$h^B_{\mu_3,\vec{s}}(x) = h'_{\vec{s}}(\mu_3(x)).$$

$\clubsuit$

We now prove the following:

Theorem 16 *For $y \in \{0,1\}^{\ell^6}$, define*

$$\mathsf{Sibs}_{\vec{s}}(y) \stackrel{\text{def}}{=} \log\left|\mathsf{siblings}_{h'_{\vec{s}}}(y)\right|, \qquad \overline{\mathsf{Sibs}} \stackrel{\text{def}}{=} \mathbf{Exp}_{\vec{s}\leftarrow S^I}[\mathsf{Sibs}_{\vec{s}}(y)]$$
$$\mathsf{Easy}_{\vec{s}}(y) \stackrel{\text{def}}{=} \log\left|\mathsf{easy}_{h'_{\vec{s}}}(y)\right|, \qquad \overline{\mathsf{Easy}} \stackrel{\text{def}}{=} \mathbf{Exp}_{\vec{s}\leftarrow S^I}[\mathsf{Easy}_{\vec{s}}(y)]$$

(as in Lemma 15, $\overline{\mathsf{Sibs}}$ and $\overline{\mathsf{Easy}}$ do not depend on the specific choice of y). Assume B satisfies

$$\frac{1}{2}\left(\overline{\mathsf{Sibs}} + \overline{\mathsf{Easy}}\right) \;\le\; B \;\le\; \frac{1}{2}\left(\overline{\mathsf{Sibs}} + \overline{\mathsf{Easy}} + \frac{\ell^4}{4}\right),$$

and fix $x \in \{0,1\}^{\ell^6-B}$. Then with all but negligible probability (over choice of $\mu_3, \vec{s}$), there does not exist an $x' \in \{0,1\}^{\ell^6-B}$ such that $\mu_3(x') \in \mathsf{easy}_{h'_{\vec{s}}}(\mu_3(x))$. In particular, then, with all but negligible probability, if x' is a sibling of x (with respect to $h^B_{\mu_3,\vec{s}}$), then $\mu_3(x')$ is a **hard** *sibling of $\mu_3(x)$ (with respect to $h'_{\vec{s}}$).*

We remark that only the lower bound on B is used in proving this theorem; the upper bound on B will be used subsequently but we find it convenient to state it here.

Proof We will show something slightly stronger: namely, that the statement of the theorem holds even when conditioned on the event $\mu_3(x) = y$, for arbitrary $y \in \{0,1\}^{\ell^6}$. (We let this conditioning be implicit in everything that follows.) Letting $\overline{\mathsf{Ratio}}$ be as in Lemma 15, we see that $\overline{\mathsf{Ratio}} = \overline{\mathsf{Sibs}} - \overline{\mathsf{Easy}}$. We can then easily derive:

$$\begin{aligned}
\frac{1}{2}\cdot\left(B - \overline{\mathsf{Easy}}\right) \;&\ge\; \frac{1}{2}\cdot\left(\frac{1}{2}\cdot\left(\overline{\mathsf{Sibs}} + \overline{\mathsf{Easy}}\right) - \overline{\mathsf{Easy}}\right)\\
&=\; \frac{1}{4}\cdot\left(\overline{\mathsf{Sibs}} - \overline{\mathsf{Easy}}\right)\\
&=\; \frac{1}{4}\cdot\overline{\mathsf{Ratio}} \;\ge\; \ell^4/8\,,
\end{aligned}$$

where the last inequality holds for ℓ large enough using Lemma 15.

Let $\mu_3(x) = y = y_1\cdots y_I$ where $I = 2\ell^5$ and $y_i \in \{0,1\}^{\ell/2}$. Let $\mathsf{Easy}^{(i)}_{s_i}(y_i) \stackrel{\text{def}}{=} \log|\mathsf{easy}_{h_{s_i}}(y_i)|$ and notice that $\mathsf{Easy}_{\vec{s}}(y) = \sum_{i\in I}\mathsf{Easy}^{(i)}_{s_i}(y_i)$ and that the $\left\{\mathsf{Easy}^{(i)}_{\vec{s}}(y_i)\right\}_{i\in I}$ are independent random variables. Furthermore, we have $0 \le \mathsf{Easy}^{(i)}_{s_i}(y_i) \le \frac{\ell}{2}$ since each $y_i \in \{0,1\}^{\ell/2}$ can have at most $2^{\ell/2}$ siblings under h_{s_i}. Similarly, $\mathsf{Easy}_{\vec{s}}(y) \le \ell^6$ (and hence the same bound holds for $\overline{\mathsf{Easy}}$). Applying Lemma 2, we thus obtain:

$$\Pr_{\vec{s} \leftarrow S^I}\left[\mathsf{Easy}_{\vec{s}}(y) - \overline{\mathsf{Easy}} \ge \frac{1}{2}\left(B - \overline{\mathsf{Easy}}\right)\right] \le \Pr_{\vec{s} \leftarrow S^I}\left[\left|\mathsf{Easy}_{\vec{s}}(y) - \overline{\mathsf{Easy}}\right| \ge \ell^4/8\right] \le 2 \cdot e^{-\Omega(\ell)},$$

and so with all but negligible probability,

$$\mathsf{Easy}_{\vec{s}}(y) - B \;\le\; \frac{1}{2}\left(\overline{\mathsf{Easy}} - B\right) \;\le\; -\ell^4/8.$$

When this is the case, the probability (over choice of μ_3) that there exists an $x' \in \{0,1\}^{\ell^6 - B}$ for which $\mu_3(x') \in \mathsf{easy}_{h'_{\vec{s}}}(y)$ is at most

$$2^{\ell^6 - B} \cdot \frac{|\mathsf{easy}_{h'_{\vec{s}}}(y)|}{2^{\ell^6}} \;=\; 2^{\mathsf{Easy}_{\vec{s}}(y) - B} \;\le\; 2^{-\ell^4/8}$$

(using pairwise independence of U_n^3), which is negligible. We conclude that with all but negligible probability there does not exist an $x' \in \{0,1\}^{\ell^6 - B}$ such that $\mu_3(x') \in \mathsf{easy}_{h'_{\vec{s}}}(\mu_3(x))$. ■

As an immediate corollary of Lemma 14 and Theorem 16, we have:

Corollary 17 *Assume B satisfies the condition stated in Theorem 16, and that $\mathcal{F}$ is a one-way function family. Then the following is negligible for all* PPT *A:*

$$\Pr[x \leftarrow A(1^n); \mu_3 \leftarrow U_n^3; \vec{s} \leftarrow S^I; \bar{x} \leftarrow A(1^n, \mu_3, \vec{s}, x) : h^B_{\mu_3,\vec{s}}(x) = h^B_{\mu_3,\vec{s}}(\bar{x}) \bigwedge x \neq \bar{x}].$$

Given the above corollary, we are almost done. However, two problems remain to be solved. The first problem is that whenever $B(n) \ge 4\ell(n)^5 \log \ell(n)$, functions in the family H_n^B do not compress their input. The second problem is that we do not, in general, know the value of $B(n)$ as required by Theorem 16. We address each of these problems in turn in the following sections.

3.5 Achieving Compression

First, we show how to ensure compression without affecting the result stated in Corollary 17.

Construction 5 Let $\mathcal{U}_4 = \{U_n^4\}_{n \in \mathbb{N}}$ be a pairwise independent function family such that $U_n^4 = \{\mu_{4,s} : \{0,1\}^{\ell^6 - 4\ell^5 \log \ell} \to \{0,1\}^{\ell^6 - B - \frac{\ell}{300}}\}$. (From now on, we drop explicit mention of the key s and simply speak of functions $\mu_4 \in U_n^4$.) Let $\mathcal{H}^B = \{H_n^B\}$ be as in Construction 4. Construct $\mathcal{G}^B = \{G_n^B\}$ where $G_n^B = \{g^B_{\mu_3,\vec{s},\mu_4} : \{0,1\}^{\ell^6 - B} \to \{0,1\}^{\ell^6 - B - \frac{\ell}{300}}\}$, and $g^B_{\mu_3,\vec{s},\mu_4}$ is defined as follows:

$$g^B_{\mu_3,\vec{s},\mu_4}(x) = \mu_4(h^B_{\mu_3,\vec{s}}(x)).$$

♣

We now show:

Theorem 18 *Assume B satisfies the condition stated in Theorem 16, and fix $x \in \{0,1\}^{\ell^6 - B}$. Then with all but negligible probability (over choice of $\mu_3, \vec{s}, \mu_4$) we have $\mathsf{siblings}_{g^B_{\mu_3,\vec{s},\mu_4}}(x) = \mathsf{siblings}_{h^B_{\mu_3,\vec{s}}}(x)$; i.e., μ_4 induces no additional collisions for x.*

Proof If we can show that $\left|\mathsf{image}(h^B_{\mu_3,\vec{s}})\right| \le 2^{\ell^6-B-\frac{\ell}{200}}$ with all but negligible probability, then the theorem follows using a simple union bound. Let $\mathsf{Ratio}_{\vec{s}}(y)$ and $\overline{\mathsf{Ratio}}$ be as in Lemma 15, and $\mathsf{Sibs}_{\vec{s}}(y)$, $\mathsf{Easy}_{\vec{s}}(y)$, $\overline{\mathsf{Sibs}}$, and $\overline{\mathsf{Easy}}$ be as in Theorem 16. Note that:

$$\begin{aligned}
\overline{\mathsf{Sibs}} - B &\ge \overline{\mathsf{Sibs}} - \frac{1}{2}(\overline{\mathsf{Sibs}} + \overline{\mathsf{Easy}}) - \frac{\ell^4}{8} \quad \text{(using the assumed upper-bound on } B\text{)} \\
&= \frac{1}{2}(\overline{\mathsf{Sibs}} - \overline{\mathsf{Easy}}) - \frac{\ell^4}{8} \\
&= \frac{1}{2}\overline{\mathsf{Ratio}} - \frac{\ell^4}{8} \\
&\ge \frac{\ell^4}{8} \quad \text{(for } \ell \text{ large enough, by Lemma 15),} \qquad (12)
\end{aligned}$$

and so, in particular, $\overline{\mathsf{Sibs}} > B$. We derive the desired bound on $\left|\mathsf{image}(h^B_{\mu_3,\vec{s}})\right|$ by separately bounding the expected sizes of $\mathsf{image}(h^B_{\mu_3,\vec{s}})$ intersected with, respectively, the sets $\bigcup_{i\ge\overline{\mathsf{Sibs}}} \mathsf{image}_i(h'_{\vec{s}})$, $\bigcup_{\overline{\mathsf{Sibs}}>i\ge B} \mathsf{image}_i(h'_{\vec{s}})$, and $\bigcup_{i<B} \mathsf{image}_i(h'_{\vec{s}})$.

Claim 19 *For ℓ large enough,* $\left|\mathsf{image}(h^B_{\mu_3,\vec{s}}) \cap \left(\bigcup_{i\ge\overline{\mathsf{Sibs}}} \mathsf{image}_i(h'_{\vec{s}})\right)\right| \le 2^{\ell^6-B-\frac{\ell^4}{8}}$.

Proof (of claim) This follows easily, since:

$$\begin{aligned}
\left|\mathsf{image}(h^B_{\mu_3,\vec{s}}) \cap \left(\bigcup_{i\ge\overline{\mathsf{Sibs}}} \mathsf{image}_i(h'_{\vec{s}})\right)\right| &\le \left|\bigcup_{i\ge\overline{\mathsf{Sibs}}} \mathsf{image}_i(h'_{\vec{s}})\right| \\
&\le \sum_{i=\overline{\mathsf{Sibs}}}^{\ell^6} 2^{-i}|\mathsf{size}_i(h'_{\vec{s}})| \\
&\le 2^{-\overline{\mathsf{Sibs}}} \sum_{i=0}^{\ell^6} |\mathsf{size}_i(h'_{\vec{s}})| \\
&= 2^{\ell^6-\overline{\mathsf{Sibs}}} \le 2^{\ell^6-B-\frac{\ell^4}{8}}
\end{aligned}$$

(for ℓ large enough), where the final inequality uses Eq. (12). □

Claim 20 *For any μ_3 and with all but negligible probability over choice of $\vec{s}$, we have*[6]

$$\left|\mathsf{image}(h^B_{\mu_3,\vec{s}}) \cap \left(\bigcup_{\overline{\mathsf{Sibs}}>i\ge B} \mathsf{image}_i(h'_{\vec{s}})\right)\right| \le 2^{\ell^6-B-\frac{\ell}{200}-2}$$

for ℓ large enough.

[6] In [9], the bound obtained is $2^{\ell^6-B+\frac{\ell}{40}-1}$ which is not sufficient for the remainder of the proof there.

Proof (of claim) Again, we have:

$$\left|\mathsf{image}(h^B_{\mu_3,\vec{s}}) \cap \left(\bigcup_{\overline{\mathsf{Sibs}}>i\geq B} \mathsf{image}_i(h'_{\vec{s}})\right)\right| \leq \left|\bigcup_{\overline{\mathsf{Sibs}}>i\geq B} \mathsf{image}_i(h'_{\vec{s}})\right| \leq \sum_{i=B}^{\overline{\mathsf{Sibs}}-1} 2^{-i}|\mathsf{size}_i(h'_{\vec{s}})|\,.$$

For arbitrary $y \in \mathsf{domain}(h'_{\vec{s}}) = \{0,1\}^{\ell^6}$, write $y = y_1 \cdots y_I$ where $I = 2\ell^5$ and $y_i \in \{0,1\}^{\ell/2}$. Let $\mathsf{Sibs}^{(i)}_{s_i}(y_i) \overset{\text{def}}{=} \log|\mathsf{siblings}_{h'_{s_i}}(y_i)|$ and notice that $\mathsf{Sibs}_{\vec{s}}(y) = \sum_{i\in I}\mathsf{Sibs}^{(i)}_{s_i}(y_i)$ and that the random variables $\{\mathsf{Sibs}^{(i)}_{s_i}(y_i)\}_{i\in I}$ are independent. Furthermore, we have $0 \leq \mathsf{Sibs}^{(i)}_{s_i}(y_i) \leq \frac{\ell}{2}$ since each $y_i \in \{0,1\}^{\ell/2}$ can have at most $2^{\ell/2}$ siblings. Similarly, $\overline{\mathsf{Sibs}} \leq \ell^6$. Applying Lemma 2 for any $a < \overline{\mathsf{Sibs}}$, we see that:

$$\Pr_{\vec{s}\leftarrow S^I}\left[|\mathsf{Sibs}_{\vec{s}}(y) - \overline{\mathsf{Sibs}}| \geq a\right] \;<\; 2\cdot e^{-\frac{2a^2}{3\ell^7}}$$

which implies

$$\Pr_{\vec{s}\leftarrow S^I}\left[\left|\mathsf{siblings}_{h'_{\vec{s}}}(y)\right| \;\leq\; 2^{\overline{\mathsf{Sibs}}-a}\right] \;<\; 2\cdot e^{-\frac{2a^2}{3\ell^7}}.$$

Since the above holds for any y, a standard calculation shows that for any $\varepsilon < 1$ at least a $(1-\varepsilon)$ fraction of $\vec{s}$ satisfy the following condition: The fraction of $y \in \{0,1\}^{\ell^6}$ such that $|\mathsf{siblings}_{h'_{\vec{s}}}(y)| < 2^{\overline{\mathsf{Sibs}}-a}$ is at most $2\cdot e^{-\frac{2a^2}{3\ell^7}}/\varepsilon$. Fix $\varepsilon = 2\cdot 2^{-\frac{\ell}{400}}$. For any $i < \overline{\mathsf{Sibs}}$, setting $a = \overline{\mathsf{Sibs}} - i - 1$ shows that with probability at least $1-\varepsilon$ (over choice of $\vec{s}$), the fraction of $y \in \{0,1\}^{\ell^6}$ with fewer than 2^{i+1} siblings under $h'_{\vec{s}}$ is at most $e^{-\frac{2}{3}(\overline{\mathsf{Sibs}}-i-1)^2\ell^{-7}}/2^{-\frac{\ell}{400}}$. Taking a union bound over all $i < \overline{\mathsf{Sibs}}$, and using the fact that $\varepsilon\cdot\overline{\mathsf{Sibs}}$ is negligible, we have that with all but negligible probability over choice of $\vec{s}$, the following holds for *all* $i < \overline{\mathsf{Sibs}}$:

The fraction of $y \in \{0,1\}^{\ell^6}$ with fewer than 2^{i+1} siblings is at most $e^{-\frac{2}{3}(\overline{\mathsf{Sibs}}-i-1)^2\ell^{-7}}/2^{-\frac{\ell}{400}}$.

An equivalent way of expressing this is that with all but negligible probability over choice of $\vec{s}$ the following holds for any $i < \overline{\mathsf{Sibs}}$:

$$\sum_{j\leq i}|\mathsf{size}_i(h'_{\vec{s}})| \leq 2^{\ell^6+\frac{\ell}{400}}e^{-\frac{2}{3}(\overline{\mathsf{Sibs}}-i-1)^2\ell^{-7}} \tag{13}$$

and so, in particular, with all but negligible probability over choice of $\vec{s}$:

$$|\mathsf{size}_i(h'_{\vec{s}})| \leq 2^{\ell^6+\frac{\ell}{400}}e^{-\frac{2}{3}(\overline{\mathsf{Sibs}}-i-1)^2\ell^{-7}} \quad \text{for all } i < \overline{\mathsf{Sibs}}. \tag{14}$$

It follows that with all but negligible probability over choice of $\vec{s}$

$$\begin{aligned}
\sum_{i=B}^{\overline{\mathsf{Sibs}}-1} 2^{-i}|\mathsf{size}_i(h'_{\vec{s}})| &\leq \sum_{i=B}^{\overline{\mathsf{Sibs}}-1} 2^{-i}\cdot 2^{\ell^6+\frac{\ell}{400}}\cdot e^{-\frac{2}{3}(\overline{\mathsf{Sibs}}-i-1)^2\ell^{-7}} \\
&\leq 2^{\ell^6+\frac{\ell}{400}} \sum_{j=0}^{\overline{\mathsf{Sibs}}-B-1} 2^{j-\overline{\mathsf{Sibs}}+1}\cdot e^{-\frac{2}{3}j^2\ell^{-7}} \quad (\text{setting } i = \overline{\mathsf{Sibs}} - j - 1) \\
&\leq 2^{\ell^6+\frac{\ell}{400}-\overline{\mathsf{Sibs}}+1} \sum_{j=0}^{\overline{\mathsf{Sibs}}-B-1} 2^{j-\frac{2}{3}j^2\ell^{-7}} \\
&\leq 2^{\ell^6+\frac{\ell}{400}-\overline{\mathsf{Sibs}}+1} \sum_{j=0}^{\overline{\mathsf{Sibs}}-B-1} 2^{\overline{\mathsf{Sibs}}-1-B-\frac{2}{3}(\overline{\mathsf{Sibs}}-1-B)^2\ell^{-7}},
\end{aligned}$$

using for the last inequality the fact that $j - \frac{1}{2}j^2\ell^{-7}$ is increasing for $j < \ell^7$ and $\overline{\mathsf{Sibs}} \leq \ell^6$. Continuing, with all but negligible probability over choice of $\vec{s}$ we have:

$$\begin{aligned}
\sum_{i=B}^{\overline{\mathsf{Sibs}}-1} 2^{-i}|\mathsf{size}_i(h'_{\vec{s}})| &\leq 2^{\ell^6+\frac{\ell}{400}-\overline{\mathsf{Sibs}}+1} \cdot \overline{\mathsf{Sibs}} \cdot 2^{\overline{\mathsf{Sibs}}-1-B-\frac{2}{3}(\overline{\mathsf{Sibs}}-1-B)^2\ell^{-7}} \\
&= \overline{\mathsf{Sibs}} \cdot 2^{\ell^6+\frac{\ell}{400}-B-\frac{2}{3}(\overline{\mathsf{Sibs}}-1-B)^2\ell^{-7}} \\
&\leq 2^{\ell^6-B+\frac{\ell}{400}+7\log\ell-\frac{1}{96}(\ell^4-8)^2\ell^{-7}} \qquad \text{(using } \overline{\mathsf{Sibs}} \leq \ell^7 \text{ and Eq. (12))} \\
&\leq 2^{\ell^6-B-\frac{\ell}{200}-2},
\end{aligned}$$

where the final inequality holds for ℓ large enough. This completes the proof of the claim. □

Claim 21 *With all but negligible probability over choice of $\mu_3, \vec{s}$, we have:*

$$\left|\mathsf{image}(h^B_{\mu_3,\vec{s}}) \cap \left(\bigcup_{i<B} \mathsf{image}_i(h'_{\vec{s}})\right)\right| \leq 2^{\ell^6-B-\frac{\ell}{200}-1}.$$

Proof (of claim) Note that

$$\left|\mathsf{image}(h^B_{\mu_3,\vec{s}}) \cap \left(\bigcup_{i<B} \mathsf{image}_i(h'_{\vec{s}})\right)\right| \quad \leq \quad \left|\left\{x' \in \{0,1\}^{\ell^6-B} : \mu_3(x') \in \bigcup_{i<B} \mathsf{size}_i(h'_{\vec{s}})\right\}\right|.$$

Now, with all but negligible probability over choice of $\vec{s}$ we have:

$$\begin{aligned}
\left|\bigcup_{i<B} \mathsf{size}_i(h'_{\vec{s}})\right| &= \sum_{i=0}^{B-1} |\mathsf{size}_i(h'_{\vec{s}})| \\
&\leq 2^{\ell^6+\frac{\ell}{400}} \cdot e^{-\frac{2}{3}(\overline{\mathsf{Sibs}}-B)^2\ell^{-7}} \qquad \text{(by Eq. (13))} \\
&\leq 2^{\ell^6+\frac{\ell}{400}} \cdot 2^{-\frac{1}{96}(\ell^4)^2\ell^{-7}} \qquad \text{(using Eq. (12) and } e > 2\text{)} \\
&\leq 2^{\ell^6-\frac{\ell}{128}},
\end{aligned}$$

where the final inequality holds for ℓ large enough. Assuming the above bound holds, the expectation (over choice of μ_3) of the number of points $x' \in \{0,1\}^{\ell^6-B}$ for which $\mu_3(x') \in \bigcup_{i<B} \mathsf{size}_i(h'_{\vec{s}})$ is at most $2^{\ell^6-B-\frac{\ell}{128}}$. Applying Corollary 4 shows that with all but negligible probability, the number of x' mapped to $\bigcup_{i<B} \mathsf{size}_i(h'_{\vec{s}})$ is less than $2^{\ell^6-B-\frac{\ell}{200}-1}$. The claim follows. □

Combining the preceding three claims (and applying a union bound), we see that with all but negligible probability over choice of $\mu_3, \vec{s}$ we have

$$\begin{aligned}
|\mathsf{image}(h^B_{\mu_3,\vec{s}})| &\leq 2^{\ell^6-B-\frac{\ell^4}{8}} + 2^{\ell^6-B-\frac{\ell}{200}-2} + 2^{\ell^6-B-\frac{\ell}{200}-1} \\
&\leq 2^{\ell^6-B-\frac{\ell}{200}},
\end{aligned}$$

for ℓ large enough. This completes the proof of the theorem, as discussed earlier. ■

Combining Corollary 17 and Theorem 18, we obtain:

Corollary 22 *Assume B satisfies the condition stated in Theorem 16, and that $\mathcal{F}$ is a one-way function family. Then the following is negligible for all* PPT *A:*

$$\Pr\left[\begin{array}{c} x \leftarrow A(1^n); \mu_3 \leftarrow U_n^3; \vec{s} \leftarrow S^I; \mu_4 \leftarrow U_n^4; \\ \bar{x} \leftarrow A(1^n, \mu_3, \vec{s}, \mu_4, x) \end{array} \; : \; g^B_{\mu_3,\vec{s},\mu_4}(x) = g^B_{\mu_3,\vec{s},\mu_4}(\bar{x}) \bigwedge x \neq \bar{x}\right].$$

Proof By Theorem 18, with all but negligible probability over choice of $\mu_3, \vec{s}$, and μ_4 it holds that any $\bar{x}$ satisfying $g^B_{\mu_3,\vec{s},\mu_4}(\bar{x}) = g^B_{\mu_3,\vec{s},\mu_4}(x)$ also satisfies $h^B_{\mu_3,\vec{s}}(\bar{x}) = h^B_{\mu_3,\vec{s}}(x)$. So if there exists a PPT A which outputs a sibling of x under $g^B_{\mu_3,\vec{s},\mu_4}$ with non-negligible probability, then there exists a PPT A' which outputs a sibling of x under $h^B_{\mu_3,\vec{s}}$ with non-negligible probability, contradicting Corollary 17. ■

3.6 Removing the Non-Uniformity

There remains one final problem to solve. Corollary 22 holds only for values of B satisfying the condition stated in Theorem 16. While this demonstrates the existence of a universal one-way hash family via a non-uniform construction (which chooses a correct value of $B = B(n)$ for each n), it does not immediately yield a *uniform* construction of a universal one-way hash family. This problem, however, is relatively easy to resolve. Recall from Theorem 16 that we only require B to be within an additive factor of $\frac{\ell^4}{16}$ from the quantity $\alpha = \frac{1}{2}\left(\overline{\mathsf{Sibs}} + \overline{\mathsf{Easy}}\right) + \frac{\ell^4}{16}$. Furthermore, we have $0 \leq \alpha < 2\ell^6$. Thus, by running sufficiently-many copies of g^B in parallel (using all relevant values of B) we will obtain a uniform construction of a universal one-way hash family. We give the details now.

First, using $\mathcal{G}^B$ and standard techniques [8, 4] we construct a family $\bar{\mathcal{G}}^B = \{\bar{G}^B_n\}$ where $\bar{G}^B_n = \{\bar{g}^B_\kappa : \{0,1\}^{\ell^6} \to \{0,1\}^{\ell^3}\}$ and such that, for B satisfying the condition stated in Theorem 16 an appropriate analogue of Corollary 22 holds. Then, we proceed as follows:

Construction 6 Let $\bar{\mathcal{G}}^B$ be as discussed above. Let $J = J(n) = 2\ell^6/(\ell^4/8) = 16\ell^2$. Construct $\bar{\mathcal{H}} = \{\bar{H}_n\}$ where $\bar{H}_n = \{\bar{h}_{\kappa_0,\ldots,\kappa_J} : \{0,1\}^{\ell^6} \to \{0,1\}^{(J+1)\ell^3}\}$ and $\bar{h}_{\vec{\kappa}}$ is defined as follows:

$$\bar{h}_{\kappa_0,\ldots,\kappa_J}(x) = \bar{g}^0_{\kappa_0}(x) \cdots \bar{g}^{i\cdot\frac{\ell^4}{8}}_{\kappa_i}(x) \cdots \bar{g}^{J\cdot\frac{\ell^4}{8}}_{\kappa_J}(x).$$

♣

Note that $\bar{\mathcal{H}}$ indeed compresses its input (for large enough n). Furthermore, an adversary who finds a value $\bar{x} \neq x$ for which $\bar{h}_{\vec{\kappa}}(\bar{x}) = \bar{h}_{\vec{\kappa}}(x)$ (for a pre-determined input x and randomly chosen $\vec{\kappa}$) also finds a value $\bar{x} \neq x$ for which $\bar{g}^{B_i}_{\kappa_i}(\bar{x}) = \bar{g}^{B_i}_{\kappa_i}(x)$ for all i (where $B_i = i \cdot \frac{\ell^4}{8}$). Since we are guaranteed that B_i satisfies the conditions of Theorem 16 for *some* $i \in \{0, \ldots, J\}$, a straightforward hybrid argument yields the main result:

Theorem 23 *Assuming $\mathcal{F}$ is a one-way function family, $\bar{\mathcal{H}}$ is a universal one-way hash family.*

References

[1] M. Bellare and S. Micali. How to sign given any trapdoor permutation. *J. ACM*, 39(1):214–233, 1992.

[2] A. De Santis and M. Yung. On the design of provably-secure cryptographic hash functions. In *Advances in Cryptology — Eurocrypt '90*, volume 473 of *Lecture Notes in Computer Science*, pages 412–431. Springer, 1991.

[3] O. Goldreich. *Foundations of Cryptography, vol. 1: Basic Tools.* Cambridge University Press, 2001.

[4] O. Goldreich. *Foundations of Cryptography, vol. 2: Basic Applications.* Cambridge University Press, 2004.

[5] S. Goldwasser, S. Micali, and R.L. Rivest. A digital signature scheme secure against adaptive chosen-message attacks. *SIAM Journal on Computing*, 17(2):281–308, 1988.

[6] S. Goldwasser, S. Micali, and A. Yao. Strong signature schemes. In *Proceedings of the 15th Annual ACM Symposium on Theory of Computing*, pages 431–439. ACM Press, 1983.

[7] R. Motwani and P. Raghavan. *Randomized Algorithms.* Cambridge University Press, 1997.

[8] 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*, pages 33–43. ACM Press, 1989.

[9] J. Rompel. One-way functions are necessary and sufficient for secure signatures. In *Proceedings of the 22nd Annual ACM Symposium on Theory of Computing*, pages 387–394. ACM Press, 1990.

[10] J.P. Schmidt, A. Siegel, and A. Srinivasan. Chernoff-Hoeffding bounds for applications with limited independence. *SIAM J. Discrete Math.*, 8(2):223–250, 1995.