The GGM Function Family is a Weakly One-Way Family of Functions

Aloni Cohen[1] and Saleet Klein[2]

[1] MIT, Cambridge, MA, USA
aloni@mit.edu
[2] Tel Aviv University, Tel Aviv, Israel
saleetklein@mail.tau.ac.il

Abstract. We give the first demonstration of the cryptographic hardness of the Goldreich-Goldwasser-Micali (GGM) function family when the secret key is exposed. We prove that for any constant $\epsilon > 0$, the GGM family is a $1/n^{2+\epsilon}$-weakly one-way family of functions, when the lengths of secret key, inputs, and outputs are equal. Namely, any efficient algorithm fails to invert GGM with probability at least $1/n^{2+\epsilon}$ – *even when given the secret key.*

Additionally, we state natural conditions under which the GGM family is strongly one-way.

1 Introduction

Pseudorandom functions (PRFs) are fundamental objects in general and in cryptography in particular. A pseudorandom function ensemble is a collection of (efficient) functions $\mathcal{F} = \{f_s\}_{s\in\{0,1\}^*}$ indexed by a *secret key* $s \in \{0,1\}^*$ with the dual properties that (1) given the secret key s, f_s is efficiently computable and (2) without knowledge of the secret key, no probabilistic polynomial-time algorithm can distinguish between oracle access to a random function from the ensemble and access to a random oracle. The security property of PRFs depends on the absolute secrecy of the key, and no security is guaranteed when the secret key is revealed. Pseudorandom functions have found wide use: in cryptography to construct private-key encryption and digital signatures [Gol04], in computational learning theory for proving negative results [Val84], and in computational complexity to demonstrate the inherent limits of using natural proofs to prove circuit lower-bounds [RR97].

The first construction of pseudorandom function families starting from any one-way functions came in 1986 by Goldreich, Goldwasser, and Micali [GGM86]. Assuming only that a function is hard to invert, the construction amplifies the secrecy of a short random secret key into an exponentially-long, randomly-accessible sequence of pseudorandom values. For about 10 years, this was the only known method to construct provably secure PRFs, even from specific number-theoretic assumptions. Almost 30 years later, it remains the only generic approach to construct PRFs from any one-way function.

Almost three decades after its conception, we are continuing to discover surprising power specific to the GGM pseudorandom function family. The basic ideas of this construction were used in constructions of broadcast encryption schemes in the early 90s [FN94]. Additionally, these same ideas were to construct function secret sharing schemes for point functions, leading to 2-server computationally-secure PIR schemes with poly-logarithmic communication [BGI15]. More recently, Zhandry exhibited the first quantum-secure PRF by demonstrating that the (classical) GGM ensemble (instantiated with a quantum-secure pseudorandom generator) is secure even against quantum adversaries [Zha12]. In [BW13,BGI14,KPTZ13], the notion of constrained pseudorandom functions was introduced. The "constrained keys" for these PRFs allow a user to evaluate the function on special subsets of the domain while retaining pseudorandomness elsewhere. The GGM ensemble (and modifications thereof) is a constrained PRF for the family of prefix-constraints (including point-puncturing), and GGM yields the simplest known construction of constrained PRFs. This family of constraints is powerful enough to enable many known applications of these families for program obfuscation [SW14].

In this work, we give the first demonstration that the GGM family enjoys some measure of security even when the secret key is revealed to an attacker. In this setting, pseudorandom functions do not necessarily guarantee *any security*. For example, the Luby-Rackoff family of pseudorandom permutations [LR88] are efficiently invertible given knowledge of the secret key. This suggests that we must examine *specific* constructions of pseudorandom functions to see if security is retained when the secret key is revealed. In this work, we ask the following question:

> *What security, if any, does the GGM ensemble provide when the secret key is known?*

A version of this question was posed and addressed by Goldreich[3] in 2002 [Gol02]. Goldreich casts the question from the angle of correlation intractability. Informally, a function ensemble $\{f_s\}_{s\in\{0,1\}^*}$ is correlation intractable if – even given the function description s – it is computationally infeasible to find an input x such that x and $f_s(x)$ satisfy some "sparse" relation. Correlation intractability was formalized in [CGH04], which proved that no such family exists for $|x| \geq |s|$.

In [Gol02], Goldreich proves that the GGM ensemble is not correlation intractable, even for $|x| < |s|$, in a very strong sense. Goldreich constructs a pseudorandom generator $G^{(0)}$ which, when used to instantiate the GGM ensemble, allows an adversary with knowledge of the secret key s to efficiently find preimages $x \in f_s^{-1}(0^n)$. This allows the inversion of f_s for a specific image 0^n, but not necessarily for random images.

[3] And posed much earlier by Micali and by Barak: see Acknowledgments of [Gol02].

1.1 Our contributions

In this work, we prove that the length-preserving[4] GGM ensemble is a weakly one-way family of functions. This means that any efficient algorithm $\mathcal{A}$, when given a random secret key s and $f_s(x)$ for a random input x, must fail to invert with non-negligible probability.

Moreover, we prove that if either a random function in $\mathcal{F}_G$ is "regular" in the sense that each image has a polynomially-bounded number of pre-images, or is "nearly surjective" in a sense made precise below, then the length-preserving GGM ensemble is strongly one-way. Formally:

Theorem 1. *Let $\mathcal{F}_G = \{f_s\}_{s\in\{0,1\}^*}$ be the length-preserving GGM function ensemble with pseudorandom generator G, where $f_s : \{0,1\}^{|s|} \to \{0,1\}^{|s|}$. Then for every constant $\epsilon > 0$, $\mathcal{F}_G$ is a $1/n^{2+\epsilon}$–weakly one-way collection of functions. That is, for every probabilistic polynomial-time algorithm $\mathcal{A}$, for every constant $\epsilon > 0$, and all sufficiently large $n \in \mathbb{N}$,*

$$\Pr_{\substack{s\leftarrow U_n \\ x\leftarrow U_n}}\left[\mathcal{A}(s, f_s(x)) \in f_s^{-1}(f_s(x))\right] < 1 - \frac{1}{n^{2+\epsilon}} \tag{1}$$

where U_n is the uniform distribution over $\{0,1\}^n$.

Theorem 2. *Let $\mathcal{F}_G$ be the GGM ensemble with pseudorandom generator G. $\mathcal{F}_G$ is a strongly one-way collection of functions if either of the following hold:*

(a) There exists a negligible function $\mathsf{negl}(\cdot)$ such that for all sufficiently large $n \in \mathbb{N}$

$$\mathop{\mathbb{E}}_{s\leftarrow U_n}\left[\frac{|\mathsf{Img}(f_s)|}{2^n}\right] \geq 1 - \mathsf{negl}(n) \tag{2}$$

(b) There exists a polynomial B such that for all sufficiently large $n \in \mathbb{N}$ and for all $s, y \in \{0,1\}^n$

$$\left|f_s^{-1}(y)\right| \leq B(n) \tag{3}$$

Remark 1. The conditions of Theorem 2 are very strong conditions. Whether a pseudorandom generator G exists which makes the induced GGM ensemble satisfy either condition is an interesting and open question. The possibility of such a generator is open even for the stronger requirement that for every secret key s, f_s is a permutation.

Remark 2. The length-preserving restriction can be somewhat relaxed to the case when $|x| = |s| \pm O(\log|s|)$, affecting the weakly one-way parameter. A partial result holds when $|x| > |s| + \omega(\log|s|)$, and nothing is currently known if $|x| < |s| - \omega(\log|s|)$. See the full version for further discussion.

[4] We consider the secret keys, inputs, and outputs to be of the same lengths. See Remark 2.

1.2 Overview of proof

Let's go into the land of wishful thinking and imagine that for each secret key $s \in \{0,1\}^n$, every string $y \in \{0,1\}^n$ occurs exactly once in the image of f_s; that is, suppose that the GGM ensemble $\mathcal{F}_G$ is a family of permutations. In this case we can prove that the GGM family is strongly one-way (in fact, this is a special case of Theorem 2).

The assumption that $\mathcal{F}_G$ is a permutation implies the following two facts.[5]

- *Fact 1*: For each secret key $s \in \{0,1\}^n$, the distributions $f_s(U_n)$ and U_n are identical.
- *Fact 2*: For each string $y \in \{0,1\}^n$, there are exactly two pairs $(b,x) \in \{0,1\} \times \{0,1\}^n$ such that $G_b(x) = y$, where G is the PRG underlying the GGM family, and $G_0(x)$ and $G_1(x)$ are the first and second halves of $G(x)$ respectively.

We may now prove that the GGM ensemble is strongly one-way in two steps:

- *Step 1*: Switch the adversary's input to uniformly random.
- *Step 2*: Construct a distinguisher for the PRG.

Step 1. For a PPT algorithm $\mathcal{A}$, let $1/\alpha(n)$ be $\mathcal{A}$'s probability of successfully inverting y with secret key s; namely:

$$\Pr_{\substack{s \leftarrow U_n \\ y \leftarrow f_s(U_n)}} [\mathcal{A}(s,y) \in f_s^{-1}(y)] = \frac{1}{\alpha(n)}$$

By Fact 1, $\mathcal{A}$ has exactly the same success probability if y is sampled uniformly from $\{0,1\}^n$:

$$\Pr_{\substack{s \leftarrow U_n \\ y \leftarrow U_n}} [\mathcal{A}(s,y) \in f_s^{-1}(y)] = \frac{1}{\alpha(n)}$$

Step 2. We now construct a PPT algorithm $\mathcal{D}$ that has advantage $1/2\alpha(n) - \mathsf{negl}(n)$ in distinguishing outputs from the PRG G from random strings (i.e., U_{2n} and $G(U_n)$). By the security of G, this implies that $1/\alpha(n) = \mathsf{negl}(n)$, completing the proof.

[5] While these are indeed facts in the land of wishful thinking, they are not generally true. In this overview we wish to highlight only the usefulness of these facts, and believe that their proofs (though elementary), do not further this goal.

The distinguisher $\mathcal{D}$ is defined as follows:

Input: (y_0, y_1) `// a sample from either` $G(U_n)$ `or` U_{2n}
Sample a secret key $s \leftarrow U_n$ and a bit $b \leftarrow U$;
Compute $x \leftarrow \mathcal{A}(s, y_b)$;
Let $\widetilde{x} = x \oplus 0^{n-1}1$ `//` $\widetilde{x}$ `differs from` x `only at the last bit`;
if $f_s(x) = y_b$ *and* $f_s(\widetilde{x}) = y_{1-b}$ **then**
| Output 1; `// Guess ''PRG''`
else
| Output 0; `// Guess ''random''`
end

Algorithm 1: The PRG distinguisher $\mathcal{D}$

Notice that if $\mathcal{D}$ outputs 1, then either (y_0, y_1) or (y_1, y_0) is in $\mathsf{Img}(G)$. If (y_0, y_1) was sampled uniformly from U_{2n}, then this happens with probability at most $2^{n+1}/2^{2n}$. Therefore,

$$\Pr[\mathcal{D}(U_{2n}) = 1] \leq 1/2^{n-1}.$$

Now we use Fact 2 from above. There are only 2 possible x's that $\mathcal{A}$ could have output in agreement with $f_s(x)$; if (y_0, y_1) was sampled from $G(U_n)$ and $f_s(x) = y_b$ (which happens with probability $1/\alpha(n)$), then with probability at least $1/2$: $f_s(\widetilde{x}) = y_{1-b}$. Therefore,

$$\Pr[\mathcal{D}(G(U_n)) = 1] \geq 1/2\alpha(n),$$

completing the proof of this special case.

Leaving the land of wishful thinking, the proof that the GGM ensemble is weakly one-way follows exactly the same two steps as the special case proved above, but the facts we used are not true in general. We carry out Step 1 in the Input Switching Proposition (Proposition 1): we more carefully analyze the relationship between the distributions $f_s(U_n)$ and U_n, losing a factor of $1 - 1/n^{2+\epsilon}$ in the adversary's probability of successfully inverting. We carry out Step 2 in the Distinguishing Lemma (Lemma 2): we analyze the success probability of the distinguisher (the same one as above) by more carefully reasoning about the number of preimages for a value y.

Organization Section 2 contains standard definitions and the notation used throughout this work. Section 3 contains the proof of Theorem 1, leaving the proof of the crucial Combinatorial Lemma (Lemma 1) to Section 4. Theorem 2 is proved in Section 5, and Section 6 concludes.

2 Preliminaries

2.1 Notation

For two strings a and b we denote by $a\|b$ their concatenation. For a bit string $x \in \{0,1\}^n$, we denote by $x[i]$ its i-th bit, and by $x[i:j]$ (for $i < j$) the sequence $x[i]\|x[i+1]\|\cdots\|x[j]$. We abbreviate 'probabilistic polynomial time' as 'PPT'.

For a probability distribution D, we use $\mathrm{Supp}(D)$ to denote the support of D. We write $x \leftarrow D$ to mean that x is a sample from the distribution D. By U_n, we denote the uniform distribution over $\{0,1\}^n$, and omit the subscript when $n = 1$. For a probabilistic algorithm A, we let $A(x)$ denote a sample from the probability distribution induced over the outputs of A on input x, though we occasionally abuse notation and let $A(x)$ denote the distribution itself. For a function $f : X \to Y$ and a distribution D over X, we denote by $f(D)$ the distribution $(f(x))_{x \leftarrow D}$ over Y.

Definition 1 (Computationally Indistinguishable). *Two ensembles* $\{X_n\}_{n\in\mathbb{N}}$, $\{Y_n\}_{n\in\mathbb{N}}$ *are* computationally indistinguishable *if for every probabilistic polynomial-time algorithm* $\mathcal{A}$*, every polynomial* $p(\cdot)$*, and all sufficiently large* $n \in \mathbb{N}$

$$\left|\Pr\left[\mathcal{A}\left(X_n\right) = 1\right] - \Pr\left[\mathcal{A}\left(Y_n\right) = 1\right]\right| \leq \frac{1}{p(n)}$$

We write $X_n \approx_c Y_n$ *to denote that* $\{X_n\}_{n\in\mathbb{N}}$ *and* $\{Y_n\}_{n\in\mathbb{N}}$ *are computationally indistinguishable.*

Definition 2 (Multiset). *A* multi-set M *over a set* S *is a function* $M : S \to \mathbb{N}$*. For each* $s \in S$*, we call* $M(s)$ *the* multiplicity of s*. We say* $s \in M$ *if* $M(s) \geq 1$*, and denote the* size *of* M *by* $|M| = \sum_S M(s)$*. For two multi-sets* M *and* M' *over* S*, we define their intersection* $M \cap M'$ *to be the multiset* $(M \cap M')(s) = \min[M(s), M'(s)]$ *containing each element with the smaller of the two multiplicities.*

2.2 Standard cryptographic notions, and the GGM ensemble

Definition 3 (One-way collection of functions; adapted from [Gol04]). *A collection of functions* $\{f_s : \{0,1\}^{|s|} \to \{0,1\}^*\}_{s\in\{0,1\}^*}$ *is called* strongly (weakly) one-way *if there exists a probabilistic polynomial-time algorithm* Eval *such that the following two conditions hold:*

- Efficiently computable: *On input* $s \in \{0,1\}^*$*, and* $x \in \{0,1\}^{|s|}$*, algorithm* Eval *always outputs* $f_s(x)$*.*
- Strongly one-way*: For every polynomial* $w(\cdot)$*, for every probabilistic polynomial-time algorithm* $\mathcal{A}$ *and all sufficiently large* n*,*

$$\Pr_{\substack{s\leftarrow U_n \\ x\leftarrow U_n}}\left[\mathcal{A}(s, f_s(x)) \in f_s^{-1}(f_s(x))\right] < \frac{1}{w(n)} \tag{4}$$

- Weakly one-way*: There exists a polynomial* $w(\cdot)$ *such that for every probabilistic polynomial-time algorithm* $\mathcal{A}$ *and all sufficiently large* n*,*

$$\Pr_{\substack{s\leftarrow U_n \\ x\leftarrow U_n}}\left[\mathcal{A}(s, f_s(x)) \in f_s^{-1}(f_s(x))\right] < 1 - \frac{1}{w(n)} \tag{5}$$

  In this case, the collection is said to be $1/w(n)$-weakly one-way*.*

We emphasize that in weakly one-way definition the polynomial $w(n)$ bounds the success probability of *every* efficient adversary. Additionally, weakly one-way collections can be easily amplified to achieve (strongly) one-way collections [Gol04].

We will use the following notation.

Definition 4 (Inverting Advantage). *For an adversary $\mathcal{A}$ and distribution D over $(s, y) \in \{0,1\}^n \times \{0,1\}^n$, we define the* inverting advantage of $\mathcal{A}$ on distribution D *as*

$$\mathsf{Adv}_{\mathcal{A}}(D) = \Pr_{(s,y)\leftarrow D}\left[\mathcal{A}(s, y) \in f_s^{-1}(y)\right] \tag{6}$$

Definition 5 (Pseudo-random generator). *An efficiently computable function $G : \{0,1\}^n \to \{0,1\}^{2n}$ is a (length-doubling)* pseudorandom generator (PRG)*, if $G(U_n)$ is computationally indistinguishable from U_{2n}. Namely for any PPT $\mathcal{D}$*

$$\left|\Pr[\mathcal{D}(G(U_n)) = 1] - \Pr[\mathcal{D}(U_{2n}) = 1]\right| = \mathsf{negl}(n)$$

Definition 6 (GGM function ensemble [GGM86]). *Let G be a deterministic algorithm that expands inputs of length n into string of length $2n$. We denote by $G_0(s)$ the $|s|$-bit-long prefix of $G(s)$, and by $G_1(s)$ the $|s|$-bit-long suffix of $G(s)$ (i.e., $G(s) = G_0(s)\|G_1(s)$. For every $s \in \{0,1\}^n$ (called the* secret key*), we define a function $f_s^G : \{0,1\}^n \to \{0,1\}^n$ such that for every $x \in \{0,1\}^n$,*

$$f_s^G(x[1], \ldots, x[n]) = G_{x[n]}(\cdots(G_{x[2]}(G_{x[1]}(s)))\cdots) \tag{7}$$

For any $n \in \mathbb{N}$, we define F_n to be a random variable over $\{f_s^G\}_{s\in\{0,1\}^n}$. We call $\mathcal{F}_G = \{F_n\}_{n\in\mathbb{N}}$ the GGM function ensemble instantiated with generator G.

We will typically write f_s instead of f_s^G.

The construction is easily generalized to the case when $|x| \neq n$. Though we define the GGM function ensemble as the case when $|x| = n$, it will be useful to consider the more general case.

2.3 Statistical distance

For two probability distributions D and D' over some universe X, we recall two equivalent definitions of their statistical distance $\mathrm{SD}(D, D')$:

$$\mathrm{SD}(D, D') := \frac{1}{2}\sum_{x\in X} |D(x) - D'(x)| = \max_{S\subseteq X} \sum_{x\in S} D(x) - D'(x)$$

For a collection of distributions $\{D(p)\}$ with some parameter p, and a distribution P over the parameter p, we write

$$(p, D(p))_P$$

to denote the distribution over pairs (p, x) induced by sampling $p \leftarrow P$ and subsequently $x \leftarrow D(p)$.[6] It follows from the definition of statistical distance (see appendix) that for distributions P, $D(P)$, and $D'(P)$:

$$\mathrm{SD}\left(\left(p, D(p)\right)_P, \left(p, D'(p)\right)_P\right) = \mathop{\mathbb{E}}_{p \leftarrow P}\left[\mathrm{SD}\big(D(p), D'(p)\big)\right] \tag{8}$$

The quantity $|\mathsf{Img}(f)|$ is related to the statistical distance between the uniform distribution U_n and the distribution $f(U_n)$. For any $f : \{0,1\}^n \to \{0,1\}^n$,

$$\mathrm{SD}(f(U_n), U_n) = 1 - \frac{|\mathsf{Img}(f)|}{2^n} \tag{9}$$

This identity can be easily shown by expanding the definition of statistical distance, or by considering the histograms of the two distributions and a simple counting argument. See the appendix for a proof.

2.4 Rényi divergences

Similar to statistical distance, the Rényi divergence is a useful tool for relating the probability of some event under two distributions. Whereas the statistical distance yields an additive relation between the probabilities in two distributions, the Rényi divergence yields a multiplicative relation. The following is adapted from Section 2.3 of [BLL+15].

For any two discrete probability distributions P and Q such that $\mathrm{Supp}(P) \subseteq \mathrm{Supp}(Q)$, we define *the power of the Rényi divergence* (of order 2) by

$$R\left(P \| Q\right) = \left(\sum_{x \in \mathrm{Supp}(Q)} \frac{P(x)^2}{Q(x)} \right) \quad . \tag{10}$$

An important fact about Rényi divergence is that for an abitrary event $E \subseteq \mathrm{Supp}(Q)$

$$Q(E) \geq \frac{P(E)^2}{R\left(P \| Q\right)} \quad . \tag{11}$$

3 The weak one-wayness of GGM

We now outline the proof of Theorem 1: that the GGM function ensemble is $1/n^{2+\epsilon}$-weakly one-way. The proof proceeds by contradiction, assuming that there exists a PPT $\mathcal{A}$ which inverts on input (s, y) with $> 1 - 1/n^{2+\epsilon}$ probability, where s is a uniform secret key and y is sampled as a uniform image of f_s.

At a high level there are two steps. The first step (captured by the Input Switching Proposition below) is to show that the adversary successfully inverts

[6] For example, the distribution $(x, \mathsf{Bernoulli}(\mathsf{x}))_{\mathsf{Uniform}[0,1]}$ is the distribution over (x, b) by drawing the parameter x uniformly from $[0, 1]$, and subsequently taking a sample b from the Bernoulli distribution with parameter x.

with some non-negligible probability, even when y is sampled uniformly from $\{0,1\}^n$, instead of as a uniform image from f_s. The second step (captured by the Distinguishing Lemma below) will then use the adversary to construct a distinguisher for the PRG underlying the GGM ensemble. The proof of Input Switching Proposition (Proposition 1) depends on the Combinatorial Lemma proved in Section 4. Together, these suffice to prove Theorem 1.

3.1 Step 1: The Input Switching Proposition

As discussed in the overview, our goal is to show that for any adversary that inverts with probability $> 1-1/n^{2+\epsilon}$ on input distribution $(s,y) \leftarrow (s, f_s(U_n))_{s \leftarrow U_n}$ will invert with non-negligible probability on input distribution $(s,y) \leftarrow (U_n, U_n)$. For convenience, we name these distributions:

- D_{owf}: This is $\mathcal{A}$'s input distribution in the weakly one-way function security game in Definition 3. Namely,

$$D_{\mathsf{owf}} = (s, f_s(U_n))_{s \leftarrow U_n}$$

- D_{rand}: This is our target distribution (needed for Step 2), in which s and y are drawn uniformly at random. Namely,

$$D_{\mathsf{rand}} = (U_n, U_n)$$

Proposition 1 (Input Switching Proposition). *For every constant $\epsilon > 0$ and sufficiently large $n \in \mathbb{N}$*

$$\mathsf{Adv}_{\mathcal{A}}(D_{\mathsf{owf}}) > 1 - 1/n^{2+\epsilon} \quad \Longrightarrow \quad \mathsf{Adv}_{\mathcal{A}}(D_{\mathsf{rand}}) > 1/\mathsf{poly}(n) \tag{12}$$

It suffices to show that for every constant $\epsilon > 0$ and sufficiently large $n \in \mathbb{N}$

$$|\mathsf{Adv}_{\mathcal{A}}(D_{\mathsf{owf}}) - \mathsf{Adv}_{\mathcal{A}}(D_{\mathsf{rand}})| < 1 - 1/n^{2+\epsilon} - 1/\mathsf{poly}(n) \tag{13}$$

If $\mathrm{SD}(D_{\mathsf{owf}}, D_{\mathsf{rand}}) < 1 - 1/n^2$, then the above follows immediately (even for an unbounded adversary).[7] If instead $\mathrm{SD}(D_{\mathsf{owf}}, D_{\mathsf{rand}}) \geq 1 - 1/n^2$, we must proceed differently.[8]

What if instead y is sampled as a random image from $f_{s'}$, where s' is a *totally independent* seed? Namely, consider the following distribution over (s,y):

- D_{mix}: This is the distribution in which y is sampled as a uniform image from $f_{s'}$ and s, s' are independent secret keys.

$$D_{\mathsf{mix}} = (s, f_{s'}(U_n))_{s,s' \leftarrow U_n \times U_n}$$

[7] Whether this indeed holds depends on the PRG used to instantiate the GGM ensemble. We do not know if such a PRG exists.

[8] If there exists a PRG, then there exists a PRG such that $\mathrm{SD}(D_{\mathsf{owf}}, D_{\mathsf{rand}}) = 1 - \mathbb{E}_{s \leftarrow U_n}[|\mathsf{Img}(f_s)|/2^n] \geq 1 - 1/n^2$. For example, if the PRG only uses the first $n/2$ bits of its input, then $|\mathsf{Img}(f_s)| < 2^{n/2+1}$.

In order to understand the relationship between $\mathsf{Adv}_{\mathcal{A}}(D_{\mathsf{owf}})$ and $\mathsf{Adv}_{\mathcal{A}}(D_{\mathsf{mix}})$ we define our final distributions, parameterized by an integer $k \in [0, n-1]$. These distributions are related to D_{owf} and D_{mix}, but instead of sampling (s, s') from $U_n \times U_n$, they are sampled from $(G(f_r(U_k)))_{r \leftarrow U_n}$. If $k = 0$, we define $f_r(U_k) = r$.

- D_0^k: Like D_{owf} but the secret key is $s = G_0(\hat{s})$ where $\hat{s}$ is sampled as $\hat{s} \leftarrow (f_r(U_k))_{r \leftarrow U_n}$. Namely,

$$D_0^k = (s, f_s(U_n))_{\substack{r \leftarrow U_n;\ \hat{s} \leftarrow f_r(U_k) \\ s = G_0(\hat{s})}}$$

- D_1^k: Like D_{mix}, but the secret keys are $s = G_0(\hat{s})$ and $s' = G_1(\hat{s})$ where $\hat{s}$ is sampled as $\hat{s} \leftarrow (f_r(U_k))_{r \leftarrow U_n}$. Namely,

$$D_1^k = (s, f_{s'}(U_n))_{\substack{r \leftarrow U_n;\ \hat{s} \leftarrow f_r(U_k) \\ (s,s') = (G_0(\hat{s}), G_1(\hat{s}))}}$$

Claim (Indistinguishability of Distributions). For every $k \in [0, n-1]$,

$$\text{(a) } D_{\mathsf{owf}} \approx_c D_0^k, \quad \text{(b) } D_1^k \approx_c D_{\mathsf{mix}}, \quad \text{(c) } D_{\mathsf{mix}} \approx_c D_{\mathsf{rand}}$$

Proof (Indistinguishability of Distributions). By essentially the same techniques as in [GGM86], the pseudorandomness of the PRG implies that for any $k \leq n$, the distribution $f_{U_n}(U_k)$ is computationally indistinguishable from U_n. Claim (c) follows immediately. By the same observation, $D_0^k \approx_c D_0^0$ and $D_1^k \approx_c D_1^0$. Finally, by the pseudorandomness of the PRG, $D_{\mathsf{owf}} \approx_c D_0^0$ and $D_1^0 \approx D_{\mathsf{mix}}$. This completes the proofs of (a) and (b).

The above claim and the following lemma (proved in Section 4) allow us to complete the proof of the Input Switching Proposition (Proposition 1).

Lemma 1 (Combinatorial Lemma). *Let D_{owf}, D_0^k, D_1^k, D_{mix} and D_{rand} be defined as above. For every constant $\epsilon' > 0$ and every $n \in \mathbb{N}$,*

- *either there exists $k^* \in [0, n-1]$ such that*

$$\mathrm{SD}\left(D_0^{k^*}, D_1^{k^*}\right) \leq 1 - \frac{1}{n^{2+\epsilon'}} \tag{L.1}$$

- *or*

$$\mathrm{SD}\left(D_{\mathsf{owf}}, D_{\mathsf{rand}}\right) < \frac{2}{n^{\epsilon'/2}} \tag{L.2}$$

We now prove (13) and thereby complete the proof of Input Switching Proposition (Proposition 1). Fix a constant $\epsilon > 0$ and $n \in \mathbb{N}$. Apply the Combinatorial Lemma (Lemma 1) with $\epsilon' = \epsilon/2$. In the case that (L.2) is true,

$$|\mathsf{Adv}_{\mathcal{A}}(D_{\mathsf{owf}}) - \mathsf{Adv}_{\mathcal{A}}(D_{\mathsf{rand}})| \leq \mathrm{SD}(D_{\mathsf{owf}}, D_{\mathsf{rand}}) < \frac{2}{n^{\epsilon/4}}$$

In the case that (L.1) is true, we use the Triangle Inequality. Let $k^* \in [0, n-1]$ be as guaranteed by (L.1):

$$\begin{aligned}
&|\mathsf{Adv}_{\mathcal{A}}(D_{\mathsf{owf}}) - \mathsf{Adv}_{\mathcal{A}}(D_{\mathsf{rand}})| \\
&\qquad \leq \left|\mathsf{Adv}_{\mathcal{A}}(D_{\mathsf{owf}}) - \mathsf{Adv}_{\mathcal{A}}(D_0^{k^*})\right| + \left|\mathsf{Adv}_{\mathcal{A}}(D_0^{k^*}) - \mathsf{Adv}_{\mathcal{A}}(D_1^{k^*})\right| \\
&\qquad\quad + \left|\mathsf{Adv}_{\mathcal{A}}(D_1^{k^*}) - \mathsf{Adv}_{\mathcal{A}}(D_{\mathsf{mix}})\right| + \left|\mathsf{Adv}_{\mathcal{A}}(D_{\mathsf{mix}}) - \mathsf{Adv}_{\mathcal{A}}(D_{\mathsf{rand}})\right| \\
&\qquad \leq \mathsf{negl}(n) + \left(1 - \frac{1}{n^{2+\epsilon'/2}}\right) + \mathsf{negl}(n) + \mathsf{negl}(n) \\
&\qquad \leq 1 - \frac{1}{n^{2+\epsilon/4}} + \mathsf{negl}(n)
\end{aligned}$$

3.2 Step 2: The Distinguishing Lemma

As discussed in the overview, in this step we show that any efficient algorithm $\mathcal{A}$ that can invert f_s on uniformly random values $y \in \{0,1\}^n$ with probability $\geq 1/\alpha(n)$ can be used to distinguish the uniform distribution from uniform images of the PRG G underlying the GGM ensemble with probability $\geq 1/\mathsf{poly}(\alpha(n))$. Formally, we prove the following lemma:

Lemma 2 (Distinguishing Lemma). *Let G be a PRG and $\mathcal{F}_G$ the corresponding GGM ensemble. For all PPT algorithms $\mathcal{A}$ and polynomials $\alpha(n)$, there exists a PPT distinguisher $\mathcal{D}$ which for all $n \in \mathbb{N}$:*

$$\begin{aligned}
&\mathsf{Adv}_{\mathcal{A}}(U_n \times U_n) \geq \frac{1}{\alpha(n)} \\
&\qquad \implies \left|\Pr\left[\mathcal{D}\left(G\left(U_n\right)\right) = 1\right] - \Pr\left[\mathcal{D}\left(U_{2n}\right) = 1\right]\right| \geq \left(\frac{1}{4\alpha(n)}\right)^5 - \mathsf{negl}(n)
\end{aligned}$$

Proof. Let $\mathcal{A}$ be a PPT algorithm such that for some polynomial $\alpha(n)$

$$\mathsf{Adv}_{\mathcal{A}}(U_n \times U_n) \geq \frac{1}{\alpha(n)} \tag{14}$$

The distinguisher $\mathcal{D}$ is defined as follows:

Input: (y_0, y_1) `// a sample from either` $G(U_n)$ `or` U_{2n}
Sample a secret key $s \leftarrow U_n$ and a bit $b \leftarrow U$;
Compute $x \leftarrow \mathcal{A}(s, y_b)$;
Let $\widetilde{x} = x \oplus 0^{n-1}1$ `//` $\widetilde{x}$ `differs from` x `only at the last bit`;
if $f_s(x) = y_b$ *and* $f_s(\widetilde{x}) = y_{1-b}$ **then**
 | Output 1; `// Guess ‘‘PRG’’`
else
 | Output 0; `// Guess ‘‘random’’`
end

Algorithm 2: The PRG distinguisher $\mathcal{D}$

Next we show that the distinguisher $\mathcal{D}$ outputs 1 given input sampled uniformly with only negligible probability, but outputs 1 with some non-negligible probability given input sampled from $G(U_n)$. This violates the security of the PRG, contradicting assumption (14).

Observe that if $\mathcal{D}$ outputs 1, then either (y_0, y_1) or (y_1, y_0) is in $\mathsf{Img}(G)$. If (y_0, y_1) was sampled uniformly from U_{2n}, then this happens with probability at most $2^{n+1}/2^{2n}$. Therefore,

$$\Pr[\mathcal{D}(U_{2n}) = 1] = \mathsf{negl}(n) \tag{15}$$

We prove that

$$\Pr[\mathcal{D}(G(U_n)) = 1] \geq \left(\frac{1}{4\alpha(n)}\right)^5 \tag{16}$$

At a very high level, the intuition is that for most $(y_0, y_1) \in \mathsf{Img}(G)$, there are not too many y_1' for which either $(y_0, y_1') \in \mathsf{Img}(G)$ or $(y_1', y_0) \in \mathsf{Img}(G)$ (similarly for y_0' and y_1). After arguing that $\mathcal{A}$ must invert even on such "thin" y's, the chance that $y_{1-b}' = y_{1-b}$ is significant. We now formalize this high level intuition.

We define the function $G_* : \{0,1\} \times \{0,1\}^n \to \{0,1\}^n$

$$G_*(b, y) = G_b(y)$$

Definition 7 (θ-thin, θ-fat). *An element $y \in \mathsf{Img}(G_*)$ is called* θ-thin *under G if $|G_*^{-1}(y)| \leq \theta$. Otherwise, it is called* θ-fat*. Define the sets*

$$\mathsf{Thin}_\theta := \{y \in \mathsf{Img}(G_*) \ : \ y \text{ is } \theta\text{-thin}\}$$
$$\mathsf{Fat}_\theta := \{y \in \mathsf{Img}(G_*) \ : \ y \text{ is } \theta\text{-fat}\}$$

Note that $\mathsf{Thin}_\theta \sqcup \mathsf{Fat}_\theta = \mathsf{Img}(G_*)$

We define an ensemble of distributions $\{Z_n\}$, where each Z_n is the following distribution over $(s, y_0, y_1, b) \in \{0,1\}^n \times \{0,1\}^n \times \{0,1\}^n \times \{0,1\}$:

$$Z_n = (U_n, G_0(r), G_1(r), U)_{r \leftarrow U_n}. \tag{17}$$

Additionally, for every $x \in \{0,1\}^n$, we define $\widetilde{x}$ to be x with its last bit flipped, namely

$$\widetilde{x} = x \oplus 0^{n-1}1.$$

We begin by expanding $\Pr[\mathcal{D}(G(U_n)) = 1]$.

$$\begin{aligned}
&\Pr[\mathcal{D}(G(U_n)) = 1] \\
&\quad = \Pr_{(s,y_0,y_1,b)\leftarrow Z_n}[f_s(x) = y_b \ \wedge \ f_s(\widetilde{x}) = y_{1-b} \mid x \leftarrow \mathcal{A}(s, y_b)] \\
&\quad \geq \Pr_{(s,y_0,y_1,b)\leftarrow Z_n}[y_b \in \mathsf{Thin}_\theta] && (18) \\
&\qquad \cdot \Pr_{(s,y_0,y_1,b)\leftarrow Z_n}\left[f_s(x) = y_b \;\middle|\; \begin{array}{c} x \leftarrow \mathcal{A}(s, y_b) \\ y_b \in \mathsf{Thin}_\theta \end{array}\right] && (19) \\
&\qquad \cdot \Pr_{(s,y_0,y_1,b)\leftarrow Z_n}\left[f_s(\widetilde{x}) = y_{1-b} \;\middle|\; \begin{array}{c} x \leftarrow \mathcal{A}(s, y_b) \\ y_b \in \mathsf{Thin}_\theta \ \wedge \ f_s(x) = y_b \end{array}\right] && (20)
\end{aligned}$$

To show that $Pr[\mathcal{D}(G(U_n)) = 1]$ is non-negligible, it's enough to show that (18), (19), and (20) are each non-negligible.

The first term can be lower-bounded by

$$\Pr_{(s,y_0,y_1,b)\leftarrow Z_n}[y \in \mathsf{Thin}_\theta] \geq \frac{1}{2\alpha(n)} - \frac{1}{\theta} \tag{21}$$

To see why, first recall that by hypothesis $\mathsf{Adv}_{\mathcal{A}}(U_n \times U_n) \geq \frac{1}{\alpha(n)}$. If $y \notin \mathsf{Img}(f_s)$, then of course $\mathcal{A}(s, y)$ cannot output a preimage of y. Therefore $2^n/\alpha(n) \leq |\mathsf{Img}(f_s)| \leq |\mathsf{Img}(G_*)|$. On the other hand, because each θ-fat y must have at least θ preimages, and the domain of G_* is of size 2^{n+1}, there cannot be too many θ-fat y's:

$$|\mathsf{Fat}_\theta| \leq \frac{2^{n+1}}{\theta} \tag{22}$$

Recalling that $\mathsf{Img}(G_*) = \mathsf{Thin}_\theta \sqcup \mathsf{Fat}_\theta$:

$$\begin{aligned}
\Pr_{y\leftarrow G_U(U_n)}[y \in \mathsf{Thin}] &= \frac{|\{(b, x) \ : \ G_b(x) \in \mathsf{Thin}_\theta\}|}{2^{n+1}} \\
&\geq \frac{|\mathsf{Thin}_\theta|}{2^{n+1}} \\
&= \frac{1}{2\alpha(n)} - \frac{1}{\theta}
\end{aligned}$$

The second term can be lower-bounded by:

$$\Pr_{(s,y_0,y_1,b)\leftarrow Z_n}\left[f_s(x) = y_b \;\middle|\; \begin{array}{c} x \leftarrow \mathcal{A}(s, y_b) \\ y_b \in \mathsf{Thin}_\theta \end{array}\right] \geq \left(\frac{1}{4\alpha(n)}\right)^3 \tag{23}$$

We now provide some intuition for the proof of the above, which is included in the appendix in full. In the course of that argument, we will set $\theta = 4\alpha(n)$.

Definition 8 (q-good). *For any $q \in [0, 1]$, an element $y \in \{0, 1\}^n$ is called q-*good with respect to *θ if it is both θ-thin and $\mathcal{A}$ finds some preimage of y for a uniformly random secret key s with probability at least q. Namely,*

$$\mathsf{Good}_q := \left\{y \in \mathsf{Thin}_\theta \ : \ \Pr_{s\leftarrow U_n}[\mathcal{A}(s, y) \in f_s^{-1}(y)] > q\right\}$$

The marginal distribution of y_b where $(s, y_0, y_1, b) \leftarrow Z_n$ is $G_U(U_n)$. To make the notation more explicit, we use the latter notation for the intuition below. In this notation, (23) can be written

$$\Pr_{\substack{s\leftarrow U_n \\ y\leftarrow G_U(U_n)}}\left[\mathcal{A}(s, y) \in f_s^{-1}(y) \;\middle|\; y \in \mathsf{Thin}_\theta\right] \geq \left(\frac{1}{4\alpha(n)}\right)^3$$

The proof of the above inequality boils down to two parts. First, we show that, by the definition of θ-thin:

$$\Pr_{\substack{s\leftarrow U_n \\ y\leftarrow U_n}}[y \in \mathsf{Good}_q \mid y \in \mathsf{Thin}_\theta] \geq \theta \cdot \Pr_{\substack{s\leftarrow U_n \\ y\leftarrow G_U(U_n)}}[y \in \mathsf{Good}_q \mid y \in \mathsf{Thin}_\theta]$$

Second, we must lower-bound the latter quantity. At a high level, this second step follows from the fact that most of the $y \in \{0,1\}^n$ are θ-thin. By assumption, $\mathcal{A}$ inverts with decent probability when $y \leftarrow U_n$, and therefore must invert with some not-too-much-smaller probability when conditioning on the event $y \in \mathsf{Thin}_\theta$.

The third term can be lower-bounded by:

$$\Pr_{(s,y_0,y_1,b)\leftarrow Z_n}\left[f_s(\widetilde{x}) = y_{1-b} \;\middle|\; \begin{array}{c} x \leftarrow \mathcal{A}(s,y_b) \\ y_b \in \mathsf{Thin}_\theta \;\wedge\; f_s(x) = y_b \end{array}\right] \geq \frac{1}{\theta} \tag{24}$$

To see why, suppose that indeed $y_b \in \mathsf{Thin}_\theta$ and $f_s(x) = y_b$. Because y_b is θ-thin, there are at most θ-possible values of $y'_{1-b} := f_s(\widetilde{x})$, where $\widetilde{x} = x \oplus 0^{n-1}1$. The true y_{1-b} is hidden from the adversary's view, and takes each of the possible values with probability at least $1/\theta$. Thus the probability that $y_{1-b} = y'_{1-b}$ is as above.

Finally, letting $\theta = 4\alpha(n)$ as required to lower-bound the second term and putting it all together implies that

$$\Pr\left[\mathcal{D}(G(U_n)) = 1\right] > \left(\frac{1}{2\alpha(n)} - \frac{1}{\theta}\right) \cdot \left(\frac{1}{4\alpha(n)}\right)^3 \cdot \frac{1}{\theta} \tag{25}$$

$$\geq \left(\frac{1}{4\alpha(n)}\right)^5 \tag{26}$$

This completes the proof of Lemma 2.

4 The Combinatorial Lemma

In the proof of the Input Switching Proposition (Proposition 1), we defined the following distributions over $(s,y) \in \{0,1\}^n \times \{0,1\}^n$, for $k \in [0, n-1]$. If $k = 0$, we define $f_r(U_k) = r$.

$$\begin{aligned}
D_{\mathsf{owf}} &= (s, f_s(U_n))_{s\leftarrow U_n} \\
D_0^k &= \left(G_0(\hat{s}), f_{G_0(\hat{s})}(U_n)\right)_{\ r\leftarrow U_n;\ \hat{s}\leftarrow f_r(U_k)} \\
D_1^k &= \left(G_0(\hat{s}), f_{G_1(\hat{s})}(U_n)\right)_{r\leftarrow U_n;\ \hat{s}\leftarrow f_r(U_k)} \\
D_{\mathsf{mix}} &= (s, f_{s'}(U_n))_{s,s'\leftarrow U_n\times U_n} \\
D_{\mathsf{rand}} &= (U_n, U_n)
\end{aligned}$$

We define two additional distributions:

$$\begin{aligned}
\widehat{D}_0^k &= \left(\hat{s}, f_{G_0(\hat{s})}(U_n)\right)_{\ r\leftarrow U_n;\ \hat{s}\leftarrow f_r(U_k)} \\
\widehat{D}_1^k &= \left(\hat{s}, f_{G_1(\hat{s})}(U_n)\right)_{r\leftarrow U_n;\ \hat{s}\leftarrow f_r(U_k)}
\end{aligned}$$

We restate the lemma stated and used in the proof Input Switching Proposition.

Lemma 1 (Combinatorial Lemma). *Let D_{owf}, D_0^k, D_1^k, D_{mix} and D_{rand} be defined as above. For every constant $\epsilon' > 0$ and every $n \in \mathbb{N}$,*

– *either there exists* $k^* \in [0, n-1]$ *such that*

$$\mathrm{SD}\left(D_0^{k^*}, D_0^{k^*}\right) \leq 1 - \frac{1}{n^{2+\epsilon'}} \tag{L.1}$$

– *or*

$$\mathrm{SD}\left(D_{\mathsf{owf}}, D_{\mathsf{rand}}\right) < \frac{2}{n^{\epsilon'/2}} \tag{L.2}$$

We will prove something slightly stronger, namely that either (L.1*) or (L.2) holds, where (L.1*) is:

$$\mathrm{SD}\left(\widehat{D}_0^{k^*}, \widehat{D}_1^{k^*}\right) \leq 1 - \frac{1}{n^{2+\epsilon'}} \tag{L.1*}$$

To see why (L.1*) implies (L.1), observe that for every k, given a sample from $\widehat{D}_0^k$ (resp. $\widehat{D}_1^k$) it is easy to generate a sample from D_0^k (resp. D_1^k). Thus an (unbounded) distinguisher for the former pair of distributions implies an (unbounded) distinguisher with at least the same advantage for the latter pair.[9]

Remark 3. By (8) and (9), $\mathrm{SD}(D_{\mathsf{owf}}, D_{\mathsf{rand}}) = 1 - \mathbb{E}_{s \leftarrow U_n}[\mathsf{Img}(f_s)/2^n]$. Using (L.1*) and this interpretation of (L.2), the lemma informally states that either:

– There is a level k^* such that for a random node $\hat{s}$ on the k^*th level, the subtrees induced by the left child $G_0(\hat{s})$ and the right child $G_1(\hat{s})$ are not too dissimilar.
– The image of f_s is in expectation, a very large subset of the co-domain.

Finally, it is worth noting that the proof of this lemma is purely combinatorial and nowhere makes use of computational assumptions. As such, it holds for and GGM-like ensemble instantiated with arbitrary length-doubling function G.

Proof (Combinatorial Lemma). Fix $n \in \mathbb{N}$ and a secret key $s \in \{0,1\}^n$. Recall that for a multi-set M, $M(x)$ is the multiplicity of the element x in M.

For every $k \in [0, n-1]$ and $v \in \{0,1\}^k$ (letting $\{0,1\}^0 = \{\varepsilon\}$, where ε is the empty string), we define two multi-sets over $\{0,1\}^n$ ('L' for 'leaves') which together contain all the leaves contained in the subtree with prefix v of the GGM tree rooted at s.

$$\begin{aligned} L^s_{v,0} &= \{f_s(x) : x = v\|0\|t\}_{t \in \{0,1\}^{n-k-1}} \\ L^s_{v,1} &= \{f_s(x) : x = v\|1\|t\}_{t \in \{0,1\}^{n-k-1}} \end{aligned} \tag{27}$$

Define $I^s_v := L^s_{v,0} \cap L^s_{v,1}$ to be their intersection.

For each $v \in \{0,1\}^k$, we define a set B^s_v of "bad" inputs x to the function f_s. For each $y \in I^s_v$, there are at least $I^s_v(y)$-many distinct x_0 (respectively, x_1) such that $f_s(x_0) = y$ and $x_0 = v\|0\|t$ begins with the prefix $v\|0$ (respectively, $v\|1$). Assign arbitrarily $I^s_v(y)$-many such x_0 and x_1 to the set B^s_v. By construction,

$$|B^s_v| = 2|I^s_v| \tag{28}$$

[9] This essentially a data-processing inequality.

Let $B^s = \bigcup_{k=0}^{n-1} \bigcup_{v \in \{0,1\}^k} B^s_v$, and let $Q^s := \{0,1\}^n \setminus B^s$ be the set of "good" inputs.

Observe that f_s is injective on Q^s. To see why, consider some $x \in Q^s$, and let $x' \neq x$ be such that $f_s(x) = f_s(x') = y$ if one exists. Suppose that the length of their longest common prefix v is maximal among all such x'. By the maximality of the prefix v, x must be in B^s_v. Therefore,

$$|\mathsf{Img}(f_s)| \geq |Q^s| \tag{29}$$

To reduce clutter we define the following additional notation: for every secret key $r \in \{0,1\}^n$ and level $\ell \in [n]$ we define

$$\Delta_{\mathsf{mix}}(r;\ell) = \mathrm{SD}(f_{G_0(r)}(U_\ell); f_{G_1(r)}(U_\ell))$$

Informally, $\Delta_{\mathsf{mix}}(r;\ell)$ is the difference between the left and right subtrees rooted at r of depth ℓ. For all $\ell < n$ and $r \in \{0,1\}^n$:

$$\Delta_{\mathsf{mix}}(r;\ell) \geq \Delta_{\mathsf{mix}}(r;n) \tag{30}$$

This can be seen by expanding the definitions, or by considering the nature of the distributions as follows. The GGM construction implies that if two internal nodes have the same label, then their subtrees exactly coincide. Thus, the fraction of nodes at level n that coincide on trees rooted at $G_0(r)$ and $G_1(r)$ is at least the fraction of nodes at level ℓ that coincide.

For every secret key $s \in \{0,1\}^n$, $k \in [0, n-1]$, and $v \in \{0,1\}^k$, it holds that:

$$\Delta_{\mathsf{mix}}(f_s(v); n-k-1) = 1 - \frac{|I^s_v|}{2^{n-k-1}} \tag{31}$$

Rearranging (31) and using (30) with $\ell = n-k$, we have that

$$\frac{|I^s_v|}{2^{n-k-1}} \leq 1 - \Delta_{\mathsf{mix}}(f_s(v); n) \tag{32}$$

Claim. For $\epsilon > 0$, $n \in \mathbb{N}$, if $\mathrm{SD}(\widehat{D}_0^{k^*}, \widehat{D}_1^{k^*}) \leq 1 - \frac{1}{n^{2+\epsilon'}}$ (i.e., if (L.1*) is false), then

$$1 - \mathop{\mathbb{E}}_{s \leftarrow U_n}\left[\frac{|Q^s|}{2^n}\right] = \mathop{\mathbb{E}}_{s \leftarrow U_n}\left[\frac{|B^s|}{2^n}\right] < \frac{2}{n^{\epsilon/2}} \tag{33}$$

See proof below. This claim implies (L.2) as follows, completing the proof:

$$\mathrm{SD}\big(D_{\mathsf{owf}}, D_{\mathsf{rand}}\big) = 1 - \mathop{\mathbb{E}}_{s \leftarrow U_n}\left[\frac{|\mathsf{Img}(f_s)|}{2^n}\right] \leq 1 - \mathop{\mathbb{E}}_{s \leftarrow U_n}\left[\frac{|Q^s|}{2^n}\right] < 1 - \frac{2}{n^{\epsilon/2}} \tag{34}$$

Proof (of Claim). We can now bound the expected size of $|B^s|$ as follows.

$$\begin{aligned}
&\mathop{\mathbb{E}}_{s\leftarrow U_n}\left[\frac{|B^s|}{2^n}\right] \qquad (35)\\
&= \Pr_{\substack{s\leftarrow U_n\\ x\leftarrow U_n}}[x\in B^s]\\
&\le \sum_{k=0}^{n-1}\sum_{v\in\{0,1\}^k}\Pr_{s,x}[x\in B^s_v] \qquad \text{by the definition of } B^s\\
&= \sum_{k=0}^{n-1}\Pr_{s,x}\left[x\in B^s_{x[1:k]}\right]\\
&\le \sum_{k=0}^{n-1} T\cdot\Pr_{s,x}\left(\frac{|B^s_{x[1:k]}|}{2^{n-k}}\le T\right)+\Pr_{s,x}\left(\frac{|B^s_{x[1:k]}|}{2^{n-k}}> T\right) \quad \text{for any } 0\le T\le 1\\
&\le \sum_{k=0}^{n-1} T+\Pr_{s,x}\left(\frac{|I^s_{x[1:k]}|}{2^{n-k-1}}> T\right) \qquad \text{by (28)}
\end{aligned}$$

Fix constant $\epsilon>0$. Suppose (L.1*) is false; namely, for all $k\in[0,n-1]$,

$$\mathrm{SD}\left(\widehat{D}_0^{k^*},\widehat{D}_1^{k^*}\right)=\mathop{\mathbb{E}}_{\substack{r\leftarrow U_n\\ \hat{s}\leftarrow f_r(U_k)}}\left[\Delta_{\mathsf{mix}}(\hat{s};n)\right]>1-\frac{1}{n^{2+\epsilon}} \qquad (36)$$

By Markov's Inequality, for any $\tau>0$:

$$\Pr_{\substack{r\leftarrow U_n\\ \hat{s}\leftarrow f_r(U_k)}}\left[1-\Delta_{\mathsf{mix}}(\hat{s};n)>\frac{\tau}{n^{2+\epsilon}}\right]<\frac{1}{\tau} \qquad (37)$$

Observe that the distributions $\left(f_s(x[1:k])\right)_{\substack{s\leftarrow U_n\\ x\leftarrow U_n}}$ and $\left(\hat{s}\right)_{\substack{r\leftarrow U_n\\ \hat{s}\leftarrow f_r(U_k)}}$ are identical. Therefore, by inequality (32) and the above Markov bound:

$$\Pr_{\substack{s\leftarrow U_n\\ x\leftarrow U_n}}\left(\frac{|I^s_{x[1:k]}|}{2^{n-k-1}}>T\right)\le\Pr_{\substack{s\leftarrow U_n\\ x\leftarrow U_n}}\left(1-\Delta_{\mathsf{mix}}(f_s(x[1:k]);n)>T\right)\le\frac{1}{Tn^{2+\epsilon}} \qquad (38)$$

Continuing the series of inequalities from (35):

$$\begin{aligned}
&\le \sum_{k=0}^{n-1}\left(T+\frac{1}{Tn^{2+\epsilon}}\right) \qquad \text{by (32)}\\
&\le n\frac{\tau}{n^{2+\epsilon}}+n\frac{1}{\tau} \qquad \text{for } T=\frac{\tau}{n^{2+\epsilon}}, \text{ by (37)}\\
&= \frac{2}{n^{\epsilon/2}} \qquad \text{for } \tau=n^{1+\epsilon/2}
\end{aligned}$$

This completes the proof of the claim.

5 When is GGM strongly one-way?

Theorem 2 shows that under some natural – albeit strong – conditions, the GGM function ensemble is strongly one-way. Whether pseudorandom generators G exist that induce these conditions in the GGM ensemble is, as yet, unknown.

Theorem 2. *Let $\mathcal{F}_G$ be the GGM ensemble with pseudorandom generator G. $\mathcal{F}_G$ is a strongly one-way collection of functions if either of the following hold:*

(a) There exists a negligible function $\mathsf{negl}(\cdot)$ such that for all sufficiently large n

$$\mathop{\mathbb{E}}_{s \leftarrow U_n}\left[\frac{|\mathsf{Img}(f_s)|}{2^n}\right] \geq 1 - \mathsf{negl}(n) \tag{39}$$

(b) There exists a polynomial $\beta(\cdot)$ such that for all sufficiently large n and for all $s, y \in \{0,1\}^n$

$$\left|f_s^{-1}(y)\right| \leq \beta(n) \tag{40}$$

Remark 4. These two conditions have some overlap, but neither is contained in the other. Additionally, a weaker – but somewhat more abstruse – condition than (b) also suffices: namely, that $\sum_{s,y}\left(\frac{|f_s^{-1}(y)|}{2^n}\right)^2$ is bounded above by some polynomial. This quantity is related to the collision entropy of the distribution $(s, f_s(U_n))_{s \leftarrow U_n}$.

Proof (Theorem 2). Suppose $\mathcal{F}_G$ satisfies one of the conditions of Theorem 2. Further suppose towards contradiction that there exists a probabilistic polynomial-time $\mathcal{A}$ and a polynomial $w(\cdot)$, such that for infinitely-many $n \in \mathbb{N}$

$$\mathsf{Adv}_{\mathcal{A}}\big((s, f_s(U_n))_{s \leftarrow U_n}\big) \geq \frac{1}{w(n)} \tag{41}$$

By the Distinguishing Lemma, to derive a contradiction it suffices to prove for some polynomial $\alpha(\cdot)$ related to w

$$\mathsf{Adv}_{\mathcal{A}}(U_n \times U_n) > \frac{1}{\alpha(n)} \tag{42}$$

Case (a) Applying equations (9) and (8) to the assumption on $\mathbb{E}_{s \leftarrow U_n}\left[\frac{\mathsf{Img}(f_s)}{2^n}\right]$ yields

$$\mathrm{SD}\left((s, f_s(U_n))_{U_n}, (U_n, U_n)\right) \leq \mathsf{negl}(n) \tag{43}$$

It follows immediately that (42) holds for $1/\alpha(n) = 1/w(n) - 1/\mathsf{poly}(n)$, for any polynomial poly (e.g. for $1/\alpha(n) = 1/2w(n)$).

Case (b) For this case, we use the facts about Rényi divergence from the Preliminaries and follow that notation closely. Let $P = D_{\mathsf{owf}} = (s, f_s(U_n))_{s \leftarrow U_n}$ and $Q = D_{\mathsf{rand}} = U_{2n}$ be probability distributions over $\{0,1\}^{2n}$.

Claim. $R(P\|Q) \leq \beta(n)^2$.

Proof (of Claim).

$$\begin{aligned}
R\left(P\|Q\right) &= \sum_{(s,y)\in\{0,1\}^{2n}} \frac{P(s,y)^2}{Q(s,y)} \\
&= 2^{2n} \sum_{s,y} P(s,y)^2 \\
&= 2^{2n} \sum_{s,y} \left(\frac{1}{2^n} \cdot \Pr_P[y|s] \right)^2 \\
&= \sum_{s,y} \Pr_P[y|s]^2 \\
&= \sum_{s,y} \left(\frac{|f_s^{-1}(y)|}{2^n} \right)^2 \\
&\leq \beta(n)^2
\end{aligned}$$

Let the event

$$E = \left\{ (s,y) \in \{0,1\}^n \times \{0,1\}^n : \Pr_{\mathcal{A}}[\mathcal{A}(s,y) \in f_s^{-1}(y)] > \frac{1}{2w(n)} \right\}$$

be the set of pairs (s,y) on which $\mathcal{A}$ successfully inverts with probability at least $1/2w(n)$. By an averaging argument:

$$\begin{aligned}
\frac{1}{w(n)} < \mathsf{Adv}_{\mathcal{A}}(P) &= \Pr_{(s,y)\leftarrow P}[\mathcal{A}(s,y) \in f_s^{-1}(y)] \\
&= \Pr_P[\mathcal{A}(s,y) \in f_s^{-1}(y) \ \wedge \ E] \\
&\quad + \Pr_P[\mathcal{A}(s,y) \in f_s^{-1}(y) \ \wedge \ \neg E] \\
&\leq \Pr_P[E] + \Pr[\mathcal{A}(s,y) \in f_s^{-1}(y) \mid \neg E] \\
&\leq P(E) + \frac{1}{2w(n)}
\end{aligned}$$

Using (11) from the Preliminaries (i.e., $Q(E) \geq \frac{P(E)^2}{R(P\|Q)}$), we get that

$$P(E) > \frac{1}{2w(n)} \quad \Longrightarrow \quad Q(E) > \frac{1}{4w(n)^2 B(n)^2} \tag{44}$$

From the definition of event E, it follows that the condition in (42) holds, completing the proof:

$$\mathsf{Adv}_{\mathcal{A}}(Q) = \Pr_{(s,y)\leftarrow U_{2n}}[\mathcal{A}(s,y) \in f_s^{-1}(y)] > \frac{Q(E)}{2w(n)} > \frac{1}{8w(n)^3 B(n)^2} \tag{45}$$

6 Conclusion

In this work, we demonstrated that the length-preserving Goldreich-Goldwasser-Micali function family is weakly one-way. This is the first demonstration that the family maintains some cryptographic hardness even when the secret key is exposed.

Open questions Two interesting open questions suggest themselves.

1. Is GGM strongly one-way for all pseudorandom generators, or does there exist a generator for which the induced GGM ensemble can be inverted some non-negligible fraction of the time? A positive answer to this question would be very interesting and improve upon this work; a negative answer would be a spiritual successor to [Gol02].
2. In the absence of a positive answer to the above, do there exist pseudorandom generators for which the induced GGM ensemble is strongly one-way? In particular, do there exist generators that satisfy the requirements of Theorem 2?

Acknowledgments We would like to thank Shafi Goldwasser, Ran Canetti, and Alon Rosen for their encouragement throughout this project. We would additionally like to thank Justin Holmgren for discussions about the proof of Lemma 1, and Krzysztof Pietrzak, Nir Bitansky, Vinod Vaikuntanathan, Adam Sealfon, and anonymous reviewers for their helpful feedback.

This work was done in part while the authors were visiting the Simons Institute for the Theory of Computing, supported by the Simons Foundation and by the DIMACS/Simons Collaboration in Cryptography through NSF grant CNS-1523467. Aloni Cohen was supported in part by the NSF GRFP, along with NSF MACS - CNS-1413920, DARPA IBM - W911NF-15-C-0236, and Simons Investigator Award Agreement Dated 6-5-12. Saleet Klein was supported in part by ISF grant 1536/14, along with ISF grant 1523/14, and the Check Point Institute for Information Security. Both authors were supported by the MIT-Israel Seed Fund.

References

BGI14. Elette Boyle, Shafi Goldwasser, and Ioana Ivan. Functional signatures and pseudorandom functions. In *Public-Key Cryptography - PKC 2014 - 17th International Conference on Practice and Theory in Public-Key Cryptography, Buenos Aires, Argentina, March 26-28, 2014. Proceedings*, pages 501–519, 2014.

BGI15. Elette Boyle, Niv Gilboa, and Yuval Ishai. Function secret sharing. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 337–367. Springer, 2015.

BLL+15. Shi Bai, Adeline Langlois, Tancrède Lepoint, Damien Stehlé, and Ron Steinfeld. Improved security proofs in lattice-based cryptography: using the rényi divergence rather than the statistical distance. In *Advances in Cryptology–ASIACRYPT 2015*, pages 3–24. Springer, 2015.

BW13. Dan Boneh and Brent Waters. Constrained pseudorandom functions and their applications. In *Advances in Cryptology - ASIACRYPT 2013 - 19th International Conference on the Theory and Application of Cryptology and Information Security, Bengaluru, India, December 1-5, 2013, Proceedings, Part II*, pages 280–300, 2013.

CGH04. Ran Canetti, Oded Goldreich, and Shai Halevi. The random oracle methodology, revisited. *Journal of the ACM (JACM)*, 51(4):557–594, 2004.

FN94. Amos Fiat and Moni Naor. Broadcast encryption. In *Advances in CryptologyCRYPTO93*, pages 480–491. Springer, 1994.

GGM86. Oded Goldreich, Shafi Goldwasser, and Silvio Micali. How to construct random functions. *Journal of the ACM (JACM)*, 33(4):792–807, 1986.

Gol02. Oded Goldreich. The ggm construction does not yield correlation intractable function ensembles. 2002.

Gol04. Oded Goldreich. *Foundations of cryptography: volume 2, basic applications*. Cambridge university press, 2004.

KPTZ13. Aggelos Kiayias, Stavros Papadopoulos, Nikos Triandopoulos, and Thomas Zacharias. Delegatable pseudorandom functions and applications. In *2013 ACM SIGSAC Conference on Computer and Communications Security, CCS'13, Berlin, Germany, November 4-8, 2013*, pages 669–684, 2013.

LR88. Michael Luby and Charles Rackoff. How to construct pseudorandom permutations from pseudorandom functions. *SIAM Journal on Computing*, 17(2):373–386, 1988.

RR97. Alexander A. Razborov and Steven Rudich. Natural proofs. *J. Comput. Syst. Sci.*, 55(1):24–35, 1997.

SW14. Amit Sahai and Brent Waters. How to use indistinguishability obfuscation: deniable encryption, and more. In *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014*, pages 475–484, 2014.

Val84. Leslie G Valiant. A theory of the learnable. *Communications of the ACM*, 27(11):1134–1142, 1984.

Zha12. Mark Zhandry. How to construct quantum random functions. In *Foundations of Computer Science (FOCS), 2012 IEEE 53rd Annual Symposium on*, pages 679–687. IEEE, 2012.

A Appendix

Proof of (8):

$$\begin{aligned}
&\mathrm{SD}\left((p, D(p))_P, (p, D'(p))_P\right) \\
&= \frac{1}{2} \sum_{(p,x)\in \mathrm{Supp}(P)\times X} \left| \Pr_{(p,D(p))_P}(p,x) - \Pr_{(p,D'(p))_P}(p,x) \right| \\
&= \sum_{p\in\mathrm{Supp}(P)} \Pr_P(p) \cdot \frac{1}{2} \sum_{x\in X} \left| \Pr_{D(p)}(x) - \Pr_{D'(p)}(x) \right| \\
&= \sum_{p\in\mathrm{Supp}(P)} \Pr_P(p) \cdot \mathrm{SD}\left(D(p), D'(p)\right) \\
&= \mathop{\mathbb{E}}_{p\leftarrow P}\left[\mathrm{SD}\left(D(p), D'(p)\right)\right]
\end{aligned}$$

Proof of (9):

$$\begin{aligned}
\mathrm{SD}(f(U_n), U_n) &= \frac{1}{2} \sum_{\alpha\in\{0,1\}^n} \left| \Pr[f(U_n)=\alpha] - \Pr[U_n=\alpha] \right| \\
&= \frac{1}{2} \sum_{\alpha} \left| \frac{|f^{-1}(\alpha)|}{2^n} - \frac{1}{2^n} \right| \\
&= \frac{1}{2} \Big(\sum_{\alpha\in\mathsf{Img}(f)} \left| \frac{|f^{-1}(\alpha)|}{2^n} - \frac{1}{2^n} \right| + \sum_{\alpha\notin\mathsf{Img}(f)} \frac{1}{2^n} \Big) \\
&= \frac{1}{2} \Big(1 - \frac{|\mathsf{Img}(f)|}{2^n} + 1 - \frac{|\mathsf{Img}(f)|}{2^n} \Big) \\
&= 1 - \frac{|\mathsf{Img}(f)|}{2^n}
\end{aligned}$$

Proof of Inequality (23): Recall the following definition.

Definition 8 (q-good) *For any $q \in [0,1]$, an element $y \in \{0,1\}^n$ is called* q-good with respect to θ *if it is both θ-thin and $\mathcal{A}$ finds some preimage of y for a uniformly random secret key s with probability at least q. Namely,*

$$\mathsf{Good}_q := \left\{ y \in \mathsf{Thin}_\theta \ : \ \Pr_{s\leftarrow U_n}[\mathcal{A}(s,y) \in f_s^{-1}(y)] > q \right\}$$

We begin with two observations:

– The distribution over y_b is equivalent to the distribution $(G_b(x))_{(b,x)\leftarrow U\times U_n}$. The number of pairs (b,x) such that $G_b(x) \in \mathsf{Good}_q$ is at least $|\mathsf{Good}_q|$,

while the number of pairs (b, x) such that $G_b(x) \in \mathsf{Thin}_\theta$ is at most $\theta|\mathsf{Thin}_\theta|$. Therefore:

$$\begin{aligned}\Pr_{\substack{s\leftarrow U_n\\ y\leftarrow G_U(U_n)}}[y \in \mathsf{Good}_q \mid y \in \mathsf{Thin}_\theta] \\ &= \Pr_{(b,x)\leftarrow U\times U_n}[G_b(x) \in \mathsf{Good}_q \mid G_b(x) \in \mathsf{Thin}_\theta] \\ &\geq \frac{1}{\theta}\cdot\frac{|\mathsf{Good}_q|}{|\mathsf{Thin}_\theta|} \\ &= \frac{1}{\theta}\cdot\Pr_{s,y\leftarrow U_n}[y \in \mathsf{Good}_q | y \in \mathsf{Thin}_\theta]\end{aligned}$$

– By definition of Good_q:

$$\Pr_{\substack{s\leftarrow U_n\\ y\leftarrow G_U(U_n)}}\left[\mathcal{A}(s, y) \in f_s^{-1}(y) \;\middle|\; y \in \mathsf{Good}_q\right] > q \tag{46}$$

Combining the above

$$\begin{aligned}&\Pr_{\substack{s\leftarrow U_n\\ y\leftarrow G_U(U_n)}}\left[\mathcal{A}(s, y) \in f_s^{-1}(y) \;\middle|\; y \in \mathsf{Thin}_\theta\right] \\ &\geq \Pr_{\substack{s\leftarrow U_n\\ y\leftarrow G_U(U_n)}}[y \in \mathsf{Good}_q \mid y \in \mathsf{Thin}_\theta] \quad \cdot \Pr_{\substack{s\leftarrow U_n\\ y\leftarrow G_U(U_n)}}\left[\mathcal{A}(s, y) \in f_s^{-1}(y) \;\middle|\; y \in \mathsf{Good}_q\right] \\ &\geq \frac{q}{\theta}\cdot\Pr_{s,y\leftarrow U_n}[y \in \mathsf{Good}_q | y \in \mathsf{Thin}_\theta] \end{aligned} \tag{47}$$

If we show that

$$\Pr_{s,y\leftarrow U_n}[y \in \mathsf{Good}_q | y \in \mathsf{Thin}_\theta] \geq \frac{1}{\alpha(n)} - \frac{2}{\theta} - q \tag{48}$$

then selecting $\theta = 4\alpha(n)$ and $q = 1/4\alpha(n)$, the value of (47) is bounded below by

$$\begin{aligned}\Pr_{\substack{s\leftarrow U_n\\ y\leftarrow G_U(U_n)}}\left[\mathcal{A}(s, y) \in f_s^{-1}(y) \;\middle|\; y \in \mathsf{Thin}_\theta\right] &\geq \frac{q}{\theta}\cdot\Pr_{s,y\leftarrow U_n}[y \in \mathsf{Good}_q | y \in \mathsf{Thin}_\theta] \\ &\geq \left(\frac{1}{4\alpha(n)}\right)^3\end{aligned}$$

The following proves inequality (48) and completes the proof of (23).

$$\begin{aligned}
\frac{1}{\alpha(n)} &< \Pr_{\substack{s\leftarrow U_n\\ y\leftarrow U_n}}[\mathcal{A}(s,y)\in f_s^{-1}(y)] \qquad \text{by (14)}\\
&= \Pr_{\substack{s\leftarrow U_n\\ y\leftarrow U_n}}[\mathcal{A}(s,y)\in f_s^{-1}(y) \ \wedge\ y\in \mathsf{Thin}_\theta]\\
&\quad + \Pr_{\substack{s\leftarrow U_n\\ y\leftarrow U_n}}[\mathcal{A}(s,y)\in f_s^{-1}(y) \ \wedge\ y\notin \mathsf{Thin}_\theta]\\
&\le \Pr_{\substack{s\leftarrow U_n\\ y\leftarrow U_n}}[\mathcal{A}(s,y)\in f_s^{-1}(y) \mid y\in \mathsf{Thin}_\theta] + \Pr_{y\leftarrow U_n}[y\notin \mathsf{Thin}_\theta]\\
&\le \Pr_{\substack{s\leftarrow U_n\\ y\leftarrow U_n}}[\mathcal{A}(s,y)\in f_s^{-1}(y) \mid y\in \mathsf{Thin}_\theta] + \frac{2^{n+1}/\theta}{2^n} \qquad \text{by (22)}
\end{aligned}$$

$$\begin{aligned}
\implies \frac{1}{\alpha(n)} - \frac{2}{\theta} &< \Pr_{\substack{s\leftarrow U_n\\ y\leftarrow U_n}}[\mathcal{A}(s,y)\in f_s^{-1}(y) \mid y\in \mathsf{Thin}_\theta]\\
&= \Pr_{\substack{s\leftarrow U_n\\ y\leftarrow U_n}}[y\in \mathsf{Good}_q \mid y\in \mathsf{Thin}_\theta]\\
&\qquad \cdot \Pr_{\substack{s\leftarrow U_n\\ y\leftarrow U_n}}[A(s,y)\in f_s^{-1}(y) \mid y\in \mathsf{Good}_q]\\
&\quad + \Pr_{\substack{s\leftarrow U_n\\ y\leftarrow U_n}}[y\notin \mathsf{Good}_q \mid y\in \mathsf{Thin}_\theta]\\
&\qquad \cdot \Pr_{\substack{s\leftarrow U_n\\ y\leftarrow U_n}}[A(s,y)\in f_s^{-1}(y) \mid y\in \mathsf{Thin}_\theta \setminus \mathsf{Good}_q]\\
&\le \Pr_{\substack{s\leftarrow U_n\\ y\leftarrow U_n}}[y\in \mathsf{Good}_q \mid y\in \mathsf{Thin}_\theta] + q
\end{aligned}$$

The final inequality is by the definition of $\mathsf{Thin}_\theta \setminus \mathsf{Good}_q$.