On the Correlation Intractability of Obfuscated Pseudorandom Functions

Ran Canetti* Yilei Chen† Leonid Reyzin‡

April 21, 2015

Abstract

A family of hash functions is called "correlation intractable" if it is hard to find, given a random function in the family, an input-output pair that satisfies any "sparse" relation, namely any relation that is hard to satisfy for truly random functions. Correlation intractability captures a strong and natural random-oracle-like property. However, it is widely considered to be unobtainable. Indeed, it was shown that correlation intractable functions do not exist for some length parameters [Canetti, Goldreich and Halevi, J.ACM 04]. Furthermore, no candidate constructions have been proposed in the literature for any setting of the parameters.

We construct a correlation intractable function ensemble that withstands all relations with a priori bounded polynomial complexity. We assume the existence of sub-exponentially secure indistinguishability obfuscators, puncturable pseudorandom functions, and input-hiding obfuscators for evasive circuits. The existence of the latter is implied by Virtual-Grey-Box obfuscation for evasive circuits [Bitansky et al, CRYPTO 14].

*Boston University & Tel Aviv University. canetti@bu.edu. Supported by NSF Grants CNS1413920 and AF-1218461, ISF grant 1523/14, and the Check Point Institute for Information Security.

†Boston University. chenyl@bu.edu. Supported by US NSF grants CNS-1012798, CNS-1012910 and AF-1218461. Part of the research conducted while at Tel Aviv University funded by the Check Point Institute for Information Security.

‡Boston University. reyzin@cs.bu.edu. Supported by US NSF grants CNS-1012910, CNS-1012798, and CNS-1422965.

Contents

1 Introduction

To what extent can we construct efficient function families that "behave like random functions" is an intriguing question in cryptography. One of the most elusive properties of random functions, that has so far remained unattainable by efficient constructions, is correlation intractability, proposed by Canetti, Goldreich and Halevi [CGH04]. Roughly speaking, correlation intractable functions guarantee that it is infeasible to find input-output pairs that satisfy some "rare" relation. A bit more precisely, a binary relation R is called *sparse*, if for each value x, only a negligible fraction of y values satisfy $(x, y) \in R$. A function ensemble is *correlation intractable* if, for any sparse relation R, it is infeasible for the adversary to find, given a random function f in family, a value x such that $(x, f(x))$ is in the relation.

The only known results regarding the existence of correlation intractable functions are negative. Specifically, for some settings of the parameters (e.g. when the key is shorter than the input), correlation intractable functions were shown not to exist. This observation was used in [CGH04] to demonstrate the uninstantiability of the random oracle model. However, whether correlation intractable functions exist for other settings of the parameters, and based on what assumptions, remains open.

We note that, beyond the foundational appeal, correlation intractability is desirable in real world applications. For example, consider the hash function used to build the block chain in the Bitcoin protocol [Nak08]. Its main security property, needed to obtain proofs of work, can be stated as correlation intractability with respect to a specific set of relations, which come from protocol-defined constraints on the input and the output (specifically, the input needs to contain appropriate transaction information and the output needs to begin with the correct number of zeros).

More generally, consider any multi-party game which uses the value returned by a random oracle, applied to the previous moves of players, as a substitute for unpredictable public randomness. Correlation intractable functions can be used to instantiate the random oracle in such a game without significant change in the properties of the game. We stress that the security properties needed here are not implied by existing notions such as one-wayness or collision resistance.

Alternative solutions towards instantiating the random oracle. Several alternative notions have been proposed in attempt to capture random-oracle-like properties of hash functions. These notions include perfect one-wayness [Can97, CMR98], entropy preservation [BLV06], seed incompressibility [HMR08], and universal computational extractors [BHK13]. Their relations to correlation intractability will be discussed later in section 1.4. To the best of our knowledge, none of the known results regarding these notions shed light on the question of the existence of correlation intractable functions.

Obfuscated pseudorandom functions. The obfuscation of pseudorandom functions (PRFs) is a natural approach to constructing functions with random-oracle-like properties. Indeed, if the obfuscation was perfect, then the adversary would be unable to take advantage of the code any more than by merely having oracle access to the function. Such a security definition of obfuscation is formalized in the work of Hada [Had00] and Barak et al. [BGI+12], and is termed as *Virtual-black-box* (VBB) Obfuscation. However, they also show that VBB obfuscation is impossible for many function families. In particular, Barak et al. [BGI+12] explicitly construct a pseudorandom function, such that given any code (regardless of how it is obfuscated) of the PRF, the adversary can find an input which evaluates to a fixed value. This certainly breaks correlation intractability. We also know that no pseudorandom function family can be VBB obfuscated with respect to auxiliary input [GK05, BCC+14].

However, these results do not rule out the possibility that there exist pseudorandom functions whose obfuscated version is correlation intractable.

A reasonable next step may thus be to consider PRFs with additional properties, such as constrained or puncturable PRFs [KPTZ13, BW13, BGI14]. Indeed, as demonstrated by multiple works, starting with the ingenious work of Sahai and Waters [SW14], puncturable PRFs are an extremely powerful tool when combined with obfuscation of general programs. In this combination, indistinguishability obfuscation (iO [BGI+12, GR14, GGH+13b], which is a relatively weak notion compared to VBB obfuscation) is sufficient to obtain interested construcitons. In particular, puncturable PRFs have been used together with iO to instantiate some random-oracle-like hash functions, including universal hardcore functions [BST14], universal computational extractors [BM14], and functions used for the full-domain-hash construction [HSW14]. Furthermore, the constructions of [BST14] and [BM14] are simply obfuscating puncturable PRFs. It is thus natural to ask:

Are obfuscated puncturable PRFs correlation intractable? If so, under what assumptions?

1.1 Our results

We make progress towards answering the above questions. Specifically, we show that puncturable pseudorandom functions, obfuscated using an indistinguishability obfuscator, satisfy bounded correlation intractability. "Bounded" means that there is an (a priori) upper bound on the computational complexity size of the relation that the adversary can choose. This result holds under the assumption of sub-exponentially secure general iO and puncturable PRFs, and also requires the existence of *Input-Hiding Obfuscation* (IHO) for evasive circuit families, which we now explain. Recall that a boolean circuit family is evasive if for any input, only negligibly many circuits in the family evaluate to 1. An obfuscator on evasive circuits achieves the "input-hiding" property, if it is infeasible for the adversary to find, given an obfuscated version of a random function in the family, a preimage of 1 for that function. Candidate IHOs for general evasive circuits are proposed by Bitansky et al. [BCKP14] and Badrinarayanan et al. [BMSZ15].

Theorem 1.1 (Bounded correlation intractability, informal)**.** Let n be a security parameter, $l(n)$ and $m(n)$ be the input and output lengths, and $p(\cdot)$ be an arbitrary polynomial. Assuming the existence of input-hiding obfuscation for all evasive circuits, sub-exponentially secure indistinguishability obfuscation, and sub-exponentially secure puncturable pseudorandom functions, there is a construction of a correlation-intractable function ensemble w.r.t. all sparse relations that are recognizable by circuits of size up to $p(n)$. The construction is iO of the puncturable PRF with padding of size dependent on $p(n)$.

Bounded correlation intractability is indeed a qualitatively weaker property than fully correlation intractability. Still, even in its bounded form, correlation intractability is a very strong and potentially useful notion that has not been constructed before. In particular, it suffices for the Bitcoin application mentioned above.

1.2 Our techniques

Failure of applying the "standard" puncturing techniques. Recall that a PRF is puncturable if for any key K and value input value x it is possible to generate a key $K\{x\}$ that is "punctured" at x, such that $F_K(x)$ remains pseudorandom even given $K\{x\}$, and yet $K\{x\}$ allows evaluating F_K at all points other than x. To use puncturable PRFs, one typically punctures the key at the "bad" points that threaten the security of the scheme. In our scenario, the adversary first chooses a relation R, and then the "bad" inputs are those x values that satisfy

$R(x, F_K(x)) = 1$, where K is randomly sampled after R is fixed. However, it is not clear how puncturing at these bad points helps here at all. In fact puncturing at these points may actually help the adversary identify these "useful" inputs, and thus facilitate breaking correlation intractability.

On a higher level, the "standard" puncturing technique succeeds when the "bad" input values are "selected" before the PRF key K is chosen, whereas for correlation intractability, the "bad" points are determined after K.

A "counterintuitive" puncturing strategy. To get around this difficulty, we start from the following observation: for any sparse relation, the "bad" inputs x (i.e., those for which $R(x, F_K(x)) = 1$) can be recognized by a circuit from an evasive circuit family. Now, assume that there is a way to decompose the PRF into two independent branches: one defined on the "bad" inputs, which form an evasive set, the other defined on the "innocent" inputs. Then we could apply the input-hiding obfuscator to the "bad" (evasive) branch, and base correlation intractability on the hardness of finding inputs that lead to a non-zero output of an obfuscated evasive circuit. We obtain such a decomposition by puncturing the key only at the points that belong to the "innocent" branch, where the input-output pairs are not in the relation.

However, instantiating this idea requires some more work. Specifically, we build an alternative pseudorandom function family $\mathcal{F}^R$ that depends on the relation R chosen by the adversary, and is in fact correlation intractable w.r.t. the specific relation. The detail of the key-switching strategy is the technical heart of the proof.

The proof in a nutshell. To better illustrate the main idea, we present an overview of the proof. The analysis goes through 3 hybrids, as will be presented by the games between the adversary and the challenger. Hybrid 0 represents the original game. Hybrid 1, 2, and 3 are intermediate games that are indistinguishable by the adversary. Finally we will show that the adversary cannot break correlation intractability in hybrid 3, therefore concluding that the adversary also fails in hybrid 0, since hybrids 0 and 3 are indistinguishable.

We note that the circuits being iOed shall be padded with the same size, which is possible in our construction if an a priori bound on the size of the relation is given. Under this limitation, our techniques suffice to prove only a bounded version of correlation intractability. For the simplicity of the overview, we postpone the details of padding to the formal proof and now present the hybrids.

0. For any sparse relations R picked by the adversary, the challenger samples a key K of puncturable PRF $\mathcal{F}$ and obfuscates it:
$$h_k^0(\cdot) = \mathsf{iO}(F_K(\cdot))$$
This is the original game.

1. Given the relation R, the challenger samples a key K of puncturable PRF $\mathcal{F}$, and embeds the relation into the description of the function:
$$h_k^1(x) = \mathsf{iO}\left(\begin{array}{lll} \text{if } R(x, F_K(x)) = 1, & \text{return } F_K(x) & ;\text{the "bad" branch} \\ \text{else}, & \text{return } F_K(x) & ;\text{the "innocent" branch} \end{array}\right)$$
Note that h^1 has the same functionality as h^0, and therefore it is indistinguishable from the original function by iO. (Recall that an iO scheme iO guarantees that $\mathsf{iO}(C) \approx \mathsf{iO}(C')$ for any two circuits C, C' that have the same size and functionality.) This is a preparation step, which enables us to partition the function as described above.

2. Puncture the key at the "innocent" branch, and replace it with a freshly generated key K' for a different puncturable PRF $\mathcal{F}^R$ parameterized by R:

$$h_k^2(x) = \mathsf{iO}\left(\begin{array}{lll} \text{if } R(x, F_K(x)) = 1, & \text{return } F_K(x) & ;\text{the "bad" branch} \\ \text{else}, & \text{return } F^R_{K'}(x) & ;\text{the "innocent" branch} \end{array}\right)$$

On a high level, $\mathcal{F}^R$ should satisfy two properties:

(a) Being puncturable, so that we can switch from the original function to the new function point-by-point;

(b) With high probability, there does not exist an x such that $(x, F^R_{K'}(x)) \in R$.

To generate a key K' for $\mathcal{F}^R$, we sample a set of independent puncturable PRF keys K_1, ..., $K_{T(n)}$ from $\mathcal{F}$. The function $F^R_{K'}$ executes in a "rejection sampling" fashion, such that for input x, it goes through the keys K_1, ..., $K_{T(n)}$ one by one, evaluates on the first key K_i for which $(x, F_{K_i}(x))$ is not in the relation. A similar construction has been proposed in [Nis99, CGH04] to achieve "relation-specific" correlation intractable functions.

To prove the indistinguishability of h^1 and h^2, we go over the inputs one by one, which requires exponentially many sub-hybrids. Each hybrid is based on the sub-exponential hardness of iO and the puncturability of $\mathcal{F}$ (upon which $\mathcal{F}^R$ is built).

3. Wrap the first "if-trigger", together with the underlying evasive function, by input-hiding obfuscation. The function h_k^3 is then generated as:

$$h_k^3(x) = \mathsf{iO}\left(\begin{array}{ll} y \leftarrow \mathsf{IHO}\left(\begin{array}{ll} \text{if } R(x, F_K(x)) = 1, & \text{return } F_K(x) \\ \text{else}, & \text{return } \perp \end{array}\right) & ;\text{the "bad" branch} \\ \text{if } y = \perp,\ y = F^R_{K'}(x) & ;\text{the "innocent" branch} \\ \text{return } y & \end{array}\right)$$

h^3 is indistinguishable from h^2 because they are functionally equivalent and obfuscated by iO.

Finally, we note that finding the x values that trigger the non-zero values on the "input-hiding-box" is hard, given R and an "innocent" function $F^R_{K'}$ generated independently (even if not obfuscated). Since the adversary cannot distinguish whether she is given the original function h^0 or the function h^3, and finding an input on h^3 that satisfies the relation is hard, it should also be infeasible for the adversary to break correlation intractability on the original function.

1.3 More on input-hiding obfuscation for evasive functions

Obfuscators that only take care of evasive functions (circuits) are considered easier to construct than general-purpose obfuscators. There are no impossibility results known for the strong black-box definitions (like average-case VBB) for evasive circuits [BBC+14], as opposed to the case for general circuits [BGI+12]. Moreover, candidate obfuscators for certain subclasses of evasive functions, including point functions [Can97, Wee05] and hyperplanes [CRV10], were constructed before the first proposal for a general-purpose obfuscator [GGH+13b]. However, the only known ways to construct obfuscators that can handle **all** evasive circuits do not appear to be simpler than general-purpose obfuscations. We mention two such constructions below.

One way to get evasive circuits obfuscators is to use obfuscators with a stronger security guarantee, called *Strong indistinguishability obfuscation* (siO), which roughly says: if two circuits C_0 and C_1 are drawn from two distributions that are *concentrated* on the same function, then $\mathsf{siO}(C_0)$ is indistinguishable from $\mathsf{siO}(C_1)$. It is shown by Bitansky et al. [BCKP14] that siO for a (general) circuit class $\mathcal{C}$ is equivalent to the worst-case VGB obfuscation for $\mathcal{C}$. They also show that siO/VGB for NC^1 circuits can be obtained from the obfuscators in the idealized graded encoding model [BR14, BGK+14, Zim14, AB15], under the assumptions that the underlying graded encoding schemes satisfy a strong form of semantic security [PST14].

We show in appendix A that siO for evasive circuit class $\mathcal{C}$ implies input-hiding obfuscation for $\mathcal{C}$. As a result, under the same assumption by Bitansky et al. [BCKP14], we obtain IHO for NC^1. In addition, one can simply assume that existing candidate obfuscators for $\mathsf{P/poly}$ are IHO. We note that, even if one only assume the existence of IHO for low-depth evasive circuits classes, our techniques suffice to obtain correlation intractable functions w.r.t. relations recognizable by approximately the same circuit class, as long as puncturable PRF exists in this circuit class. In particular, assuming the existence of puncturable PRFs in NC^1 [BLMR13, HKW14], we obtain correlation intractable functions w.r.t. relations recognizable by NC^1 circuits. This is already interesting and suffices for applications including the Bitcoin protocol.

An alternative construction of an evasive circuits obfuscator was proposed by Badrinarayanan et al. [BMSZ15]. This construction is protected against the devastating zeroing attack [CHL+14] on the candidate graded encodings [GGH13a, CLT13]. Currently, this construction is analyzed only in a relaxed version of idealized graded encoding model.

Proposing simpler constructions of input-hiding obfuscation without going through the full-fledged VGB, or basing IHO on simpler assumptions, is an interesting open problem.

1.4 More on related work

Correlation intractability and constant-round public-coin zero-knowledge proofs. Hada and Tanaka show that the existence of correlation intractable hash functions (w.r.t. relations that are not necessarily efficient) implies 3 round public-coin auxiliary-input zero-knowledge proofs exist only for languages in BPP [HT06]. The key observation is based on a relation $R_{x\notin\mathcal{L}}$ defined as

$$(\alpha,\beta)\in R_{x\notin\mathcal{L}} \Leftrightarrow x\notin\mathcal{L}\wedge\exists\gamma,\Pr[\mathsf{Ver}(x,\alpha,\beta,\gamma)=\mathsf{Accept}]\geq\mathsf{non.negl}.$$

where x is the instance, α,β,γ are the 3 messages in the protocol. The relation is sparse due to the statistical soundness of the underlying proof. Given the fact that the bounded simulator cannot break the correlation intractability, it should be able to decide the membership of the instance.

However, deciding the membership in the relation $R_{x\notin\mathcal{L}}$ requires (at least) an auxiliary string γ in addition to the input α and output β, whereas the construction of correlation intractable function proposed in this paper can only handle relations that takes exactly one input and one output. An alternative way of describing the relation is proposed by Halevi et al. [HMR08] who define the relation with multiple invocations, and set γ as part of the inputs of the additional invocations. Our construction hasn't been proved to work for relations with multiple invocations.

Entropy-preserving hashing. The notion of "entropy-preserving hashing", formalized by Barak, Lindell and Vadhan [BLV06] as being sufficient to achieve Fiat-Shamir heuristics for proofs [FS86], is closely related to

correlation intractability. Roughly speaking, the definition requires that after the adversary is given the key and chooses the input, the output conditioned on the input has high entropy.

We show (in appendix B) that entropy preservation and correlation intractability implies each other. However, the connections are shown w.r.t. relations that are not necessarily decidable by poly-size circuits. Therefore, our construction is not necessarily entropy-preserving. The existence of entropy-preserving hash functions remains open. In fact Bitansky et al. show that entropy preservation is impossible to prove from black-box reduction to falsifiable assumptions [BDSG+13]. As a corollary, correlation intractability w.r.t. possibly inefficient relations is impossible to obtain from black-box reduction to falsifiable assumptions. We don't know if the same impossibility holds for CI w.r.t. efficiently recognizable relations.

Alternative approaches to instantiating random oracles. Several alternative definitions have been proposed in order to capture the random-oracle-like properties, including but not limited to "perfect one-wayness" [Can97, CMR98], "seed-incompressibility" [HMR08], and "universal computation extractor" (UCE) [BHK13]. These definitions are quite different from correlation intractability. In particular, the later two model the security game in two stages, where the adversary in the first stage doesn't get full access to the description of the function, to avoid the impossibility results in [CGH04]. It turns out that one can separate correlation intractability and these notions. An example is given in appendix B that separates UCE and correlation intractability.

Separations, of course, do not show incompatibility: indeed, a construction may naturally satisfy many security definitions simultaneously. For example, essentially the same construction as in this paper (obfuscated puncturable PRFs) was shown to also satisfy a subclass of UCE by Brzuska and Mittelbach [BM14]. Further exploring constructions that satisfy multiple definitions simultaneously (and, in particular, gaining a better understanding of puncturable PRFs) is an interesting future direction.

Additional related work. A canonical construction of a PRF from a pseudorandom generator (PRG), now known as the GGM PRF, was given by Goldreich, Goldwasser and Micali [GGM86]. Suppose we simply publish a GGM PRF seed in the clear to allow public evaluation, without any obfuscation. Is such a function correlation intractable? This questions was posed in the 1990s and answered negatively by Goldreich [Gol02]. He constructed a specialized PRG, such that the GGM PRF built on this PRG is not correlation intractable. In fact one can find a preimage of $0^{m(n)}$ with non-negligible probability.

Correlation intractability is a natural criterion for designing efficient ciphers and hash functions. For example, it is used by Mandal et al. [MPS12] to analyze the 6-round Feistel construction. In particular, they show that the 6-round Feistel construction is sequentially indifferentiable from a random invertible permutation, which implies that it is correlation intractable under an idealized assumption on the Feistel round function.

Organization of the rest of the paper. Conventions of notations, definitions of puncturable pseudorandom functions, obfuscators are presented in section 2. The definition of correlation intractability is presented in section 3. The formal construction and proof is given in section 4. In appendix A we show input-hiding obfuscations for evasive circuits are implied by siO. In appendix B we compare correlation intractability with other random-oracle-like notions.

2 Preliminaries

2.1 Function families

A function ensemble $\mathcal{F}$ has a key generation function $g : S \to K$; on seeds s of length $\sigma(n)$, g produces a key k of length $\kappa(n)$ for a function with input length $l(n)$ and output length $m(n)$:

$$\mathcal{F} = \{f_k : \{0,1\}^{l(n)} \to \{0,1\}^{m(n)}, k = g(s), s \in \{0,1\}^{\sigma(n)}\}_{n\in\mathbb{N}}$$

By default we denote $k \xleftarrow{\$} \mathcal{F}_n$ (sometimes abbreviated as k in the equations) as sampling a key k uniformly random from $\mathcal{F}_n$.

Definition 2.1 (Evasive function family)**.** Let $\mathcal{F} = \{f_k : \{0,1\}^{l(n)} \to \{0,1\}\}_{n\in\mathbb{N}}$ be a function ensemble, we say $\mathcal{F}_n$ is *evasive* if there is a negligible function $\mathsf{negl}(\cdot)$ such that for all $x \in \{0,1\}^{l(n)}$:

$$\Pr_k[f_k(x) \neq 0] \leq \mathsf{negl}(n)$$

2.2 Puncturable pseudorandom functions

Definition 2.2 (Puncturable PRF [KPTZ13, BW13, BGI14, SW14])**.** Let $l(n)$ and $m(n)$ be the input and output lengths. A family of puncturable pseudorandom functions $\mathcal{F} = \{F_K\}$ is given by a triple of efficient functions (Gen, Eval, $\mathsf{Puncture}$), where $\mathsf{Gen}(1^n)$ generates the key K, such that F_K maps from $\{0,1\}^{l(n)}$ to $\{0,1\}^{m(n)}$; $\mathsf{Eval}(K, x)$ takes a key K, an input x, outputs $F_K(x)$; $\mathsf{Puncture}(K, x^*)$ takes a key and an input x^*, outputs a punctured key $K\{x^*\}$.

It satisfies the following conditions:

Functionality preserved over unpunctured points: For all x^* and keys K, if $K\{x^*\} = \mathsf{Puncture}(K, x^*)$, then for all $x \neq x^*$, $\mathsf{Eval}(K, x) = \mathsf{Eval}(K\{x^*\}, x)$.

Pseudorandom on the punctured points: For every p.p.t adversary A who chooses an input x^*, the value of F on x^* is indistinguishable from random in the presence of the key punctured at x^*. That is, the following two distributions are indistinguishable for A: $(x^*, K\{x^*\}, F_K(x^*))$ and $(x^*, K\{x^*\}, r^*)$, where r^* is uniform in $\{0,1\}^{m(n)}$.

Theorem 2.3 ([GGM86, KPTZ13, BW13, BGI14])**.** If one-way function exists, then for all length parameters $l(n)$, $m(n)$, there is a puncturable PRF family that maps from $l(n)$ bits to $m(n)$ bits.

The XOR patch We augment the puncturable PRFs by XORing the output by a uniform random string. The patched function is then 1-universal, which is a useful property in the analysis. In addition, we also rely on the XOR construction explicitly in the proof of lemma 4.9.

Definition 2.4 (1-Universality)**.** A function family $\mathcal{F}$ is *1-universal* if for all $x \in \{0,1\}^{l(n)}$, for all $y \in \{0,1\}^{m(n)}$:

$$\Pr_k[F_k(x) = y] = 2^{-m(n)}$$

Construction 2.5 (XOR-patched puncturable PRF)**.** Given any puncturable pseudorandom function family $\mathcal{F} = \{F_K : \{0,1\}^{l(n)} \to \{0,1\}^{m(n)}\}_{n\in\mathbb{N}}$, we construct $\mathcal{F}^U$ as:

$$F^U_{K,u}(x) = F_K(x) \oplus u$$

where $K \xleftarrow{\$} \mathcal{F}$, $u \xleftarrow{\$} \{0,1\}^{m(n)}$.

To puncture K, u on x^*, output $K\{x^*\}, u$.

For the simplicity of presentation, in the rest of the paper, we use the same name and notation for 1-universal puncturable PRF and standard puncturable PRF. Readers could assume all the puncturable PRFs used in this paper are XOR-patched without loss of generality.

2.3 Obfuscation

In this work we use indistinguishability obfuscation for all circuits, and input-hiding obfuscation for all evasive circuit families. Both obfuscators considered in this paper perfectly preserve the functionality, and cause a polynomial blow-up on the size of the function description. To be precise, for the function family $\mathcal{F} = \{f : \{0,1\}^{l(n)} \to \{0,1\}^{m(n)}\}_{f\in\mathcal{F}_n}$, a probabilistic algorithm Obf is an obfuscator, if

1. The string $\mathsf{Obf}(f)$ describes a function that computes the same function as f;
2. There is a polynomial $B(\cdot)$ such that $|\mathsf{Obf}(f)| \leq B(|f|)$.

The difference lies in the security properties: indistinguishability obfuscation guarantees that the obfuscation of any functionally equivalent circuits cannot be distinguished; whereas input-hiding obfuscation only applies on evasive circuits, and promises to hide all non-zero inputs.

Definition 2.6 (Indistinguishability Obfuscation [BGI+12])**.** Obf is an indistinguishability Obfuscator (iO) for $\mathcal{F}$ if for any feasible adversary A, there is a negligible function $\mathsf{negl}(\cdot)$ such that for all functions f_0 and f_1 that have identical functionalities, and are of the same size, it holds that

$$|\Pr[A(\mathsf{iO}(f_0)) = 1] - \Pr[A(\mathsf{iO}(f_1)) = 1]| \leq \mathsf{negl}(n)$$

Definition 2.7 (Input-hiding Obfuscation for evasive functions [BBC+14])**.** An obfuscator for a evasive function collection $\mathcal{F}$ is *input-hiding* (IHO) if for every p.p.t. adversary A there exist a negligible function $\mathsf{negl}(\cdot)$ s.t. for every auxiliary input $z \in \{0,1\}^{\mathsf{poly}(n)}$:

$$\Pr_{k\xleftarrow{\$}\mathcal{F}_n}[f_k(A(\mathsf{IHO}(f_k), z)) = 1] \leq \mathsf{negl}(n)$$

We will show in appendix A that IHO is implied by Virtual-Grey-Box obfuscation, or equivalently, strong indistinguishability obfuscation, as a corollary of the results by Barak et al. [BBC+14] and Bitansky et al. [BCKP14].

3 Correlation Intractability

We recall the definitions of correlation intractability, initially proposed in [CGH98, CGH04].

Definition 3.1 (Sparse relations)**.** A binary relation R is sparse[1] with respect to length parameters $l(n)$, $m(n)$, if there is a negligible function $\delta(\cdot)$ such that for every $x \in \{0,1\}^{l(n)}$:

$$\Pr_{y \in \{0,1\}^{m(n)}}[R(x,y)=1] \leq \delta(n)$$

In some cases, we quantitatively describes the relations as $\delta(n)$-sparse, and even more precisely, $\delta_x(n)$-sparse when specifying the density on the input x.

Definition 3.2 (Correlation Intractability)**.** A family of functions $\mathcal{H} = \{h_k : \{0,1\}^{l(n)} \to \{0,1\}^{m(n)}\}_{n \in \mathbb{N}}$ is correlation intractable (CI) if for all (non-uniform, p.p.t.) adversary A, for all sparse relations R, there's a negligible function $\mathsf{negl}(\cdot)$ such that:

$$\Pr_{k \xleftarrow{\$} \mathcal{H}_n}[x \leftarrow A(k) : R(x, h_k(x)) = 1] < \mathsf{negl}(n)$$

In the definition above, the sparse relations may not be efficiently recognizable. A reasonable weakening on definition 3.2 is to restrict the relations to be recognizable by poly-size circuits:

Definition 3.3 (CI-P/poly[2])**.** The definition is same as definition 3.2 except that we restrict the relations to be recognizable by poly-size circuits $C : \{0,1\}^{l(n)+m(n)} \to \{0,1\}$ s.t. $C(x,y) = 1$ iff $R(x,y) = 1$.

This definition can be further weakened by giving an a priori bound $p(n)$ on the size of the circuit that defines the relation, instead of allowing circuits of arbitrary polynomial size.

Definition 3.4 (Bounded CI)**.** The definition is same with definition 3.3 except that we further restrict the relations to be recognizable by circuits of size bounded by $p(n)$.

On the length parameters It is shown in [CGH04] that a function family cannot be correlation intractable when the key length $\kappa(n)$ of the function is short compared to the input length $l(n)$:

Claim 3.5 ([CGH04])**.** $\mathcal{H}_n$ is not correlation intractable w.r.t. poly-size relations when $\kappa(n) \leq l(n)$.

Proof sketch: Consider the diagonalization relation $R = \{(k, h_k(k)) | k \in K\}$. The attacker output k. □

For the extensions of the impossibility result, we refer the readers to [CGH04] for the details.

As opposed to the relation between input and key lengths, the relation between input and outputs lengths is not restricted. Although CI is meant to model cryptographic hash functions (which have short outputs), the definition of CI is also meaningful (and non-trivial) for the functions whose output is longer than input. In fact, our construction works for both cases. The only requirement is that the output length $m(n)$ shall be super-logarithmic, i.e. $m(n) \geq \omega(\log(n))$.

[1] This is called $(l(n), m(n))$-restricted sparse relation in [CGH04], as opposed to the "unrestricted" version where the input length is not prescribed. In this paper we remove the "restriction" in the term, since the case where the input length is unbounded is shown to be impossible (cf. claim 3.5), and the "restricted" definition is indeed a natural and interesting setting. Also, in [CGH04] and subsequently in [HT06, HMR08, MPS12], they also define "evasive" relations, which is equivalent to sparse for relations with 1-invocation, and with non-uniform adversaries. Throughout this paper, we only define and use "sparse" relations, since we focus on 1-invocation relations. The term "evasive" only serves the definition of "evasive circuit collections" [BBC+14] (cf. def. 2.1) to avoid confusion.

[2] This notion is called "Weak Correlation Intractability" in [CGH04].

4 Bounded Correlation Intractability from Obfuscating Puncturable PRF

In this section we give the construction of correlation intractable function ensembles with respect to all the sparse relations recognizable by circuits of size up to a given polynomial $p(\cdot)$.

Construction 4.1 (Bounded CI). Let $\mathcal{F} = \{F_K : \{0,1\}^{l(n)} \to \{0,1\}^{m(n)}\}_{n\in\mathbb{N}}$ be a puncturable PRF ensemble (XOR-patched, cf. construction 2.5). Let the function ensemble $\mathcal{H} = \{h_k : \{0,1\}^{l(n)} \to \{0,1\}^{m(n)}\}_{n\in\mathbb{N}}$ be constructed as

$$h_k(\cdot) = \mathsf{iO}(F_K(\cdot), \mathsf{padding}(n))$$

where $K \xleftarrow{\$} \mathcal{F}_n$, for some length of $\mathsf{padding}$.

Theorem 4.2 (Bounded CI). Fix an arbitrary polynomial $p(\cdot)$ and denote by $\mathcal{R} = \{R : \{0,1\}^{l(n)+m(n)} \to \{0,1\}, |R| \leq p(n)\}_{n\in\mathbb{N}}$ the class of all sparse relations that are recognizable by circuits of size up to $p(n)$. Assuming the existence of input-hiding obfuscation for all evasive circuits, sub-exponentially secure indistinguishability obfuscation for $\mathsf{P/poly}$, and sub-exponentially secure puncturable PRF, there is an appropriate polynomial size of $\mathsf{padding}$ such that the family $\mathcal{H}$ is correlation intractable w.r.t. $\mathcal{R}$.

The minimum size of padding will be discussed at the end of the proof (cf. remark 4.12). In short, it depends on p and the blow-up due to input-hiding obfuscation. In the proof below, we drop the explicit mention of padding from the construction in order to simplify notation.

Proof of Theorem 4.2: The proof goes through 3 hybrids. From the original game which captures the security definition of correlation intractability, we move to intermediate games 1, 2, and 3 that are indistinguishable by the adversary. Finally we will show that the adversary cannot win in game 3, concluded that the adversary also fails in game 0, since the adversary cannot distinguish game 0 and game 3.

Game 0: The original game. The adversary chooses a $\delta(n)$-sparse relation R, then receives the description of the function h_k constructed by the challenger:

$$h_k(\cdot) = \mathsf{iO}(F_K(\cdot)) \tag{0}$$

The adversary wins if he outputs an x such that $R(x, h_k(x)) = 1$.

Game 1: Embed the relation into the description without changing the functionality. The adversary picks a sparse relation R. The challenger samples a puncturable key K, then generates h_k which has the relation R embedded:

$$h_k(x) = \mathsf{iO}\left(\begin{array}{ll} \text{if } R(x, F_K(x)) = 1, & \text{return } F_K(x) \\ \text{else}, & \text{return } F_K(x) \end{array}\right) \tag{1}$$

The h_k in game 0 and game 1 have identical functionalities, therefore indistinguishable by any p.p.t. adversary by iO.

Game 2: Keep the "bad" branch, puncture the "innocent" one, and replace it with an independently generated puncturable function. The adversary picks a sparse relation R. The challenger generates h_k as:

$$h_k(x) = \mathsf{iO}\left(\begin{array}{ll} \text{if } R(x, F_K(x)) = 1, & \text{return } F_K(x) \\ \text{else}, & \text{return } F^R_{K'}(x) \end{array}\right) \tag{2}$$

where $K \xleftarrow{\$} \mathcal{F}_n$. $F^R_{K'}$ is constructed as follows:

Construction 4.3 ($\mathcal{F}^R$)**.** Let $\mathcal{F}^R = \{F^R_{K'} : \{0,1\}^{l(n)} \to \{0,1\}^{m(n)}\}_n$ be a function family, where each $F^R_{K'}$ is constructed as:

$$F^R_{K'}(x) = \left(\begin{array}{l} \underline{K' = \{R, T(n), K_i, i \in [T(n)]\}} \\ \text{for } i = 1 \text{ to } T(n): \\ \quad \text{if } R(x, F_{K_i}(x)) = 0, \text{ return } F_{K_i}(x) \\ \quad \text{if } R(x, F_{K_i}(x)) = 1, \text{ continue} \\ \text{return } \perp \end{array}\right) \tag{2.e}$$

in which $T(n) = O\left(\frac{l(n)}{\log(n)}\right)$, $K_i, i \in [T(n)]$, are sampled independently from the XOR-patched puncturable PRF family $\mathcal{F}$.

The functionality of $F^R_{K'}$ is, roughly, on input x, output $F_{K_i}(x)$ where K_i is the first key among $K_1, ..., K_{T(n)}$ that satisfies $R(x, F_{K_i}(x)) = 0$. The iteration time $T(n)$ is set large enough to make sure that $F^R_{K'}$ output $\perp$ with probability less than $2^{-2 \cdot l(n)}$. The next lemma shows that $T(\cdot)$ is a polynomial.

Lemma 4.4. Let $\mathcal{F} = \{F_K : \{0,1\}^{l(n)} \to \{0,1\}^{m(n)}\}_{n \in \mathbb{N}}$ be a 1-universal function ensemble. For any sparse relation R, for $K_i \xleftarrow{\$} \mathcal{F}$, $i = 1, ..., T(n)$, where $T(n) = O\left(\frac{l(n)}{\log(n)}\right)$, the following event

$$E = \text{“}\exists x \in \{0,1\}^{l(n)} : R(x, F_{K_1}(x)) = 1 \wedge ... \wedge R(x, F_{K_{T(n)}}(x)) = 1\text{”}$$

happens with probability $2^{-2 \cdot l(n)}$.

Proof of Lemma 4.4: Let $\delta(\cdot)$ be arbitrary negligible functions. For any $\delta(n)$-sparse relation R, for every $x^* \in \{0,1\}^{l(n)}$, the probability that $(x^*, F_{K_i}(x^*)) \in R$ holds for all the $T(n)$ keys F_{K_i} that are independently sampled, $i \in [1, ..., T(n)]$, is $\delta(n)^{T(n)}$, due to 1-universality. If this holds for all the inputs with probability less than $2^{-3 \cdot l(n)}$, then the probability of E is less or equal to $2^{-2 \cdot l(n)}$ by union bound. Since $\delta(n)$ is negligible, which means exists constant c such that $\delta(n) < n^{-c}$, we derive the following bound for $T(n)$:

$$\delta(n)^{T(n)} \leq \left(n^{-c}\right)^{T(n)} \leq 2^{-3 \cdot l(n)} \Rightarrow T(n) \geq O\left(\frac{l(n)}{\log(n)}\right)$$

□

Next we give a puncturing algorithm for $\mathcal{F}^R$:

Algorithm 4.5 (Puncturing[3] $\mathcal{F}^R$)**.** The algorithm $\mathsf{Puncture}(K', x^*)$ works as follows: given x^*, it runs program (2.e) on x^*, and records the index J (in the loop "for $i = 1$ to $T(n)$") where the program returns (it

[3]There's a "lazier" algorithm to puncture $\mathcal{F}^R$: for $K' = \{K_i\}_{i \in [T(n)]}$,simply puncture all the $K_i, i \in [T(n)]$, on x^*. This algorithm is correct, and all the security proofs preserve as well. However the lazier algorithm unnecessarily generates a punctured key $K'\{x^*\}$ with bigger size.

implies that $R(x^*, F_{K_j}(x^*)) = 1$, $j = 1, ..., J-1$). Then, it punctures K_J on x^*, produces $K_J\{x^*\}$, and reconstructs the program in the same way except that it replace K_J with $K_J\{x^*\}$.

$$F^R_{K'\{x^*\}}(x) = \left(\begin{array}{l} \underline{K'\{x^*\} = \{R, T(n), J, K_i, i \in [T(n)] \setminus \{J\}, K_J\{x^*\}\}} \\ \text{if } x = x^*, \text{ return } \perp \\ \text{else, for } i = 1 \text{ to } T(n): \\ \quad \text{if } i = J, \text{ return } F_{K_J\{x^*\}}(x) \\ \quad \text{else if } R(x, F_{K_i}(x)) = 0, \text{ return } F_{K_i}(x) \\ \quad \text{else if } R(x, F_{K_i}(x)) = 1, \text{ continue} \\ \text{return } \perp \end{array} \right) \tag{2.p}$$

This doesn't change the functionality on any other points. Furthermore, revealing the other keys will not leak additional information of the real value of $F^R_{K'}(x^*)$, a.k.a. $F_{K_J}(x^*)$.

To show the indistinguishability of game 1 and game 2, we introduce $2^{l(n)}$ intermediate hybrids, one for each input. Between the adjacent hybrids, we switch the evaluation key on the corresponding input from the original key K, to the freshly generated key K', if the input lives in the "innocent" branch. The indistinguishability of adjacent hybrids is proved based on sub-exponential hardness assumptions of iO and the puncturability of $\mathcal{F}$. For the coherence of the presentation, the details of the proof are presented after we go through all the hybrids.

Game 3: Wrap the "bad" branch by input-hiding obfuscation, without changing the functionality. The adversary picks a relation R. The challenger generates the h_k that is functionally equivalent to the one from game 2. The difference is, in game 3, he wraps the function of the first if-trigger with input-hiding obfuscation[4], and then iO the entire function:

$$h_k(x) = \mathsf{iO}\left(\begin{array}{l} y \leftarrow \mathsf{IHO}\left(\begin{array}{ll} \text{if } R(x, F_K(x)) = 1, & \text{return } F_K(x) \\ \text{else,} & \text{return } \perp \end{array} \right) \\ \text{if } y = \perp, \ y = F^R_{K'}(x) \\ \text{return } y \end{array} \right) \tag{3}$$

Let $E^i_{R,K}(x)$ denote the i-th bit of the output of $\left(\begin{array}{ll} \text{if } R(x, F_K(x)) = 1, & \text{return } F_K(x) \\ \text{else,} & \text{return } \perp \end{array} \right)$. We observe that $\mathcal{E}_R = \{E^i_{R,K} : \{0,1\}^{l(n)} \to \{0,1\}\}_{i \in [m(n)], n \in \mathbb{N}}$ is an evasive function family: If not, the (non-uniform) adversary distinguishes the PRF from a truly random function, by always querying the "dense" input.

Note that the h_k described in construction (2) and (3) are functionally equivalent. The adversary cannot distinguish game 2 and game 3 following indistinguishability obfuscation.

Finally, in Game 3: The adversary pick a relation R, got back h_k as described in construction (3), inside which, there's an evasive function wrapped in an input-hiding obfuscator, and an independently generated function $F^R_{K'}$, where **no** input-output pairs on $F^R_{K'}$ satisfies the relation with high probability.

Following the property of input-hiding obfuscation, for any p.p.t. adversary A:

$$\Pr_{K \xleftarrow{\$} \mathcal{F}_n, K' \leftarrow \mathcal{F}^R_n} [A(\mathsf{IHO}(E_{R,K}(\cdot)), R, F^R_{K'}) \to x : \ E_{R,K}(x) \neq \perp] < \mathsf{negl}(n)$$

[4] We split the circuit $E_{R,K}(\cdot)$ by output bits and obfuscate $E^i_{R,K}(\cdot)$, $i \in [m(n)]$ individually.

Together with the "innocent branch" which is designed to be statistically separated from the relation:

$$\Pr_{K' \leftarrow \mathcal{F}_n^R}[\exists x : \ R(x, F_{K'}^R(x)) = 1] < 2^{-2 \cdot l(n)}$$

the overall advantage of the adversary is bounded by

$$\Pr_{K \xleftarrow{\$} \mathcal{F}_n, K' \leftarrow \mathcal{F}_n^R}[A(h_k(\cdot)) \rightarrow x : R(x, h_k(x)) = 1] \leq \mathsf{negl}(n) + 2^{-2 \cdot l(n)}$$

which is negligible.

Since any p.p.t. adversary cannot distinguish whether he is in game 3 or game 0, we complete the proof of correlation intractability of $\mathcal{H}$. □

4.1 Proof of indistinguishability of game 1 and game 2

A few experiments are introduced below to describe and analyze the puncturability of $\mathcal{F}$ and $\mathcal{F}^R$ w.r.t. some "biased" distribution, which refers to the uniform distributions "truncated" by the sparse relation R picked by the adversary.

We show that the indistinguishability for the biased puncturability experiments follows the indistinguishability of the "standard" puncturability game. In particular, the reductions are *sub-exponential hardness preserving*, in the sense that if one assumes sub-exponential hardness of the "standard" puncturability game for $\mathcal{F}$, then the security game in the biased puncturing experiments for $\mathcal{F}$ and $\mathcal{F}^R$ (which is constructed from $\mathcal{F}$) are also sub-exponentially hard.

4.1.1 The biased puncturing experiments

First we introduce the algorithms and notations for sampling strings and PRF keys from the uniform distribution truncated by a sparse relation:

Algorithm 4.6 (Sampling strings from the truncated distribution)**.** Denote the following rejection sampling procedure as $r^* \leftarrow U^{m(n)} \backslash R(x^*, \cdot)$ (abbreviation: $r^* \leftarrow U \backslash R$): Given a $\delta(n)$-sparse relation $R : \{0,1\}^{l(n)+m(n)} \rightarrow \{0,1\}$, a string $x^* \in \{0,1\}^{l(n)}$, considering the sampling procedure for a string $r^* \in \{0,1\}^{m(n)}$:

$$r^* = \left[\begin{array}{l} \text{Repeat} : \ r \xleftarrow{\$} \{0,1\}^{m(n)}, \ \text{until } R(x^*, r) = 0 \\ \text{return } r \end{array} \right]$$

The sampling is efficient, with expected rounds of sampling being $\frac{1}{1-\delta_{x^*}(n)}$.

Algorithm 4.7 (Sampling PRF keys from the truncated distribution)**.** Denote the following rejection sampling procedure as $K \leftarrow \mathcal{F}_n \setminus R(x^*, \cdot)$ (abbreviation: $K \leftarrow \mathcal{F} \setminus R$): Let $\mathcal{F} = \{F_K : \{0,1\}^{l(n)} \rightarrow \{0,1\}^{m(n)}\}_{n \in \mathbb{N}}$ be a 1-universal PRF family. Given a $\delta(n)$-sparse relation $R : \{0,1\}^{l(n)+m(n)} \rightarrow \{0,1\}$, a string $x^* \in \{0,1\}^{l(n)}$, considering the sampling procedure for a PRF key $K \in \mathcal{F}_n$:

$$K = \left[\begin{array}{l} \text{Repeat} : \ K \xleftarrow{\$} \mathcal{F}_n, \ \text{until } R(x^*, F_K(x^*)) = 0 \\ \text{return } K \end{array} \right]$$

The sampling is efficient, with expected rounds of sampling being $\frac{1}{1-\delta_{x^*}(n)}$.

Experiment 4.8 (Biased puncturing experiment for $\mathcal{F}$)**.** Let $\mathcal{F} = \{F_K : \{0,1\}^{l(n)} \to \{0,1\}^{m(n)}\}_{n\in\mathbb{N}}$ be an XOR-patched puncturable PRF ensemble, consider the following experiment between the adversary and the challenger:

1. The adversary picks an input x^* and a $\delta(n)$-sparse relation R, send to the challenger.
2. The challenger samples $K \leftarrow \mathcal{F}_n \setminus R(x^*, \cdot)$, then punctures K on x^*, producing $K\{x^*\}$. He also samples $r^* \leftarrow U^{m(n)} \setminus R(x^*, \cdot)$.
3. The challenger tosses a coin b, if $b = 0$, outputs $K\{x^*\}$, $F_K(x^*)$; if $b = 1$, outputs $K\{x^*\}$, r^*. Denote these two cases as B_0 and B_1.
4. The adversary chooses $g \in \{0,1\}$, wins if $g = b$.

Lemma 4.9 (Biased puncturing lemma for $\mathcal{F}$)**.** Let $\epsilon_{\mathsf{Puncture}}(n)$ be the maximum distinguishing advantage for any p.p.t. adversary in the standard puncturability game of $\mathcal{F}$, then for any p.p.t. adversary A in the experiment 4.8,

$$| \Pr[A(B_0) = 1] - \Pr[A(B_1) = 1] | \leq 2 \cdot \epsilon_{\mathsf{Puncture}}(n)$$

where the probabilities are taken over $K \leftarrow \mathcal{F}_n \setminus R(x^*, \cdot)$, $r^* \leftarrow U^{m(n)} \setminus R(x^*, \cdot)$.

Proof of lemma 4.9 For any p.p.t. adversary A that distinguishes the two cases in experiment 4.8 with advantage $\eta(n)$, we build an adversary D for the standard puncturability game of $\mathcal{F}$ as follows:

1. Upon receiving the relation R and input x^* picked by A, D queries the challenger of puncturability game on x^*, got back $K\{x^*\}$, and y^* equals to either $F_K(x^*)$, or $r^* \xleftarrow{\$} \{0,1\}^{m(n)}$.
2. D runs the following decision procedure: if $R(x^*, y^*) = 1$, then tosses a coin; else if $R(x^*, y^*) = 0$, then D sends $K\{x^*\}$, y^* to A, following the decision of A.

The analysis of the advantage of D:

$$\begin{aligned}
& \left| \Pr_{K,r^*}[D(x^*, K\{x^*\}, F_K(x^*)) = 1] - \Pr_{K,r^*}[D(x^*, K\{x^*\}, r^*) = 1] \right| \\
= & \Big| \Pr_{K,r^*}[D(x^*, K\{x^*\}, F_K(x^*)) = 1 | R(x^*, F_K(x^*)) = 0] \cdot \Pr_K[R(x^*, F_K(x^*)) = 0] && \text{(B.a)} \\
& + \Pr_{K,r^*}[D(x^*, K\{x^*\}, F_K(x^*)) = 1 | R(x^*, F_K(x^*)) = 1] \cdot \Pr_K[R(x^*, F_K(x^*)) = 1] && \text{(B.b)} \\
& - \Pr_{K,r^*}[D(x^*, K\{x^*\}, r^*) = 1 | R(x^*, r^*) = 0] \cdot \Pr_{r^*}[R(x^*, r^*) = 0] && \text{(B.c)} \\
& - \Pr_{K,r^*}[D(x^*, K\{x^*\}, r^*) = 1 | R(x^*, r^*) = 1] \cdot \Pr_{r^*}[R(x^*, r^*) = 1] \Big| && \text{(B.d)} \\
\geq & \Big| \Pr_{K \leftarrow \mathcal{F}\setminus R, r^* \leftarrow U\setminus R}[D(x^*, K\{x^*\}, F_K(x^*)) = 1] && \text{(B.e)} \\
& - \left[\Pr_{K \leftarrow \mathcal{F}\setminus R, r^* \leftarrow U\setminus R}[D(x^*, K\{x^*\}, r^*) = 1] + \epsilon_{\mathsf{Puncture}}(n) \right] \Big| \cdot (1 - \delta_{x^*}(n)) && \text{(B.f)} \\
= & \ | \Pr[A(B_0) = 1] - \Pr[A(B_1) = 1] - \epsilon_{\mathsf{Puncture}}(n) | \cdot (1 - \delta_{x^*}(n)) \\
\geq & \ (\eta(n) - \epsilon_{\mathsf{Puncture}}(n)) \cdot (1 - \delta_{x^*}(n)) \geq \eta(n)/2
\end{aligned}$$

where $\Pr_K[R(x^*, F_K(x^*)) = 1] = 1 - \delta_{x^*}(n)$ follows the 1-universality of $\mathcal{F}$; (B.b) and (B.d) cancel out since $\Pr_K[R(x^*, F_K(x^*)) = 1] = \Pr_{r^*}[R(x^*, r^*) = 1] = \delta_{x^*}(n)$, and D tosses a coin in both cases; (B.a) is equal to (B.e) since the distribution of r^* doesn't affect the distinguisher's view at all.

It is left to show that the decisions of any adversary in (B.c) and (B.f) are $\epsilon_{\mathsf{Puncture}}(n)$-close. Here we make explicit use of the XOR-patched puncturable PRF $\mathcal{F}$ ($\mathcal{F}^U$ in construction 2.5): suppose by contradiction, there's an adversary A_2 that given $K^U\{x^*\} = (K\{x^*\}, u)$, $r^* \leftarrow U^{m(n)} \setminus R(x^*, \cdot)$, behaves differently (with probability $\eta'(n)$) under $K^U \xleftarrow{\$} \mathcal{F}_n$ and $K^U \leftarrow \mathcal{F}_n \setminus R(x^*, \cdot)$, we build a distinguisher D_2 that breaks the standard puncturability of the XOR-patched puncturable PRF $\mathcal{F}$:

1. D_2 chooses x^*, got $K^U\{x^*\} = (K\{x^*\}, u)$, y^*, where $K^U \xleftarrow{\$} \mathcal{F}$, y^* equals to either $F_K(x^*) \oplus u$ or $r^* \xleftarrow{\$} \{0,1\}^{m(n)}$.
2. Then D_2 samples u_2 such that $y^* \oplus u_2 \leftarrow U^{m(n)} \setminus R(x^*, \cdot)$, and creates a new punctured key as $K^{U_2}\{x^*\} = (K\{x^*\}, u \oplus u_2)$. This sampling trick induces different distributions for $K^{U_2}\{x^*\}$ depends on y^*: if $y^* = F_K(x^*) \oplus u$, then $K^{U_2} \leftarrow \mathcal{F}_n \setminus R(x^*, \cdot)$; if $y^* = r^*$, then $K^{U_2} \xleftarrow{\$} \mathcal{F}_n$.
3. Furthermore, D_2 samples $r_2 \leftarrow U^{m(n)} \setminus R(x^*, \cdot)$ independently, and sends $K^{U_2}\{x^*\}, r_2$ to the adversary A_2, follows his decisions.

Then the advantage of D_2 equals to the advantage of A_2 in the standard puncturability game. □

Experiment 4.10 (Biased puncturing experiment for $\mathcal{F}^R$)**.** Let $\mathcal{F} = \{F_K : \{0,1\}^{l(n)} \to \{0,1\}^{m(n)}\}_{n\in\mathbb{N}}$ be an XOR-patched puncturable PRF ensemble (which the $\mathcal{F}^R$ in the experiment is based on), consider the following experiment between the adversary and the challenger:

1. The adversary picks an input x^* and a $\delta(n)$-sparse relation R, send to the challenger.
2. The challenger samples $K' \leftarrow \mathcal{F}^R$ according to construction 4.3, then punctures K' on x^* using algorithm 4.5, producing $K'\{x^*\}$.
3. The challenger tosses a coin b', if $b' = 0$, outputs $K'\{x^*\}$, $F^R_{K'}(x^*)$; if $b' = 1$, outputs $K'\{x^*\}$, $r^* \leftarrow U^{m(n)} \setminus R(x^*, \cdot)$. Denote these two cases as B'_0 and B'_1.
4. Adversary chooses g', wins if $g' = b'$.

Lemma 4.11 (Biased puncturing lemma for $\mathcal{F}^R$)**.** Let $\epsilon_{\mathsf{Puncture}}(n)$ be the maximum distinguishing advantage for any p.p.t. adversary in the puncturability game of $\mathcal{F}$, from which $\mathcal{F}^R$ is built, then for any p.p.t. adversary A' in the experiment 4.10:

$$\mid \Pr[A'(B'_0) = 1] - \Pr[A'(B'_1) = 1] \mid \leq 2 \cdot \epsilon_{\mathsf{Puncture}}(n) + 2^{-2 \cdot l(n)}$$

where the probabilities are taken over $K' \xleftarrow{\$} \mathcal{F}^R_n$, $r^* \leftarrow U^{m(n)} \setminus R(x^*, \cdot)$.

Proof of lemma 4.11 According to the construction and the puncturing algorithm of $\mathcal{F}^R$, the advantage of A' for experiment 4.10, is the sum of advantages for distinguishing the biased puncturability experiments of $K_J \xleftarrow{\$} \mathcal{F}$ w.r.t. R, conditioned on the events that $J = 1, ..., T(n)$:

$$
\begin{aligned}
& \left| \Pr[A'(B'_0) = 1] - \Pr[A'(B'_1) = 1] \right| \\
\leq & \sum_{J=1}^{T(n)} \left[\Pr_{K_1,...,K_J \xleftarrow{\$} \mathcal{F}_n} [R(x^*, F_{K_1}(x^*)) = 1 \wedge ... \wedge R(x^*, F_{K_{J-1}}(x^*)) = 1 \wedge R(x^*, F_{K_J}(x^*)) = 0] \cdot \right. \\
& \left| \Pr_{K_J \xleftarrow{\$} \mathcal{F}_n, r^* \leftarrow U \setminus R} [D_J(x^*, K_J\{x^*\}, F_{K_J}(x^*)) = 1 | R(x^*, F_{K_J}(x^*)) = 0] \right. \\
& \left. \left. - \Pr_{K_J \xleftarrow{\$} \mathcal{F}_n, r^* \leftarrow U \setminus R} [D_J(x^*, K_J\{x^*\}, r^*) = 1 | R(x^*, F_{K_J}(x^*)) = 0] \right| \right] + \Pr_{K' \xleftarrow{\$} \mathcal{F}^R} [F^R_{K'}(x^*) = \perp] \\
\leq & \sum_{J=1}^{T(n)} \delta_{x^*}^{J-1}(n) \cdot (1 - \delta_{x^*}(n)) \cdot 2 \cdot \epsilon_{\mathsf{Puncture}}(n) + 2^{-2 \cdot l(n)} \\
= & \ (1 - \delta_{x^*}^{T(n)}) \cdot 2 \cdot \epsilon_{\mathsf{Puncture}}(n) + 2^{-2 \cdot l(n)} \leq 2 \cdot \epsilon_{\mathsf{Puncture}}(n) + 2^{-2 \cdot l(n)}
\end{aligned}
$$

where $\Pr_{K_J}[R(x^*, F_{K_J}(x^*)) = 1] = \delta_{x^*}(n)$ follows the 1-universality of $\mathcal{F}$; the fact that

$$
\begin{aligned}
& \left| \Pr_{K_J \xleftarrow{\$} \mathcal{F}_n, r^* \leftarrow U \setminus R} [D_J(x^*, K_J\{x^*\}, F_{K_J}(x^*)) = 1 | R(x^*, F_{K_J}(x^*)) = 0] \right. \\
& \left. - \Pr_{K_J \xleftarrow{\$} \mathcal{F}_n, r^* \leftarrow U \setminus R} [D_J(x^*, K_J\{x^*\}, r^*) = 1 | R(x^*, F_{K_J}(x^*)) = 0] \right| \leq 2 \cdot \epsilon_{\mathsf{Puncture}}(n)
\end{aligned}
$$

follows lemma 4.9, for $J = 1, ..., T(n)$. □

4.1.2 Indistinguishability of game 1 and game 2

The analysis goes through $2^{l(n)}$ hybrids, one for each input.

Descriptions of Hybrids H_0, ..., $H_{2^{l(n)}}$: For $z \in \{0, 1, ..., 2^{l(n)}\}$, the hybrid H_z is described as follows: for the $\delta(n)$-sparse relation R picked by the adversary, the challenger samples a 1-universal puncturable PRF key $K \xleftarrow{\$} \mathcal{F}$, and constructs $K' \xleftarrow{\$} \mathcal{F}^R$ by algorithm (4.3), then produces the function h_k as:

$$
h_k(x) = \mathsf{iO}\left(\begin{array}{l} \text{if } R(x, F_K(x)) = 1, \text{ return } F_K(x) \\ \text{else, return } \left(\begin{array}{ll} \text{if } x \geq z, & \text{return } F_K(x) \\ \text{if } x < z, & \text{return } F^R_{K'}(x) \end{array} \right) \end{array} \right) \tag{2.z}
$$

By the description of h_k in each hybrid, h_k in H_0 has the same functionality with function (1) in game 1; whereas h_k in $H_{2^{l(n)}}$ has the same functionality with function (2) in game 2.

Let $\epsilon_{\mathsf{Puncture}}$ be the adversary's advantage of winning the puncturability game of $\mathcal{F}$, and ϵ_{iO} be the advantage of distinguishing the iO of two identical functions. We prove indistinguishability of H_0 and $H_{2^{l(n)}}$ by showing

the sub-exponential hardness of distinguishing H_{z^*} and H_{z^*+1}, for all $z^* \in \{0, ..., 2^{l(n)} - 1\}$, based on the sub-exponential hardness of puncturability and iO. Specifically, let

$$\epsilon_{\mathsf{Puncture}} = \epsilon_{\mathsf{iO}} = 2^{-l(n)} \cdot \mathsf{negl}(n)$$

For the adversary between the hybrids H_{z^*} and H_{z^*+1}, consider the following two cases:

(1) If $R(z^*, F_K(z^*)) = 1$, then the h_k functions sampled according to algorithm (2.z) in H_{z^*} and H_{z^*+1} have identical functionality, so that they are indistinguishable by iO.

(2) If $R(z^*, F_K(z^*)) = 0$, then we prove the indistinguishability of H_{z^*} and H_{z^*+1} by further introducing intermediate hybrids $H_{z^*.1}$, $H_{z^*.2}$ and $H_{z^*.3}$ - In each of them, the challenger punctures K and K' on z^*, produces $K\{z^*\}$, $K'\{z^*\}$. y^* is obtained differently in the 3 hybrids:

1. $H_{z^*.1}$: $y^* = F_K(z^*)$;
2. $H_{z^*.2}$: $y^* \leftarrow U^{m(n)} \setminus R(z^*, \cdot)$;
3. $H_{z^*.3}$: $y^* = F^R_{K'}(z^*)$.

The challenger then constructs h_k as:

$$h_k(x) = \mathsf{iO}\left(\begin{array}{l} \text{if } R(x, F_{K\{z^*\}}(x)) = 1, \text{ return } F_{K\{z^*\}}(x) \\ \text{else, return } \left(\begin{array}{ll} \text{if } x > z^*, & \text{return } F_{K\{z^*\}}(x) \\ \text{if } x = z^*, & \text{return } y^* \\ \text{if } x < z^*, & \text{return } F^R_{K'\{z^*\}}(x) \end{array}\right) \end{array}\right) \tag{2.z.p}$$

Indistinguishability of hybrids H_{z^*} and $H_{z^*.1}$: By the description of h_k in $H_{z^*.1}$ (cf. function (2.z.p)), if $y^* = F_K(z^*)$, then it is functionally equivalent to h_k in H_{z^*} (cf. function (2.z)), therefore indistinguishable following iO;

Indistinguishability of hybrids $H_{z^*.1}$ and $H_{z^*.2}$: Suppose by contradiction there's a p.p.t. adversary A_1 that distinguishes hybrids $H_{z^*.1}$ and $H_{z^*.2}$ with probability $\eta(n) \cdot 2^{-l(n)}$, where η is a non-negligible function over n. We build an adversary A for the biased puncturability experiment 4.8:

1. Upon receiving the relation R picked by A_1, A queries the challenger of the biased puncturability experiment for $\mathcal{F}$ on R and z^*, got back $K\{z^*\}$ and y^*.
2. In addition, A samples $K' \xleftarrow{\$} \mathcal{F}^R_n$ according to construction 4.3, punctures K' on z^* using algorithm 4.5, then constructs h_k with R, z^*, $K\{z^*\}$, y^* and $K'\{z^*\}$ according to function (2.z.p).
3. A sends h_k to A_1. If A_1 chooses hybrid $H_{z^*.1}$, A outputs 0; if A_1 chooses hybrid $H_{z^*.2}$, A outputs 1.

By the description of experiment 4.8, K is sampled such that $R(z^*, F_K(z^*)) = 0$, y^* equals to either $F_K(z^*)$, or a value $r^* \leftarrow U^{m(n)} \setminus R(z^*, \cdot)$, so that A simulates the exact same distribution of h_k that A_1 gets in hybrids $H_{z^*.1}$ and $H_{z^*.2}$:

$$\begin{aligned} &\left| \Pr_{K \leftarrow \mathcal{F} \setminus R, r^* \leftarrow U \setminus R}[A(z^*, K\{z^*\}, F_K(z^*)) = 1] - \Pr_{K \leftarrow \mathcal{F} \setminus R, r^* \leftarrow U \setminus R}[A(z^*, K\{z^*\}, r^*) = 1] \right| \\ = &\left| \Pr_{K \leftarrow \mathcal{F} \setminus R, r^* \leftarrow U \setminus R, K' \xleftarrow{\$} \mathcal{F}^R}[A_1(H_{z^*.1}) = 1] - \Pr_{K \leftarrow \mathcal{F} \setminus R, r^* \leftarrow U \setminus R, K' \xleftarrow{\$} \mathcal{F}^R}[A_1(H_{z^*.2}) = 1] \right| \end{aligned}$$

Therefore, the advantage of A is $2^{-l(n)} \cdot \eta(n)$. By lemma 4.9, there is an adversary D for the standard puncturability game for $\mathcal{F}$ with advantage at least $2^{-l(n)} \cdot \eta(n)/2$. It violates the $\mathsf{negl}(n) \cdot 2^{-l(n)}$-hardness assumption of puncturing for $\mathcal{F}_n$.

Indistinguishability of hybrids $H_{z^*.2}$ and $H_{z^*.3}$: Suppose by contradiction, there's a p.p.t. adversary A_2 that distinguishes hybrids $H_{z^*.2}$ and $H_{z^*.3}$ with probability $\eta(n) \cdot 2^{-l(n)}$, where η is a non-negligible function over n. We build an adversary A' for the biased puncturability experiment 4.10:

1. Upon receiving the relation R picked by A_2, A' queries the challenger of the biased puncturability experiment for $\mathcal{F}^R$ (cf. experiment 4.10) with R and z^*, got back $K'\{z^*\}$, y^*.
2. In addition, A' samples $K \leftarrow \mathcal{F}_n \setminus R(z^*, \cdot)$, punctures K on z^*, then constructs h_k with R, z^*, $K\{z^*\}$, y^* and $K'\{z^*\}$ according to construction (2.z.p).
3. A' sends h_k to A_2. If A_2 chooses hybrid $H_{z^*.3}$, A' outputs 0; if A_2 chooses hybrid $H_{z^*.2}$, A' outputs 1.

By the description of experiment 4.10, y^* equals to either $F^R_{K'}(z^*)$, or a value $r^* \leftarrow U^{m(n)} \setminus R(z^*, \cdot)$, so that A' simulates the exact same distribution of h_k that A_2 gets in hybrids $H_{z^*.3}$ and $H_{z^*.2}$:

$$
\begin{aligned}
& \left| \Pr_{K' \xleftarrow{\$} \mathcal{F}^R, r^* \leftarrow U \setminus R} [A'(z^*, K'\{z^*\}, F^R_{K'}(z^*)) = 1] - \Pr_{K' \xleftarrow{\$} \mathcal{F}^R, r^* \leftarrow U \setminus R} [A'(z^*, K'\{z^*\}, r^*) = 1] \right| \\
= & \left| \Pr_{K \leftarrow \mathcal{F} \setminus R, r^* \leftarrow U \setminus R, K' \xleftarrow{\$} \mathcal{F}^R} [A_2(H_{z^*.3}) = 1] - \Pr_{K \leftarrow \mathcal{F} \setminus R, r^* \leftarrow U \setminus R, K' \xleftarrow{\$} \mathcal{F}^R} [A_2(H_{z^*.2}) = 1] \right|
\end{aligned}
$$

Therefore, the advantage of A' is $2^{-l(n)} \cdot \eta(n)$, which means that there is an adversary for the standard puncturability game for $\mathcal{F}$ with advantage at least $2^{-l(n)} \cdot \eta(n)/2$, by lemma 4.11. It violates the $\mathsf{negl}(n) \cdot 2^{-l(n)}$-hardness assumption of puncturing for $\mathcal{F}_n$.

Indistinguishability of hybrids $H_{z^*.3}$ and H_{z^*+1}: By the description of h_k in $H_{z^*.3}$ (cf. function (2.z.p)), if $y^* = F^R_{K'}(z^*)$, then it is functionally equivalent to h_k in H_{z^*+1} (cf. function (2.z)), therefore indistinguishable following iO.

To conclude, the adversary's advantage of distinguishing hybrids H_{z^*} and H_{z^*+1} is bounded by:

$$
\begin{aligned}
& | \Pr[A(H_{z^*}) = 1] - \Pr[A(H_{z^*+1}) = 1] | \\
< & | \Pr[A(H_{z^*}) = 1 | R(z^*, F_K(z^*)) = 1] - \Pr[A(H_{z^*+1}) = 1 | R(z^*, F_K(z^*)) = 1] | \\
& + | \Pr[A(H_{z^*}) = 1 | R(z^*, F_K(z^*)) = 0] - \Pr[A(H_{z^*+1}) = 1 | R(z^*, F_K(z^*)) = 0] | \\
\leq & \ \epsilon_{\mathsf{iO}} + | \Pr[A(H_{z^*}) = 1 | R(z^*, F_K(z^*)) = 0] - \Pr[A(H_{z^*.1}) = 1 | R(z^*, F_K(z^*)) = 0] | \\
& + | \Pr[A(H_{z^*.1}) = 1 | R(z^*, F_K(z^*)) = 0] - \Pr[A(H_{z^*.2}) = 1 | R(z^*, F_K(z^*)) = 0] | \\
& + | \Pr[A(H_{z^*.2}) = 1 | R(z^*, F_K(z^*)) = 0] - \Pr[A(H_{z^*.3}) = 1 | R(z^*, F_K(z^*)) = 0] | \\
& + | \Pr[A(H_{z^*.3}) = 1 | R(z^*, F_K(z^*)) = 0] - \Pr[A(H_{z^*+1}) = 1 | R(z^*, F_K(z^*)) = 0] | \\
\leq & \ \epsilon_{\mathsf{iO}} + \epsilon_{\mathsf{iO}} + 2 \cdot \epsilon_{\mathsf{Puncture}} + 2 \cdot \epsilon_{\mathsf{Puncture}} + 2^{-2 \cdot l(n)} + \epsilon_{\mathsf{iO}} \\
\leq & \ 8 \cdot \mathsf{negl}(n) \cdot 2^{-l(n)}
\end{aligned}
$$

The proof completes by taking the sum of all the probability of distinguishing H_{z^*} and H_{z^*+1} over $z^* \in \{0, ..., 2^{l(n)} - 1\}$. Assuming $2^{-l(n)} \cdot \mathsf{negl}(n)$-hardness of iO and puncturability of $\mathcal{F}$, the adversary can distinguish whether he is in H_0 (a.k.a. game 1) or $H_{2^{l(n)}}$ (a.k.a. game 2) with negligible probability. □

4.2 Discussion

Remark 4.12 (The size of padding)**.** Let $\kappa_{\mathcal{F}}(n)$ be the key size of $\mathcal{F}_n$, $\kappa^*_{\mathcal{F}}(n)$ be the punctured key size of $\mathcal{F}_n$, $B(\cdot)$ be the maximum blow-up of the input-hiding obfuscation. The size of $F^R_{K'}$ is $T(n) \cdot (p(n) + \kappa_{\mathcal{F}}(n))$. The maximum size of $\mathsf{IHO}(E_{R,K})$ is $B((p(n) + 2 \cdot \kappa_{\mathcal{F}}(n)) \cdot m(n))$. The size of padding is bounded by

$$|\mathsf{padding}(n)| \leq B((p(n) + 2 \cdot \kappa_{\mathcal{F}}(n)) \cdot m(n)) + T(n) \cdot (p(n) + 2 \cdot \kappa_{\mathcal{F}}(n)) + 2 \cdot \kappa^*_{\mathcal{F}}(n) = \mathsf{poly}(n)$$

As the analysis suggests, the key size of the function inherently exceeds the maximum size of R. The existence of correlation intractable functions with a prescribed description size that works for all poly-size relations (i.e. CI-P/poly) remains an open problem.

Another limitation of our result is that we don't know if obfuscated puncturable PRFs are correlation intractable w.r.t. relations with multiple invocations (mCI). The definitions and conditional impossibility results for mCI are discussed in [CGH04, Nis99]. Further studying the existence of mCI is an interesting future direction.

Acknowledgment

We are grateful to Nir Bitansky, Cheng Chen, Omer Paneth, and Oxana Poburinnaya for their enlightening discussions in the early stage of this work. We also thank Ethan Heilman for discussions on the Bitcoin protocol.

References

[AB15] Benny Applebaum and Zvika Brakerski. Obfuscating circuits via composite-order graded encoding. In Yevgeniy Dodis and Jesper Buus Nielsen, editors, *Theory of Cryptography - 12th Theory of Cryptography Conference, TCC 2015, Warsaw, Poland, March 23-25, 2015, Proceedings, Part II*, volume 9015 of *Lecture Notes in Computer Science*, pages 528–556. Springer, 2015.

[BBC+14] Boaz Barak, Nir Bitansky, Ran Canetti, Yael Tauman Kalai, Omer Paneth, and Amit Sahai. Obfuscation for evasive functions. In Lindell [Lin14], pages 26–51.

[BC10] Nir Bitansky and Ran Canetti. On strong simulation and composable point obfuscation. In Tal Rabin, editor, *Advances in Cryptology - CRYPTO 2010, 30th Annual Cryptology Conference, Santa Barbara, CA, USA, August 15-19, 2010. Proceedings*, volume 6223 of *Lecture Notes in Computer Science*, pages 520–537. Springer, 2010.

[BCC+14] Nir Bitansky, Ran Canetti, Henry Cohn, Shafi Goldwasser, Yael Tauman Kalai, Omer Paneth, and Alon Rosen. The impossibility of obfuscation with auxiliary input or a universal simulator. In Garay and Gennaro [GG14], pages 71–89.

[BCKP14] Nir Bitansky, Ran Canetti, Yael Tauman Kalai, and Omer Paneth. On virtual grey box obfuscation for general circuits. In Garay and Gennaro [GG14], pages 108–125.

[BDSG+13] Nir Bitansky, Dana Dachman-Soled, Sanjam Garg, Abhishek Jain, Yael Tauman Kalai, Adriana López-Alt, and Daniel Wichs. Why "fiat-shamir for proofs" lacks a proof. In *TCC*, pages 182–201, 2013.

[BGI+12] Boaz Barak, Oded Goldreich, Russell Impagliazzo, Steven Rudich, Amit Sahai, Salil P. Vadhan, and Ke Yang. On the (im)possibility of obfuscating programs. *J. ACM*, 59(2):6, 2012.

[BGI14] Elette Boyle, Shafi Goldwasser, and Ioana Ivan. Functional signatures and pseudorandom functions. In Hugo Krawczyk, editor, *Public Key Cryptography*, volume 8383 of *Lecture Notes in Computer Science*, pages 501–519. Springer, 2014.

[BGK+14] Boaz Barak, Sanjam Garg, Yael Tauman Kalai, Omer Paneth, and Amit Sahai. Protecting obfuscation against algebraic attacks. In Nguyen and Oswald [NO14], pages 221–238.

[BHK13] Mihir Bellare, Viet Tung Hoang, and Sriram Keelveedhi. Instantiating random oracles via uces. In Ran Canetti and Juan A. Garay, editors, *Advances in Cryptology - CRYPTO 2013 - 33rd Annual Cryptology Conference, Santa Barbara, CA, USA, August 18-22, 2013. Proceedings, Part II*, volume 8043 of *Lecture Notes in Computer Science*, pages 398–415. Springer, 2013.

[BLMR13] Dan Boneh, Kevin Lewi, Hart William Montgomery, and Ananth Raghunathan. Key homomorphic prfs and their applications. In Canetti and Garay [CG13], pages 410–428.

[Blu86] Manuel Blum. How to prove a theorem so no one else can claim it, August 1986. Invited 45 minute address to the International Congress of Mathematicians, 1986. To appear in the Proceedings of ICM 86.

[BLV06] Boaz Barak, Yehuda Lindell, and Salil P. Vadhan. Lower bounds for non-black-box zero knowledge. *J. Comput. Syst. Sci.*, 72(2):321–391, 2006.

[BM14] Christina Brzuska and Arno Mittelbach. Using indistinguishability obfuscation via uces. In Sarkar and Iwata [SI14], pages 122–141.

[BMSZ15] Saikrishna Badrinarayanan, Eric Miles, Amit Sahai, and Mark Zhandry. Post-zeroizing obfuscation: The case of evasive circuits. Cryptology ePrint Archive, Report 2015/167, 2015. http://eprint.iacr.org/.

[BR14] Zvika Brakerski and Guy N. Rothblum. Virtual black-box obfuscation for all circuits via generic graded encoding. In Lindell [Lin14], pages 1–25.

[BST14] Mihir Bellare, Igors Stepanovs, and Stefano Tessaro. Poly-many hardcore bits for any one-way function and a framework for differing-inputs obfuscation. In Sarkar and Iwata [SI14], pages 102–121.

[BW13] Dan Boneh and Brent Waters. Constrained pseudorandom functions and their applications. In Kazue Sako and Palash Sarkar, editors, *ASIACRYPT (2)*, volume 8270 of *Lecture Notes in Computer Science*, pages 280–300. Springer, 2013.

[Can97] Ran Canetti. Towards realizing random oracles: Hash functions that hide all partial information. In Burton S. Kaliski Jr., editor, *Advances in Cryptology - CRYPTO '97, 17th Annual International Cryptology Conference, Santa Barbara, California, USA, August 17-21, 1997, Proceedings*, volume 1294 of *Lecture Notes in Computer Science*, pages 455–469. Springer, 1997.

[CG13] Ran Canetti and Juan A. Garay, editors. *Advances in Cryptology - CRYPTO 2013 - 33rd Annual Cryptology Conference, Santa Barbara, CA, USA, August 18-22, 2013. Proceedings, Part I*, volume 8042 of *Lecture Notes in Computer Science*. Springer, 2013.

[CGH98] Ran Canetti, Oded Goldreich, and Shai Halevi. The random oracle methodology, revisited (preliminary version). In Vitter [Vit98], pages 209–218.

[CGH04] Ran Canetti, Oded Goldreich, and Shai Halevi. The random oracle methodology, revisited. *J. ACM*, 51(4):557–594, 2004.

[CHL+14] Jung Hee Cheon, Kyoohyung Han, Changmin Lee, Hansol Ryu, and Damien Stehle. Cryptanalysis of the multilinear map over the integers. Cryptology ePrint Archive, Report 2014/906, 2014. http://eprint.iacr.org/.

[CLT13] Jean-Sébastien Coron, Tancrède Lepoint, and Mehdi Tibouchi. Practical multilinear maps over the integers. In Canetti and Garay [CG13], pages 476–493.

[CMR98] Ran Canetti, Daniele Micciancio, and Omer Reingold. Perfectly one-way probabilistic hash functions (preliminary version). In Vitter [Vit98], pages 131–140.

[CRV10] Ran Canetti, Guy N. Rothblum, and Mayank Varia. Obfuscation of hyperplane membership. In Daniele Micciancio, editor, *Theory of Cryptography, 7th Theory of Cryptography Conference, TCC 2010, Zurich, Switzerland, February 9-11, 2010. Proceedings*, volume 5978 of *Lecture Notes in Computer Science*, pages 72–89. Springer, 2010.

[FS86] Amos Fiat and Adi Shamir. How to prove yourself: Practical solutions to identification and signature problems. In *CRYPTO*, pages 186–194, 1986.

[GG14] Juan A. Garay and Rosario Gennaro, editors. *Advances in Cryptology - CRYPTO 2014 - 34th Annual Cryptology Conference, Santa Barbara, CA, USA, August 17-21, 2014, Proceedings, Part II*, volume 8617 of *Lecture Notes in Computer Science*. Springer, 2014.

[GGH13a] Sanjam Garg, Craig Gentry, and Shai Halevi. Candidate multilinear maps from ideal lattices. In Thomas Johansson and Phong Q. Nguyen, editors, *EUROCRYPT*, volume 7881 of *Lecture Notes in Computer Science*, pages 1–17. Springer, 2013.

[GGH+13b] Sanjam Garg, Craig Gentry, Shai Halevi, Mariana Raykova, Amit Sahai, and Brent Waters. Candidate indistinguishability obfuscation and functional encryption for all circuits. In *FOCS*, pages 40–49. IEEE Computer Society, 2013.

[GGM86] Oded Goldreich, Shafi Goldwasser, and Silvio Micali. How to construct random functions. *J. ACM*, 33(4):792–807, 1986.

[GK05] Shafi Goldwasser and Yael Tauman Kalai. On the impossibility of obfuscation with auxiliary input. In *FOCS*, pages 553–562. IEEE Computer Society, 2005.

[Gol02] Oded Goldreich. The ggm construction does not yield correlation intractable function ensembles. *IACR Cryptology ePrint Archive*, 2002:110, 2002.

[GR14] Shafi Goldwasser and Guy N. Rothblum. On best-possible obfuscation. *J. Cryptology*, 27(3):480–505, 2014.

[Had00] Satoshi Hada. Zero-knowledge and code obfuscation. In Tatsuaki Okamoto, editor, *ASIACRYPT*, volume 1976 of *Lecture Notes in Computer Science*, pages 443–457. Springer, 2000.

[HKW14] Susan Hohenberger, Venkata Koppula, and Brent Waters. Adaptively secure puncturable pseudorandom functions in the standard model. *IACR Cryptology ePrint Archive*, 2014:521, 2014.

[HMR08] Shai Halevi, Steven Myers, and Charles Rackoff. On seed-incompressible functions. In Ran Canetti, editor, *TCC*, volume 4948 of *Lecture Notes in Computer Science*, pages 19–36. Springer, 2008.

[HSW14] Susan Hohenberger, Amit Sahai, and Brent Waters. Replacing a random oracle: Full domain hash from indistinguishability obfuscation. In Nguyen and Oswald [NO14], pages 201–220.

[HT06] Satoshi Hada and Toshiaki Tanaka. Zero-knowledge and correlation intractability. *IEICE Transactions*, 89-A(10):2894–2905, 2006.

[KPTZ13] Aggelos Kiayias, Stavros Papadopoulos, Nikos Triandopoulos, and Thomas Zacharias. Delegatable pseudorandom functions and applications. In Ahmad-Reza Sadeghi, Virgil D. Gligor, and Moti Yung, editors, *ACM Conference on Computer and Communications Security*, pages 669–684. ACM, 2013.

[Lin14] Yehuda Lindell, editor. *Theory of Cryptography - 11th Theory of Cryptography Conference, TCC 2014, San Diego, CA, USA, February 24-26, 2014. Proceedings*, volume 8349 of *Lecture Notes in Computer Science*. Springer, 2014.

[MPS12] Avradip Mandal, Jacques Patarin, and Yannick Seurin. On the public indifferentiability and correlation intractability of the 6-round feistel construction. In Ronald Cramer, editor, *Theory of Cryptography - 9th Theory of Cryptography Conference, TCC 2012, Taormina, Sicily, Italy, March 19-21, 2012. Proceedings*, volume 7194 of *Lecture Notes in Computer Science*, pages 285–302. Springer, 2012.

[Nak08] Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system. *Consulted*, 1(2012):28, 2008.

[Nis99] Kobbi Nissim. Two results regarding correlation intractability. *Manuscript*, 1999.

[NO14] Phong Q. Nguyen and Elisabeth Oswald, editors. *Advances in Cryptology - EUROCRYPT 2014 - 33rd Annual International Conference on the Theory and Applications of Cryptographic Techniques, Copenhagen, Denmark, May 11-15, 2014. Proceedings*, volume 8441 of *Lecture Notes in Computer Science*. Springer, 2014.

[PST14] Rafael Pass, Karn Seth, and Sidharth Telang. Indistinguishability obfuscation from semantically-secure multilinear encodings. In Juan A. Garay and Rosario Gennaro, editors, *CRYPTO (1)*, volume 8616 of *Lecture Notes in Computer Science*, pages 500–517. Springer, 2014.

[SI14] Palash Sarkar and Tetsu Iwata, editors. *Advances in Cryptology - ASIACRYPT 2014 - 20th International Conference on the Theory and Application of Cryptology and Information Security, Kaoshiung, Taiwan, R.O.C., December 7-11, 2014, Proceedings, Part II*, volume 8874 of *Lecture Notes in Computer Science*. Springer, 2014.

[SW14] Amit Sahai and Brent Waters. How to use indistinguishability obfuscation: deniable encryption, and more. In David B. Shmoys, editor, *STOC*, pages 475–484. ACM, 2014.

[Vit98] Jeffrey Scott Vitter, editor. *Proceedings of the Thirtieth Annual ACM Symposium on the Theory of Computing, Dallas, Texas, USA, May 23-26, 1998*. ACM, 1998.

[Wee05] Hoeteck Wee. On obfuscating point functions. In Harold N. Gabow and Ronald Fagin, editors, *Proceedings of the 37th Annual ACM Symposium on Theory of Computing, Baltimore, MD, USA, May 22-24, 2005*, pages 523–532. ACM, 2005.

[Zim14] Joe Zimmerman. How to obfuscate programs directly. *IACR Cryptology ePrint Archive*, 2014:776, 2014.

Appendices

A Input-hiding obfuscation for evasive functions

In this section we introduce one of the known approaches to designing input-hiding obfuscation for evasive circuits. As a corollary of the results from [BBC+14] and [BCKP14], IHO is implied by Virtual-Grey-Box (VGB) obfuscation, or equivalently, strong indistinguishability obfuscation (siO).

Definition A.1 (Concentrated / Evasive function distribution)**.** Let $\mathcal{F} = \{f_k : \{0,1\}^{l(n)} \to \{0,1\}\}_{n\in\mathbb{N}}$ be a function ensemble, $\tilde{\mathcal{F}}_n$ be a distribution on $\mathcal{F}_n$. Let $\mathsf{maj}_{\tilde{F}_n}(x) = \mathbb{E}_{f\leftarrow\tilde{\mathcal{F}}_n} f(x)$ be the common output on x for functions drawn from $\tilde{\mathcal{F}}_n$.

1. $\tilde{\mathcal{F}}_n$ is *concentrated* if there is a negligible function $\mathsf{negl}(\cdot)$ that
$$\max_{x\in\{0,1\}^{l(n)}} \Pr_{f\leftarrow\tilde{\mathcal{F}}_n}[f(x) \neq \mathsf{maj}_{\tilde{\mathcal{F}}_n}(x)] \leq \mathsf{negl}(n)$$
2. (Rephrasing definition 2.1) $\tilde{\mathcal{F}}_n$ is *evasive* if it is concentrated, and $\forall x \in \{0,1\}^{l(n)}$, $\mathsf{maj}_{\tilde{F}_n}(x) = 0$

Definition A.2 (Strong indistinguishability Obfuscator [BCKP14])**.** An obfuscator is a Strong indistinguishability Obfuscator (siO) for $\mathcal{F}$ if for any two concentrated distribution ensembles $\tilde{\mathcal{F}}^0_n$, $\tilde{\mathcal{F}}^1_n$ on $\mathcal{F}_n$ s.t. $\mathsf{maj}_{\tilde{\mathcal{F}}^0_n} \equiv \mathsf{maj}_{\tilde{\mathcal{F}}^1_n}$, and for any p.p.t. adversary A, there is a negligible function $\mathsf{negl}(\cdot)$:
$$\left|\Pr_{f_0\leftarrow\tilde{\mathcal{F}}^0_n}[A(\mathsf{siO}(f_0)) = 1] - \Pr_{f_1\leftarrow\tilde{\mathcal{F}}^1_n}[A(\mathsf{siO}(f_1)) = 1]\right| \leq \mathsf{negl}(n)$$

Definition A.3 (Virtual-Grey-Box Obfuscation [BC10])**.** Obf is a Virtual-Grey-Box (VGB) Obfuscator for $\mathcal{F}$ if for any feasible adversary A, there is a simulator S, and a negligible function $\mathsf{negl}(\cdot)$ such that for all $f \in \mathcal{F}$:
$$|\Pr[A(\mathsf{Obf}(f)) = 1] - \Pr[S^f(1^{|f|}) = 1]| \leq \mathsf{negl}(|f|)$$
where the running time of S is computationally unbounded, but only sends polynomially many queries to f (such a simulator is usually called "semi-bounded").

Theorem A.4 ([BCKP14])**.** An obfuscator is siO for $\mathcal{F}$ iff it is worst-case VGB obfuscator for $\mathcal{F}$.

Theorem A.5 (SiO implies IHO for evasive functions)**.** Let $\mathcal{F} = \{f_k : \{0,1\}^{l(n)} \to \{0,1\}\}_{n\in\mathbb{N}}$ be an evasive function ensemble, Obf be a strong iO for $\mathcal{F}$, then Obf is an input-hiding obfuscator for $\mathcal{F}$.

Proof sketch: $\tilde{\mathcal{F}}^0_n$ be the uniform distribution on $\mathcal{F}$, $\tilde{\mathcal{F}}^1_n$ be an arbitrary distribution on zero function, then $\mathsf{maj}_{\tilde{\mathcal{F}}^0_n} \equiv \mathsf{maj}_{\tilde{\mathcal{F}}^1_n} \equiv 0$. Therefore
$$\Pr_{f_0\leftarrow\tilde{\mathcal{F}}^0_n}[f_0(A(\mathsf{siO}(f_0), z)) = 1] \leq \Pr_{f_1\leftarrow\tilde{\mathcal{F}}^1_n}[f_1(A(\mathsf{siO}(f_1), z)) = 1] + \mathsf{negl}(n) = \mathsf{negl}(n)$$

□

B Correlation intractability versus other notions

We explore the relation between correlation intractability and other notions for cryptographic hash functions. Correlation intractability immediately implies other definitions which could be directly translated to "hiding a sparse relation". The notion of "entropy preserving hashing" is closely related to correlation intractability. In fact we show that they imply each other under certain conditions.

Some other related notions are not necessarily implied by or containing correlation intractability. As mentioned in the introduction, showing separations does not mean that they are fundamentally unrelated with correlation intractability. In fact one can easily patch the definitions to include both one and the other. The purpose of proving the separations is to try to demonstrate the weakness in each of the definitions alone.

B.1 Relations with entropy-preserving hashing

Recall the definition of *Entropy Preserving* (EP) from [BLV06]:

Definition B.1 (Entropy preservation)**.** A family of hash function $\mathcal{H} = \{h_k : \{0,1\}^{l(n)} \to \{0,1\}^{m(n)}, k = g(s), s \in \{0,1\}^{\sigma(n)}\}_{n\in\mathbb{N}}$ ensures conditional entropy[5] greater than $\delta(n)$ if for all (non-uniform, p.p.t.) adversary A:

$$H(h_k(A(k))|A(k)) > \delta(n)$$

Equivalently:

$$\mathbb{E}_{k,A}[H(h_k(X)|_{X=A(k)})] > \delta(n)$$

Notice that in order to get meaningful (i.e. non-zero) conditional entropy, the length of the key $\kappa(n)$ must be bigger then the length of the input $l(n)$, otherwise the adversary could always output the key (i.e. $A(k) \to k$) so that the conditional entropy will be zero (same to the diagonalization attack of correlation intractability [CGH04]). In other words, we hope that there are multiple choices of keys that could lead the adversary to return the same input, and $h_k(x)$ on these candidate seeds and fixed input has different values.

[BLV06] proposed 3 bounds for $\delta(n)$, each being interested on its own:

- (Best possible) $\delta(n) > m(n) - O(\log n)$. If achievable, would imply that constant-round public-coin auxiliary-input zero-knowledge proofs exist only for languages in BPP.
- (Somewhat) $\delta(n) > 1/\mathsf{poly}(n)$, also interesting, and imply that 3-round public-coin auxiliary-input optimally sound zero-knowledge proofs exist only for languages in BPP.
- (Minimum/Weakest) $\delta(n) > 0$, still interesting, and imply that the parallel composition of some classic protocols (e.g. Blum's protocol [Blu86]) are not auxiliary-input zero-knowledge.

An equivalent formalization of the minimum conjecture:

Conjecture B.2 ([BLV06])**.** There is a polynomial $p(\cdot)$ such that the following holds: For every *non-uniform deterministic* polynomial-time algorithm A and all sufficiently large n, there are circuits C_1, C_2 of size at most $p(n)$ such that $\alpha = A(C_1) = A(C_2)$ but $C_1(\alpha) \neq C_2(\alpha)$.

[5] The entropy of a random variable X is defined as $H(X) = \mathbb{E}_{x \xleftarrow{\$} X}[\log \frac{1}{\Pr[X=x]}]$. For jointly distributed random variables (X, Y), the conditional entropy of Y given X is defined to be $\mathbb{E}_{y \xleftarrow{\$} Y}[H(X|_{Y=y})]$, where $X|_{Y=y}$ denotes the conditional distribution of X given that $Y = y$.

Note that even the construction of the weakest notion of entropy-preservation is unknown. In fact it is shown by Bitansky et al. to be impossible to obtain from black-box reduction to falsifiable assumptions [BDSG+13].

The connections We show that entropy preservation and correlation intractability (where the sparse relations are not necessarily efficient recognizable) are equivalent. To focus on the concepts, we simplify the definition of correlation intractability by assuming that if the adversary exists, it breaks correlation intractability with probability 1. With such a simplification in mind, it is easy to connect the density of the sparse relations to the conditional entropy.

Theorem B.3 (Entropy preservation implies correlation intractability)**.** If a function family $\mathcal{H}$ guarantees the best-possible entropy-preserving, i.e. for all p.p.t. adversary A:

$$H(h_k(A(k))|A(k)) > m(n) - O(\log(n))$$

then it is correlation intractable.

Proof sketch: If $\mathcal{H}$ is not correlation intractable, in the sense that there's a sparse relation R, an adversary A that:

$$\Pr_k[x \leftarrow A(k) : (x, h_k(x)) \in R] = 1$$

Since R is sparse, which means for all x, the possible y values form a negligibly small subset of the range. Therefore the conditional entropy is:

$$H(h_k(A(k))|A(k)) < m(n) - \omega(\log(n))$$

which forms a contradiction. □

Theorem B.4 (Correlation intractability implies entropy preservation)**.** If a function family $\mathcal{H}$ is correlation intractable, then it is also entropy-preserving, i.e. for all p.p.t. adversary A:

$$H(h_k(A(k))|A(k)) > m(n) - O(\log(n))$$

Proof sketch: If it is not entropy-preserving, then there's an Adv A, such that

$$H(h_k(A(k))|A(k)) < m(n) - \omega(\log(n))$$

We define a relation by enumerating the keys, and query A on each key to get x, and the corresponding $y = h_k(x)$, then adding (x, y) into the relation. Formally, let R be:

$$R = \{(x, h_k(x)) \mid x = A(k),\ k = g(s),\ s \in \{0,1\}^{\sigma(n)}\}$$

R is sparse since the adversary can always break entropy-preservation, which means the portion of the possible outputs conditioned on the adversary's choice of the input is negligible. □

Notice that this relation is not likely to be efficiently recognizable, therefore our construction of CI w.r.t. efficiently recognizable relations is not necessarily entropy-preserving.

B.2 Separations between correlation intractability and other notions

Several random-oracle-like notions are defined in an "indistinguishability" fashion, which attempt to capture the intuition that, given only limited access to or partial information from the function, it is hard for the adversary to distinguish whether the information is obtained from the hash function or a truly random function. The notions defined in this way include but are not limited to seed-incompressibility[6] [HMR08] and universal computational extractor (UCE) [BHK13]. Since these notions are diverse, and each has different variants on its own, we refer the reader to the original papers for the details.

We show that the property of correlation intractability alone does not imply the "indistinguishability" definitions in general. The construction which demonstrates the separation is to append a fixed bit (say '1') after any correlation intractable functions. This construction is inspired by the ideas from section 4.4 of [BHK13], where UCE is separated from other notions including collision resistance.

Construction B.5. Let $\mathcal{H} = \{h_k : \{0,1\}^{l(n)} \to \{0,1\}^{m(n)-1}, k = g(s), s \in \{0,1\}^{\sigma(n)}\}_{n\in\mathbb{N}}$ be a correlation intractable function ensemble, we construct $\mathcal{H}'$ by padding an 1-bit at the end of the output:

$$h'_{k'}(x) = h_k(x)||1$$

Theorem B.6. $\mathcal{H}'$ is correlation intractable.

Proof sketch: If there is an attacker A', a sparse relation $R' : \{0,1\}^{l(n)+m(n)} \to \{0,1\}$, a non-negligible function $\eta(\cdot)$ such that

$$\Pr_{k'}[x \leftarrow A'(k') : R'(x, h'_{k'}(x)) = 1] > \eta(n)$$

then we build an adversary A and a sparse relation $R : \{0,1\}^{l(n)+m(n)-1} \to \{0,1\}$ against $\mathcal{H}$: the relation R is defined as

$$R = \{(x,y) \mid R'(x, y||1) = 1,\ x \in \{0,1\}^{l(n)},\ y \in \{0,1\}^{m(n)-1}\}$$

The density of R is at most twice as much as the density of R', so it is sparse. Given the key k, A construct $h'_{k'}$ by padding a bit '1' at the end of the output of h_k, send $h'_{k'}$ to A', follows the answer of A'. The probability that the output of A breaks R is exactly the probability that A' breaks R':

$$\Pr_{k}[x \leftarrow A(k) : R(x, h_k(x)) = 1] \geq \eta(n)$$

which contradicts to the assumption that $\mathcal{H}$ is correlation intractable. □

Note that this transformation works regardless of the efficiency of checking the relation.

[6] [HMR08] discussed both indistinguishability-style and correlation intractability-style definitions, when the adversary is only given partial information of the key (e.g. with an a priori bound on the length).