There is no Indistinguishability Obfuscation in Pessiland

Tal Moran* Alon Rosen†

October 7, 2013

Abstract

We show that if $\mathsf{NP} \neq \mathrm{co}-\mathsf{RP}$ then the existence of efficient indistinguishability obfuscation ($i\mathcal{O}$) implies the existence of one-way functions. Thus, if we live in "Pessiland", where NP problems are hard on the average but one-way functions do not exist, or even in "Heuristica", where NP problems are hard in the worst case but easy on average, then $i\mathcal{O}$ is impossible. Our result makes it redundant to explicitly assume the existence of one-way functions in most "cryptographically interesting" applications of $i\mathcal{O}$.

1 Introduction

Program obfuscation, the task of making code unintelligible while preserving its functionality, was first rigorously studied by Barak et al. [2]. In that work, they defined a notion of virtual black box (VBB) obfuscation, and proved that it is impossible to realize in general. In addition, they proposed a weaker notion of *indistinguishability obfuscation* ($i\mathcal{O}$), whose applicability was not clear at the time, but nevertheless avoided their impossibility results. VBB obfuscation requires that access to the obfuscated program gives no more power than access to an impenetrable black box with the same input-output functionality. $i\mathcal{O}$ is weaker in that it guarantees that for any two circuits C_0, C_1 of same size that compute the same function, it is hard to distinguish an obfuscation of C_0 from an obfuscation of C_1. Barak et al. showed that $i\mathcal{O}$ is always realizable, albeit inefficiently: the $i\mathcal{O}$ can simply canonicalize the input circuit C by outputting the lexicographically first circuit that computes the same function.

The interest in $i\mathcal{O}$ has gained considerable momentum following the works of Garg et al. [9] who proposed an *efficient* candidate construction of $i\mathcal{O}$ for all circuits (along with a candidate construction of functional encryption), and of Sahai and Waters [12], who demonstrated the wide applicability of $i\mathcal{O}$ for the construction of many powerful cryptographic primitives. The security of the Garg et al. construction is based on a specific family of intractability assumptions (different for any obfuscated function) closely related to multilinear maps [7, 6]. Being introduced only recently, these assumptions are still not well-understood, though several recent works have provided support to the security (in the even stronger VBB sense) of related constructions in idealized algebraic models [5, 4, 1].

$i\mathcal{O}$ is a weaker primitive than VBB obfuscation. In fact, it is not hard to see that we cannot even hope to prove that $i\mathcal{O}$ implies one-way functions: Indeed, if $\mathsf{P} = \mathsf{NP}$ then one-way functions do not exist but $i\mathcal{O}$ does exist (since the lexicographically first circuit that computes the same function can be found efficiently). Therefore, we do not expect to build many "cryptographically interesting" tools just from $i\mathcal{O}$, but usually need to combine it with other assumptions. (One exception is witness encryption [10], which can be constructed from $i\mathcal{O}$ alone.) It is known that $i\mathcal{O}$ can be combined with one-way functions (OWFs)

*Efi Arazi School of Computer Science, IDC Herzliya, Israel. Email: talm@idc.ac.il

†Efi Arazi School of Computer Science, IDC Herzliya, Israel. Email: alon.rosen@idc.ac.il. Supported by ISF grant no. 1255/12 and by the ERC under the EU's Seventh Framework Programme (FP/2007-2013) ERC Grant Agreement n. 307952..

to construct many powerful primitives such as public-key encryption, identity-based encryption, attribute-based encryption (via witness encryption), as well as NIZKs, CCA encryption, deniable encryption [12], and two message multi-party computation [8]. However, it is still not clear whether assuming one-way functions is actually necessary for any of the above applications.

In this short note, we observe that the existence of $i\mathcal{O}$, combined with the assumption that $\mathsf{NP} \neq \mathrm{co}-\mathsf{RP}$ (a worst-case hardness assumption) implies the existence of one-way functions (an average-case hardness assumption). To the best of our knowledge, this observation has not been previously made in the literature. Indeed, in all of the above "cryptographically interesting" applications of $i\mathcal{O}$, the existence of one-way functions is explicitly assumed.

Borrowing from Impagliazzo's terminology [11], we can summarize the state of affairs as follows. If $i\mathcal{O}$ is achievable, then Impagliazzo's five worlds collapse to two: we are either in *Algorithmica*, where $\mathsf{P} = \mathsf{NP}$, or in *Cryptomania*, where public-key encryption is possible. *Minicrypt*, where one-way functions exist but public-key encryption is not possible, is ruled out due to the existing constructions of public-key crypto from $i\mathcal{O}$ and one-way functions. Our new result rules out *Pessiland*, where $\mathsf{BPP} \neq \mathsf{NP}$, but one-way functions do not exist, and *Heuristica*, where NP problems are hard in the worst case but easy on average.

The idea behind our result is very simple. Given an indistinguishability obfuscation scheme $i\mathcal{O}(C; r)$ (that uses randomness r to obfuscate a circuit C), our candidate one-way function is defined as

$$f(x) = i\mathcal{O}(Z; x),$$

where Z is a circuit of appropriate size and input length that always outputs zero. Assuming that $i\mathcal{O}$ satisfies both functionality and indistinguishability, we show how to use an adversary A that can invert the function f with non-negligible advantage (over the choice of a random input x) in order to (one-sided) probabilistically decide the circuit (un)satisfiability of a given circuit C. This is done by simply observing whether A succeeds in inverting or not. The key observations in our argument are the following:

- If C is unsatisfiable, then it always outputs zero. Thus, by indistinguishability of the $i\mathcal{O}$ scheme A inverts f with non-negligible advantage even if we replace $f(x) = i\mathcal{O}(Z; x)$, with $f(x) = i\mathcal{O}(C; x)$.
- If C is satisfiable, then by functionality of the $i\mathcal{O}$ scheme, $i\mathcal{O}(C; x)$ cannot be a circuit that always outputs zero. Thus, A can *never* invert f when we replace $f(x) = i\mathcal{O}(Z; x)$, with $f(x) = i\mathcal{O}(C; x)$.

We note that our result makes strong use of the "perfect functionality" required by the definition of $i\mathcal{O}$ (see Def. 2.2). While this is indeed satisfied by the candidate constructions, it is an interesting question whether an approximate version of $i\mathcal{O}$ also implies one-way functions.

2 Definitions

We start by giving the definitions of one-way functions and of indistinguishability obfuscation. A function $\varepsilon(k)$ is said to be *negligible* if for all polynomial $p(k)$ and sufficiently large $k \in \mathbb{N}$ it holds that $\varepsilon(k) < 1/p(k)$.

Definition 2.1 (One-way function)**.** A family, $\mathcal{F} = \{f_k\}_{k\in\mathbb{N}}$, of efficiently computable functions $f_k : \{0,1\}^k \to \{0,1\}^*$ is said to be *one-way* if for all PPT adversaries A, there exists a negligible function $\varepsilon(k)$ such that for every security parameter $k \in \mathbb{N}$:

$$\Pr_{x,s}\left\{A(1^k, f_k(x), s) \in f_k^{-1}(f_k(x))\right\} < \varepsilon(k),$$

where x and s are uniformly chosen in their corresponding domains.

Indistinguishability obfuscation was introduced in [2] and given a candidate construction in [9], and subsequently in [4, 1, 5].

Definition 2.2 (Indistinguishability obfuscation [3])**.** A PPT algorithm $i\mathcal{O}(1^k, C; r)$ is said to be an *indistinguishability obfuscator* ($i\mathcal{O}$) for $\mathcal{C}$, if it satisfies:

1. **Functionality:** For any $C \in \mathcal{C}$,
$$\Pr_r \left\{ \forall x : i\mathcal{O}(1^k, C; r)(x) = C(x) \right\} = 1 \ .$$
2. **Indistinguishability:** For any (not necessarily uniform) PPT distinguisher D, there exists a negligible function $\varepsilon(k)$ such that the following holds: For all security parameters $k \in \mathbb{N}$, for all pairs of circuits $C_0, C_1 \in \mathcal{C}_k$, we have that if $C_0(x) = C_1(x)$ for all inputs x, then
$$\left| \Pr_{r,s} \left\{ D(i\mathcal{O}(1^k, C_0; r); s) = 1 \right\} - \Pr_{r,s} \left\{ D(i\mathcal{O}(1^k, C_1; r); s) = 1 \right\} \right| \leq \varepsilon(k)$$

3 From Indistinguishability Obfuscation to One-Way Functions

Let $i\mathcal{O}(1^k, C; r)$ be an Indistinguishability Obfuscator, where C is the input circuit and r the randomness used in obfuscation. Let $Z_{k;n}$ be a canonical constant zero circuit with inputs of k bits padded to n gates. For every $k \in \mathbb{N}$ define

$$f_k(x) \doteq i\mathcal{O}(1^k, Z_{k;n}; x),$$

and let $\mathcal{F} = \left\{ f_k : \{0,1\}^k \to \{0,1\}^* \right\}_{k \in \mathbb{N}}$ be the corresponding (efficiently computable) family of functions.

Theorem 3.1. *If* $\mathsf{NP} \neq \mathsf{co}-\mathsf{RP}$ *then* $\mathcal{F}$ *is a family of one-way functions.*

Proof. Suppose, in contradiction, that f is not one-way. Then there exists a PPT adversary A who can invert f_k (for all k) with probability $p(k)$ such that p is some inverse polynomial in k. Let $f = f_k$ and define

$$\delta(C, Z_{k,n}) \doteq \left| \Pr_{x,s} \left\{ A(1^k, i\mathcal{O}(C; x), s) \in f^{-1}(x) \right\} - \Pr_{x,s} \left\{ A(1^n, f(x), s) \in f^{-1}(x) \right\} \right|$$

This is the difference in the probability that A successfully finds a preimage for y with respect to f when it is given a random obfuscation of the circuit Z (i.e., a random element in the image of f) and when A is given a random obfuscation of the circuit C. Note that obfuscations of C might not even be in the image of f. However, the following claim asserts that this happens with negligible probability for Cs that implement the zero function:

Claim 3.2. *For all n, every PPT A and every circuit C' with n gates that implements the constant zero function for inputs of k bits, $\delta(C', Z_{k,n}) \leq \varepsilon(k)$, where ε is a negligible function.*

Proof. This follows immediately from the security of the obfuscation scheme. Since C' and Z have identical functionality and size, it must hold that for every PPT B:

$$\left| \Pr_{x,s} \left\{ B(1^k, i\mathcal{O}(C'; x), s) = 1 \right\} - \Pr_{x,s} \left\{ B(1^k, i\mathcal{O}(C'; x), s) = 1 \right\} \right| < \varepsilon(k) \ .$$

Taking $B(1^k, y, s)$ to be the algorithm that runs $x' \leftarrow A(1^k, y, s)$ and outputs 1 iff $f(x') = y$, we get a distinguisher for obfuscations of C' and Z with advantage δ. □

On the other hand, for any circuit that does not implement the constant zero function, there will *never* be a preimage under f.

Claim 3.3. *For all n, every PPT A and every circuit C' with n gates such that $\exists x : C'(x) \neq 0$, it holds that*

$$\Pr_{x,r}\left\{A(1^k, i\mathcal{O}(C'; x), s) \in f^{-1}(x)\right\} = 0 \,.$$

Proof. This claim follows immediately from the functionality property of the obfuscation algorithm: Since the output of the obfuscator is a circuit that has identical functionality to the input circuit, the output of $i\mathcal{O}(C'; x)$ cannot be a circuit that implements the constant zero function. Thus, it cannot be in the image of f. (Note that for this argument to hold, it is critical that the obfuscator *perfectly* preserve functionality). □

Given a Circuit-SAT instance C^* on k variables with n gates, we will now use A to (one-sided) probabilistically decide if C^* is satisfiable with inverse polynomial advantage:

1: **for** $i := 1$ to $t = 2k/p$ **do**
2: Choose x uniformly at random.
3: Compute $y \leftarrow i\mathcal{O}(C^*; x)$.
4: Run $x^* \leftarrow A(1^k, y, s)$ (s is chosen uniformly at random)
5: **if** $f(x^*) = y$ **then**
6: **return** "Unsatisfiable"
7: **end if**
8: **end for**
9: **return** "Satisfiable"

If C^* is unsatisfiable, then it implements the constant zero function. Hence, by Claim 3.2, it follows that the condition in line 5 will be true with probability p, independently in each iteration (since we choose x and s uniformly at random and independently in each iteration of the loop). The probability that it fails in all t iterations is $(1 - p)^t$ which is negligible. Thus, in this case the algorithm will return "Unsatisfiable" with all but negligible probability.

If C^* is satisfiable, by Claim 3.3 the inverting adversary will never succeed, hence with probability 1 the algorithm will output "Satisfiable". □

Acknowledgements. We thank Nir Bitansky for helpful suggestions.

References

[1] B. Barak, S. Garg, Y. Tauman-Kalai, O. Paneth, and A. Sahai. Protecting obfuscation against algebraic attacks. *IACR Cryptology ePrint Archive*, 2013:631, 2013.

[2] B. Barak, O. Goldreich, R. Impagliazzo, S. Rudich, A. Sahai, S. P. Vadhan, and K. Yang. On the (im)possibility of obfuscating programs. In *CRYPTO*, pages 1–18, 2001.

[3] B. Barak, O. Goldreich, R. Impagliazzo, S. Rudich, A. Sahai, S. P. Vadhan, and K. Yang. On the (im)possibility of obfuscating programs. *J. ACM*, 59(2):6, 2012.

[4] Z. Brakerski and G. Rothblum. Virtual black-box obfuscation for all circuits via generic graded encoding. *IACR Cryptology ePrint Archive*, 2013:563, 2013.

[5] R. Canetti and V. Vaikuntanathan. Obfuscating branching programs using black-box pseudo-free groups. *IACR Cryptology ePrint Archive*, 2013:500, 2013.

[6] J.-S. Coron, T. Lepoint, and M. Tibouchi. Practical multilinear maps over the integers. In *CRYPTO (1)*, pages 476–493, 2013.

[7] S. Garg, C. Gentry, and S. Halevi. Candidate multilinear maps from ideal lattices. In *EUROCRYPT*, pages 1–17, 2013.

[8] S. Garg, C. Gentry, S. Halevi, and M. Raykova. Two-round secure mpc from indistinguishability obfuscation. *IACR Cryptology ePrint Archive*, 2013:601, 2013.

[9] S. Garg, C. Gentry, S. Halevi, M. Raykova, A. Sahai, and B. Waters. Candidate indistinguishability obfuscation and functional encryption for all circuits. Cryptology ePrint Archive, Report 2013/451, 2013. `http://eprint.iacr.org/2013/451`.

[10] S. Garg, C. Gentry, A. Sahai, and B. Waters. Witness encryption and its applications. In *STOC*, pages 467–476, 2013.

[11] R. Impagliazzo. A personal view of average-case complexity. In *Structure in Complexity Theory Conference*, pages 134–147. IEEE Computer Society, 1995.

[12] A. Sahai and B. Waters. How to use indistinguishability obfuscation: Deniable encryption, and more. Cryptology ePrint Archive, Report 2013/454, 2013. `http://eprint.iacr.org/2013/454`.