A Note on the Impossibility of Obfuscation with Auxiliary Inputs

Shafi Goldwasser* Yael Tauman Kalai†

October 17, 2013

Abstract

In this note we revisit the problem of obfuscation with auxiliary inputs. We show that the existence of indistinguishablity obfuscation (iO) implies that all functions with sufficient "pseudo-entropy" cannot be obfuscated with respect to a virtual box definition (VBB) in the presence of (dependent) auxiliary input. Namely, we show that for any candidate obfuscation $\mathcal{O}$ and for any function family $\mathcal{F} = \{f_s\}$ with sufficient pseudo-entropy, there exists an (efficiently computable) auxiliary input aux, that demonstrates the insecurity of $\mathcal{O}$. This is true in a strong sense: given $\mathcal{O}(f_s)$ and aux one can efficiently recover the seed s, whereas given aux and oracle access to f_s it is computationally hard to recover s.

A similar observation was pointed out in a recent work of Goldwasser *et. al.* (Crypto 2013), assuming *extractable* witness encryption. In this note we show that the extractability property of the witness encryption is not needed to get our negative result, and all that is needed is the existence of witness encryption, which in turn can be constructed from iO obfuscation.

*MIT and the Weizmann Institute.

†Microsoft Research.

1 Introduction

The study of *Program Obfuscation* — a method that transforms a program (say described as a Boolean circuit) into a form that is executable, but otherwise completely unintelligible — has been a longstanding research direction in cryptography. It was formalized by Barak *et. al.* [BGI$^+$01], who showed that there exist (contrived) function families which are not obfuscatable under a very natural definition of obfuscation (VBB) and various relaxations. The virtual-black-box (VBB) definition essentially requires that anything that can be efficiently computable given an obfuscation of a program, could be efficiently computable from black box access to the program.

Following this work, much effort has been devoted to show the existence of obfuscators for natural classes of programs. However until the recent proposed construction result of Garg *et. al.* [GGH$^+$13], all known obfuscation candidates were for very simple classes of functions, such as point functions [Can97], hyper-planes [CRV10], conjunctions [BR13b], and d-CNFs [BR13a]. The recent breakthrough work of Garg *et. al.* gave the first candidate for general-purpose obfuscation, and conjectured that it is a *best possible* [GR07] or *indistinguishability* obfuscator (iO-obf). Namely, given any two circuits C_1, C_2 of the same size for a functionality f, no polynomial time adversary can distinguish between the obfuscation of C_1 and the obfuscation of C_2. Followup works [BR13c, BGK$^+$13] proved security of (variants) of their scheme in the generic multi-linear group model. Subsequent to Garge's work, a flood of results have appeared showing that the existence of iO-obf suffices for many applications, previously considered outcome of VBB applications, such as the constructions of public-key encryption from private-key encryption, the existence of deniable encryption, and much more.

In this note, in contrast, we show that the existence of iO-obf shows, in a somewhat strange twist, limitations on the possibility of VBB obfuscation. In particular, on the possibility of VBB obfuscation with auxiliary input. The latter is a strengthening of VBB obfuscation, introduced by Goldwasser and Kalai [GK05], which corresponds to the setting where the adversary, which is given the obfuscated circuit, may have some additional a priori information. This is essentially the case of interest in any cryptographic usage of obfuscation imaginable. Goldwasser and Kalai prove the existence of many "natural" classes of functions that are not obfuscatable w.r.t. (contrived) auxiliary input. In particular, two types of auxiliary inputs were considered in [GK05]: *dependent* auxiliary inputs, where the auxiliary input may depend on the function being obfuscated, and *independent* auxiliary inputs, which is independent of the function being obfuscated. For the case of dependent auxiliary inputs, they proved that every function family with super-polynomial "pseudo-entropy" cannot be obfuscated w.r.t. auxiliary inputs, *assuming the class of point-filer functions are obfuscatable w.r.t. auxiliary inputs*.[1] Thus, their result is a conditional one.

Recently, the notion of witness encryption was put forth by Garg *et. al.* [GGSW13]. It was observed by Goldwasser *et. al.* [GKP$^+$13] that an extractable version of witness encryption can be used to obfuscate the class of point-filer functions w.r.t. auxiliary inputs. Thus, together with [GK05], this implies that the existence of an extractable witness encryption scheme implies that *any* function with super-polynomial pseudo-entropy cannot be obfuscated w.r.t. auxiliary inputs.

Here we show that the proof of [GK05] actually implies that witness encryption *without* the extractability property, suffices in order to prove that all functions with super-polynomial pseudo-entropy are not obfuscatable w.r.t. auxiliary inputs. We note that the (general-purpose) iO-obf, of Garg *et. al.* [GGH$^+$13], implies the existence of witness encryption. Thus, our observation implies that iO-obf implies that all functions with super-polynomial pseudo-entropy are not obfuscatable w.r.t. auxiliary inputs.

We refer the reader to Definition 2.3 for the precise definition of circuit families with super-polynomial pseudo entropy, but note that such families include all pseudo-random function families, as well as every semantically secure secret-key and public key encryption scheme, or any secure digital signature scheme where randomness is generated by using a (secret) pseudo-random function.

[1] We refer the reader to [GK05] for the definition of a point-filter function family.

We emphasize that in this note, we consider auxiliary inputs that may depend on the function being obfuscated, but as in [GK05], we consider only auxiliary inputs that are efficiently computable and do not depend on the randomness used by the obfuscator.

2 Preliminaries

Definition 2.1 (VBB Obfuscation with auxiliary inputs). *Let $\mathcal{F} = \{f_s\}$ be a family of poly-size circuits. We say that $\mathcal{O}$ is an obfuscation of $\mathcal{F}$ with (dependent) auxiliary inputs if the following holds:*

- ***Correctness:*** *For every function $f_s \in \mathcal{F}$, and every possibly input x,*

$$\Pr[\mathcal{O}(f_s)(x) = f_s(x)] = 1.$$

- ***Polynomial slowdown:*** *There exists a polynomial p such that $|\mathcal{O}(f_s)| \leq p(|f_s|)$.*

- ***Security with auxiliary input:*** *For every PPT $\mathcal{A}$ there exists a PPT S such that for every auxiliary input $\mathsf{aux} = \mathsf{aux}(\mathsf{s})$ (that is efficiently computable from s) and every predicate π*

$$\Pr[\mathcal{A}(\mathcal{O}(f_s), \mathsf{aux}) = \pi(\mathsf{s}, \mathsf{aux})] - \Pr[\mathsf{S}^{\mathsf{f_s}}(\mathsf{aux}(\mathsf{s})) = \pi(\mathsf{s}, \mathsf{aux})] = \mathrm{negl}(\mathsf{k}),$$

where the probability is over randomly chosen $s \in_R \{0,1\}^k$.

Remark. Our impossibility result for VBB obfuscation with auxiliary inputs, holds even if we restrict the auxiliary input to be efficiently computable given *oracle access* to f_s.

Definition 2.2 (Witness encryption). *A witness encryption scheme for an NP language $\mathcal{L}$ with corresponding witness relation $\mathcal{R}_{\mathcal{L}}$ is a pair of PPT algorithms $(\mathsf{Enc}, \mathsf{Dec})$ such that the following holds.*

- ***Correctness:*** *For all $(x, w) \in \mathcal{R}_{\mathcal{L}}$, for every $b \in \{0,1\}$,*

$$\Pr[\mathsf{Dec}(\mathsf{Enc}_{\mathsf{x}}(1^{\mathsf{k}}, \mathsf{b}), \mathsf{w}) = \mathsf{b}] = 1 - \mathrm{negl}(\mathsf{k}).$$

- ***Semantic Security:*** *For every $x \notin \mathcal{L}$ and for every PPT adversary $\mathcal{A}$,*

$$\Pr[\mathcal{A}(\mathsf{Enc}_{\mathsf{x}}(1^{\mathsf{k}}, \mathsf{b})) = \mathsf{b}] \leq \frac{1}{2} + \mathrm{negl}(\mathsf{k}),$$

where the probability is over $b \in_R \{0,1\}$ and over the random coin tosses of Enc and $\mathcal{A}$.

Definition 2.3 (Pseudo-entropy of a circuit class). *Let $p = p(k)$ be a polynomial. We say that a class of circuits $\mathcal{C} = \{\mathcal{C}_k\}_{k \in \mathbb{N}}$ has pseudo-entropy at least $p = p(k)$, if there exists a polynomial $t = t(k)$ and a subset $I \subseteq \{0,1\}^k$ of size $t(k)$, and for every $C \in C_k$ there exists a random variable $Y^C = (Y_1, \ldots, Y_t)$, such that the following holds:*

1. *Y^C has statistical min-entropy at least $p(k)$.*
2. *For every PPT oracle machine $\mathcal{D}$ there is a negligible function μ such that for every $k \in \mathbb{N}$*

$$\left| \Pr\left[\mathcal{D}^{C|_{\bar{I}}}(Y^C) = 1\right] - \Pr\left[\mathcal{D}^{C|_{\bar{I}}}(C(I)) = 1\right] \right| \leq \mu(k),$$

where $C(I) \triangleq \{C(x)\}_{x \in I}$,[2] and where the circuit $C|_{\bar{I}}$ agrees with C on every $x \notin I$, and outputs $\perp$ on every $x \in I$. The probability above is over $C \leftarrow \mathcal{C}_k$, over the random variable Y^C, and over the randomness of the distinguisher $\mathcal{D}$.

We say that $\mathcal{C}$ has super-polynomial pseudo-entropy if it has entropy at least p for every polynomial p.

[2] There is a slight abuse of notations here. We use $C(I)$ to denote both a set and a list (or a vector).

3 Impossibility for Obfuscation with Auxiliary Inputs

As was mentioned in the introduction, Goldwasser and Kalai [GK05] proved that either point-filter functions are not obfuscatable with auxiliary inputs or *all* functions with sufficient "pseudo-entropy" are not obfuscatable with auxiliary inputs. It was recently observed by Goldwasser *et. al.* [GKP+13] that extractable witness encryption implies that point-filter functions are obfuscatable with auxiliary inputs, and thus, that any function with sufficient "pseudo-entropy" is not obfuscatable with auxiliary inputs. We now show that the same impossibility result (with essentially the same proof as in [GK05]) can be obtained assuming the existence of witness encryption (without any extractability property).

Theorem 3.1. *Assume the existence of a witness encryption scheme. Then any function with super-polynomial pseudo-entropy cannot be obfuscated w.r.t. auxiliary input.*

In what follows, for the sake of simplicity, let us prove Theorem 3.1 for any pseudo-random function family. Then we show how the proof extends to any function with super-polynomial pseudo-entropy.

Proof. Assume the existence of a witness encryption scheme for some NP-complete language $\mathcal{L}$. Let $\mathcal{F} = \{f_s\}$ be any family of pseudo-random functions.

Suppose for the sake of contradiction that there exists an obfuscator $\mathcal{O}$ for $\mathcal{F}$ that takes as input $s \in \{0,1\}^k$ and outputs an obfuscated circuit $\mathcal{O}(f_s)$ of size $t = t(k)$ (for some polynomial t). Let $\mathcal{L}'$ be the NP language defined as follows:

$$\mathcal{L}' = \{x = (z_1, \ldots, z_{2t}) : \text{there exists a circuit } C \text{ of size } |C| \leq t \text{ s.t. } C(i) = z_i \; \forall i \in [2t]\}$$

Set $x = (f_s(1), \ldots, f_s(2t))$ and let $\mathsf{aux(s)} = \mathsf{Enc_x}(1^\mathsf{k}, \mathsf{b})$, where $b \leftarrow \{0,1\}$ is a random bit, and where Enc is a witness encryption for $\mathcal{L}'$. Note that the fact that there is a witness encryption for an NP-complete language implies that there is a witness encryption for *any* NP language, and in particular for $\mathcal{L}'$.

Given $\mathcal{O}(f_s)$ and $\mathsf{aux(s)} = \mathsf{Enc_x}(1^\mathsf{k}, \mathsf{b})$, one can efficiently decrypt b with probability $1 - \mathrm{negl}(k)$, since $\mathcal{O}(f_s)$ is a valid witness of x.

On the other hand, we prove the following claim, which contradicts the security of $\mathcal{O}$.

Claim 3.1. *For any PPT adversary S which takes as input $\mathsf{aux(s)} = \mathsf{Enc_x}(1^\mathsf{k}, \mathsf{b})$, and has black-box access to f_s,*

$$\Pr[S^{f_s}(\mathsf{Enc_x}(1^\mathsf{k}, \mathsf{b})) = \mathsf{b}] \leq \frac{1}{2} + \mathrm{negl}(\mathsf{k}).$$

Proof of Claim 3.1 Suppose for the sake of contradiction that there exists a PPT adversary S such that

$$\Pr[S^{f_s}(\mathsf{Enc_x}(1^\mathsf{k}, \mathsf{b})) = \mathsf{b}] \geq \frac{1}{2} + \epsilon(\mathsf{k}),$$

for some non-negligible function ϵ, where the probability is over random $(s, b) \leftarrow \{0,1\}^{k+1}$ and over the randomness of Enc.

The fact that f_s is a pseudo-random function implies that

$$\Pr[S^{R}(\mathsf{Enc_{x^*}}(1^\mathsf{k}, \mathsf{b})) = \mathsf{b}] \geq \frac{1}{2} + \frac{\epsilon(\mathsf{k})}{2}, \tag{1}$$

where R is a truly random function, and $x^* = (R(1), \ldots, R(2t))$. This is the case, since otherwise the PPT adversary S can be used to distinguish R from f_s, contradicting the pseudo-randomness of f_s.

Note that $x^* \notin \mathcal{L}'$ and therefore Equation (1) contradicts the semantic security of the underlying witness-encryption scheme. □

□

3.1 Extending the proof of Theorem 3.1 to any function family with super-polynomial pseudo-entropy.

Proof. Let $\mathcal{C}$ be any circuit class of polynomial size with super-polynomial pseudo-entropy. Suppose for the sake of contradiction that $\mathcal{C}$ has an obfuscator with auxiliary inputs, denoted by $\mathcal{O}$. Let $p = p(k)$ be a polynomial such that for every $C_k \in \mathcal{C}_k$, it holds that $|\mathcal{O}(C_k)| \leq p(k)$.

The fact that $\mathcal{C}$ has super-polynomial pseudo-entropy implies that it has pseudo-entropy at least $2p(k)$. In particular, (recalling Definition 2.3) this implies that there exists a polynomial $t = t(k)$ and a subset $I \subseteq \{0,1\}^k$ of size $t(k)$ such that for every C there exists a random variable $Y^C = (Y_1, \ldots, Y_t)$ such that

1. Y^C has statistical min-entropy at least $2p(k)$.
2. For every PPT oracle machine $\mathcal{D}$ there is a negligible function μ such that for every $k \in \mathbb{N}$

$$\left|\Pr[\mathcal{D}^{C|_{\bar{I}}}(Y^C) = 1] - r[\mathcal{D}^{C|_{\bar{I}}}(C(I)) = 1]\right| \leq \mu(k),$$

where $C(I) \triangleq \{C(x)\}_{x \in I}$, and where the circuit $C|_{\bar{I}}$ agrees with C on every $x \notin I$, and outputs $\perp$ on every $x \in I$. The probability above is over $C \leftarrow \mathcal{C}_k$, over the random variable Y^C, and over the randomness of the distinguisher $\mathcal{D}$.

We define an NP language $\mathcal{L}'$ similarly to above.

$$\mathcal{L}' = \{(z_1, \ldots, z_t) : \text{there exists a circuit } C \text{ of size } |C| \leq p \text{ s.t. } C(i) = z_i \; \forall i \in [t]\}$$

Set $x = C(I) = (C(x))_{x \in I}$ and let $\mathsf{aux}(\mathsf{C}) = \mathsf{Enc}_{\mathsf{x}}(1^{\mathsf{k}}, \mathsf{b})$, where $b \leftarrow \{0,1\}$ is a random bit, and where Enc is a witness encryption for the language $\mathcal{L}'$.

Note that given $\mathcal{O}(C)$ and $\mathsf{aux}(\mathsf{C}) = \mathsf{Enc}_{\mathsf{x}}(1^{\mathsf{k}}, \mathsf{b})$, one can efficiently decrypt b with probability $1 - \mathrm{negl}(k)$, since $\mathcal{O}(C)$ is a valid witness of x. It remains to prove the following claim, which is analogous to Claim 3.1.

Claim 3.2. *For any PPT adversary S which takes as input $\mathsf{aux}(\mathsf{s}) = \mathsf{Enc}_{\mathsf{x}}(1^{\mathsf{k}}, \mathsf{b})$, and has black-box access to C,*

$$\Pr[S^C(\mathsf{Enc}_{\mathsf{x}}(1^{\mathsf{k}}, \mathsf{b})) = \mathsf{b}] \leq \frac{1}{2} + \mathrm{negl}(\mathsf{k}).$$

Proof of Claim 3.2.

Suppose for the sake of contradiction that there exists a PPT adversary S such that

$$\Pr[S^C(\mathsf{Enc}_{\mathsf{x}}(1^{\mathsf{k}}, \mathsf{b})) = \mathsf{b}] \geq \frac{1}{2} + \epsilon(\mathsf{k}),$$

for some non-negligible function ϵ, where the probability is over random $C \leftarrow \mathcal{C}_k$ and over the randomness of Enc.

By the definition of x, this implies that there exists a PPT oracle machine S_2 such that

$$\Pr[S_2^{C|_{\bar{I}}}(\mathsf{Enc}_{\mathsf{x}}(1^{\mathsf{k}}, \mathsf{b})) = \mathsf{b}] \geq \frac{1}{2} + \epsilon(\mathsf{k}).$$

The fact that $\mathcal{C}$ has super-polynomial pseudo-entropy (see Definition 2.3) implies that

$$\Pr[S_2^{C|_{\bar{I}}}(\mathsf{Enc}_{\mathsf{x}^*}(1^{\mathsf{k}}, \mathsf{b})) = \mathsf{b}] \geq \frac{1}{2} + \frac{\epsilon(\mathsf{k})}{2}, \tag{2}$$

where $x^* = Y^C$.

Note however that x^* has min-entropy $2p(k)$ and thus is not in $\mathcal{L}'$. Thus, Equation (2) contradicts the semantic security of the underlying witness-encryption scheme. □

□

References

[BGI+01] Boaz Barak, Oded Goldreich, Russell Impagliazzo, Steven Rudich, Amit Sahai, Salil P. Vadhan, and Ke Yang. On the (im)possibility of obfuscating programs. In *CRYPTO*, pages 1–18, 2001.

[BGK+13] Boaz Barak, Sanjam Garg, Yael Tauman Kalai, Omer Paneth, and Amit Sahai. Protecting obfuscation against algebraic attacks. Cryptology ePrint Archive, Report 2013/631, 2013. `http://eprint.iacr.org/`.

[BR13a] Zvika Brakerski and Guy N. Rothblum. Black-box obfuscation for d-cnfs. Cryptology ePrint Archive, Report 2013/557, 2013. `http://eprint.iacr.org/`.

[BR13b] Zvika Brakerski and Guy N. Rothblum. Obfuscating conjunctions. In *CRYPTO*, pages 416–434, 2013.

[BR13c] Zvika Brakerski and Guy N. Rothblum. Virtual black-box obfuscation for all circuits via generic graded encoding. Cryptology ePrint Archive, Report 2013/563, 2013. `http://eprint.iacr.org/`.

[Can97] Ran Canetti. Towards realizing random oracles: Hash functions that hide all partial information. In *CRYPTO*, pages 455–469, 1997.

[CRV10] Ran Canetti, Guy N. Rothblum, and Mayank Varia. Obfuscation of hyperplane membership. In *TCC*, pages 72–89, 2010.

[GGH+13] Sanjam Garg, Craig Gentry, Shai Halevi, Mariana Raykova, Amit Sahai, and Brent Waters. Candidate indistinguishability obfuscation and functional encryption for all circuits. Cryptology ePrint Archive, Report 2013/451, 2013. `http://eprint.iacr.org/`.

[GGSW13] Sanjam Garg, Craig Gentry, Amit Sahai, and Brent Waters. Witness encryption and its applications. In *STOC*, 2013.

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

[GKP+13] Shafi Goldwasser, Yael Kalai, Raluca Ada Popa, Vinod Vaikuntanathan, and Nickolai Zeldovich. Succinct functional encryption and applications: Reusable garbled circuits and beyond. In *STOC*, 2013.

[GR07] Shafi Goldwasser and Guy N. Rothblum. On best-possible obfuscation. In *TCC*, pages 194–213, 2007.