A Note on Black-Box Complexity of Indistinguishability Obfuscation

Mohammad Mahmoody[*] Ameer Mohammed[†]
Soheil Nematihaji[‡] Rafael Pass[§] abhi shelat[¶]

March 19, 2016

Abstract

Mahmoody et al. (TCC 2016-A) showed that basing indistinguishability obfuscation (IO) on a wide range of primitives in a semi-black-box way is *as hard as* basing public-key cryptography on one-way functions. The list included any primitive $\mathcal{P}$ that can be realized relative to random trapdoor permutations or degree-$O(1)$ graded encoding model for any finite ring secure against computationally unbounded polynomial-query attackers.

In this note, we rely on the recent result of Brakerski, Brzuska, and Fleischhacker (ePrint 2016/226) and rule out *fully* black-box constructions of IO from any such primitive P, assuming the existence of one-way functions and $\mathbf{NP} \not\subseteq \mathbf{coAM}$.

At a technical level, we show that attacks in idealized (randomized) oracle models that succeed with *constant* advantage over the trivial bound (e.g., guessing the obfuscated circuit with probability $1/2 + 1/10$) would remain successful for an infinite sequence of security parameters for a *non-zero* measure of the oracles.

Keywords: Indistinguishability Obfuscation, Black-Box Separations.

1 Introduction

Let $\mathcal{P}$ be any primitive that can be realized relative to the random trapdoor permutation oracle or the degree-$O(1)$ graded encoding model in a way that is secure against polynomial-query attackers. Examples of $\mathcal{P}$ include CCA-secure public-key encryption [NY90, BR93], hierarchical identity based encryption [GS02, HL02], non-interactive zero-knowledge proofs for **NP** [BDSMP91, BY96, Gol11],

[*] University of Virginia, `mohammad@cs.virginia.edu`. Supported by NSF CAREER award CCF-1350939. The work was done in part while the author was visiting the Simons Institute for the Theory of Computing, supported by the Simons Foundation and by the DIMACS-Simons Collaboration in Cryptography through NSF grant CNS-1523467.

[†] University of Virginia, `am8zv@virginia.edu`. Supported by University of Kuwait.

[‡] University of Virginia, `sn8fb@virginia.edu`. Supported by NSF award CCF-1350939.

[§] Cornell University, `rafael@cs.cornell.edu`. Work supported in part by a Microsoft Faculty Fellowship, Google Faculty Award, NSF Award CNS-1217821, NSF Award CCF-1214844, AFOSR Award FA9550-15-1-0262 and DARPA and AFRL under contract FA8750-11-2-0211.

[¶] University of Virginia, `shelat@cs.virginia.edu`. Work supported by NSF CAREER Award 0845811, NSF TC Awards 1111781 and 0939718, DARPA and AFRL under contract FA8750-11-C-0080, Microsoft New Faculty Fellowship, SAIC Scholars Research Award, and Google Research Award.

etc. This short note rules out fully-black-box constructions [RTV04] of indistinguishability obfuscators (IO) [BGI$^+$01, GGH$^+$13] from any such primitive $\mathcal{P}$ assuming the mildest of complexity assumptions, namely existence of one-way functions and that $\mathbf{NP} \not\subseteq \mathbf{coAM}$.[1]

Previously, Mahmoody et al. [MMN$^+$16] showed the weaker separation result that realizing a black-box construction of IO from $\mathcal{P}$ is as hard as solving the notoriously difficult problem of constructing public-key cryptography from one-way functions.[2] Their result (which relies on the work of [CKP15, Pas15, MMN15]) starts from a construction of IO in an idealized model $\mathcal{I}$, realizing the primitive $\mathcal{P}$, and shows how to "compile-out" the use of the oracle $\mathcal{I}$ (and consequently the primitive $\mathcal{P}$) to produce an *approximate*, i.e., sometimes incorrect, albeit *statistically secure* IO. Subsequently, this small correctness error can be eliminated in the cases of primitives such as public-key encryption [DNR04, Hol06].

Whereas Goldwasser and Rothblum [GR07, GR14] already rule out the possibility of statistically-secure IO under $\mathbf{NP} \not\subseteq \mathbf{coAM}$, at the time of the result of [MMN$^+$16] nothing more was known about the implications of statistically-secure *approximate* IO. Recently, Brakerski, Brzuska, and Fleischhacker [BBF16] provide an elegant argument that resolves this gap—in fact, they show a stronger result by ruling out *δ-statistically secure ε-approximately correct* IO (based on standard assumptions) as long as $2\varepsilon + 3\delta < 1$.[3] Using [BBF16], we can extend the result of [MMN$^+$16] to rule out fully-black-box constructions of IO from $\mathcal{P}$ based on the same assumptions used in [BBF16].

Theorem 1.1 (Main–Informal)**.** *Assuming the existence of one-way functions and* $\mathbf{NP} \not\subseteq \mathbf{coAM}$*, there is no fully-black-box construction of IO from any primitive that can be constructed relative to random trapdoor permutation oracle or degree-$O(1)$ graded encoding model for any finite ring.*

At first glance, such a result would seem to follow obviously from [MMN$^+$16] and [BBF16] as given a putative construction Q of IO from $\mathcal{P}$, one can apply the "compiling-out" procedure from [MMN$^+$16] to produce an ε-approximate statistical IO scheme which cannot exist by the result of [BBF16]. However, a closer look at this argument reveals some subtle challenges. More formally, the above argument consists of two steps. Step one is a composition theorem: assuming there is a fully-black-box construction Q of IO from $\mathcal{P}$ and assuming the existence of $\mathcal{P}$ relative to $\mathcal{I}$, we need to show that there is a black-box construction of approximate IO relative to $\mathcal{I}$. Then, the second step would be to rule out the existence of black-box constructions of approximate IO relative to the ideal model $\mathcal{I}$ (which follows from the works of [MMN$^+$16, BBF16]).

Weak vs. strong constructions in idealized models. We first note that the composition theorem needed in the first step above does not hold for the "typical" proofs of security in idealized models. Such proofs of security in idealized models, which we refer to as *weak* constructions, bound the success probability of the adversary by $\text{negl}(n)$ where this probability is also over the randomness of the oracle. To get the desired composition theorem, we would need to fix the oracle $\mathcal{I}$ and be secure (with high probability) with respect to the *fixed* oracle $\mathcal{I}$. Here we refer to such constructions as *strong* constructions in idealized models. As observed by [MMN$^+$16] the desired composition theorem does indeed hold with respect to strong constructions in idealized models (see Lemma 3.5). Moreover, the primitives of interest to our separation from IO (e.g., CCA-secure encryption, HIBE, etc.) all have *strong* black-box constructions in their corresponding idealized

[1] Note that $\mathbf{NP} \subseteq \mathbf{coAM}$ would imply that the polynomial-time hierarchy collapses.

[2] However, the result of [MMN$^+$16] is stronger as it also applies to *semi*-black-box constructions [RTV04] of IO.

[3] Surprisingly, they show that the primitive does exist when $2\varepsilon + \delta > 1$.

models. However, in order to get the desired separation we would also need to rule out *strong* black-box constructions of IO in certain idealized models.

Ruling out strong constructions of IO. Weak constructions of IO in such models are already ruled out by [MMN$^+$16, BBF16]. Thus all we need to show is that such result (i.e., ruling out weak constructions) also rules out strong constructions of IO in idealized models. Unfortunately, as it turns out, these two notions are not comparable in general (see Remark 2.5). At a technical level, in this note we show that ruling out weak constructions in idealized models also rules out the existence of strong constructions so long as the attacker succeeds with $\Omega(1)$ advantage over the trivial bound (as opposed to just the seemingly sufficient $1/\operatorname{poly}$ advantage). Fortunately the result of [BBF16] shows how to break approximate statistical IO with a *constant* advantage over $1/2$, which together with our reduction between the strong and weak constructions in idealized models would finish the proof. To prove our reduction, we show that any attacker that succeeds with constant probability over an infinite number of security parameters remains a successful attacker (over an infinite number of security parameters) with at least a *constant measure* over the choice of the full oracle $\mathcal{I}$.

An alternative to Borel-Cantelli lemma. The standard technique in seemingly similar situations in black-box separation results is to use the Borel-Cantelli lemma (Lemma 2.8), but that lemma does not apply in our case since the attacker's success probability is *not* sufficiently close to one. Rather, we show that an alternative lemma (see Lemma 2.9) could be used as an alternative tool to the Borel-Cantelli lemma in scenarios in which an infinite sequence of "good" events each happen with a constant probability (as opposed to $\approx 1 - 1/n^2$) to show that with constraint measure over the probability space an infinite number of good events hold simultaneously.

2 Preliminaries

2.1 Definitions

Definition 2.1 (Fully black-box constructions [RTV04])**.** A *fully-black-box* construction of a primitive $\mathcal{Q}$ from a primitive $\mathcal{P}$ consists of two PPT algorithms (Q, S) as follows:

- **Implementation:** if oracle P implements $\mathcal{P}$, then Q^P implements $\mathcal{Q}$.
- **Security reduction:** for any oracle P implementing $\mathcal{P}$ and for any (computationally unbounded) oracle adversary A breaking the security of Q^P, $S^{P,A}$ breaks the security of P.

Reingold, Trevisan and Vadhan [RTV04] also defined other (more relaxed) notions of black-box constructions, and Baecher, Brzuska, and Fischlin [BBF13] further studied those notions in more details. We refer the readers to [RTV04, BBF13] for those extensions. We will, however, assume one general property about the primitives that we deal with in this work: function P implementing $\mathcal{P}$ will be partitioned into sub-domains indexed by "security parameter" n and any adversary A who successfully breaks P would have to "win" over an infinite number of security parameters for a "noticeable" advantage.

We skip defining IO and approximate IO and directly define the generalized notion of approximate computational IO. We first recall a statistical variant of this notion defined by [BBF16].

Definition 2.2 ([BBF16] Approximate Statistical Correlation IO)**.** A PPT O is an (ε, δ)-*approximate statistical computational IO* (CIO for short) if:

- **Approximate correctness:** $\Pr[O(C)(x) \neq C(x)] \leq \varepsilon(|C|)$ where the probability is over the randomness of the obfuscator and the input x.
- **Statistical correlation:** For every pair of circuits $C_1 \equiv C_2$ of the same size n, the statistical distance between $O(C_1)$ and $O(C_2)$ (both defined over the randomness of O) is at most $\delta(n)$.

A computational variant of Definition 2.2 can be defined analogously:

Definition 2.3 (Approximate Computational Correlation IO)**.** A PPT O is an (ε, δ)-approximate *computational* CIO if it satisfies the same correctness condition as approximate statistical CIO and:

- **Computational correlation:** For every poly-time adversary A and for every pair of circuits $C_1 \equiv C_2$ of equal size n, it holds that $\Pr[A(O(C_1)) = 1] - \Pr[A(O(C_2))] \leq \delta(n)$.

Fully-black-box constructions of IO. A fully-black-box construction of approximate computational CIO from primitive $\mathcal{P}$ could be defined through a combination of Definitions 2.1 and 2.3. Here we emphasize that the input circuits do not have any oracle gates while the obfuscation algorithm and the final circuits could use the oracle implementing $\mathcal{P}$. This seemingly restricted model is in fact sufficient for all known applications (see [MMN+16] for more discussions).

Idealized Models. An idealized model $\mathcal{I}$ is a randomized oracle; examples include the random oracle, random trapdoor permutation oracle, generic group model, graded encoding model, etc. An $I \leftarrow \mathcal{I}$ can (usually) be represented as a sequence $(I_1, I_2, \dots)$ where I_n is the part of I that is defined for "security parameter" n. The distribution over the infinite object $I \leftarrow \mathcal{I}$ could naturally be defined through finite distributions $\mathcal{D}_i$ over the finite space of I_i. Caratheodory's extension theorem shows that such finite probability distributions could always be extended consistently to a measure space over the full infinite space of $I \leftarrow \mathcal{I}$ (see Theorem 4.6 of [Hol15] for a proof).

Definition 2.4 (Strong Black-Box Constructions in Idealized Models [MMN+16])**.** We say a primitive $\mathcal{P}$ has a strong black-box construction in the idealized model $\mathcal{I}$ if there is an oracle-aided algorithm P such that:

- **Completeness:** P^I implements $\mathcal{P}$ correctly for every $I \leftarrow \mathcal{I}$.
- **Black-box security:** Let A be an oracle-aided adversary $A^{\mathcal{I}}$ where the *query complexity* of A is bounded by the specified complexity of the attacks for primitive $\mathcal{P}$. For example if $\mathcal{P}$ is polynomially secure (resp., quasi-polynomially secure), then A only asks a polynomial (resp., quasi-polynomial) number of queries but is computationally unbounded otherwise. Then, for any such A, with measure one over the choice of $I \overset{\$}{\leftarrow} \mathcal{I}$, it holds that A does *not* break P^I.

Remark 2.5 (Weak vs. Strong Constructions)**.** We called the constructions of Definition 2.4 "strong" because many constructions in idealized models use a "weaker" security variant. In a weak construction P of a primitive $\mathcal{P}$ in an idealized model $\mathcal{I}$, the completeness is defined similarly to Definition 2.4, but when it comes to security, the advantage of A in breaking the scheme is calculated *also over the randomness of* $\mathcal{I}$. It can be shown that the strong and weak (black-box) constructions in idealized models are *not* comparable in general.[4] However, looking ahead, our

[4] For "weak$\not\Rightarrow$strong" consider a construction of (un-keyed) collision resistant hash functions in the random oracle model $\mathcal{R}$ in which $h(x)$ is equal to the first $|x| - 1$ bits of $\mathcal{R}(x)$. For "strong$\not\Rightarrow$weak" consider a trivial primitive in the Boolean random oracle model $\mathcal{B}$ in which a trivial attacker A succeeds in its attack over security parameter n if $\mathcal{B}$ is equal to 0 over the first $\log(n)$ queries. Then the only oracle for which A succeeds in its attack for an infinite sequence of security parameters is the constant zero oracle, which has a measure zero of being sampled.

proof of Theorem 1.1 shows that when the attacker achieves constant $\Omega(1)$ advantage over the trivial bound, a strong black-box construction would is also a weak black-box construction.

In what follows, unless specified otherwise, by constructions in idealized models we refer to strong black-box constructions.

2.2 Borrowed Results

Theorem 2.6 ([BBF16]). *Suppose one-way functions exist,* $\mathbf{NP} \not\subseteq \mathbf{coAM}$*, and* $\delta, \varepsilon \colon \mathbb{N} \mapsto [0,1]$ *are such that* $2\varepsilon(n) + 3\delta(n) < 1 - 1/\operatorname{poly}(n)$*, then there is no* (ε, δ) *approximate statistical CIO for all poly-size circuits.*

Theorem 2.7 ([Pas15, MMN15]). *Suppose* O' *is an approximately correct obfuscation algorithm with error at most* ε' *in idealized model* $\mathcal{I}$ *where* $\mathcal{I}$ *is random trapdoor permutation oracle or the degree-*$O(1)$ *graded encoding model for finite rings. Suppose* $\varepsilon'' \geq 1/\operatorname{poly}(n)$*. Then there is another obfuscation algorithm* O *in the plain model such that:*

- *The running time of* O *is* $\operatorname{poly}(n/\varepsilon''(n))$ *where* n *is the size of the input circuit and it is approximately correct with error at most* $\varepsilon = \varepsilon' + \varepsilon''$*.*
- *There is a simulator* Sim *in the idealized model that runs in time* $\operatorname{poly}(n/\varepsilon''(n))$*, and for any circuit* C*, the distributions* $\mathsf{Sim}^{\mathcal{I}}(O'^{\mathcal{I}}(C))$ *and* $O(C)$ *have statistical distance* $\operatorname{negl}(|C|)$*.*

2.3 Measure Theoretic Tools

By a *probability space* we mean a *measure space* with total measure equal to one, and by $\Pr[E]$ we denote the measure of the measurable set E. For a sequence of measurable sets $\mathcal{E} = (E_1, E_2, \dots)$ defined over some measure space, the limit supremum of $\mathcal{E}$ is defined as $\operatorname{limSup}(\mathcal{E}) = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} E_m$. It can be shown that $\operatorname{limSup}(\mathcal{E})$ is measurable if E_i is so for all i.

Lemma 2.8 (Borel–Cantelli [Bor09, Can17]). *Let* $\mathcal{E} = (E_1, E_2, \dots)$ *be a sequence of measurable sets over some probability space, and* $\sum_{n=1}^{\infty} \Pr[E_i] = O(1)$*. Then* $\operatorname{limSup}(\mathcal{E})$ *has measure zero.*

Lemma 2.9. *If* $\mathcal{E} = (E_1, E_2, \dots)$ *is a sequence of measurable sets over some probability space, and* $\Pr[E_i] \geq \delta$ *for all* $i \in \mathbb{N}$*, then* $\Pr[\operatorname{limSup}(\mathcal{E})] \geq \delta$*.*

Proof. We use the following proposition.

Proposition 2.10 ([Ale03] Proposition 37, Part (iii)). *Let* $B_1 \supseteq B_2 \supseteq \dots$ *be a sequence of measurable sets over some measure space, and* $\Pr[B_1] < \infty$*. Then* $\Pr\left[\bigcap_{n=1}^{\infty} B_n\right] = \lim_{n\to\infty} \Pr[B_n]$*.*

Now let $B_n = \bigcup_{m=n}^{\infty} E_m$, and so $\operatorname{limSup}(\mathcal{E}) = \bigcap_{n=1}^{\infty} B_n$. Since the measure space is a probability space, thus we have $\Pr[B_1] \leq 1$, and we can apply the above proposition to conclude that

$$\lim_{n\to\infty} \Pr[B_n] = \Pr\Big[\bigcap_{n=1}^{\infty} B_n\Big] = \Pr[\operatorname{limSup}(\mathcal{E})].$$

Finally, because $\Pr[B_n] \geq \Pr[E_i] \geq \delta$ for every n, we get $\delta \leq \lim_{n\to\infty} \Pr[B_n] = \Pr[\operatorname{limSup}(\mathcal{E})]$. □

3 Proving the Main Separation

In this section we prove Theorem 1.1. First we formalize the statement by specifying the way $\mathcal{P}$ is constructed in the idealized models.

Theorem 3.1 (Main–Formal)**.** *Assuming the existence of one-way functions and* **NP** $\not\subseteq$ **coAM***, there is no fully-black-box construction of IO from any primitive* $\mathcal{P}$ *that has a strong black-box construction in the random trapdoor permutation oracle or the degree-*$O(1)$ *graded encoding model for any finite ring.*

In fact, we prove a stronger separation that holds for approximate computational CIO as well.

Theorem 3.2. *Assuming there is no* (ε, δ) *approximate statistical CIO, there is no fully-black-box construction of* (ε', δ') *approximate computational CIO for any* $\varepsilon' \leq \varepsilon - n^{-\Omega(1)}, \delta' \leq \delta - \Omega(1)$ *from any of the primitives listed in Theorem 3.1.*

Proving Theorem 3.1 using Theorems 2.6 and 3.2. Theorem 2.6 rules out (ε, δ) approximate statistical CIO (assuming OWFs and **NP** $\not\subseteq$ **coAM**) for some $\varepsilon = 1/\operatorname{poly}(n)$ and $\delta = 2.99$. Thus, if we choose $\varepsilon' = \varepsilon/2$ and $\delta' = \delta/2$, then Theorem 3.1 follows from Theorems 3.2 and 2.6.

In the following we will focus on proving Theorem 3.2.

Remark 3.3 (The need for constant δ.)**.** Our proof of Theorem 3.2 crucially relies on the fact that $\delta - \delta' \geq \Omega(1)$ which in turn requires $\delta \geq \Omega(1)$. Thus, the separation holds because the attacker of [BBF16] could achieve $\delta \approx 1/3$ (as opposed to just $1/\operatorname{poly}(n)$). More technically, our proof will make use of Lemma 2.9 rather than the Borel-Cantelli lemma, and that is the source of our need for $\delta \geq \Omega(1)$. However, in case one can improve the result of [BBF16] to cover the setting of $\varepsilon = 1/\operatorname{poly}(n)$ and $\delta = 1 - \alpha$ for arbitrary small $\alpha = 1/\operatorname{poly}(n)$, then our Theorem 3.2 could be improved to any $\delta' = \delta - 1/\operatorname{poly}(n)$. In fact the proof will be simple and will not use our Lemma 2.9 and could be based on the Borel-Cantelli lemma (see the end of this section for a sketch).

Remark 3.4 (Ruling out relativizing constructions)**.** In Theorem 3.1 we focus on ruling out fully-black-box constructions. However, the proof can be extended to rule out relativizing constructions (of IO from the set of listed primitives) using standard techniques and the fact that an optimal statistical distinguisher can be implemented in **PSPACE**. In particular, the separating oracle would be a random sample from the idealized oracle $I \leftarrow \mathcal{I}$ and an oracle for a **PSPACE**-complete oracle. However, interestingly, in our case the sampled $I \leftarrow \mathcal{I}$ would only work with *constant* measure (which is enough since it is still a positive measure) due to using Lemma 2.9 as opposed to measure one, which is typically the case in black-box separations.

Proof of Theorem 3.2. In the following, let $\mathcal{Q}$ denote the primitive of (ε', δ') approximate computational CIO. Also let $\mathcal{P}$ be any primitive that can be constructed in the idealized models listed in Theorem 3.1 (according to Definition 2.4), and let P be the implementation of $\mathcal{P}$ relative to $\mathcal{I}$.

For sake of contradiction, in the following we let Q be the fully-black-box construction of $\mathcal{Q}$ from $\mathcal{P}$. First we recall a composition lemma from [MMN+16] showing that $\mathcal{Q}$ could also be implemented relative to $\mathcal{I}$ as well.[5] Then we rule out the existence of black-box constructions of $\mathcal{Q}$ from $\mathcal{I}$ to conclude that Q could not exist.

[5] [MMN+16] proved a variant of Lemma 3.5 for semi-black-box constructions, and sketched the proof for fully-black-box case. For sake of completeness here we recall the proof for fully-black-box constructions.

Lemma 3.5 (Composition lemma [MMN+16]). *Suppose Q is a fully-black-box construction of $\mathcal{Q}$ from $\mathcal{P}$, and suppose P is (a strong black-box) implementation of $\mathcal{P}$ relative to $\mathcal{I}$. Then Q^P is a (strong black-box) implementation of $\mathcal{Q}$ relative to the same idealized model $\mathcal{I}$.*

Proof. It is easy to see that Q^P is an implementation of $\mathcal{Q}$ relative to $\mathcal{I}$ (by completeness of the constructions P and Q), and so the completeness holds. The proof of security follows. For sake of contradiction, let $A^{\mathcal{I}}$ be any efficient query successful attacker against the implementation Q^P (of $\mathcal{Q}$) in the idealized model $\mathcal{I}$ which rules out its strong black-box property. Namely, there is a non-zero measure fraction of $I \xleftarrow{\$} \mathcal{I}$ it holds that A^I breaks the security of Q^{P^I}. For any such fixed I, the security reduction $S^{A^I, I}$ (of the fully-black-box construction Q of P) would break the security of P^I. By combining the algorithms S and A we get that the efficient query attacker $(S^A)^I = B^I$ breaks the security of P^I with non-zero measure over the sampled oracle $I \xleftarrow{\$} \mathcal{I}$. But this contradicts the assumption that $\mathcal{P}$ is securely realized in $\mathcal{I}$ in a strong black-box way. Therefore Q^P is also a *strong black-box* construction of $\mathcal{Q}$ relative to $\mathcal{I}$. □

In the following we will use Theorems 2.7 and 2.6 to rule out the possibility of any *strong black-box* construction of $\mathcal{Q}$ relative to $\mathcal{I}$ which (with Lemma 3.5) shows that Q could not exist.

Let $\varepsilon'' = \varepsilon - \varepsilon' \geq 1/\operatorname{poly}(n)$ and $\delta'' = \delta - \delta' \geq \Omega(1)$. Since P is a construction of $\mathcal{P}$ relative to $\mathcal{I}$, thus $O'^{\mathcal{I}} = (Q^P)^{\mathcal{I}}$ is an ε' approximate obfuscation mechanism relative to $\mathcal{I}$. Let O be the ε approximate obfuscator in the plain model that exists due to Theorem 2.7. The assumption in Theorem 3.2 is that O cannot be an (ε, δ) approximate statistical CIO. Therefore, there is a computationally unbounded adversary A and an infinite sequence of circuits $(C_0^1, C_1^1), \dots, (C_0^i, C_1^i), \dots$ such that for all i: $|C_0^i| = |C_1^i|$, $C_0^i \equiv C_1^i$, and $\Pr_{b \leftarrow \{0,1\}}[A(O(C_b^i)) = b] \geq 1/2 + \delta(n)/2$.

Now consider another attacker A' in the idealized model $\mathcal{I}$ which, given a circuit B' as input, runs the simulator of Theorem 2.7 to get the circuit $B = \mathsf{Sim}^{\mathcal{I}}(B')$ and then runs A over B to output whatever A does. By the property of the simulator Sim we conclude that A' is an efficient query (computationally unbounded) attacker in the idealized model $\mathcal{I}$ that achieves

$$\Pr_{b \leftarrow \{0,1\}, I \leftarrow \mathcal{I}}[A'^I(O'^I(C_b^i)) = b] \geq 1/2 + \delta(n)/2 - \operatorname{negl}(n)$$

where $|C_0^i| = |C_1^i| = n$.

A crucial point is that the above probability is *also over the randomness of the oracle* $I \leftarrow \mathcal{I}$ for every i, while we are interested in *fixing* $I \leftarrow \mathcal{I}$ and getting a successful attack for infinitely many circuits at the same time. By a simple averaging argument we can get:

$$\Pr_{I \leftarrow \mathcal{I}}\left[\Pr_{b \leftarrow \{0,1\}}[A'^I(O'^I(C_b^i)) = b] \geq 1/2 + \delta'(n)/2\right] \geq \delta''(n)/2 - \operatorname{negl}(n).$$

Thus, if we define the event E_i over the sampled oracle $I \leftarrow \mathcal{I}$ as:

$$E_i \text{ holds if: } \Pr_{b \leftarrow \{0,1\}}[A'^I(O'^I(C_b^i)) = b] \geq 1/2 + \delta'/2$$

then we get $\Pr[E_i] \geq \delta''(n)/2 - \operatorname{negl}(n) \geq \delta''/3$ for every $i \in \mathbb{N}$. Now we can apply Lemma 2.9 to conclude that with probability at least $\delta''/3$ over the choice of $I \leftarrow \mathcal{I}$ an infinite number of the events E_i's would happen at the same time for I. We call $I \leftarrow \mathcal{I}$ a good oracle if it is indeed the case that infinitely many of the events E_i's happen over I. By definition, for any good oracle I, the attacker A' successfully breaks $(Q^P)^I$ (as an implementation of $\mathcal{Q}$ in model $\mathcal{I}$) over infinitely many pairs of circuits while asking only an efficient number of oracle queries to I. The existence of such A' who breaks $(Q^P)^I$ for non-zero (in fact $\geq \delta''/3$) measure of the choice of the oracles $I \leftarrow \mathcal{I}$ prevents Q^P from being a strong black-box construction of $\mathcal{Q}$ relative to $\mathcal{I}$. □

Case of $\delta' \approx 1 - 1/\operatorname{poly}(n)$. Theorem 3.2 was sufficient for us to derive Theorem 3.1, however that is not the strongest separation one can imagine for approximate computational CIO as it does not cover the case of $1 - 1/\operatorname{poly}(n)$. The work of [BBF16] shows that whenever $2\varepsilon + \delta > 1$ then there is in fact a way to achieve (ε, δ) approximate statistical CIO. Thus one can imagine the possibility that the result of [BBF16] could ultimately be improved to rule out (ε, δ) approximate statistical CIO for $O(\varepsilon) + \delta < 1 - 1/\operatorname{poly}(n)$. Below, we show that such result, if proved, could be used to derive lower bounds on complexity of (ε', δ') approximate computational CIO large $\delta' \approx 1 - 1/\operatorname{poly}(n)$.

Theorem 3.6. *If there is no (ε, δ) approximate statistical CIO for $\delta = 1 - \rho$ for sufficiently small $\rho = 1/\operatorname{poly}(n)$ (e.g., $\rho = 1/n^4$ suffices), then there is no fully-black-box construction of $(\varepsilon', \delta' = 1 - \sqrt{\rho})$ approximate computational CIO for any $\varepsilon' \leq \varepsilon - n^{-\Omega(1)}$ from the primitives of Theorem 3.1.*

Thus, the main difference between Theorem 3.2 and Theorem 3.6 is that in Theorem 3.6 we cover the case of $\delta' = 1 - 1/\operatorname{poly}(n)$, but we also rely on stronger assumption that $\delta = 1 - 1/\operatorname{poly}(n)$.

Proof of Theorem 3.6. The proof is identical to that of Theorem 3.2 except for the following. Since the attackers A and A' will succeed in guessing the correct circuit with probability $1 - 1/\operatorname{poly}(1) \approx 1$ we can do a better averaging argument to get a better attack after fixing the oracle. Namely, define the event E_i as:

$$E_i \text{ holds if: } \Pr_{b \leftarrow \{0,1\}}[A'^I(O'^I(C_b^i)) = b] \geq 1 - \sqrt{\rho(n)/2}$$

where n is the size of the circuits C_0^i, C_1^i. Then we can conclude that $\Pr[E_i] \geq 1 - 10\sqrt{\rho(n)}$. Now, since the events E_i happen with large probability and that $\sum_n 10\sqrt{\rho(n)} < \infty$ we can apply the Borel-Cantelli lemma (Lemma 2.8) to conclude that with measure *one* over the choice of the oracle $I \leftarrow \mathcal{I}$ all but finitely many of E_i's would happen. The rest of the proof is remains unchanged. □

References

[Ale03] Michelle Alexopoulos. Notes on set theory and probability theory, 2003. http://www.biostat.umn.edu/~dipankar/pubh7440/ProbSets.pdf. 5

[BBF13] Paul Baecher, Christina Brzuska, and Marc Fischlin. Notions of black-box reductions, revisited. In *Advances in Cryptology-ASIACRYPT 2013*, pages 296–315. Springer, 2013. 3

[BBF16] Zvika Brakerski, Christina Brzuska, and Nils Fleischhacker. On statistically secure obfuscation with approximate correctness. Cryptology ePrint Archive, Report 2016/226, 2016. http://eprint.iacr.org/. 2, 3, 5, 6, 7, 8

[BDSMP91] Manuel Blum, Alfredo De Santis, Silvio Micali, and Giuseppe Persiano. Noninteractive zero-knowledge. *SIAM Journal on Computing*, 20(6):1084–1118, 1991. 1

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

[Bor09] Émile Borel. Les probabilités dénombrables et leurs applications arithmétiques. *Rendiconti del Circolo Matematico di Palermo (1884-1940)*, 27(1):247–271, 1909. 5

[BR93] Mihir Bellare and Phillip Rogaway. Random Oracles are Practical: A Paradigm for Designing Efficient Protocols. In *ACM Conference on Computer and Communications Security*, pages 62–73, 1993. 1

[BY96] Mihir Bellare and Moti Yung. Certifying permutations: Noninteractive zero-knowledge based on any trapdoor permutation. *J. Cryptology*, 9(3):149–166, 1996. 1

[Can17] Francesco Paolo Cantelli. Sulla probabilita come limite della frequenza. *Atti Accad. Naz. Lincei*, 26(1):39–45, 1917. 5

[CKP15] Ran Canetti, Yael Tauman Kalai, and Omer Paneth. On obfuscation with random oracles. Cryptology ePrint Archive, Report 2015/048, 2015. `http://eprint.iacr.org/`. 2

[DNR04] Cynthia Dwork, Moni Naor, and Omer Reingold. Immunizing encryption schemes from decryption errors. In *Advances in Cryptology-EUROCRYPT 2004*, pages 342–360. Springer, 2004. 2

[GGH$^+$13] Sanjam Garg, Craig Gentry, Shai Halevi, Mariana Raykova, Anant Sahai, and Brent Waters. Candidate indistinguishability obfuscation and functional encryption for all circuits. In *IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 40–49. IEEE, 2013. 1

[Gol11] Oded Goldreich. Basing non-interactive zero-knowledge on (enhanced) trapdoor permutations: The state of the art. In *Studies in Complexity and Cryptography. Miscellanea on the Interplay between Randomness and Computation*, pages 406–421. Springer, 2011. 1

[GR07] Shafi Goldwasser and Guy N Rothblum. On best-possible obfuscation. In *Theory of Cryptography*, pages 194–213. Springer, 2007. 2

[GR14] Shafi Goldwasser and Guy N Rothblum. On best-possible obfuscation. *Journal of Cryptology*, 27(3):480–505, 2014. 2

[GS02] Craig Gentry and Alice Silverberg. Hierarchical id-based cryptography. In *Proceedings of the 8th International Conference on the Theory and Application of Cryptology and Information Security: Advances in Cryptology*, ASIACRYPT '02, pages 548–566, London, UK, UK, 2002. Springer-Verlag. 1

[HL02] Jeremy Horwitz and Ben Lynn. Toward hierarchical identity-based encryption. In LarsR. Knudsen, editor, *Advances in Cryptology EUROCRYPT 2002*, volume 2332 of *Lecture Notes in Computer Science*, pages 466–481. Springer Berlin Heidelberg, 2002. 1

[Hol06] Thomas Holenstein. *Strengthening key agreement using hard-core sets*. PhD thesis, ETH ZURICH, 2006. 2

[Hol15] Thomas Holenstein. Complexity theory, 2015. `http://www.complexity.ethz.ch/education/Lectures/ComplexityFS15/skript_printable.pdf`. 4

[MMN15] Mohammad Mahmoody, Ameer Mohammed, and Soheil Nematihaji. More on impossibility of virtual black-box obfuscation in idealized models. Cryptology ePrint Archive, Report 2015/632, 2015. `http://eprint.iacr.org/`. 2, 5

[MMN$^+$16] Mohammad Mahmoody, Ameer Mohammed, Soheil Nematihaji, Rafael Pass, and Abhi Shelat. Lower bounds on assumptions behind indistinguishability obfuscation. In *Theory of Cryptography*, pages 49–66. Springer, 2016. 2, 3, 4, 6

[NY90] Moni Naor and Moti Yung. Public-key cryptosystems provably secure against chosen ciphertext attacks. In *In Proc. of the 22nd STOC*, pages 427–437. ACM Press, 1990. 1

[Pas15] Rafael Pass and abhi shelat. Impossibility of vbb obfuscation with ideal constant-degree graded encodings. Cryptology ePrint Archive, Report 2015/383, 2015. `http://eprint.iacr.org/`. 2, 5

[RTV04] Omer Reingold, Luca Trevisan, and Salil P. Vadhan. Notions of reducibility between cryptographic primitives. In *Theory of Cryptography Conference, TCC 2004*, volume 2951 of *Lecture Notes in Computer Science*, pages 1–20. Springer, 2004. 1, 2, 3