Electronic Colloquium on Computational Complexity, Report No. 85 (2021)

The Final Nail in the Coffin of Statistically-Secure Obfuscator.

Ilya Volkovich *

June 16, 2021

Abstract

We present an elementary, self-contained proof of the result of Goldwasser and Rothblum [GR07] that the existence of a (perfect) statistically secure obfuscator implies a collapse of the polynomial hierarchy. In fact, we show that an existence of a weaker object implies a somewhat stronger statement. In addition, we extend the result of [GR07] to the case of *imperfect* statistically secure obfuscator.

1 Introduction

An *indistinguishability obfuscator* ($\mathcal{IO}$) is a an efficient (potentially) randomized procedure that maps Boolean circuits into "somewhat" larger, "unintelligible" circuits, yet preserving the original circuit input-output functionality. This procedure has served as an ingredient in many cryptographic applications (see e.g. [GGHR14, SW14] and references within). More formally speaking, $\mathcal{IO}$ maps a circuit C into another circuit $\hat{C}$ with the same functionally, such that given two circuits C_1, C_2 of the same size and functionality, the distributions of $\hat{C}_1$ and $\hat{C}_2$ are *indistinguishable* (for a formal definition, see Definition 2.3). The notion of indistinguishability has been studied in two settings: the computational setting - when the distributions are required to be indistinguishable only by polynomial-time algorithms, and the information-secure (a.k.a statistical) setting - when distributions should be indistinguishable by *any* (even very inefficient) algorithm.

Several previous results exhibited negative consequences of the existence of obfuscators in various regimes. In [GR07], Goldwasser and Rothblum showed that the existence of statistically-secure obfuscator implies that $\mathsf{NP} \subseteq \mathsf{coAM}$. Brakerski et al. [BBF16] also considered the statistical setting, yet allowing the obfuscator to output an "approximately" correct circuit. That is, with high probability the obfuscator should output a circuit which is functionally close to the original circuit[1]. They have shown that the existence of approximately correct statistically-secure obfuscator[2] implies that either $\mathsf{NP} \subseteq \mathsf{coAM}$ or that one-way functions do not exists. It is to be noted that the containment $\mathsf{NP} \subseteq \mathsf{coAM}$ is *believed* to be very unlikely as it results in a collapse of the polynomial hierarchy (see e.g.[BHZ87]).

In the computational setting, Komargodski et al. [KMN+14] gave an easy argument that the existence of computationally-secure obfuscator implies existence of one-way functions, unless $\mathsf{NP} \subseteq \mathsf{BPP}$. They have also extended the result to "imperfect" computationally-secure obfuscators. That is, obfuscators that output the correct circuit with high probability.

Indeed, approximately correct obfuscators are weaker than imperfect, since in the former case, the output circuit $\hat{C}$ should, with high probably, agree with C on many (but not, necessarily all) inputs , whereas in the latter case, $\hat{C}$ should agree with C on **all** inputs. Yet, the above negative results are incomparable.

*Department of Computer Science, Boston College, Chestnut Hill, MA 02467. Email: ilya.volkovich@bc.edu

[1] The formal definition used in [BBF16] is somewhat different, however equivalent.

[2] In fact, they that shown that even a "correlated" obfuscator is sufficient. See Definition 2.3 for more details.

ISSN 1433-8092

1.1 Our Results

In this short note we put the final nail in the coffin of statistically-secure obfuscators, by presenting an elementary, self-contained proof of the result of that the existence of a perfect statistically-secure obfuscator implies a collapse of the polynomial hierarchy.

Theorem 1. *If there exists a perfect statistical obfuscator for polynomial-size circuits or even* 3CNF *formulas then* NP $\subseteq$ coNP.

In addition, we show a simple extension of the result of [GR07] to the case of *imperfect* statistically-secure obfuscators.

Theorem 2. *If there exists an* imperfect *statistical obfuscator for polynomial-size circuits or even* 3CNF *formulas then* NP $\subseteq$ coAM.

The proofs of Theorems 1 and 2 is given in Section 3. We also remark that the result in Theorem 2 is incomparable with [BBF16]. While, as was discussed earlier, the preconditions of [BBF16] are weaker, their consequences also involve one-way functions.
We further observe that Theorem 1 extends to an intermediate indistinguishability setting when the obfuscator is required to be secure against polynomial-time algorithms with an oracle access to the MCSP. Here MCSP denotes the Minimal Circuit Size Problem. For more details see Section 4.

Theorem 3. *If there exists a perfect obfuscator for polynomial-size circuits or even* 3CNF *formulas that is secure against polynomial-time algorithms with an oracle to* MCSP *then* NP $\subseteq$ coNP.

1.2 Techniques

We follow the approach initiated by [GR07]. To provide more intuition, we sketch their result.

Let $\mathcal{IO}$ be a perfect statistically-secure obfuscator and let C be a circuit of size s. Rather than just obfuscating C once, we consider the distribution on circuits associated with C by $\mathcal{IO}$. To this end, we define the function $D_C(r) := \mathcal{IO}(C; r)$, where the input r denotes the random string used by $\mathcal{IO}$. The corresponding (induced) distribution D_C is the output of $D_C(r)$ on r chosen uniformly at random. Given the above, the result of [GR07] shows how to distinguish between satisfiable and unsatisfiable circuits C via the statistical distances between their distributions D_C (see Definition 2.1 for more details).

Let $\perp_s$ denote a canonical unsatisfiable circuit of size s and let $D_s(r) := \mathcal{IO}(\perp_s; r)$. If C is satisfiable, then by the correctness requirement D_C and D_s have completely disjoint supports and hence $\Delta(D_C, D_s)$ is "large". Otherwise, if C is unsatisfiable, then by the security requirement $\Delta(D_C, D_s)$ is "small". Next, they observe that the problem of distinguishing between a pair of distributions with a "large" statistical distance and a pair with a "small" one was shown to lie in AM $\cap$ coAM (see Lemma 2.9 for more details). In conclusion, the existence of a perfect statistically-secure obfuscator implies that NP $\subseteq$ coAM, which in turn results in a collapse of the polynomial hierarchy (see e.g.[BHZ87]).

We simplify the result by observing that it is actually sufficient to determine whether a pair of distributions D_C and D_s have a disjoint support. This task can be carried out easily in coNP. Consequently, we obtain an elementary, self-contained proof that the existence of a perfect statistically-secure obfuscator implies NP $\subseteq$ coNP. In fact, we show that a somewhat stronger claim holds. See Lemma 3.1 for the formal statement.

Next, we extend the result of [GR07] to the case of *imperfect* statistically-secure obfuscator. As before, if C is unsatisfiable, then by the security requirement $\Delta(D_C, D_s)$ is "small". However, if C is satisfiable, then the support of D_C and D_s may no longer be disjoint. We observe that, nonetheless, $\Delta(D_C, D_s)$ is still "sufficiently large" (see Lemma 2.7 for the exact statement).

2 Preliminaries

A function $\mathrm{negl}(n)$ is *negligible* if for any $k \in \mathbb{N}$ there exists $n_k \in \mathbb{N}$ s.t. for all $n > n_k$, $\mathrm{negl}(n) < 1/n^k$.
Let X and Y be two random variables taking values in some finite domain $\mathcal{U}$.

Definition 2.1 (Statistical Distance and Support). *We define the* support *of a random variable X as*

$$\mathrm{Supp}(X) := \{u \in \mathcal{U} \mid \Pr[X = u] > 0\}.$$

The Statistical Distance *between X and Y is defined as*

$$\Delta(X, Y) := \max_A \Pr_{u \sim X}[A(u) = 1] - \Pr_{u \sim Y}[A(u) = 1],$$

where $A : \mathcal{U} \to \{0, 1\}$ is (possibly, inefficient and/or probabilistic) algorithm.
Note that the equality is attained for A such that $A(u) = 1 \iff \Pr[X = u] \geq \Pr[Y = u]$.

Our main argument will rely on the following simple observation:

Observation 2.2. $\Delta(X, Y) < 1 \iff \mathrm{Supp}(X) \cap \mathrm{Supp}(Y) \neq \emptyset$.

We now recall the definition of our main object of study - Indistinguishability Obfuscator.

Definition 2.3 (Indistinguishability Obfuscator [BGI+12, KMN+14, BBF16]). *We say that a polynomial-time procedure $\mathcal{IO}(C; r)$ is an* Indistinguishability Obfuscator *for a circuit class $\mathcal{C}$ with the following:*

1. ***(Perfect/Imperfect) Correctness:*** *We say that $\mathcal{IO}$ is ε-*imperfect *if for every circuit $C \in \mathcal{C}$: $\Pr_r[C \equiv \mathcal{IO}(C; r)] \geq 1 - \varepsilon(|C|)$. If $\varepsilon = 0$, then we say that $\mathcal{IO}$ is* perfect.
2. ***Polynomial slowdown:*** *There is $k \in \mathbb{N}$ s.t. for every circuit $C \in \mathcal{C}$ and every r: $|\mathcal{IO}(C; r)| \leq |C|^k$.*
3. ***Security:*** *We say that $\mathcal{IO}$ is δ-*uncorrelated *if for all pairs of circuits $C_1, C_2 \in \mathcal{C}$ such that $C_1 \equiv C_2$ and $|C_1| = |C_2| = s$ it holds: $\Delta(D_{C_1}, D_{C_2}) \leq \delta(s)$.*
   We say that $\mathcal{IO}$ is statistically secure*, if $\delta(s) = \mathrm{negl}(s)$, for some negligible function $\mathrm{negl}(s)$.*
   We say that $\mathcal{IO}$ is correlated*, if $\forall s : \delta(s) < 1$. (**Remark:** it is still possible that $\delta(s) \to 1$).*

Remark 2.4. Some definitions also contain a security parameter. In the above definition it is incorporated in the circuit size. Any reasonable encoding scheme for Boolean circuits allows to represent/regard a circuit of size s as a circuit of larger size.

Definition 2.5 (Obfuscation Distributions). *For a circuit C, we define the function $D_C(r) := \mathcal{IO}(C; r)$.*

The corresponding (induced) distribution D_C is the output of $D_C(r)$ on r chosen uniformly at random.
The following is an immediate and useful consequence of Part of 1 of Definition 2.3:

Observation 2.6. *Let $\mathcal{IO}$ be a perfect obfuscator for a circuit class $\mathcal{C}$ and let $C_1, C_2 \in \mathcal{C}$ be such that $C_1 \not\equiv C_2$. Then $\Delta(D_{C_!}, D_{C_2}) = 1$.*

Nonetheless, it is easy to see that the above is no longer true for (even slightly) imperfect obfuscators. Next is our key, but simple to prove lemma that extends the intuition behind the above observation to the imperfect case. In particular, it shows that while the distributions of imperfect obfuscations of functionally different circuits might not be disjoint, their statistical distance is still "large".

Lemma 2.7. *Let $\mathcal{IO}$ be an ε-imperfect obfuscator for a circuit class $\mathcal{C}$ and let $C_1, C_2 \in \mathcal{C}$ be such that $C_1 \not\equiv C_2$. Then $\Delta(D_{C_1}, D_{C_2}) \geq 1 - 2\varepsilon$.*

Proof. Let A be an algorithm that given an obfuscated circuit $\hat{C}$, outputs 1 iff $\hat{C} \equiv C_1$. Then

$$\Delta(D_{C_1}, D_{C_2}) \geq \Pr_{\hat{C} \sim D_{C_1}}[A(\hat{C}) = 1] - \Pr_{\hat{C} \sim D_{C_2}}[A(\hat{C}) = 1] \geq (1 - \varepsilon) - \varepsilon = 1 - 2\varepsilon.$$

Definition 2.8 (Statistical Difference [SV03]). *Let $\alpha(n) : \mathbb{N} \to \mathbb{N}$ and $\beta(n) : \mathbb{N} \to \mathbb{N}$ be computable functions, such that $\alpha(n) > \beta(n)$.*
Then $\mathrm{GapSD}^{(\alpha(n)\,,\,\beta(n))} := (\mathrm{GapSD}_{YES}^{(\alpha(n)\,,\,\beta(n))}, \mathrm{GapSD}_{NO}^{(\alpha(n)\,,\,\beta(n))})$*, where*

$$\mathrm{GapSD}_{YES}^{(\alpha(n)\,,\,\beta(n))} = \{(C_1, C_2) \mid \Delta(C_1, C_2) \geq \alpha(n)\},$$
$$\mathrm{GapSD}_{NO}^{(\alpha(n)\,,\,\beta(n))} = \{(C_1, C_2) \mid \Delta(C_1, C_2) \leq \beta(n)\}.$$

Here, C_1 and C_2 are Boolean circuits $C_1, C_2 : \{0,1\}^n \to \{0,1\}^m$ of size $\mathrm{poly}(n)$.

Lemma 2.9 ([SV03]). *Suppose $\alpha(n)^2 > \beta(n)$. Then* $\mathrm{GapSD}^{(\alpha(n)\,,\,\beta(n))}$ *is* SZK*-complete.*
$\mathsf{SZK} \subseteq \mathsf{AM} \cap \mathsf{coAM}$.

3 Proofs of the Main Results

In this section we prove our main results: Theorems 1 and 2. In fact, we prove somewhat technically stronger versions of these results. We will use the following definition throughout this section: $D_s(r) := \mathcal{IO}(\perp_s; r)$, where $\perp_s$ is a canonical unsatisfiable circuit from $\mathcal{C}$ of size s.

Lemma 3.1. *Let $\mathcal{IO}$ be a perfect, correlated obfuscator for a circuit class $\mathcal{C}$. Then $\mathcal{C}$-*SAT $\in$ coNP.

Here $\mathcal{C}$-SAT denotes the satisfiability (SAT) problem for a circuit class $\mathcal{C}$.

Proof. Let $C \in \mathcal{C}$ be a circuit given as an input and let $s = |C|$. The algorithm will non-deterministically guess r and r', and accept if and only if $D_C(r) = D_s(r')$. Now, if C is satisfiable, then by perfect correctness, for all $r, r' : D_C(r) \neq D_s(r')$ (Observation 2.6). On the other hand, if C is unsatisfiable, then by the security requirement $\Delta(D_C, D_s) < 1$ and hence by Observation 2.2, there exist r and r' as required. □

The following lemma extends the result of [GR07] to imperfect obfuscators.

Lemma 3.2. *Let $\mathcal{IO}$ an ε-imperfect, δ-uncorrelated obfuscator for a circuit class $\mathcal{C}$, such that $(1 - 2\varepsilon)^2 > \delta$. Then $\mathcal{C}$-*SAT $\in \mathsf{AM} \cap \mathsf{coAM}$.

Proof. Similarly to [GR07], we claim that $\mathcal{C}$-SAT reduces to $\mathrm{GapSD}^{(1-2\varepsilon\,,\,\delta)}$ and hence the claim follows from Lemma 2.9. Let $C \in \mathcal{C}$ be a circuit given as an input and let $s = |C|$. If C is satisfiable then by Lemma 2.7, $\Delta(D_C, D_s) \geq 1 - 2\varepsilon$. On other hand, if C is unsatisfiable, then by the security requirement $\Delta(D_C, D_s) \leq \delta$. □

4 Extensions

In this section we extend the conclusion of Theorem 1 to a weaker, intermediate indistinguishability setting. While statistical security requires indistinguishability against all (possibly even very inefficient) algorithms, we observe that the same conclusion still holds true even if we relax this requirement.

The *Minimal Circuit Size Problem (*MCSP*)*, asks to decide, for a given truth table of a Boolean function f and a parameter s, whether f can be computed by a Boolean circuit of size at most s. While it is easy to see that $\mathsf{MCSP} \in \mathsf{NP}$, the exact complexity of the problem remains unknown, despite a large body of work. In [IKV18], an $\mathcal{IO}$-based approach was proposed. Specifically, it was shown that if there exists an obfuscator $\mathcal{IO}$ that secure against efficient (randomized) algorithms with an oracle access to MCSP then $\mathsf{NP} \subseteq \mathsf{BPP}^{\mathsf{MCSP}}$. We show that the same hypothesis, in fact, leads to the conclusion of Theorem 1 and is, therefore, unlikely to hold. In particular, we show that the aforementioned security requirement implies that such an $\mathcal{IO}$ must be correlated (see Corollary 4.4 for the formal statement). In order to formalize our result, we extend the notion of security (Part 3 from Definition 2.3.)

Definition 4.1 (Security Against Class of Algorithms). *Let $\mathcal{A}$ be a class of algorithms. We say that $\mathcal{IO}$ is* secure against *$\mathcal{A}$, if for all pairs of circuits $C_1, C_2 \in \mathcal{C}$ such that $C_1 \equiv C_2$ and $|C_1| = |C_2| = s$ and for any algorithm $A \in \mathcal{A}$ it holds:*

$$\left| \Pr_{\hat{C} \sim D_{C_1}} [A(\hat{C}) = 1] - \Pr_{\hat{C} \sim D_{C_2}} [A(\hat{C}) = 1] \right| \leq \mathrm{negl}(s)$$

for some negligible function $\mathrm{negl}(s)$.

We remark that in the special case when $\mathcal{A}$ is the class of all (possibly, inefficient and/or probabilistic) algorithms, the notion of security against $\mathcal{A}$ is, in fact, equivalent to statistical security. Hence, in general security against $\mathcal{A}$ can be seen as a relaxation of statistical security. Another interesting special case is the case when $\mathcal{A}$ is the class of all efficient (randomized) algorithms. The notion of security against $\mathcal{A}$ in this case is referred to as *computational security*.

Let us now consider the security requirement from a different perspective. Fix two circuits $C_1, C_2 \in \mathcal{C}$. Given an obfuscated circuit $\hat{C}$, which results from either C_1 or C_2, the goal of a "distinguisher" $A \in \mathcal{A}$ is to distinguish between $\hat{C}$-s originating from C_1 and those originating from C_2. As $D_{C_1}(r)$ and $D_{C_2}(r)$ can be computed efficiently, we can assume wlog that A has access to both $D_{C_1}(r)$ and $D_{C_2}(r)$. Motivated by the study of $D_C(r)$ as a candidate for one-way function (see e.g. [KMN+14]) we ask the natural question "would A benefit from (oracle) access to their inverters"? To this end, we define an inverter formally.

Definition 4.2 (Inverters). *Fix an obfuscator $\mathcal{IO}$ for a circuit class $\mathcal{C}$ and let $C \in \mathcal{C}$. We say that M is an* inverter *for D_C if*

$$\Pr_{\hat{C} \sim D_C, \tau} \left[D_C(M(\hat{C}, \tau)) = \hat{C} \right] \geq 1/\operatorname{poly}(|C|)$$

here τ denotes the randomness of M.

We show that an oracle access just to one of the inverters already implies that the obfuscator must be correlated and hence unlikely to exist by previous results.

Lemma 4.3. *Let $\mathcal{IO}$ be an obfuscator secure against efficient algorithms with an oracle access to an inventor of one of the obfuscated circuits. Then $\mathcal{IO}$ is correlated.*

Proof. Let $C_1, C_2 \in \mathcal{C}$ s.t. $C_1 \equiv C_2$ and $|C_1| = |C_2| = s$. Assume for contradiction $\Delta(D_{C_1}, D_{C_2}) = 1$. WLOG, let M be an inverter for D_{C_1} and let $p = \Pr_{\hat{C}\sim D_{C_1},\tau}\left[D_{C_1}(M(\hat{C},\tau)) = \hat{C}\right]$ denote the success probability of M. Consider the following algorithm:

Given an obfuscated circuit $\hat{C}$ as an input:

1. Run M on $\hat{C}$ to obtain r; If $D_{C_1}(r) = \hat{C}$, output 1
2. Otherwise, output 0 or 1 uniformly at random

We observe the following:

- By definition: $\Pr_{\hat{C}\sim D_{C_1}}[A(\hat{C}) = 1] = p + \frac{1-p}{2} = \frac{1+p}{2}$
- By assumption and Observation 2.2, as D_{C_1} and D_{C_2} are disjoint, M will always fail to invert $\hat{C}$ produced by D_{C_2}. Therefore, $\Pr_{\hat{C}\sim D_{C_2}}[A(\hat{C}) = 1] = \frac{1}{2}$.
- Consequently: $\Pr_{\hat{C}\sim D_{C_1}}[A(\hat{C}) = 1] - \Pr_{\hat{C}\sim D_{C_2}}[A(\hat{C}) = 1] \geq \frac{p}{2}$.

As $p = 1/\operatorname{poly}(s)$, this contradicts the security requirement. Therefore, $\Delta(D_{C_1}, D_{C_2}) < 1$. □

Finally, as was shown in [ABK+06], oracle access to *any* inventor can be simulated given oracle access to MCSP. Therefore, we obtain the following corollary:

Corollary 4.4. *Let $\mathcal{IO}$ be an obfuscator secure against efficient (randomized) algorithms with an oracle access to* MCSP. *Then $\mathcal{IO}$ is correlated.*

Proof. Using a result of [ABK+06] as instantiated in [IKV18]: there exists a polynomial-time probabilistic oracle Turing machine M and $k \in \mathbb{N}$ such that for any circuit C:

$$\Pr_{\hat{C}\sim D_C,\tau}\left[D_C(M^{\mathsf{MCSP}}(C,\hat{C},\tau)) = \hat{C}\right] \geq 1/|C|^k$$

where τ denotes the randomness of M. Consequently, there exists an efficient algorithm with an oracle access to MCSP that can simulate an inventor. The claim follows from the previous lemma. □

5 Discussion & Open Questions

In this note we presented an elementary, self-contained proof that the existence of a perfect statistically-secure obfuscator implies that NP = coNP. Could we extend the conclusion to P = NP (under the same hypothesis)? Indeed, as was noted in [BGI+12], if P = NP, then there exists a trivial 0-uncorrelated obfuscator. Therefore, such an extension would establish an equivalence between the two.

In terms of further simplification, can one use Lemma 2.7 to provide a simpler proof for the result of [KMN+14]? In particular, combining with the result of [Gol90]? Finally, can we show that $\mathsf{NP} \subseteq \mathsf{BPP}^{\mathsf{MCSP}}$ under the assumption that computational $\mathcal{IO}$ exist?

Acknowledgement

The author would like to thank Chris Peikert for many useful discussions.

References

[ABK+06] E. Allender, H. Buhrman, M. Koucký, D. van Melkebeek, and D. Ronneburger. Power from random strings. *SIAM J. Comput.*, 35(6):1467–1493, 2006. doi:10.1137/050628994.

[BBF16] Z. Brakerski, C. Brzuska, and N. Fleischhacker. On statistically secure obfuscation with approximate correctness. In *CRYPTO*, pages 551–578. 2016.

[BGI+12] 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):6:1–6:48, 2012. doi:10.1145/2160158.2160159.

[BHZ87] R. B. Boppana, J. Håstad, and S. Zachos. Does co-NP have short interactive proofs? *Inf. Process. Lett.*, 25(2):127–132, 1987.

[GGHR14] S. Garg, C. Gentry, S. Halevi, and M. Raykova. Two-round secure MPC from indistinguishability obfuscation. In *Theory of Cryptography - 11th Theory of Cryptography Conference, TCC*, pages 74–94. 2014.

[Gol90] O. Goldreich. A note on computational indistinguishability. *Inf. Process. Lett.*, 34(6):277–281, 1990. doi:10.1016/0020-0190(90)90010-U.

[GR07] S. Goldwasser and G. N. Rothblum. On best-possible obfuscation. In *Theory of Cryptography, 4th Theory of Cryptography Conference, TCC*, pages 194–213. 2007.

[IKV18] R. Impagliazzo, V. Kabanets, and I. Volkovich. The power of natural properties as oracles. In *33rd Computational Complexity Conference CCC*, pages 7:1–7:20. 2018. doi:10.4230/LIPIcs.CCC.2018.7.

[KMN+14] I. Komargodski, T. Moran, M. Naor, R. Pass, A. Rosen, and E. Yogev. One-way functions and (im)perfect obfuscation. In *55th IEEE Annual Symposium on Foundations of Computer Science, FOCS*, pages 374–383. 2014. doi:10.1109/FOCS.2014.47.

[SV03] A. Sahai and S. P. Vadhan. A complete problem for statistical zero knowledge. *J. ACM*, 50(2):196–249, 2003.

[SW14] A. Sahai and B. Waters. How to use indistinguishability obfuscation: deniable encryption, and more. In *Symposium on Theory of Computing, STOC*, pages 475–484. 2014.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il