Lower Bounds on Assumptions behind Indistinguishability Obfuscation

Mohammad Mahmoody[1,⋆], Ameer Mohammed[1,⋆⋆], Soheil Nematihaji[1,⋆⋆⋆], Rafael Pass[2,†], and abhi shelat[1,‡]

[1] University of Virginia, {mohammad, am8zv,sn8fb,shelat}@virginia.edu
[2] Cornell University rafael@cs.cornell.edu

Abstract. Since the seminal work of Garg et al. (FOCS'13) in which they proposed the first candidate construction for indistinguishability obfuscation (iO for short), iO has become a central cryptographic primitive with numerous applications. The security of the proposed construction of Garg et al. and its variants are proved based on multi-linear maps (Garg et al. Eurocrypt'13) and their idealized model called the graded encoding model (Brakerski and Rothblum TCC'14 and Barak et al. Eurocrypt'14). Whether or not iO could be based on standard and well-studied hardness assumptions has remain an elusive open question. In this work we prove *lower bounds* on the assumptions that imply iO in a black-box way, based on computational assumptions. Note that any lower bound for iO needs to somehow rely on computational assumptions, because if $\mathbf{P} = \mathbf{NP}$ then statistically secure iO does exist. Our results are twofold:

1. There is no fully black-box construction of iO from (exponentially secure) collision-resistant hash functions unless the polynomial hierarchy collapses. Our lower bound extends to (separate iO from) any primitive implied by a random oracle in a black-box way.
2. Let $\mathcal{P}$ be any primitive that exists relative to random trapdoor permutations, the generic group model for any finite abelian group, or degree-$O(1)$ graded encoding model for any finite ring. We show that

⋆ 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.

⋆⋆ Supported by University of Kuwait.

⋆⋆⋆ Supported by NSF award CCF-1350939.

† 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. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the US Government.

‡ Work performed while visiting Cornell Tech, and supported by NSF CAREER Award 0845811, NSF TC Award 1111781, NSF TC Award 0939718, DARPA and AFRL under contract FA8750-11-C-0080, Microsoft New Faculty Fellowship, SAIC Scholars Research Award, and Google Research Award.

achieving a black-box construction of iO from $\mathcal{P}$ is *as hard as* basing public-key cryptography on one-way functions.
In particular, for any such primitive $\mathcal{P}$ we present a constructive procedure that takes any black-box construction of iO from $\mathcal{P}$ and turns it into a a construction of semantically secure public-key encryption form any one-way functions. Our separations hold even if the construction of iO from $\mathcal{P}$ is *semi*-black-box (Reingold, Trevisan, and Vadhan, TCC'04) and the security reduction could access the adversary in a non-black-box way.

Keywords: Indistinguishability Obfuscation, Black-Box Separations.

1 Introduction

The celebrated work of Barak et al. [3] initiated a formal study of the notion of program obfuscation which is the process of making programs unintelligible while preserving their functionalities. The main result of [3] was indeed a negative one showing that a strong form of obfuscation, called virtual black-box obfuscation, is indeed impossible for general circuits. The same work [3] also defined a weaker notion of obfuscation, called *indistinguishability* obfuscation (iO). The security of iO only requires that the obfuscation of two equivalent and same-size circuits C_1, C_2 should be computationally indistinguishable in the eyes of efficient adversaries.

The first candidate construction for iO was presented in the breakthrough work of Gentry et al. [12]. [12] showed how to construct iO for $\mathbf{NC}_1$ circuits based on multi-linear assumptions [11], and also showed how to boost iO for $\mathbf{NC}_1$ to iO for general circuits based on the learning with error (LWE) assumption [33]. The work of [12] led to an active area with a long list of results using iO as a "central hub" [36] for cryptographic tasks/primitives and basing them on iO together with one-way functions or other (relatively weak) standard assumptions. Interestingly, as shown by [23] the one-way function itself could be based on iO and the *worst-case* assumption that $\mathbf{NP} \neq \mathbf{BPP}$, leading to tens of applications solely based on iO and $\mathbf{NP} \neq \mathbf{BPP}$.[3]

Assumptions behind iO. Since the first candidate construction of iO was presented by [12] a few other variants of this constructions with different assumptions have been presented. Brakerski and Rothblum [8] showed that in an idealized model based on multilinear maps, known as the *graded encoding model* one can achieve iO (or even VBB obfuscation) assuming the bounded speedup hypothesis. Barak et al. [2] improved the result of [8] by making the construction unconditionally secure in the graded encoding model. Miles et al. [27] took another step in this direction by making the construction even more secure by allowing unlimited additions across different encoding "levels". Pass et al. [32] con-

[3] Note that we cannot hope to get OWFs from iO alone without any hardness for class $\mathbf{NP}$ since iO exist unconditionally if $\mathbf{NP} = \mathbf{P}$.

structed iO for $\mathbf{NC}_1$ circuits based on (subexponentially secure) semantically-secure multilinear encodings. Their work was the first result basing iO on falsifiable assumptions [29], however they relied on super-polynomial assumptions. Gentry et al. [15] construct iO based on subgroup elimination assumptions.

Despite all the efforts mentioned above to base iO on hardness assumptions, the assumptions behind the constructions of iO seem to be qualitatively different compared to other cryptographic primitives, even in comparison with very powerful primitives such as fully homomorphic encryption [35, 14] which could be constructed from LWE [9]. The recent beautiful work of [1] proved the first *limitation* on the power of iO by ruling out constructions of collision-resistant hash functions from iO and OWF (even if iO uses OWF in a non-black-box way).

To the best of our knowledge, no lower-bounds on the complexity of the assumptions behind iO are proved yet. The same work of [1] ruled out fully black-box constructions of iO from private-key functional encryption schemes (PFE), but this result requires the iO to also handle circuits with PFE gates. We note, however, that obfuscating circuits in *plain* model with no oracle gates is in fact sufficient for all applications of iO. Moreover, since iO exists if $\mathbf{NP} = \mathbf{P}$, any lower bound on the complexity of (standard definition of) iO in which we only aim at obfuscating circuits in the plain model *needs* to rely on computational assumptions (unless we first prove $\mathbf{NP} \neq \mathbf{P}$).

In this work we initiate a formal study on the assumption complexity of iO from a lower-bound perspective.[4] We prove our results in the black-box framework of Impagliazzo and Rudich [21] and its refinements by Reingold et al. [34]. Lower bounds against black-box constructions/reductions are considered fundamental due to the abundance of black-box techniques as well as their (typical) efficiency advantage over their non-black-box counterparts.

Since applications of iO almost always lead to non-black-box constructions, it could be argued that a black-box separation for iO is not meaningful. Note, however, that in this work we are not studying which primitives could be constructed from iO in a black-box way. We are instead looking at the complexity of assumptions behind iO. An instructive analogy is zero knowledge proofs for **NP** (ZKP). Using ZKPs for general **NP** statements also makes constructions non-black-box (since some piece of code is used as witness used by the prover) yet, we can indeed construct ZKPs for **NP** from one-way functions in a *fully* black-box way [17, 28, 18].[5] Therefore, even though iO leads to non-back-box constructions, the construction of iO itself could very much be black-box, and so separating it from classical primitives is also meaningful.

[4] As mentioned above, we do not allow oracle gates for the circuits that are going to be obfuscated.

[5] It is also instructive to note that even though ZKP for **NP** could be constructed from OWFs in a fully black-box way, it is conceivable that a separation would hold if we require a proof system for satisfiability of circuits with oracle gates.

1.1 Our Results

Our first lower bound holds for any primitive implied by a random oracle (e.g., exponentially secure one-way functions or collision-resistant hash functions) and it is proved for *fully* black-box constructions that treat the primitive and the adversary in a black-box way (see Definition 5).

Theorem 1 (Fully black-box separation from primitives implied by random oracle). *Unless the polynomial hierarchy collapses, there is no fully black-box construction of iO from collision-resistant hash functions or more generally any primitive implied by a random oracle in a black-box way.*

Intuition behind the proof. To prove theorem 1 we will first prove a useful lemma (see Lemma 17) which, roughly speaking, asserts that for any pair of circuits C_1, C_2, either (1) a (computationally unbounded) polynomial-query attacker can guess which one is obfuscated in the random oracle model with a probability close to one, or that (2) there is a way to obfuscate them into the same output circuit B. The latter could be used as a witness that C_1 and C_2 compute the same function, assuming that the obfuscation is an iO. Now consider the set of equivalent and same-size circuit $\mathcal{C} = \{(C_0, C_1) \mid C_0 \equiv C_1 \wedge |C_0| = |C_1|\}$. If Case (1) happens for an infinite subset of $\mathcal{C}$, we get a poly-query attacker against iO in the random oracle model which is sufficient for deriving the black-box separations of Theorem 1. On the other hand, if Case (1) happens only for a finite subset of $\mathcal{C}$, we get an efficient procedure to certify the equivalence of two given circuits, implying $\mathbf{NP} \neq \mathbf{P}$. We prove Lemma 17 by reducing it to a result by Mahmoody and Pass [25] who ruled out the existence of non-interactive commitments from one-way functions. Roughly speaking, we construct a non-interactive commitment scheme based on common input (C_1, C_2) in the random oracle model, and we show that: the cheating receiver strategy of [25] implies our Case (1), and the cheating sender strategy of [25] implies our Case (2). The result of [25] shows that either of these strategies always exist. See Section 3 for the formal and detailed proof.

Our second lower bound does not rule out black-box construction of iO based on believable assumptions, but shows that achieving such constructions for a large variety of primitives is as hard has solving another long standing open question in cryptography; namely, basing public-key encryption on one-way functions. It also captures a larger class of security reductions known as semi-black-box reductions [34] that allow the security reduction to access the adversary in a non-black-box way (see Definition 6).

Theorem 2 (Hardness of semi-black-box construction). *Let $\mathcal{P}$ be a primitive that provably exists relative to random trapdoor permutation oracle, the generic group model (for any finite abelian group) or the degree-$O(1)$ graded encoding model (for any finite ring). Any semi-black-box construction of iO from $\mathcal{P}$ (constructively) implies a construction of semantically secure public-key encryption from one-way functions.*

Primitives captured by Theorem 2. Theorem 2 captures a large set of powerful cryptographic primitives that could be constructed in idealized models. For example trapdoor permutations (and any primitive implied by TDPs in a black-box way) trivially exist relative to the idealized model of random TDPs. Even primitives that we do not know how to construct from TDPs in a black-box way (e.g., CCA secure public key encryption) are known to exist in the random TDP model [5]. The generic group model defined by Shoup [37] (see Definition 12) is an idealized model in which (a black-box form of) the DDH assumption holds unconditionally. Therefore, our separation of Theorem 2 covers any primitive that could be constructed from DDH assumption in a black-box way. The same holds for *bilinear* assumptions in the graded encoding model (see Definition 13) of degree 2. Namely, primitives that could be constructed from bilinear assumptions (in a black-box way) exist in the degree $O(1)$ graded encoding model unconditionally. This includes one-round 3-party key-agreement [22], identity based encryption [7], etc.

Intuition behind the proof. Our main tool in proving Theorem 2 is the following theorem which is implicit in the recent work of the authors [31, 24]. Even though the focus of [31, 24] is on virtual black-box obfuscation, the same construction presented in [31, 24] for the case of VBB implies the following theorem for iO.

Theorem 3 (Implicit in [31, 24]). *The existence of iO in any of the idealized models of: random trapdoor permutation oracle, generic group model for finite abelian groups, or the degree-$O(1)$ graded encoding model for finite rings, implies $1/p(n)$-approximate iO in the plain model for any polynomial $p(n)$.*

We then show that the existence of (1/6)-approximate iO and any one-way functions imply the existence of "approximately correct" and "approximately secure" public-key encryption schemes. In order to prove this we employ a construction of Sahai and Waters [36] using which which they showed that iO and OWF imply PKE. Here we show that the very same construction, when instantiated using *approximate* iO, leads to "approximately correct" and "approximately secure" public-key encryption. Finally we use a result of Holenstein [19] who showed how to amplify any approximately-correct and approximately-secure PKE into a full fledged (semantically secure) PKE for sufficiently good approximation! See Section 4 for the formal and detailed proof.

Remark 4. Our proof of Theorem 1 relies on perfect completeness of iO. Theorem 2 above holds even if with negligible probability over the obfuscator's randomness the obfuscated circuit does not compute the same function. Extending Theorem 1 to allow negligible error over the randomness of the obfuscator remains an interesting open questions.

Previous work on hardness *of black-box constructions.* Theorem 2 has the same spirit as the result by Impagliazzo and Rudich [21] who showed any semi-black-box construction of key agreement from one-way functions would implies $\mathbf{P} \neq \mathbf{NP}$. Therefore, the fact that we are far from proving $\mathbf{P} \neq \mathbf{NP}$ implies that we

as far from basing key agreements on one-way functions in a black-box way.[6] Similarly, our Theorem 2 shows that as long as we are not able to base public-key encryption on one-way functions, we cannot base iO on a variety of strong primitives in a semi-black-box way. Other results of the same flavor exist in connection with program checkers [6] for **NP**. Mahmoody and Xiao [26] showed that any construction of one-way functions based on worst-case hardness of **NP** implies program checkers for **NP** whose existence is one of the long standing open questions in complexity theory.

Falsifiability of iO. An intriguing open question regarding assumption complexity of iO is whether iO could be based on any "falsifiable" assumption [29]. A falsifiable assumption is one with an efficient challenger security game. The question is raised since an adversary attacking an iO scheme starts with proposing two *equivalent* circuits and and efficient challenger has no direct way to verify this. Since our primitives used in the theorems above are falsifiable, a separation of iO from falsifiable assumptions would imply our results for the case of *polynomially secure* primitives. However, constructions of iO based on exponentially secure falsifiable assumptions are indeed known [32]. Therefore, our results are interesting even if one can prove that iO cannot be based on falsifiable assumptions. Moreover, the known lower bounds against falsifiable assumptions [30, 16] are proved only for *black-box* proofs of security in which the adversary is used in a black-box way. Our Theorem 2 holds even for semi-black-box constructions in which the security reduction could use the adversary in a non-black-box manner.

2 Preliminaries

For circuits C, D, we denote by $C \equiv D$ that they compute the same function. By $|C|$ we denote the size of the bit representation of C. By a *partial* oracle f we denote an oracle that is only defined for a subset of possible queries. For random variables X, Y, by $X \approx Y$ we denote the fact that X and Y are distributed identically. We call a function $\epsilon(n)$ negligible if $\epsilon(n) < 1/p(n)$ for all polynomial $p(\cdot)$ and sufficiently large n. We call $\rho(n)$ overwhelming, if $1 - \rho(n)$ is negligible.

2.1 Black-Box Constructions

Definition 5 (Fully black-box constructions [34]). *A* fully black-box *construction of a primitive $\mathcal{Q}$ from a primitive $\mathcal{P}$ consists of two PPT algorithms (Q, S) as follows:*

1. *Implementation: For any oracle P that implements $\mathcal{P}$, Q^P implements $\mathcal{Q}$.*
2. *Security reduction: for any oracle P implementing $\mathcal{P}$ and for any (computationally unbounded) oracle adversary A successfully breaking the security of Q^P, it holds that $S^{P,A}$ breaks the security of P.*

[6] Formalizing semi-black-box constructions interpreting the result of [21] in this context is due to [34].

Primitives with stronger hardness. The above definition is for polynomially secure primitives. When the used primitive $\mathcal{P}$ is s-secure for a more quantitative bound $s(n) \gg \text{poly}(n)$, the security reduction S could potentially run in longer running time as well so long as it holds that: when P, A are polynomial time, the total running time of the composed algorithm $S^{P,A}$ is also small enough to be considered a legal attack against the implementation P of $\mathcal{P}$.

In the following more relaxed form of constructions, the security reduction can depend arbitrarily on the adversary but it still treats the implementation of the used primitive in a black-box way.

Definition 6 (Semi-black-box constructions [34]). *A* semi-black-box *construction of a primitive $\mathcal{Q}$ from a primitive $\mathcal{P}$ is defined similarly to the fully black-box Definition of 5 with the following difference in the security reduction:*

- *For any oracle P implementing $\mathcal{P}$ and any* efficient *oracle-aided adversary A^P who breaks the security of Q^P it holds that $S^P(A)$ breaks the security of P. Note that since A's description is efficient it could indeed be given to S in a non-black-box way.*

Remark 7. The work of Reingold et al. [34] also defines a "$\forall\exists$" variant of the semi-black-box constructions in which S^P can arbitrarily depend on A^P (rather than depending on it in a unified way). In this work we work with the basic default variant that we also find more natural.

Efficiency of adversary. We used the term *efficient* in an unspecified way so that it could be applied to complexity classes beyond polynomial time. For example, using a quasi-polynomially secure primitive $\mathcal{P}$ to construct a polynomially secure primitive $\mathcal{Q}$ would require a security primitive that is more relaxed and could lead to a quasi-polynomial (as opposed to polynomial) time attack against P using any polynomial-time attacker against Q^P.

2.2 Indistinguishability Obfuscation

Definition 8 ([3]). *A PPT algorithm O is an* indistinguishability obfuscator *(iO) if the following two hold:*

- *Correctness: For every circuit C, it holds that $\Pr_r[O_r(C) \equiv C] = 1$.*
- *Soundness: For every PPT adversary A there exists a negligible function $\alpha(\cdot)$ such that for every pair of equivalent circuits $C_1 \equiv C_2$ with the same size $|C_1| = |C_2| = n$ it holds that:*

$$\Pr_{r,s,b}[A_s(C_1, C_2, B) = b \colon b \xleftarrow{\$} \{0,1\}, B = O_r(C_b)] \leq 1/2 + \alpha(n)$$

where the probability is over the random seeds of the obfuscator O, adversary A and the random bit b.

Definition 9 (Approximate iO). *A PPT O is called an* ϵ-approximate *iO if it satisfies the same soundness condition and the following modified correctness condition.*

– *Approximate correctness:* $\Pr_{r,x}[O_r(C)(x) \neq C(x)] = \epsilon(|C|)$ *where the probability is over the randomness of the obfuscator as well as the randomly selected input.*

Definition 10 (Fully and semi-black-box constructions of iO). *A fully black-box construction of iO from primitive $\mathcal{P}$ consists of two oracle algorithms (O, S) such that*

– *Implementation (correctness): For every oracle P implementing $\mathcal{P}$, every circuit C, and every randomness r for O it holds that $B = O_r^P(C)$ is an oracle-aided circuit B such that $B^P \equiv C$.*
– *Soundness: For any oracle P implementing $\mathcal{P}$, any $\epsilon \geq 1/\operatorname{poly}(n)$ and any oracle adversary A who ϵ-breaks O^P, it holds that $S^{P,A}(1^{1/\epsilon(n)})$ breaks the security of P.*
We say that A ϵ-breaks O^P if for an infinite number of pairs of equivalent circuits $C_0 \equiv C_1$ of equal lengths n it holds that

$$\Pr_{r,s}[A_s(C_1, C_2, B) \colon b \xleftarrow{\$} \{0,1\}, B = O_r^P(C_b)] \geq 1/2 + \epsilon(n)$$

where the probability is over the random seeds of the obfuscator O, adversary A and the random bit b.

A semi-black-box construction of iO from $\mathcal{P}$ is defined similarly, with its soundness defined along the line of Definition 6. Namely, we require that for any efficient adversary A who ϵ-breaks O^P, there is also an efficient adversary breaking the security of P.

2.3 Generic/Idealized Models

Definition 11 (Random Oracle Model). *In the random oracle model, all parties have access to a randomized oracle f such that for each input x, the answer $f(x)$ is uniformly (and independently of the rest of the oracle) distributed over $\{0,1\}^{|x|}$.*

Definition 12 (Generic Group Model [37]). *Let $(G, \odot)$ be any group of size N and let S be any set of size at least N. The* generic *group oracle $\mathcal{I}[G \mapsto S]$ (or simply $\mathcal{I}$) is as follows. At first an injective random function $\sigma \colon G \mapsto S$ is chosen, and two type of queries are answered as follows.*

– **Labeling Queries.** *Given $g \in G$ oracle returns $\sigma(g)$.*
– **Addition Queries.** *Given y_1, y_2, if there exists x_1, x_2 such that $\sigma(x_1) = y_1$ and $\sigma(x_2) = y_2$, it returns $\sigma(x_1 \odot x_2)$. Otherwise it returns $\perp$.*

Definition 13 (Degree-d Ideal Graded Encoding Model). *The oracle $\mathcal{M}_R^d = (\mathrm{enc}, \mathrm{zero})$ is stateful and is parameterized by a ring R and a degree d and works in two phases. For each l the oracle $\mathrm{enc}(\cdot, l)$ is a random injective function from the ring R to the set of labels S.*

1. *Initialization phase: In this phase the oracle answers* $\mathrm{enc}(v, l)$ *queries and for each query it stores* (v, l, h) *in a list* $\mathcal{L}_O$.
2. *Zero testing phase: Suppose* $p(\cdot)$ *is a polynomial whose coefficients are explicitly represented in* R *and its monomials are represented with labels* $h_1, \ldots, h_m$ *obtained through* $\mathrm{enc}(\cdot, \cdot)$ *oracle in phase 1. Given any such query* $p(\cdot)$ *the oracle answers as follows:*
   (a) If any h_i *is not in* $\mathcal{L}_O$ *(i.e., it is not obtained in phase 1) return* false.
   (b) If the degree of $p(\cdot)$ *is more than* d *then return* false.
   (c) Let $(v_i, l_i, h_i) \in \mathcal{L}_O$. *If* $p(v_1, \ldots, v_m) = 0$ *return* true, *otherwise* false.

Generic Algorithms. A generic algorithm in the generic group model (resp. graded encoding model) is an algorithm in which no label s is used in an addition (resp. zero testing) query unless it is previously obtained through the oracle itself. In this work we only use *sparse* encodings in which $|S|/|G| = n^{\omega(1)}$ (resp. $|S|/|R| = n^{\omega(1)}$ in the graded encoding model) where n is the security parameter. Therefore, the execution of poly-time algorithms in this model will be statistically close to being generic.

Definition 14 (Primitives in Idealized Models). *We say a primitive* $\mathcal{P}$ *exists relative to the* randomized *oracle (or idealized model)* $\mathcal{I}$ *if there is an oracle-aided algorithm* P *such that:*

1. **Completeness:** *For every instantiation* I *of* $\mathcal{I}$, *it holds that* P^I *implements* $\mathcal{P}$ *correctly.*
2. **Security:** *Let* A *be an oracle-aided adversary* $A^{\mathcal{I}}$ *where the 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 *runs in in polynomial time (resp., quasi-polynomial time). For every such oracle aided* A, *with measure one over the sampling of the idealized oracle* $I \xleftarrow{\$} \mathcal{I}$, *it holds that* A *does* not *break the security of* P^I.

We call P *a* black-box *construction of* $\mathcal{P}$ *relative to* $\mathcal{I}$ *if the security property holds also in a "black-box" way defined as follows:*

- *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.*
  For every such oracle aided A, *with measure one over the sampling of the idealized oracle* $I \xleftarrow{\$} \mathcal{I}$, *it holds that* A *does* not *break the security of* P^I.

In the definition above, we only require the scheme to be secure after the adversary is fixed. This is along the line of the way the random oracle model is used in cryptography [5], and lets us easily derive certain primitives in idealized models. For example it is easy to see that a random trapdoor permutation, with measure one, is a secure TDP against any fixed adversary of polynomial query complexity. Therefore, TDPs exist in the idealized model of random TDP in

a black-box way. In fact stronger results are proved in the literature for other primitives. Impagliazzo and Rudich [21] and Gennaro and Trevisan [13] showed that one-way functions exist relative to the idealized model of random oracle, even if we sample the oracle first and then go over enumerating possible attacks. Chung et al. [10] proved a similar result for collision resistant hash functions.

The following lemma could be verified by inspection.

Lemma 15. *If there is a semi-black-box construction of $\mathcal{Q}$ from $\mathcal{P}$ then:*

1. *If $\mathcal{P}$ exists relative to idealized model $\mathcal{I}$, $\mathcal{Q}$ exists relative to $\mathcal{I}$ as well.*
2. *If in addition the construction of $\mathcal{P}$ from $\mathcal{I}$ is black-box, then a black-box construction of $\mathcal{Q}$ relative to $\mathcal{I}$ exists as well.*

Proof. Let Q be the semi-black-box construction of $\mathcal{Q}$ from $\mathcal{P}$. Let P be the implementation of $\mathcal{P}$ relative to $\mathcal{I}$. It is easy to see that Q^P is an implementation of $\mathcal{Q}$ relative to $\mathcal{I}$. Now we prove the security properties of Lemma 15.

1. Let A be any successful attacker against the implementation of Q^P in the idealized model $\mathcal{I}$. Then by non-zero measure over the choice of $I \xleftarrow{\$} \mathcal{I}$ it holds that A breaks the security of Q^{P^I}. For any such I, the security reduction $S^I(A)$ also breaks the security of P^I. This means that the attacker $S(A) = B$ breaks the security of P^I with non-zero measure over the sampled oracle $I \xleftarrow{\$} \mathcal{I}$. This contradicts the assumption that $\mathcal{P}$ is securely realized in $\mathcal{I}$.
2. A similar proof holds for the black-box constructions in idealized models. The only modification is that now we use $B = S^A$ (rather than $S(A)$).

3 Separating iO from Random Oracle Based Primitives

In this section we prove the following formalization of Theorem 1.

Theorem 16 (Theorem 1 formalized). *If* **NP** $\neq$ **co-NP** *then there is no fully black-box construction of iO from any primitive $\mathcal{P}$ that exists relative to a random oracle in a black-box way. This includes exponentially secure one-way functions and collision-resistant hash functions.*

To prove Theorem 16 we will first prove a useful lemma (see Lemma 17) which, roughly speaking, asserts that for any pair of circuits C_1, C_2, either an attacker can guess which one is obfuscated in the random oracle model with a probability close to one, or that there is a way to obfuscate them into the same output circuit B. The latter could be used as a witness that C_1 and C_2 compute the same function, assuming that the obfuscation is an iO.

Lemma 17 (Distinguish or Witness). *Let O be an oracle aided randomized polynomial-time algorithm taking circuits as input such that for every length-preserving oracle f and every randomness r it holds that $O_r^f(C) \equiv C$ (i.e., O always outputs circuits with the same input/output functionality as the input circuit C). Then, at least one of the following holds:*

1. *There is an* infinite *sequence of circuits* $(C_0^1, C_1^1), \ldots, (C_0^i, C_1^i), \ldots$ *such that* $|C_0^i| = |C_1^i|$ *for all* i*, and there exists a (computationally unbounded) poly*(n)*-query* A *such that the following holds for all* i*:*

$$\Pr_{r,s,f,b}[A_s^f(B) = b \colon b \xleftarrow{\$} \{0,1\}, B = O_r^f(C_b^i)] \geq 1 - 1/n^2$$

where n *is the bit size of the circuits:* $|C_0^i| = |C_1^i| = n$*.*
2. $\mathbf{NP} = \mathbf{co\text{-}NP}$.

We will first prove Theorem 16 using Lemma 17, and then we will prove Lemma 17.

Proof (of Theorem 16). In what follows we will always assume $\mathbf{NP} \neq \mathbf{co\text{-}NP}$. We will describe the proof for one-way functions, but it can be verified that the very same proof holds for any primitive that holds relative to random oracles in a black-box way (see Definition 14).

Suppose (O, S) is a fully black-box construction of iO from one-way functions. We use a random oracle f to implement the one-way function required by O. By Lemma 17 and the assumption that $\mathbf{NP} \neq \mathbf{co\text{-}NP}$ we know that there is a computationally unbounded attacker A and an infinite sequence of equivalent and same-size circuits $(C_0^1, C_1^1), \ldots, (C_0^i, C_1^i), \ldots$ such that A breaks the security of iO over the challenge circuits (C_0^i, C_1^i) of length $|C_0^i| = n = |C_1^i|$ by guessing which one of them is being obfuscated with probability $\geq 1 - 1/n^2$. Let $\epsilon = 1/4$. By an averaging argument, with probability at least $1 - O(1/n^2)$ over the choice of oracle f, it holds that the probability that A correctly guesses which one of (C_0^i, C_1^i) is being obfuscated is at least $1/2 + \epsilon$. Since the summation $\sum_i 1/i^2 = O(1)$ converges, by Borel-Cantelli lemma, for measure one of the random oracles f it holds that A ϵ-breaks the implemented iO O^f.

Now that A is a "legal" adversary, by definition of fully black-box iO, the security reduction $S^{f,A}$ shall break the one-way property of f. Algorithms A and S are both poly(n)-query attackers, and so the combination $B = S^A$ also asks only a polynomial number of queries to f and succeeds in breaking the one-wayness of f for nonzero measure of samples for f.

The existence of such B, however, is impossible since a random oracle f, with measure one, is secure against attackers who ask only a polynomial number of queries [21, 13].[7]

Now we prove Lemma 17. To prove Lemma 17 we will use the following lemma from [25].

Lemma 18 ([25]). *Suppose* S *is an oracle-aided PPT algorithm that calls oracle* f *and takes private input* $b \in \{0,1\}$*, randomness* r*, and common input*

[7] The works of [21, 13] work with polynomial time Turing machines or circuits, however their goal is to *fix* the random oracle f before enumerating the attackers. However, if the attacker is fixed before the sampling of f, the proofs of [21, 13] imply the one-wayness of f with measure one even if the fixed attacker is computationally unbounded.

$z \in \{0,1\}^n$ *(where n is the security parameter) and outputs $c = S_r^f(z,b)$. For any $\delta = \delta(n) \leq 1/100$, there is a (computationally unbounded) oracle-aided algorithm R such that for all $z \in \{0,1\}^n$ at least one of the following holds.*

1. *If f is the random oracle, $R^f(z,c)$ asks $\mathrm{poly}(n/\delta)$ queries and correctly guesses the random bit b that $S_r^f(z,b)$ used to generate c with probability $\geq 1-\delta(n)$. Namely:*

$$\Pr_{f,r,b}[R^f(z,c) = b \colon b \xleftarrow{\$} \{0,1\}, c = S_r^f(z,b)] \geq 1-\delta(n).$$

2. *There is a partial oracle f' of size $\mathrm{poly}(n)$ and two random seeds r_0, r_1 and a message c such that $S_{r_0}^{f'}(z,0) = c = S_{r_1}^{f'}(z,1)$. In other words, there is a message c that could be opened into both $b=0$ and $b=1$ using random seeds r_0, r_1, and the queries asked by S during these two possible executions are all described by the partial function f'.*

Remark 19. Mahmoody and Pass [25] proved a more general lemma ruling out (even "somewhere binding") non-interactive commitment schemes in the random oracle model. Lemma 18 is a special case of their result which is still sufficient for us. In the setting of [25] the security parameter is given to the parties in the form of 1^n, but their proof handles parties who in addition receive some $z \in \{0,1\}$ and the parties' behavior could also depend on the given z. For the sake of completeness, we have provided a self contained sketch of the proof of Lemma 18 in Appendix A.

Proof (of Lemma 17).

Consider the set of circuit pairs that are equivalent and of the same size: $\mathcal{C} = \{(C_0, C_1) \mid C_0 \equiv C_1 \wedge |C_0| = |C_1|\}$. We apply Lemma 18 for $\delta = 1/n^2$ as follows. Use $(C_0, C_1) = z \in \mathcal{C}$ as the common input given to both parties. Let S be a sender strategy that, given input bit b, obfuscates C_b and sends out the obfuscated circuit B.

By Lemma 18 for each $(C_0, C_1) = z \in \mathcal{C}$ either of the following holds:

1. $A^f((C_0, C_1), B)$ can guess the random b in the random oracle model correctly with probability at least $1 - 1/n^2$.
2. There is a partial oracle f' of polynomial size and two random strings r_0, r_1 such that $O_{r_b}^{f'}(C_b) = B$ for both $b \in \{0,1\}$.

Note that if $C_0 \not\equiv C_1$ then Case (2) cannot happen as no such (f', r_0, r_1) can exist by perfect completeness of iO. Therefore, if Case (2) happens, the existence of (f', r_0, r_1) serves as an efficiently verifiable proof that $C_0 \equiv C_1$.

Now let $\mathcal{C}_a$ be the subset of $\mathcal{C}$ for which Case (1) holds. There are two cases:

1. $\mathcal{C}_a$ is not finite, in which case we have shown that Case 1 of the lemma holds.
2. If $\mathcal{C}_a$ is finite, then for all (except a finite number) of $(C_0, C_1) \in \mathcal{C}$ we can efficiently prove that C_0, C_1 are equivalent circuits. This would give a proof system for proving the equivalency of two given circuits, but this problem is **co-NP**-complete. Thus, **co-NP** = **NP**.

4 Hardness of Semi-Black-Box Constructions of iO

In this section, we prove Theorem 2. We will first show that approximate iO is still powerful enough to base public-key cryptography on private-key cryptography. We will use this result and results of [31, 24] to derive Theorem 2.

Theorem 20. *The existence of* $(1/6)$*-approximate iO and any one-way functions imply the existence of semantically secure public-key encryption schemes.*

We first prove Theorem 2 using Theorems 3 and 20.

Proof (of Theorem 2 using Theorems 3 and 20). Let $\mathcal{P}$ be any such primitive with implementation P relative to the idealized model $\mathcal{I}$, and suppose O is any such semi-black-box construction of iO from $\mathcal{P}$. By Lemma 15, we conclude that $O' = O^P$ is a construction of iO in the idealized model $\mathcal{I}$. This, together with Theorem 3 imply that there is a $(1/6)$-approximate iO in the plain model. Finally, by Theorem 20 and the existence of $(1/6)$-approximate iO implies that we can construct semantically secure public-key encryption from one-way functions.

4.1 Proving Theorem 20

In this section we prove that $(1/6)$-approximate iO and one-way functions imply semantically secure public-key encryption. Therefore, any provably secure construction of $(1/6)$-approximate iO would enable us to take any one-way functions and construct a secure public-key encryption scheme from it. In the terminology of [20] it means that Cryptomania collapses to Minicrypt if $(1/6)$-approximate iO exists.

Intuition. Sahai and Waters [36] showed that iO and OWF imply PKE. Here we show that the very same construction, when instantiated using *approximate* iO, leads to "approximately correct" and "approximately secure" public-key encryption. Then, using a result of [19] we amplify the soundness and correctness to get a full fledged semantically secure public key encryption scheme.

Definition 21 (Approximate correctness and security for PKE). *We call a public-key bit-encryption scheme* $(\mathrm{Gen}, \mathrm{Enc}, \mathrm{Dec})$ *for message space* $\{0,1\}$ $\epsilon(n)$*-correct if*

$$\Pr[\mathrm{Dec}_{dk}(\mathrm{Enc}_{ek}(b)) = b \colon (ek, dk) \leftarrow \mathrm{Gen}(1^n), b \xleftarrow{\$} \{0,1\}] \geq 1 - \epsilon(n)$$

where the probability is over the randomness of the key generation, encryption, decryption, and the bit b*. We call* $(\mathrm{Gen}, \mathrm{Enc}, \mathrm{Dec})$ $\delta(n)$*-secure if for any PPT adversary* A*, it holds that*

$$\Pr[A(pk, \mathrm{Enc}_{pk}(b)) = b] \leq 1/2 + \delta(n)$$

where the probability is over the randomness of generation, encryption, the adversary, and bit b*.*

Holenstein [19] showed how to amplify any ϵ-correct and ϵ-secure PKE into a full fledged (semantically secure) PKE for sufficiently small ϵ.

Theorem 22 (Implied by Corollary 7.8 in [19]). *Suppose* $(\mathrm{Gen}, \mathrm{Enc}, \mathrm{Dec})$ *is* ϵ*-correct and* δ*-secure for* constants ϵ, δ *such that* $(1-2\epsilon)^2 > 2\delta$*. Then there exists a semantically secure PKE.*

Theorem 23 below asserts that approximate iO and one-way functions imply approximately correct and approximately secure PKE.

Theorem 23 (Approximate iO + OWF $\Rightarrow$ Approximate PKE). *If* ϵ*-approximate iO and one-way functions exist, then there is an* ϵ*-correct and* $(\epsilon + \mathrm{negl}(n))$*-secure public-key bit encryption scheme.*

We first prove Theorem 20 using Theorem 23 and then will prove 23.

Proof (of Theorem 20). Because $(1 - 2 \cdot 1/6)^2 > 2 \cdot 1/6$, Theorem 20 follows immediately from Theorem 22 and the following Theorem 23 using $\epsilon = 1/6$.

In the rest of this section we prove Theorem 23.

Proof (of Theorem 23). We show that the very same construction of PKE from iO and OWF presented by Sahai and Waters [36], when instantiated with an ϵ-approximate iO, has the demanded properties of Theorem 23.

Properties of the construction of [36]. We first describe the abstract properties of the construction of [36] (for PKE using iO and OWFs) and its security proof that we need to know.

- Construction/correctness:
  1. The key generation process generates a circuit C and publishes $O(C) = B$ as public key where O is an iO scheme.
  2. The encryption simply runs B on (r, b) where r is the encryption randomness and b is the bit to be encrypted.
  3. The scheme has completeness 1.
- Security: [36] proves the security of the construction above by showing that no PPT algorithm can distinguish between the following two random variables X_0, X_1 defined as:
  - $X_b \approx (O_s(C), C(r, b)) \approx (B, C(r, b))$ where s is the randomness for O. When clear from the context we drop the randomness s and simply write $O(C)$ denoting it as a random variable over the randomness of O.

  It can be verified by inspection that the proof of [36] for indistinguishability of X_0 and X_1 does *not* rely on completeness of the obfuscation O and only relies on its indistinguishability (when applied to circuits with the same functions). We will rely on this feature of the proof of [36] in our analysis.

Below we analyze the correctness and security of the construction of [36] when O is an ϵ-approximate iO.

Correctness. By the definition of ϵ-approximate iO and ϵ-correct bit encryption, and the fact that the [36] construction has perfect completeness when O is iO, it follows that the completeness of the new scheme (when b is also chosen at random) is at least $1 - \epsilon$. Thus the scheme is ϵ-correct.

Security. First recall that for the basic construction of [36] using (perfect) iO, no PPT attacker A can guess b with probability better than $1/2 + \text{negl}(n)$ when b is chosen at random and A is given a sample from the random variable X_b (for random b). As we mentioned above, the proof of this statement does not rely on the correctness of the used iO and only relies on its indistinguishability.

Now we want to bound the distinguishing advantage of PPT adversaries between the following random variables Y_0, Y_1:

- $Y_b \approx (B, B(r, b)) \approx (O_s(C), O_s(C)(r, b))$ where s is the randomness of the obfuscator O.

The difference between Y_b's and X_b's stems from the fact that the public-key $B = O(C)$ no longer computes the same exact function as the circuit C as the obfuscation only guarantees *approximate* correctness. We reduce the analysis of the new scheme to the original analysis of [36].

By the analysis of [36] we already know that if any PPT A is given a sample from X_b for a random b it has at most $1/2 + \text{negl}(n)$ chance of correctly guessing b. Also note that the distributions X_b and Y_b for a *random* b are ϵ-close due to the ϵ-correctness of the obfuscation. More formally, the distributions X_b and Y_b could be defined over the same sampling space using: random seeds of key generation, obfuscation, encryption, and bit b. This way with probability $\geq 1 - \epsilon$ (and by the ϵ-approximate correctness of the obfuscation) the actual sampled values of X_b and Y_b will be equal, and this implies that they are ϵ-close. As a result, when we switch the distribution of the challenge given to the adversary and give a sample of Y_b (for random b) instead of a sample from X_b, the adversary's chance of guessing b correctly can increase at most by ϵ and reach at most $1/2 + \text{negl} + \epsilon$. Therefore, the new scheme is $(\epsilon + \text{negl})$-secure according to Definition 21.

References

1. Gilad Asharov and Gil Segev. Limits on the power of indistinguishability obfuscation and functional encryption. Cryptology ePrint Archive, Report 2015/341, 2015. `http://eprint.iacr.org/`.
2. Boaz Barak, Sanjam Garg, Yael Tauman Kalai, Omer Paneth, and Amit Sahai. Protecting obfuscation against algebraic attacks. In *Advances in Cryptology–EUROCRYPT 2014*, pages 221–238. Springer, 2014.
3. Boaz Barak, Oded Goldreich, Russell Impagliazzo, Steven Rudich, Amit Sahai, Salil Vadhan, and Ke Yang. On the (im) possibility of obfuscating programs. *Journal of the ACM (JACM)*, 59(2):6, 2012.
4. Boaz Barak and Mohammad Mahmoody-Ghidary. Lower bounds on signatures from symmetric primitives. In *IEEE Symposium on Foundations of Computer Science (FOCS)*, 2007.

5. 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.
6. Manuel Blum and Sampath Kannan. Designing programs that check their work. *Journal of the ACM (JACM)*, 42(1):269–291, 1995.
7. Dan Boneh and Matthew Franklin. Identity-based encryption from the Weil pairing. *SIAM Journal on Computing*, 32(3):586–615, June 2003. Preliminary version in CRYPTO '01.
8. Zvika Brakerski and Guy N Rothblum. Virtual black-box obfuscation for all circuits via generic graded encoding. In *Theory of Cryptography Conference, TCC*, pages 1–25. Springer, 2014.
9. Zvika Brakerski and Vinod Vaikuntanathan. Efficient fully homomorphic encryption from (standard) lwe. *SIAM Journal on Computing*, 43(2):831–871, 2014.
10. Kai-Min Chung, Huijia Lin, Mohammad Mahmoody, and Rafael Pass. On the power of nonuniformity in proofs of security. In *Proceedings of the 4th conference on Innovations in Theoretical Computer Science*, pages 389–400. ACM, 2013.
11. Sanjam Garg, Craig Gentry, and Shai Halevi. Candidate multilinear maps from ideal lattices. In *Eurocrypt*, volume 7881, pages 1–17. Springer, 2013.
12. 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.
13. Rosario Gennaro and Luca Trevisan. Lower Bounds on the Efficiency of Generic Cryptographic constructions. In *IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 305–313, 2000.
14. Craig Gentry. *A fully homomorphic encryption scheme*. PhD thesis, Stanford University, 2009. `crypto.stanford.edu/craig`.
15. Craig Gentry, Allison B Lewko, Amit Sahai, and Brent Waters. Indistinguishability obfuscation from the multilinear subgroup elimination assumption. *IACR Cryptology ePrint Archive*, 2014:309, 2014.
16. Craig Gentry and Daniel Wichs. Separating succinct non-interactive arguments from all falsifiable assumptions. In Lance Fortnow and Salil P. Vadhan, editors, *ACM Symposium on Theory of Computing (STOC)*, pages 99–108, 2011.
17. Oded Goldreich, Silvio Micali, and Avi Wigderson. Proofs that yield nothing but their validity and a methodology of cryptographic protocol design (extended abstract). In *IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 174–187. IEEE, 1986.
18. Johan Håstad, Russell Impagliazzo, Leonid A. Levin, and Michael Luby. A pseudorandom generator from any one-way function. *SIAM Journal on Computing*, 28(4):1364–1396, 1999.
19. Thomas Holenstein. *Strengthening key agreement using hard-core sets*. PhD thesis, ETH ZURICH, 2006.
20. Russell Impagliazzo. A personal view of average-case complexity. In *Structure in Complexity Theory Conference*, pages 134–147, 1995.
21. Russell Impagliazzo and Steven Rudich. Limits on the provable consequences of one-way permutations. In *ACM Symposium on Theory of Computing (STOC)*, pages 44–61. ACM Press, 1989.
22. Antoine Joux. A one round protocol for tripartite diffie–hellman. In *Algorithmic number theory*, pages 385–393. Springer, 2000.

23. Ilan Komargodski, Tom Moran, Moni Naor, Rafael Pass, Arye Rosen, and Eylon Yogev. One-way functions and (im) perfect obfuscation. In *IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 374–383. IEEE, 2014.
24. 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/`.
25. Mohammad Mahmoody and Rafael Pass. The curious case of non-interactive commitments - on the power of black-box vs. non-black-box use of primitives. In Reihaneh Safavi-Naini and Ran Canetti, editors, *CRYPTO*, volume 7417 of *Lecture Notes in Computer Science*, pages 701–718. Springer, 2012.
26. Mohammad Mahmoody and David Xiao. On the power of randomized reductions and the checkability of sat. In *IEEE Conference on Computational Complexity*. IEEE Computer Society, 2010.
27. Eric Miles, Amit Sahai, and Mor Weiss. Protecting obfuscation against arithmetic attacks. Cryptology ePrint Archive, Report 2014/878, 2014. `http://eprint.iacr.org/`.
28. Moni Naor. Bit commitment using pseudorandomness. *Journal of Cryptology*, 4(2):151–158, 1991.
29. Moni Naor. On cryptographic assumptions and challenges. In *CRYPTO 2003*, pages 96–109. Springer, 2003.
30. Rafael Pass. Limits of provable security from standard assumptions. In Lance Fortnow and Salil P. Vadhan, editors, *ACM Symposium on Theory of Computing, (STOC)*, pages 109–118. ACM, 2011.
31. 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/`.
32. Rafael Pass, Karn Seth, and Sidharth Telang. Indistinguishability obfuscation from semantically-secure multilinear encodings. In *CRYPTO 2014*, pages 500–517. Springer, 2014.
33. Oded Regev. On lattices, learning with errors, random linear codes, and cryptography. In *ACM Symposium on Theory of Computing (STOC)*, 2005.
34. 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.
35. Ronald L. Rivest, Len Adleman, and Michael L. Dertouzos. On data banks and privacy homomorphisms. In *Foundations of secure computation (Workshop, Georgia Inst. Tech., Atlanta, Ga., 1977)*, pages 169–179. Academic, New York, 1978.
36. Amit Sahai and Brent Waters. How to use indistinguishability obfuscation: Deniable encryption, and more. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, pages 475–484. ACM, 2014.
37. Victor Shoup. Lower bounds for discrete logarithms and related problems. In Walter Fumy, editor, *Advances in Cryptology EUROCRYPT 97*, volume 1233 of *Lecture Notes in Computer Science*, pages 256–266. Springer Berlin Heidelberg, 1997.

A Omitted Proofs

For sake of completeness we sketch the proof of Lemma 18.

Proof (Sketch of Lemma 18). Let ϵ be a parameter to be chosen later. Let R be an attacker who maintains a list of "learned" oracle queries $\mathcal{L}$ and, given c sent by the sender for $b \xleftarrow{\$} \{0,1\}$ and common input z, it adaptively asks the lexicographically first oracle query $x \notin \mathcal{L}$ that has at least ϵ chance of being asked by sender S conditioned on the knowledge of $(\mathcal{L}, z)$. After asking such x from f, A adds $(x, f(x))$ to $\mathcal{L}$. As long as such query x exists, R asks them. It was shown in [4] that this learning algorithm asks, on average, at most m/ϵ number of queries where $m = \text{poly}(|z|)$ is the number of queries asked by the sender. So as long as $\epsilon = \text{poly}(n/\delta)$ this learning algorithm is efficient.

Now, let $\mathcal{L}$ be the final learned set by R. If conditioned on $\mathcal{L}$ it holds that both of $b = 0$ and $b = 1$ have at least ρ probability of being used by S, then by conditioning on the distribution of the sender's view on $b = 0$ or $b = 1$ all the unlearned queries remain *at most* $\epsilon/\rho = \sigma$-heavy. Now it is easy to see that if we sample a random view for S conditioned on $\mathcal{L}, b = 0$ and $\mathcal{L}, b = 1$ and call them V_0 and V_1, the probability that queries of V_0 and V_1 collide out of $\mathcal{L}$ is at most $m \cdot \sigma \leq m \cdot \epsilon/\rho$. For $\rho > m \cdot \epsilon$ this probability is less than one, which means that if $\rho > m \cdot \epsilon$, then there exists a consistent pair of views for S that he can use to output c for both cases of $b = 0$ and $b = 1$. This means that Case 2 happens.

Now let us assume that Case 2 does not happen. It means that for all executions of the algorithm A, when A is done with learning the ϵ heavy queries, the probability of either $b = 1$ or $b = 0$ conditioned on $\mathcal{L}$ is at most ϵ. This means that A can guess b correctly with probability $1 - \epsilon$.

If we can choose $\rho = O(m/\epsilon)$ and $\epsilon = \delta$ in the argument above (assuming that Case 2 does not happen) we get an attacker A that asks $O(m \cdot \epsilon/\epsilon) = O(m)$ queries. We can alternatively choose smaller ρ and cut A's execution after it asks $O(m/\delta)$ number of queries and use $\epsilon = \delta/10$. By an application of the Markov inequality A will ask more than $100(m/\delta)$ number of queries with probability at most ϵ, and so A will ask at most $O(m/\delta)$ number of queries and will guess b correctly with probability at least $2\epsilon < \delta$.