On Statistically Secure Obfuscation with Approximate Correctness

Zvika Brakerski[⋆,1], Christina Brzuska[⋆⋆,⋆⋆⋆,2], and Nils Fleischhacker[⋆⋆⋆,†,3]

[1] Weizmann Institute of Science, Rehovot, Israel
[2] Technical University of Hamburg, Hamburg, Germany
[3] CISPA, Saarland University, Saarbrücken, Germany

Abstract. Goldwasser and Rothblum (TCC '07) prove that statistical indistinguishability obfuscation (iO) cannot exist if the obfuscator must maintain perfect correctness (under a widely believed complexity theoretic assumption: $\mathcal{NP} \not\subseteq \mathcal{SZK} \subseteq \mathcal{AM} \cap \mathbf{co}\mathcal{AM}$). However, for many applications of iO, such as constructing public-key encryption from one-way functions (one of the main open problems in theoretical cryptography), *approximate* correctness is sufficient. It had been unknown thus far whether statistical approximate iO (saiO) can exist.

We show that saiO does not exist, even for a minimal correctness requirement, if $\mathcal{NP} \not\subseteq \mathcal{AM} \cap \mathbf{co}\mathcal{AM}$, and if one-way functions exist. A simple complementary observation shows that if one-way functions do not exist, then average-case saiO exists. Technically, previous approaches utilized the behavior of the obfuscator on *evasive* functions, for which saiO always exists. We overcome this barrier by using a PRF as a "baseline" for the obfuscated program.

We broaden our study and consider relaxed notions of *security* for iO. We introduce the notion of *correlation obfuscation*, where the obfuscations of equivalent circuits only need to be mildly correlated (rather than statistically indistinguishable). Perhaps surprisingly, we show that correlation obfuscators exist via a trivial construction for some parameter regimes, whereas our impossibility result extends to other regimes. Interestingly, within the gap between the parameters regimes that we show possible and impossible, there is a small fraction of parameters that still allow to build public-key encryption from one-way functions and thus deserve further investigation.

⋆ Supported by the Israel Science Foundation (Grant No. 468/14), the Alon Young Faculty Fellowship, Binational Science Foundation (Grant No. 712307) and Google Faculty Research Award.

⋆⋆ Christina Brzuska is grateful to NXP for supporting her chair for IT Security Analysis.

⋆⋆⋆ Part of this work was done while Christina Brzuska and Nils Fleischhacker were working for Microsoft Research, Cambridge.

† Supported by the German Federal Ministry of Education and Research (BMBF) through funding for the Center for IT-Security, Privacy and Accountability (CISPA – `www.cispa-security.org`) and the German research foundation (DFG) through funding for the collaborative research center 1223.

1 Introduction

Constructing public-key cryptography (e.g. public-key encryption) from private-key cryptography (such as one-way functions) is one of the most fundamental questions in theoretical cryptography, going back to the seminal paper of Diffie and Hellman [9]. Diffie and Hellman suggested that *program obfuscators* with sufficiently strong security properties would allow to realize this transformation. A program obfuscator is a compiler that takes as input a program, and outputs another program with equivalent functionality, but which is harder to reverse engineer. Diffie and Hellman suggested to obfuscate the encryption circuit of a symmetric-key encryption scheme, and use the obfuscated program as a public key so as to obtain a public-key encryption scheme. An additional hint that obfuscation may be instrumental in solving this riddle was provided by Impagliazzo and Rudich [21,20], who proved that a transformation from symmetric to public-key must make *non black-box* use of the underlying symmetric primitive. Indeed, program obfuscation is one of very few non black-box techniques known in cryptography.

Modern research showed that the Diffie-Hellman transformation requires obfuscators with security guarantees that do not exist in general [16,1,2]. However, recent years have seen incredibly prolific study of weak notions of obfuscation, following the introduction of a candidate *indistinguishability obfuscator* (iO) by Garg et al. [10]. The security guarantee of iO is that the obfuscation of two functionally equivalent circuits should result in indistinguishable output distributions. That is, that reverse engineering could not detect which of two equivalent implementations had been the source of the obfuscated program. Sahai and Waters [30] showed that even this seemingly weak notion suffices for private-key to public-key transformation (via a clever construction that does not resemble the Diffie-Hellman suggestion).

One would have hoped that a weak notion such as iO may be realizable with *statistical* security, i.e. that reverse engineering (to the limited extent required by iO) will not be possible even to an attacker with unlimited computational power. The existence of such *statistical indistinguishability obfuscator* (siO) would resolve the question of constructing public key cryptography from one-way functions, as well as would allow to construct one-way functions based on the hardness of $\mathcal{NP}$ [23]. Alas, Goldwasser and Rothblum [14,15] proved that siO cannot exist unless the polynomial hierarchy collapses (in particular that it implies $\mathcal{NP} \subseteq \mathcal{SZK}$, and it is known that $\mathcal{SZK} \subseteq \mathcal{AM} \cap \mathbf{co}\mathcal{AM}$), which is considered quite unlikely in computational complexity, and at any rate way beyond the current understanding of complexity theory. This seems to put a damper on our hopes to achieve statistically secure obfuscation.

However, the [14,15] negative result crucially relies on the *correctness* of the obfuscator. That is, it only rules out such obfuscators that perfectly preserve the functionality of the underlying primitive (at least with high probability over the coins of the obfuscator). In contrast, the symmetric to public key transformation can be made to work with only *approximate* correctness, i.e. a non-negligible correlation between the functionality of the input circuit and that of the output

circuit (where the probability is taken over the randomness of the obfuscator and the input domain). The question of whether statistical approximate iO (saiO) exists was therefore the new destination in the quest for understanding obfuscation. Interestingly, it turns out that ruling out *computational* notions of iO in some idealized models also boils down to the question of whether saiO exists (see Section 1.2 below). The study of this notion is the objective of this paper.

Our Results. We show that statistical approximate iO (saiO) does not exist if one-way functions exist (under the assumption that $\mathcal{NP} \not\subseteq \mathcal{AM} \cap \mathbf{co}\mathcal{AM}$). Thus, in particular, that saiO cannot be used for the transformation from symmetric to public-key cryptography. We show that if one-way functions exist, then any non-negligible correlation between the output of the obfuscator and the input program would imply an $\mathcal{SZK}$ algorithm for unique SAT (USAT). As SAT reduces to USAT via a randomized reduction [32], a result of Mahmoody and Xiao [27] shows that this implies that SAT is in $\mathcal{AM} \cap \mathbf{co}\mathcal{AM}$.

To complement our result, we observe that if one-way functions do not exist, then an average-case notion of saiO exists for any distribution. Specifically, for any efficiently samplable distribution over circuits, there exists an saiO obfuscator whose correctness holds with high probability over the circuits in that distribution (inverting the order of quantifiers would imply a worst-case saiO).

A Study of Correlation Obfuscation. Our impossibility results extend beyond the case of saiO. In fact, the result applies even when the *security* of the obfuscator is approximate. Namely, when we are only guaranteed that the obfuscation of functionally equivalent circuits results in distributions that have mild statistical distance (as opposed to negligible). This motivated us to explore the properties of this new kind of obfuscators, that as far as we know have not been studied in the literature before.

We consider statistical approximate *correlation* obfuscation sacO. A sacO obfuscator is characterized by two parameters $\epsilon \in [0, 1/2)$ and $\delta \in [0, 1)$. The requirement is that correctness holds with probability $1 - \epsilon$ (with respect to the randomness of the obfuscator and a random choice of input), and that obfuscating two functionally equivalent circuits results in distributions with statistical distance δ. The case of negligible δ is exactly saiO, discussed above, and the case of $\epsilon = 0$ corresponds to perfect correctness.

We observe that our impossibility result degrades gracefully and holds so long as $2\epsilon + 3\delta < 1$. We found this state of affairs unsatisfactory, and tried to extend the result to hold for the entire parameter range. However, it turns out that sacO exists via an almost trivial construction whenever $2\epsilon + \delta > 1$ (e.g. $\epsilon = \delta = 0.4$). We do not know if sacO exists in the intermediate parameter regime.

Lastly, we conduct a study of whether sacO is sufficient to construct public-key encryption from one-way functions. We present an amplified version of the Sahai-Waters construction using an amplification technique due to Holenstein. Interestingly, it appears that there is a region in the parameter domain that would allow to construct public-key encryption from one-way functions, but is

not ruled out by our current technique. See Figure 1 for the landscape of sacO parameters. We leave it as an intriguing open problem to close the gap between the various parameter regimes.

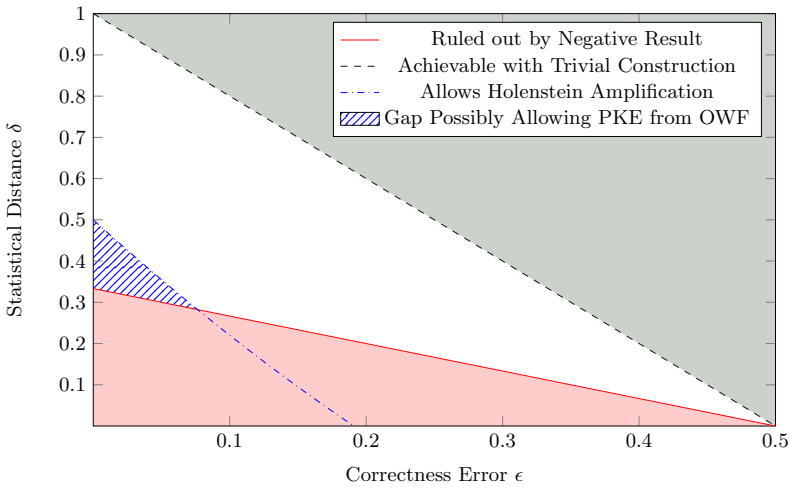


Fig. 1. The graph gives an overview over the possible range of parameters for sacO. In the upper right are parameter regimes that can be achieved using the construction described in Appendix A. In the lower left are the strong parameter regimes ruled out by our negative result in Section 3. The graph shows nicely the gap between the parameters that can be ruled out and those that can be used to construct public key encryption using the construction of Sahai and Waters as well as the amplification technique of Holenstein.

1.1 Our Techniques

Our starting point is the Goldwasser-Rothblum impossibility result. Consider a statistical iO obfuscator such that for any pair of functionally equivalent circuits, the obfuscator generates statistically indistinguishable distributions, and in addition the output circuit of the obfuscator is always *functionally equivalent* to the input circuit (this can be relaxed to hold only with high probability over the random coins of the obfuscator). Goldwasser-Rothblum observe that an unsatisfiable SAT formula Ψ is functionally equivalent to the all-zero function $\mathbf{0}$ and therefore the distributions produced by a siO obfuscator in both cases should be statistically indistinguishable. Slightly more formally, let $X[C]$ denote the distribution output by the obfuscator on input circuit C, then we get that $X[\Psi] \equiv X[\mathbf{0}]$, where $\equiv$ denotes statistical indistinguishability. On the contrary, if Ψ is a satisfiable formula, then it has a different functionality than $\mathbf{0}$ and therefore the support of $X[\Psi]$ and $X[\mathbf{0}]$ will be disjoint (and thus obviously not

statistically indistinguishable). It follows that in order to solve SAT, it suffices to tell whether $X[\Psi]$ is close to $X[\mathbf{0}]$. As we know due to Sahai and Vadhan [29], there is an $\mathcal{SZK}$ protocol that takes two polynomial-time samplers, and decides whether they sample from distributions that are ϵ_1-statistically close or ϵ_2-statistically far, so long as $(\epsilon_2 - \epsilon_1)$ is a noticeable function. The conclusion is that an siO obfuscator implies an $\mathcal{SZK}$ protocol for SAT which in turn implies that $\mathcal{NP} \subseteq \mathcal{SZK}$.

To sum up the core argument, to show that an siO obfuscator does not exists unless $\mathcal{NP} \subseteq \mathcal{SZK}$, Goldwasser-Rothblum built the formula-indexed distribution $X[\Psi]$ that samples an siO obfuscation of Ψ and has the properties that it is (i) efficiently sampleable, (ii) if Ψ is not satisfiable, then $X[\Psi]$ and $X[\mathbf{0}]$ are close, while (iii) if Ψ not satisfiable, then $X[\Psi]$ and $X[\mathbf{0}]$ are far.

Allowing the obfuscator to have approximate correctness thwarts this approach completely. Hard SAT instances are obviously ones where the density of accepting inputs is sub-polynomial, since otherwise random sampling would yield a satisfying assignment with non-negligible probability. Therefore a satisfiable and unsatisfiable SAT formulae will have almost identical functionality. One could consider an saiO obfuscator that on any SAT formula that is not trivially satisfiable, would just produce an obfuscation of $\mathbf{0}$. This means that $X[\Psi]$ will have the same distribution whether Ψ is satisfiable or not and thus, property (iii) is not satisfied anymore.

In order to overcome this issue, we construct a different distribution on formula-indexed circuits $C_X[k, \Psi]$ (where k is some uniformly random key k) such that if Ψ is not satisfiable, then $C_X[k, \Psi]$ and $C_X[k, \mathbf{0}]$ have the same functionality, and if Ψ is satisfiable, then $C_X[k, \Psi]$ and $C_X[k, \mathbf{0}]$ differ on a single point. Then, assuming one-way functions exist, we show that, although these two circuits differ on a single point only, the obfuscator saiO of $C_X[k, \Psi]$ has to produce a distribution that is statistically far from saiO of $C_X[k, \mathbf{0}]$. To do this, we rely on the fact that the obfuscator itself is computationally efficient, and therefore it cannot break the hardness of one-way functions and derived cryptographic objects such as pseudorandom functions (PRFs) or *puncturable* PRFs (see below). This way, we construct a new formula-indexed distribution $X[\Psi]$ that satisfies properties (i), (ii) and (iii) as discussed above.

Puncturable PRFs were introduced simultaneously in [6,7,22] and were utilized as an essential building block for indistinguishability obfuscation in [30]. A standard PRF is a function that can be efficiently computable using a key k, but is indistinguishable from a random function via oracle access. A puncturable PRF is a PRF where one can generate a *punctured key* $k\{x_0\}$ which allows to compute the PRF at all points except x_0, but the value at x_0 is still indistinguishable from uniform, even given the punctured key. Punctured PRFs can be constructed from any one-way function.

Based on a puncurable PRF and an saiO obfuscator O, we now construct a distribution on pairs of circuits (for now not indexed by a formula) such that the two circuits differ on a single point only and yet, an saiO obfuscator will produce distributions that are far. Let k be a key for a puncturable

PRF, let x_0 be a random point in the domain, let $k\{x_0\}$ be a key punctured at x_0 and consider the function $f_{k\{x_0\},y}$ that outputs $\mathsf{PRF}(k\{x_0\}, x) = \mathsf{PRF}(k, x)$ for all $x \neq x_0$, and outputs y on input x_0. Then by definition $f_{k\{x_0\},y}$ for a random y and $f_{k\{x_0\},y_0} = \mathsf{PRF}(k, \cdot)$ for $y_0 = \mathsf{PRF}(k, x_0)$ are identical in functionality except maybe at point x_0. However, using puncturing, we can guarantee that the distributions $\mathsf{O}(f_{k\{x_0\},y})$ and $\mathsf{O}(f_{k\{x_0\},y_0})$, where k, x_0, y are chosen uniformly at random are *statistically far*. To see this, it is enough to show that $\mathsf{O}(f_{k\{x_0\},y})$ and $\mathsf{O}(\mathsf{PRF}(k, \cdot))$ are statistically far since $f_{k\{x_0\},y_0} = \mathsf{PRF}(k, \cdot)$ and thus $\mathsf{O}(f_{k\{x_0\},y_0}) \equiv \mathsf{O}(\mathsf{PRF}(k, \cdot))$. Consider the predicate that checks whether $\mathsf{O}(\mathsf{PRF}(k, \cdot))(x_0) = \mathsf{PRF}(k, x_0)$. This predicate must have non-negligible bias towards holding true, and is efficiently checkable, which also implies that $\mathsf{O}(f_{k\{x_0\},y})(x_0) = f_{k\{x_0\},y}(x_0)$ holds true with noticeable bias, since otherwise we will have an efficient distinguisher from $f_{k\{x_0\},y_0} = \mathsf{PRF}(k, \cdot)$ in contradiction to the puncturable PRF security. Finally, since $y \neq y_0$ with high probability (assume for simplicity that the PRF and the obfuscator have long outputs and keys of half the size), this implies that $\mathsf{O}(f_{k\{x_0\},y})$ and $\mathsf{O}(f_{k\{x_0\},y_0})$ have noticeable statistical distance, since they will have noticeable probability mass on circuits that respect the functionality on x_0. Note that we used a *computational* argument, the security of punctured PRFs, to derive a *statistical* statement about the output distribution of the obfuscator.

We would like to use the aforementioned distributions to distinguish between satisfiable and unsatisfiable formulae. Let us restrict our attention to Unique-SAT formulae that are either unsatisfiable or have only one satisfying assignment. Unique-SAT is known to be $\mathcal{NP}$-Hard via a randomized reduction [32], and a result of Mahmoody and Xiao [27] shows that if Unique-SAT is in $\mathcal{SZK} \subseteq \mathcal{AM} \cap \mathbf{co}\mathcal{AM}$, then SAT is in $\mathcal{AM} \cap \mathbf{co}\mathcal{AM}$ (See Section 2.1).

Let Ψ be a formula that has a unique satisfying assignment, then one can randomize the satisfying assignment (if it exists) to be uniformly distributed over the input space (e.g. by XORing all variables with a random string). Now, consider the function $f_{k,y,\Psi}$ defined s.t. $f_{k,y,\Psi}(x) = \mathsf{PRF}(k, x)$ if x does not satisfy Ψ, and $f_{k,y,\Psi}(x) = y$ otherwise. By definition, if Ψ is unsatisfiable then $f_{k,y,\Psi} = \mathsf{PRF}(k, \cdot)$ and if Ψ is satisfiable by some x_0 (which is uniformly distributed) then $f_{k,y,\Psi} = f_{k\{x_0\},y}$. Therefore $\mathsf{O}(f_{k,y,\Psi})$ is guaranteed to have a noticeable statistical distance in the case where Ψ is unsatisfiable (in which case it is close to $\mathsf{O}(f_{k,y,\mathbf{0}})$) and in the case where it is uniquely satisfiable (in which case it is far from $\mathsf{O}(f_{k,y,\mathbf{0}})$). This will allow us to produce an $\mathcal{SZK}$ protocol to distinguish the two possibilities.

In a World without OWFs. We recall that if OWFs do not exist then for any efficiently computable function f and with overwhelming probability over a y sampled from the output distribution of f, it is possible to efficiently sample (almost) uniformly (up to negligible error) from the set $f^{-1}(y) = \{x : f(x) = y\}$ [19]. Given an efficiently sampleable distribution over circuits, we can construct an average-case obfuscator for this family as follows. Let sampC be a sampler for this distribution of circuits and consider the function $f(r, x_1, \ldots, x_m)$ for a large

polynomial m such that $f(r, x_1, \ldots, x_m) = (x_1, \ldots, x_m, C(x_1), \ldots, C(x_m))$, for $C = \mathsf{sampC}(r)$.

Now, to obfuscate a circuit C, sample $x_1, \ldots, x_m$ and compute $y_i = C(x_i)$. Then sample $(r, x_1, \ldots, x_m)$ from $f^{-1}(x_1, \ldots, x_m, y_1, \ldots, y_m)$ and finally output $C' = \mathsf{sampC}(r)$. This is clearly a perfect indistinguishability obfuscator (i.e. two circuits with the same functionality will produce identical distributions). It is also approximately correct on the average, because on average, if two circuits agree on a randomly chosen set of points, then they will have a large agreement altogether.

We note that a similar and even simpler argument shows that if all efficiently computable functions are PAC learnable [31], even allowing membership queries, then saiO with perfect indistinguishability exists. This follows immediately by definition by giving the learner (black-box) access to C, and outputting its hypothesis C' as the output of the obfuscator. In such case OWFs trivially do not exist.

The Landscape of Correlation Obfuscation. Extending our techniques to rule out sacO with $2\epsilon + 3\delta < 1$ follows from carefully analyzing the parameters in the proof outlined above (one can get $2\epsilon + 4\delta < 1$ by straightforward analysis, and the slight improvement comes from properly defining the random variables in the problem). We can show a trivial sacO obfuscator for $2\epsilon + \delta > 1$ as follows. Given an input circuit C, use random sampling to find the majority value of the truth table of C (if C is approximately balanced, then any value works). Then output the constant function taking the majority value with probability 2ϵ, and output C itself with probability $1 - 2\epsilon$. Correctness will hold with probability $1 - \epsilon$, since if C is output then correctness is perfect, and if the constant function is output then correctness is approximately $1/2$. The correlation between two functionally equivalent circuits is at least 2ϵ since the calculation of the majority value only depends on the truth table. We provide a more formal analysis in Appendix A. It seems that such a trivial obfuscator cannot imply any non-trivial results.

We notice that a sacO obfuscator can be plugged into the Sahai-Waters construction, and would imply weak notions of security and correctness for the resulting public-key encryption scheme. Holenstein [18] shows that, for some parameters, this weak notion can be amplified to standard security and correctness. Plugging in our parameters, we get that roughly when $\frac{1}{2} - 3\epsilon + 2\epsilon^2 > \delta$, sacO would imply symmetric to public key transformation using this method. This leaves a small region of parameters where sacO is not known to be impossible, and if it is possible it will imply highly non-trivial results. It is not clear whether other parameter regimes can also be useful, or whether our impossibility can be extended to rule out the entire useful regime. We refer to Figure 1 again for a visual characterization of the parameter regimes.

1.2 Consequences of Our Result

Our result strengthens previous negative results for proving the existence of iO in several ideal models. Previous works show that a construction of statistically

secure (perfectly correct) iO in any of those ideal models implies the existence of saiO in the standard model. Actually, one can generalize these results to also hold for saiO. Combined with our result, we now yield that a construction of iO or saiO in these ideal models implies that $\mathcal{NP} \subseteq \mathcal{AM} \cap \mathbf{co}\mathcal{AM}$ or the non-existence of one-way functions.

This line of research was initiated by Canetti et al. [8] who show that given a VBB obfuscator in the random oracle model, one can remove the random oracle at the cost of relaxing the correctness of the obfuscator. Pass and shelat [28] show an analogous result for VBB obfuscators in the ideal constant-degree encoding model, and Mahmoody, Mohammed, and Nematihaji [25] show analogous results for the generic group model and the generic trapdoor permutation model. All these results transform a VBB obfuscator in an oracle world into an approximately correct VBB obfuscator in the standard model. They yield an impossibility result for VBB obfuscation in the ideal models, as approximately correct VBB is known not to exist, assuming trapdoor permutations, see [8,3]. The crucial insight of Mahmoody et al. [26] is that all these oracle removal procedures are actually oblivious to the exact notion of obfuscation. The reason is that all proofs proceed by showing that the oracle-free obfuscation is as secure as the oracle-based obfuscation, i.e., the oracle-free obfuscated circuit can be simulated by an adversary in the oracle world, given the oracle-based obfuscated circuit. Therefore, if one has an iO obfuscator in any of the ideal models, via the oracle removal procedures, one obtains an saiO obfuscator in the standard model. Mahmoody et al. [26] conclude that, as an saiO obfuscator in the standard model allows to resolve the long-standing open problem of building public-key encryption from symmetric-key encryption, it seems very hard to construct such an object. In other words, their result rules out saiO assuming that building public-key encryption from symmetric-key encryption is impossible. Our result strengthens[4] their result by ruling out saiO based on the accepted complexity postulate that $\mathcal{NP} \not\subseteq \mathcal{AM} \cap \mathbf{co}\mathcal{AM}$ and the fundamental assumption of cryptography that one-way functions exist. Therefore, based on the same assumptions, iO in all aforementioned idealized models cannot exist.

1.3 Open Problems

The main question that we leave open is the set of parameters for sacO that are useful and that are (im)possible. Note that it is desirable to have more positive results not only for sacO, but also for acO, the *computational* variant of sacO, in the spirit of Bitansky-Vaikuntanathan [4] who give an assumption-based transformations from aiO to standard iO. Even if sacO for useful parameters turns out to be impossible, it might still be easier to build acO for useful parameters and then use amplification rather than to build fully secure fully correct iO directly.

[4] Note that our result is only a "stronger" result in a moral sense, but not in a formal sense. While the non-existence of one-way function would allow us to build a reduction from public-key encryption to symmetric-key encryption (as in this case, both do not exist), it is not known that $\mathcal{NP} \subseteq \mathcal{AM} \cap \mathbf{co}\mathcal{AM}$ implies that we can build a public-key encryption scheme from a one-way function.

In particular, note that for a certain parameter range of sacO, we do not know of any impossibility results of building sacO in ideal models. The oracle removal procedures that we discuss in Section 1.2 maintain security and only weaken correctness. Therefore, a variant of the oracle removal procedures can also be proven for sacO (losing some amount of correctness). As not all useful parameters for sacO are ruled out by our results, one might aim for building sacO in an ideal model for these parameters. Note that one can use our result as a sanity check for any potential oracle construction: If the construction would also work for parameters that we rule out, then it is probably better to pursue a different approach.

Another direction for building useful statistical variants of iO is to relax the computational efficiency of the obfuscator in which case the distributions $X[\Psi]$ that we considered before are not efficiently sampleable anymore (condition (i)) and thus, the $\mathcal{SZK}$ argument fails. Interestingly, Lin et al. [24] recently showed that such a notion of iO that they call XiO has indeed useful applications to transformations on functional encryption.

2 Preliminaries

We first introduce some general notation. By $n \in \mathbb{N}$, we denote the security parameter that we give to all algorithms implicitly in unary representation 1^n. By $\{0,1\}^\ell$ we denote the set of all bit-strings of length ℓ. For a finite set S, we denote the action of sampling x uniformly at random from S by $x \leftarrow\!\!\$\, S$, and denote the cardinality of S by $|S|$. Algorithms are assumed to be randomized, unless otherwise stated. We call an algorithm efficient or PPT if it runs in time polynomial in the security parameter. If $\mathcal{A}$ is randomized then by $y \leftarrow \mathcal{A}(x;r)$ we denote that $\mathcal{A}$ is run on input x and with random coins r and produced output y. If no randomness is specified, then we assume that $\mathcal{A}$ is run with freshly sampled uniform random coins, and write this as $y \leftarrow\!\!\$\, \mathcal{A}(x;\mathcal{U})$ or in shorthand $y \leftarrow\!\!\$\, \mathcal{A}(x)$. For a circuit C we denote by $|C|$ the size of the circuit. We say a function $\mathsf{negl}(n)$ is negligible if for any positive polynomial $\mathsf{poly}(n)$, there exists an $N \in \mathbb{N}$, such that for all $n > N$, $\mathsf{negl}(n) \leq \frac{1}{\mathsf{poly}(n)}$. To define statistically secure variants of obfuscation we will use the following definition of statistical distance.

Definition 1 (Statistical Distance). *For two probability distributions X, Y we define the statistical distance $\mathsf{SD}(X,Y)$ as*

$$\mathsf{SD}(X,Y) = \max_{\mathcal{A}}(\Pr_{x \leftarrow\!\!\$\, X}[\mathcal{A}(x)=1] - \Pr_{y \leftarrow\!\!\$\, Y}[\mathcal{A}(y)=1])$$

where $\mathcal{A}$ ranges over all probabilistic algorithms including inefficient ones.

2.1 Complexity Theory

We refer the reader to Goldreich's book [11] for a detailed exposition of complexity theory. We now discuss a few object that are most relevant to our proof.

We let SAT denote the set of all satisfiable CNF formulae, we let USAT denote the set of CNF formulae that have exactly one satisfying assignment, and UNSAT denote the set of CNF formulae that have no satisfying assignment. Given a formula Ψ, deciding whether $\Psi \in \mathsf{SAT}$ is an $\mathcal{NP}$-Complete problem. We recall that a *promise problem* $\Pi = (\Pi_{\mathsf{Yes}}, \Pi_{\mathsf{No}})$ is a pair of disjoint subsets of $\{0,1\}^*$. Of particular interest to us is the *unique SAT* (promise) problem $\mathsf{UniqueSAT} = (\mathsf{USAT}, \mathsf{UNSAT})$. Total problems (a.k.a languages) are a special case of promise problems, e.g. $(\mathsf{SAT}, \mathsf{UNSAT})$ is exactly the SAT problem. In such a case, it suffices to specify Π_{Yes} in order to completely define the problem.

We consider the notion of *randomized polynomial time Turing reductions* between problems. A *promise oracle* to a problem $\Pi = (\Pi_{\mathsf{Yes}}, \Pi_{\mathsf{No}})$, is one that always answers 1 on inputs in Π_{Yes} and always answers 0 on inputs in Π_{No}, but otherwise can answer arbitrarily, and even inconsistently between calls. We define the class $\mathcal{BPP}^{\Pi}$ as the class of problems solvable using a probabilistic polynomial time algorithm with access to a Π oracle. In other words, $\mathcal{BPP}^{\Pi}$ is the class of problems that are *reducible* to Π. One can verify that this class indeed composes, i.e. if $\widetilde{\Pi} \in \mathcal{BPP}^{\Pi}$ then $\mathcal{BPP}^{\widetilde{\Pi}} \subseteq \mathcal{BPP}^{\Pi}$. Valiant and Vazirani [32] showed that SAT is reducible to unique SAT.

Theorem 1 (Valiant-Vazirani). $\mathsf{SAT} \in \mathcal{BPP}^{\mathsf{UniqueSAT}}$.

An additional promise problem which will be of interest to us is the GapSD problem, defined by Sahai and Vadhan [29]. This problem essentially captures the hardness of distinguishing between efficient samplers for statistically close distributions and ones for statistically far distributions. We recall that for a circuit C (which we regard as a sampler from a distribution), $C(\mathcal{U})$ denotes the distribution generated by running C on a random input.

Definition 2 (GapSD Problem). *The problem* $\mathsf{GapSD} = (\mathsf{GapSD}_{\mathsf{Yes}}, \mathsf{GapSD}_{\mathsf{No}})$ *is defined as follows. Consider tuples of the form* $(C_0, C_1, \nu, 1^\ell)$*, where* C_0, C_1 *are circuits,* ν *is a threshold value and* 1^ℓ *is a unary encoding of a probability gap. Define*

$$\mathsf{GapSD}_{\mathsf{Yes}} = \{(C_0, C_1, \nu, 1^\ell) : \mathsf{SD}(C_0(\mathcal{U}), C_1(\mathcal{U})) < \nu\} \ ,$$

and

$$\mathsf{GapSD}_{\mathsf{No}} = \{(C_0, C_1, \nu, 1^\ell) : \mathsf{SD}(C_0(\mathcal{U}), C_1(\mathcal{U})) > \nu + 1/\ell\} \ .$$

Combining results by Mahmoody and Xiao [27] and by Bogdanov and Lee [5] as follows implies that $\mathcal{BPP}^{\mathsf{GapSD}}$ is contained in $\mathcal{AM} \cap \mathbf{co}\mathcal{AM}$.[5]

Theorem 2. $\mathcal{BPP}^{\mathsf{GapSD}} \subseteq \mathcal{AM} \cap \mathbf{co}\mathcal{AM}$.

Proof. It follows from [5, Theorem 9] that $\mathsf{GapSD} \in \mathcal{AM} \cap \mathbf{co}\mathcal{AM}$. This means that both $(\mathsf{GapSD}_{\mathsf{Yes}}, \mathsf{GapSD}_{\mathsf{No}})$ and its complement $(\mathsf{GapSD}_{\mathsf{No}}, \mathsf{GapSD}_{\mathsf{Yes}})$ have $\mathcal{AM}$ protocols, say with completeness 9/10 and soundness 1/10. Consider the

[5] In fact, by applying [27] we get that $\mathcal{BPP}^{\mathcal{SZK}} \in \mathcal{AM} \cap \mathbf{co}\mathcal{AM}$, which is almost what we need. However, it is only known that $\mathsf{GapSD} \in \mathcal{SZK}$ under a somewhat weaker definition of the GapSD problem.

protocol that takes $(C_0, C_1, \nu, 1^\ell)$ and does the following. First, execute the $\mathcal{AM}$ protocol for $(\mathsf{GapSD}_{\mathsf{Yes}}, \mathsf{GapSD}_{\mathsf{No}})$ on input $x_1 = (C_0, C_1, \nu + 1/(4\ell), 1^{(4\ell)})$. Then, execute the $\mathcal{AM}$ protocol for $(\mathsf{GapSD}_{\mathsf{No}}, \mathsf{GapSD}_{\mathsf{Yes}})$ (note the reverse order) on $x_2 = (C_0, C_1, \nu - 1/(2\ell), 1^{(4\ell)})$. Accept only if the two executions accepted. Now, assume that $\nu = \mathsf{SD}(C_0, C_1)$. Then it holds that $x_1 \in \mathsf{GapSD}_{\mathsf{Yes}}$ and $x_2 \in \mathsf{GapSD}_{\mathsf{No}}$ and therefore our new protocol accepts with probability at least 8/10. However, if $|\nu - \mathsf{SD}(C_0, C_1)| > 1/\ell$ then either $x_1 \in \mathsf{GapSD}_{\mathsf{No}}$ or $x_2 \in \mathsf{GapSD}_{\mathsf{Yes}}$ and therefore our new protocol accepts with probability at most 2/10. This means that our protocol is an $\mathcal{AM}$ protocol that, for any ϵ, can decide given (C_0, C_1), $1^{\lceil 1/\epsilon \rceil}$ and ν whether $\nu = \mathsf{SD}(C_0(\mathcal{U}), C_1(\mathcal{U}))$ or whether $|\nu - \mathsf{SD}(C_0(\mathcal{U}), C_1(\mathcal{U}))| > \epsilon$.

Consider the class $\mathbb{R}$-**TFAM** as defined in [27, Definition 3.1] and consider the real valued function $f_{\mathsf{SD}} : \{0,1\}^* \to \mathbb{R}$ defined as $f_{\mathsf{SD}}(C_0, C_1, 1^k) = \mathsf{SD}(C_0(\mathcal{U}), C_1(\mathcal{U}))$ (note that the third parameter is ignored and is used only for padding purposes). Our protocol above implies, by definition, that $f_{\mathsf{SD}} \in \mathbb{R}$-**TFAM**.

Furthermore, it holds that $\mathcal{BPP}^{\mathsf{GapSD}} \subseteq \mathcal{BPP}^{\mathcal{O}_{f_{\mathsf{SD}}}}$, for any oracle $\mathcal{O}_{f_{\mathsf{SD}}}$ that on input $x \in \{0,1\}^n$ outputs a value y such that $|y - f_{\mathsf{SD}}(x)| \leq 1/n$. To see this, we notice that we can answer GapSD queries of the form $(C_0, C_1, \nu, 1^\ell)$ as follows: First compute $y = \mathcal{O}_{f_{\mathsf{SD}}}(C_0, C_1, 1^{2\ell})$, then if $y < \nu + 1/(2\ell)$ return Yes, otherwise return No. This implies that $\mathcal{BPP}^{\mathsf{GapSD}} \subseteq \mathcal{BPP}^{\mathbb{R}\text{-}\mathbf{TFAM}}$ by [27, Definition 3.2] (when choosing $\epsilon(n) = 1/n$).

Finally, [27, Theorem 1.1] states that $\mathcal{BPP}^{\mathbb{R}\text{-}\mathbf{TFAM}} \subseteq \mathcal{AM} \cap \mathbf{co}\mathcal{AM}$, which implies that $\mathcal{BPP}^{\mathsf{GapSD}} \subseteq \mathcal{AM} \cap \mathbf{co}\mathcal{AM}$ as desired.

We now state an important corollary of Theorem 2 which shows that there would be unlikely consequences if $\mathsf{UniqueSAT} \in \mathcal{BPP}^{\mathsf{GapSD}}$.

Corollary 3. *If* $\mathsf{UniqueSAT} \in \mathcal{BPP}^{\mathsf{GapSD}}$*, then* $\mathcal{NP} \subseteq \mathcal{AM} \cap \mathbf{co}\mathcal{AM}$*.*

Proof. By definition it holds that $\mathcal{NP} \subseteq \mathcal{BPP}^{\mathsf{SAT}}$. Theorem 1 implies that $\mathcal{BPP}^{\mathsf{SAT}} \subseteq \mathcal{BPP}^{\mathsf{UniqueSAT}}$. If $\mathsf{UniqueSAT} \in \mathcal{BPP}^{\mathsf{GapSD}}$ then $\mathcal{BPP}^{\mathsf{UniqueSAT}} \subseteq \mathcal{BPP}^{\mathsf{GapSD}}$. Together with $\mathcal{BPP}^{\mathsf{GapSD}} \subseteq \mathcal{AM} \cap \mathbf{co}\mathcal{AM}$ from Theorem 2, we get

$$\mathcal{NP} \subseteq \mathcal{BPP}^{\mathsf{SAT}} \subseteq \mathcal{BPP}^{\mathsf{UniqueSAT}} \subseteq \mathcal{BPP}^{\mathsf{GapSD}} \subseteq \mathcal{AM} \cap \mathbf{co}\mathcal{AM} \ ,$$

and the corollary follows.

2.2 Obfuscation

In this subsection, we define the statistically secure variant of approximately correct indistinguishability obfuscation (saiO) and its generalization that we call statistically secure *Approximately Correct Correlation Obfuscation* (sacO). We start with the generalized variant sacO first and then define saiO as a special case. The notion of correlation obfuscation, in contrast to standard indistinguishability obfuscation, does not require that the output of the obfuscator is *indistinguishable* for functionally equivalent circuits. Rather, it only requires that there is a noticeable correlation between the outputs.

Definition 3 (Approximately Correct Correlation Obfuscation). *Let* O *be a* PPT *algorithm that takes boolean circuits (with a single output bit) as inputs and produces boolean circuits as output. For a circuit* C*, we let* $\mathsf{O}(C;r)$ *denote the output of running* O *on* C *with randomness* r*, and we let* $\mathsf{O}(C)$ *denote the distribution* $\mathsf{O}(C;r)$ *with uniform* r*.*

We say that O *is a* $(1-\epsilon)$*-approximately correct and* $(1-\delta)$*-secure correlation obfuscator sacO if the following conditions hold:*

Approximate Correctness. *For any circuit* C *it holds that*

$$\Pr_{r,x}[\mathsf{O}(C;r)(x) = C(x)] \geq 1 - \epsilon(|C|, n).$$

Correlation. *For any pair of circuits* C_1, C_2 *which compute the same function and such that* $|C_1| = |C_2|$ *it holds that* $\mathsf{SD}(\mathsf{O}(C_1), \mathsf{O}(C_2)) \leq \delta(|C_1|, n)$*.*

The definition of statistically secure approximately correct indistinguishability obfuscation (saiO) follows by requiring negligible statistical distance δ.

Definition 4 (Approximately Correct Indistinguishability Obfuscation). *Let* O *be a* $(1-\epsilon)$*-approximately correct and* $(1-\delta)$*-secure correlation obfuscator. We say that* O *is also a* $(1-\epsilon)$*-approximately correct statistically secure indistinguishability obfuscator (saiO) if there exists a negligible function* $\mathsf{negl}(|C|, n)$ *such that for all circuits* C *it holds that* $\delta(|C|, n) \leq \mathsf{negl}(|C|, n)$*.*

2.3 Puncturable Pseudorandom Functions

We use a weak notion of puncturable pseudorandom function. This notion suffices for our results and follows trivially from the stronger standard definition.

Definition 5 (Puncturable Pseudorandom Functions). *A pair of* PPT *algorithms* $(\mathsf{PRF}, \mathsf{Puncture})$ *is a puncturable pseudorandom function with one-bit output if, on input a key* $k \in \{0,1\}^n$ *or a punctured key* k^* *and an input value* $x \in \{0,1\}^n$*,* PRF *deterministically outputs a bit* b *and on input a key* $k \in \{0,1\}^n$ *and an input value* x_0*,* $\mathsf{Puncture}$ *outputs a punctured key* k^* *such that the following two properties are satisfied.*

Functionality Preserved Under Puncturing. For all keys k*, all input values* x_0*, all punctured keys* $k^* \leftarrow_\$ \mathsf{Puncture}(k, x_0)$*, and all input values* $x \neq x_0$*, it holds that*

$$\mathsf{PRF}(k^*, x) = \mathsf{PRF}(k, x).$$

Security For every PPT *adversary* $(\mathcal{A}_1, \mathcal{A}_2)$ *such that* $\mathcal{A}_1(1^n; r_1)$ *outputs an input value* x_0 *and state* st*, consider an experiment where* $\mathsf{k} \leftarrow_\$ \{0,1\}^n$*,* $\mathsf{k}^* = \mathsf{Puncture}(\mathsf{k}, x_0; t)$*, and* $b \leftarrow_\$ \{0,1\}$*. Then we have*

$$\begin{aligned} &|\Pr_{k,r_1,t,r_2}[\mathcal{A}_2(\mathsf{st}, \mathsf{k}^*, x_0, \mathsf{PRF}(k, x_0); r_2) = 1] \\ &\qquad - \Pr_{k,b,r_1,t,r_2}[\mathcal{A}_2(\mathsf{st}, \mathsf{k}^*, x_0, b; r_2) = 1] \leq \mathsf{negl}(n)\,. \end{aligned}$$

As observed by [6,7,22] puncturable PRFs can, for example, be constructed from pseudorandom generators (and thereby one-way functions [17]) via the GGM tree-based construction [12,13].

3 Negative Results for sacO and saiO

We now prove our main theorem that sacO for a large class of parameters, in particular the saiO parameters, is impossible assuming one-way functions and $\mathcal{NP} \not\subseteq \mathcal{AM} \cap \mathbf{co}\mathcal{AM}$.

Theorem 4 (Impossibility of sacO). *If $(1-\epsilon)$-approximately correct, $(1-\delta)$-secure sacO for $\mathcal{P}$ exists, and there exists some polynomial $\mathsf{poly}(|C|, n)$ such that $\delta(|C|, n) \leq \frac{1}{3} - \frac{2}{3}\epsilon(|C|, n) - \frac{1}{\mathsf{poly}(|C|,n)}$, then one-way functions do not exist or $\mathcal{NP} \subseteq \mathbf{co}\mathcal{AM} \cap \mathcal{AM}$.*

By setting δ to be some negligible function, impossibility of saiO follows immediately as a corollary.

Corollary 5 (Impossibility of saiO). *If $(1-\epsilon)$-approximately correct, saiO for $\mathcal{P}$ exists, and there exists some polynomial $\mathsf{poly}(|C|, n)$ such that $\epsilon(|C|, n) \leq \frac{1}{2} - \frac{1}{\mathsf{poly}(|C|,n)}$, then one-way functions do not exist or $\mathcal{NP} \subseteq \mathbf{co}\mathcal{AM} \cap \mathcal{AM}$.*

Proof (Theorem 4). We define an efficiently samplable distribution $X[\Psi]$ that is parametrized by a formula Ψ, and we define a reference distribution Y that should be parametrized by the size of Ψ and the number of variables in Ψ, but we omit the dependency on Ψ for readability. We note that in the introduction, we discussed to use $Y = X[\mathbf{0}]$, where $\mathbf{0}$ is a canonical representation of an unsatisfiable formula of the same size as Ψ. It is intuitive to think of Y as being indeed equal to $X[\mathbf{0}]$. However, for the sake of tightness, jumping ahead, we will use a slightly different distribution and note that this allows us to gain an additive term of δ in Claim 11.

As in the proof by Goldwasser and Rothblum [14,15] that we sketched in the introduction, we want to define $X[\Psi]$ (and Y) in a way such that properties (1), (2) and (3) are satisfied, assuming one-way functions and sacO. If we manage to do so, then we suceed in showing that these assumptions imply the collapse of the polynomial hierarchy.

Our proof will rely on the promise problem $(\mathsf{USAT}, \mathsf{UNSAT})$ rather than the language SAT (See Subsection 2.1) and therefore, instead of using the gap statistical distance problem GapSD directly as Goldwasser-Rothblum, we will consider $\mathcal{BPP}^{\mathsf{GapSD}}$ to be able to accommodate the randomized reduction from SAT to USAT (See Theorem 1).

Our proof does not rely on complexity-theoretic techniques, except for proving the following claim and showing that the theorem follows from it.

Claim 6. *Assume that there is a formula-indexed distribution $X[\Psi]$, a reference distribution Y, a function ν, and a polynomial $\mathsf{poly}(n)$ such that the following three conditions are satisfied.*

(1) There is a uniform polynomial-time algorithm $\mathcal{A}$, that on input Ψ, constructs two polynomial-size randomized circuits that sample from $X[\Psi]$ and Y respectively.

(2) If Ψ is in UNSAT*, then $X[\Psi]$ is has statistical distance at most $\nu(n)$ from Y.*

(3) If Ψ is in USAT*, then $X[\Psi]$ has statistically distance at least $\nu(n) + \frac{1}{\mathsf{poly}(n)}$ from Y.*

Then USAT *is in $\mathcal{BPP}^{\mathsf{GapSD}} \subseteq \mathcal{AM} \cap \mathbf{co}\mathcal{AM}$.*

Proof. Given that conditions (1), (2) and (3) are satisfied, we construct an algorithm $\mathcal{B}$ such that for all GapSD oracles and all formulae Ψ, $\mathcal{B}^{\mathsf{GapSD}}(\Psi)$ outputs 1 with probability 1 if $\Psi \in \mathsf{USAT}$ and 0 with probability 1 if $\Psi \in \mathsf{UNSAT}$. On input Ψ, the algorithm $\mathcal{B}$ runs $\mathcal{A}$ to get circuits for $X[\Psi]$ and Y and queries $(X[\Psi], Y, \nu(n), 1^{\mathsf{poly}(n)})$ to the GapSD oracle. $\mathcal{B}$ returns whatever the oracle returns. By properties (1), (2) and (3), the query that $\mathcal{B}$ makes is in $\mathsf{GapSD}_{\mathsf{Yes}}$ if $\Psi \in \mathsf{USAT}$ and in $\mathsf{GapSD}_{\mathsf{No}}$ if $\Psi \in \mathsf{UNSAT}$. Hence, $\mathcal{B}$ is correct and USAT is in $\mathcal{BPP}^{\mathsf{GapSD}}$. Moreover, due to Theorem 2 by Mahmoody and Xiao, $\mathcal{BPP}^{\mathsf{GapSD}} \subseteq \mathcal{AM} \cap \mathbf{co}\mathcal{AM}$.

To obtain the main theorem, we need to show that USAT is in $\mathcal{BPP}^{\mathsf{GapSD}}$ implies that $\mathcal{NP}$ is in $\mathcal{AM} \cap \mathbf{co}\mathcal{AM}$ which directly follows from Corollary 3 of Theorem 2 by Mahmoody and Xiao. Thus, if we can show that a distributions as described in conditions (1), (2) and (3) exist, then the theorem follows.

We now define $X[\Psi]$ and Y and then show that they satisfy (1), (2) and (3) assuming the existence of one-way functions and sacO with suitable correctness and security.

Definition 6 (Distribution). *Let $\ell(n)$ be a sufficiently large polynomial designating the size to which all circuits are padded before being obfuscated. Let Ψ be a formula, let* $(\mathsf{PRF}, \mathsf{Puncture})$ *be a puncturable pseudorandom function, and let* O *be a $(1-\epsilon)$-correct, statistically $(1-\delta)$-secure approximate correlation obfuscator, where $\delta(|C|, n) \leq \frac{1}{3} - \frac{2}{3}\epsilon(|C|, n) - \frac{1}{\mathsf{poly}(|C|,n)}$. We now define the distribution $X[\Psi]$ and Y, where the circuits $\mathsf{C}_X[k, b, s, \Psi]$ and $\mathsf{C}_{prf}[k]$ are defined to the right of the distributions.*

$X[\Psi](1^n)$	$\mathsf{C}_X[k, s, \Psi](x)$	$Y(1^n)$	$\mathsf{C}_{prf}[k](x)$
$k \leftarrow_\$ \{0,1\}^n$	**if** $\Psi(x \oplus s) = 1$	$k \leftarrow_\$ \{0,1\}^n$	**return** $\mathsf{PRF}(k, x)$
$s \leftarrow_\$ \{0,1\}^n$	**return** $\mathsf{PRF}(k, x) \oplus 1$	$s \leftarrow_\$ \{0,1\}^n$	
$C := \mathsf{C}_X[k, s, \Psi]$	**else**	$C := \mathsf{C}_{prf}[k]$	
$C' \leftarrow_\$ \mathsf{O}(C)$	**return** $\mathsf{PRF}(k, x)$	$C' \leftarrow_\$ \mathsf{O}(C)$	
return (k, s, C')		**return** (k, s, C')	

Claim 7 (Distribution). *The distributions defined in Definition 6 satisfy the conditions demanded in Claim 6. I.e., there exists a function ν and a polynomial* $\mathsf{poly}(n)$ *such that they satisfy the following:*

(1) There is a uniform polynomial-time algorithm $\mathcal{A}$, that on input Ψ, constructs two polynomial-size randomized circuits that sample from $X[\Psi]$ and Y respectively.

(2) If Ψ is in UNSAT, then $X[\Psi]$ is has statistical distance at most $\nu(n)$ from Y.

(3) If Ψ is in USAT, then $X[\Psi]$ has statistically distance at least $\nu(n) + \frac{1}{\mathsf{poly}(n)}$ from Y.

We will first state two claims and a lemma that will allow us to prove Claim 7. We will then prove Claim 7 and afterwards prove the claims and the lemma.

Claim 8 (Efficient Sampling). *There is a uniform polynomial-time algorithm $\mathcal{A}$, that on input Ψ, constructs two polynomial-size randomized circuits that sample from $X[\Psi]$ and Y respectively..*

Claim 9 (Statistical Proximity). *For all formulae $\Psi \in \mathsf{UNSAT}$, $X[\Psi]$ has statistical distance at most $\delta(\ell(n), n)$ from Y.*

Lemma 10 (Statistical Distance). *There exists a negligible function $\mathsf{negl}(n)$, such that for all formulae $\Psi \in \mathsf{USAT}$, $X[\Psi]$ has statistical distance at least $1 - 2\epsilon(\ell(n), n) - 2\delta(\ell(n), n) - \mathsf{negl}(n)$ from Y.*

Proof (Claim 7). Condition (1) follows immediately from Claim 8. Condition (2) follows from Claim 9 for a function $\nu(n) = \delta(\ell(n), n)$. From Lemma 10, it follows that, if Ψ is in USAT, then $X[\Psi]$ has statistically distance at least $1 - 2\epsilon(\ell(n), n) - 2\delta(\ell(n), n) - \mathsf{negl}(n)$ from Y. Combining this with the $\nu(n)$ obtained from Claim 9 we get that condition (3) holds, if there exists a polynomial $\mathsf{poly}(n)$, such that

$$\begin{aligned} & \delta(\ell(n), n) + \tfrac{1}{\mathsf{poly}(n)} \leq 1 - 2\epsilon(\ell(n), n) - 2\delta(\ell(n), n) - \mathsf{negl}(n) \\ \Leftrightarrow \quad & 3\delta(\ell(n), n) \leq 1 - 2\epsilon(\ell(n), n) - \tfrac{1}{\mathsf{poly}(n)} - \mathsf{negl}(n) \\ \Leftrightarrow \quad & \delta(\ell(n), n) \leq \frac{1}{3} - \frac{2}{3}\epsilon(\ell(n), n) - \tfrac{1}{\mathsf{poly}(n)} - \mathsf{negl}(n)\,. \end{aligned} \tag{1}$$

And, since $\mathsf{negl}(n)$ is dominated by an inverse polynomial, Equation 1 is already ensured by Definition 6, condition (3) holds, and the claim follows.

Proof (Claim 8). Sampling k and s is efficient and so is constructing $\mathsf{C}_X[k, s, \Psi]$ and $\mathsf{C}_{\mathsf{prf}}[k]$. Finally, from the efficiency of the obfuscator, it follows that $X[\Psi]$ and Y are efficiently samplable by polynomial-size randomized circuits.

Proof (Claim 9). For all unsatisfiable formulae Ψ, the circuits $\mathsf{C}_X[k, s, \Psi]$ and $\mathsf{C}_{\mathsf{prf}}[k]$ are functionally equivalent and of same size $\ell(n)$. Hence, by statistical security of the obfuscator, the distributions $(k, s, \mathsf{O}(\mathsf{C}_X[k, s, \Psi]))$ and $(k, s, \mathsf{O}(\mathsf{C}_{\mathsf{prf}}[k]))$ have statistical distance at most $\delta(\ell(n), n)$.

We now turn to the most involved part of the proof, which is to show that Lemma 10 holds. In order to show that for all formulae $\Psi \in \mathsf{USAT}$, $X[\Psi]$ is statistically far from Y, we show that, if $\Psi \in \mathsf{USAT}$, then the distribution $X[\Psi]$ has a property that Y does not have. We state the property in two claims.

Claim 11. *For all x_0, it holds that*

$$\Pr_{(k,s,C') \leftarrow\$ Y(1^n)}[C'(x_0 \oplus s) \neq \mathsf{PRF}(k, x_0 \oplus s)] \leq \epsilon(\ell(n), n).$$

Claim 12. *If $\Psi \in \mathsf{USAT}$, then there exists x_Ψ, such that*

$$\begin{aligned}\Pr_{(k,s,C') \leftarrow\$ X[\Psi](1^n)}&[C'(x_\Psi \oplus s) \neq \mathsf{PRF}(k, x_\Psi \oplus s)] \\ &\geq 1 - \epsilon(\ell(n), n) - 2\delta(\ell(n), n) - 2\mathsf{negl}(n)\,.\end{aligned}$$

Proof (Lemma 10). Lemma 10 follows directly from Claim 11 and Claim 12, because the stated properties are statistical properties, i.e., we can give an inefficient distinguisher as follows: The distinguisher determines x_Ψ through exhaustive search and then, given a sample (k, s, C') from either $X[\Psi]$ or Y, checks whether $\mathsf{PRF}(k, \cdot)$ and C' differ on input $x_\Psi \oplus s$. If the sample is from $X[\Psi]$, they will differ with probability greater than $1-\epsilon(\ell(n), n)-2\delta(\ell(n), n)-\mathsf{negl}(n)$. If on the other hand the sample is from Y, then they will differ only with probability less than $\epsilon(\ell(n), n)$. This concludes the proof of Lemma 10, subject to proving the claims.

It now remains to prove Claim 11 and Claim 12. The proof of the first property is relatively straightforward, while the proof of the second property contains the technical key arguments that we discussed above.

Proof (Claim 11). To prove the claim, we will argue that the following equalities hold:

$$\begin{aligned}&\Pr_{(k,s,C') \leftarrow\$ Y(1^n)}[C'(x_0 \oplus s) \neq \mathsf{PRF}(k, x_0 \oplus s)] && (2)\\ =&\Pr_{k,s \leftarrow\$ \{0,1\}^n, C' \leftarrow\$ \mathsf{O}(\mathsf{C}_{\mathsf{prf}}[k])}[C'(x_0 \oplus s) \neq \mathsf{PRF}(k, x_0 \oplus s)] && (3)\\ =&\Pr_{k,s \leftarrow\$ \{0,1\}^n, C' \leftarrow\$ \mathsf{O}(\mathsf{C}_{\mathsf{prf}}[k])}[C'(s) \neq \mathsf{PRF}(k, s)] && (4)\\ \leq&\epsilon(\ell(n), n) && (5)\end{aligned}$$

Equation 3 is simply a restatement of the claim. Given that s is uniformly and independently distributed, s and $x_0 \oplus s$ are distributed identically and therefore, also Equation 4 holds. Finally, Equation 4 simply checks whether an obfuscated circuit does not agree with the original circuit on a uniformly chosen input. This happens by definition of correctness with probability at most $\epsilon(\ell(n), n)$, yielding Equation 5 and concluding the proof.

Proof (Claim 12). Let x_Ψ denote the accepting assignment of Ψ. We first define the following game

$\mathsf{Game}_1(n)$

$(k, s, C') \leftarrow\$ X[\Psi]$
$x_0 := x_\Psi \oplus s$
$b := \mathsf{PRF}(k, x_0) \oplus 1$
return $(C'(x_0) \stackrel{?}{=} b)$

and observe that

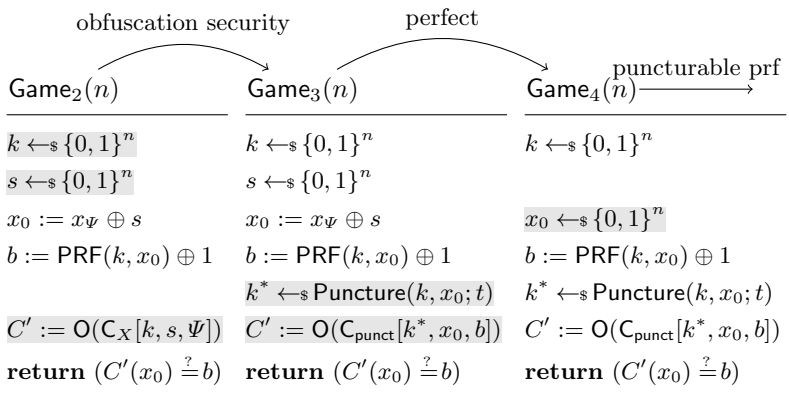
$$\Pr_{(k,s,C') \leftarrow\$ X[\Psi](1^n)}[C'(x_\Psi \oplus s) \neq \mathsf{PRF}(k, x_\Psi \oplus s)] = \Pr[\mathsf{Game}_1(n) = 1]\,.$$

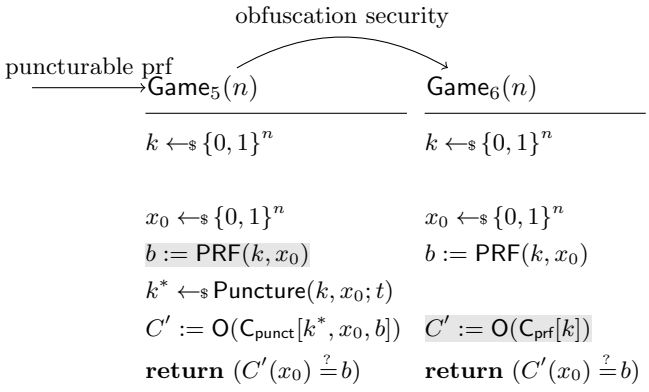
We will now bound this probability using a series of game hops. To specify the game hops, we need to specify an additional circuit $\mathsf{C}_{\mathsf{punct}}[k^*, x_0, b](x)$, that is parametrized by a punctured PRF key k^*, an input x_0 , and a bit b.

$\mathsf{C}_{\mathsf{punct}}[k^*, x_0, b](x)$

if $x = x_0$
  return b
else
  return $\mathsf{PRF}(k^*, x)$

Note that Game_2 is a re-write of Game_1 by making $X[\Psi]$ explicit.





We will first bound the differences between each pair of consecutive games and then prove a bound for $\Pr[\mathsf{Game}_6(n) = 1]$.

Hop from Game_1 *to* Game_2. The changes between the two games are purely syntactic. I.e., the definition of the sampling process from $X[\Psi]$ is explicitely written down in Game_2. Therefore, the two games are perfectly equivalent, and it holds that

$$\Pr[\mathsf{Game}_1(n) = 1] = \Pr[\mathsf{Game}_2(n) = 1]\,. \tag{6}$$

Hop from Game_2 *to* Game_3. Here it is critical to observe that $\mathsf{C}_X[k, s, \Psi]$ and $\mathsf{C}_{\mathsf{punct}}[k^*, x_0, b]$ are functionally equivalent. Even though the key is punctured on $x_0 = x_\Psi \oplus s$ in $\mathsf{C}_{\mathsf{punct}}$, this makes no difference, since PRF is never invoked on x_0 in the circuit. Instead the circuit outputs the hardcoded value $b = \mathsf{PRF}(k, x_0) \oplus 1$ on input x_0, which is the same value output by $\mathsf{C}_X[k, s, \Psi]$. Therefore, the two circuits are functionally equivalent and it follows from the statistical security of the obfuscator that the statistical difference between the distributions of C' in the two games is at most $\delta(\ell(n), n)$. It follows, that also the distribution of the outputs of Game_2 and Game_3 have a statistical distance of at most $\delta(\ell(n), n)$. I.e.,

$$|\Pr[\mathsf{Game}_3(n) = 1] - \Pr[\mathsf{Game}_2(n) = 1]| \leq \delta(\ell(n), n). \tag{7}$$

Hop from Game_3 *to* Game_4. Since s is no longer known to the obfuscator in Game_3, $x_0 := x_\Psi \oplus s$ is simply a uniformly distributed value. Thus, x_0 is distributed identically in Game_3 and Game_4 and it follows that

$$\Pr[\mathsf{Game}_3(n) = 1] = \Pr[\mathsf{Game}_4(n) = 1]\,. \tag{8}$$

Hop from Game_4 *to* Game_5. Note that x_Ψ is no longer required to evaluate Game_4 and Game_5. Therefore, the two games can be evaluated efficiently. This allows us to bound the difference between the two games by the security of the puncturable pseudorandom function. To bound the difference between games $\mathsf{Game}_4(n)$ and $\mathsf{Game}_5(n)$, we construct a distinguisher $(\mathcal{A}_1, \mathcal{A}_2)$ with advantage

$$\tfrac{1}{2} \cdot |\Pr[\mathsf{Game}_4(n) = 1] - \Pr[\mathsf{Game}_5(n) = 1]|$$

against the puncturable PRF as follows:

$\mathcal{A}_1(1^n; r_1)$	$\mathcal{A}_2(\mathsf{st}, k^*, x_0, b; r_2)$
$x_0 \leftarrow\!\!\$\ \{0,1\}^n$	$C' := \mathsf{O}(\mathsf{C}_{\mathsf{punct}}[k^*, x_0, b])$
return $(\bot, x_0)$	**return** $(C'(x_0) \stackrel{?}{=} b)$

Observe, that in the case where $\mathcal{A}_2$ receives the PRF value, it holds that

$$\Pr_{k, r_1, t, r_2}[\mathcal{A}_2(\mathsf{st}, \mathsf{k}^*, x_0, \mathsf{PRF}(k, x_0); r_2) = 1] = \Pr[\mathsf{Game}_5(n) = 1]\,. \tag{9}$$

If on the other hand, $\mathcal{A}_2$ receives a b chosen uniformly at random, then b is equal to $\mathsf{PRF}(k, x_0)$ and $\mathsf{PRF}(k, x_0) \oplus 1$ with probability $\frac{1}{2}$ respectively, and it holds that

$$\Pr_{k, b, r_1, t, r_2}[\mathcal{A}_2(\mathsf{st}, \mathsf{k}^*, x_0, b; r_2) = 1] = \frac{1}{2}\Pr[\mathsf{Game}_4(n) = 1] + \frac{1}{2}\Pr[\mathsf{Game}_5(n) = 1] \tag{10}$$

By security of the puncturable PRF, it must hold that

$$\begin{aligned}&|\Pr_{k,r_1,t,r_2}[\mathcal{A}_2(\mathsf{st},\mathsf{k}^*,x_0,\mathsf{PRF}(k,x_0);r_2)=1]\\&\qquad-\Pr_{k,b,r_1,t,r_2}[\mathcal{A}_2(\mathsf{st},\mathsf{k}^*,x_0,b;r_2)=1]\,|\le \mathsf{negl}(n)\end{aligned}$$

Combining this with Equation 9 and Equation 10 yields

$$\begin{aligned}&\left|\Pr[\mathsf{Game}_5(n)=1]-\frac{1}{2}\Pr[\mathsf{Game}_4(n)=1]-\frac{1}{2}\Pr[\mathsf{Game}_5(n)=1]\right|\le\mathsf{negl}(n)\\\implies&\frac{1}{2}\left|\Pr[\mathsf{Game}_5(n)=1]-\Pr[\mathsf{Game}_4(n)=1]\right|\le\mathsf{negl}(n)\\\implies&\left|\Pr[\mathsf{Game}_5(n)=1]-\Pr[\mathsf{Game}_4(n)=1]\right|\le 2\mathsf{negl}(n)\,.\end{aligned}\tag{11}$$

Hop from Game_5 *to* Game_6. Here it is critical to observe that $\mathsf{C}_{\mathsf{punct}}[k^*,x_0,b]$ and $\mathsf{C}_{\mathsf{prf}}[k]$ are functionally equivalent. Even though the key is punctured on x_0 in $\mathsf{C}_{\mathsf{punct}}$, this makes no difference, since PRF is never invoked on x_0 in the circuit. Instead the circuit outputs the hardcoded value $b=\mathsf{PRF}(k,x_0)$ on input x_0. Therefore, the two circuits are functionally equivalent and it follows from the statistical security of the obfuscator that the statistical difference between the distributions of C' in the two games is at most $\delta(\ell(n),n)$. It follows, that also the distribution of the outputs of Game_5 and Game_6 have a statistical distance of at most $\delta(\ell(n),n)$. I.e.,

$$\left|\Pr[\mathsf{Game}_5(n)=1]-\Pr[\mathsf{Game}_6(n)=1]\right|\le\delta(\ell(n),n).\tag{12}$$

It remains to bound the probability $\Pr[\mathsf{Game}_6(n)=1]$. Observe, that x_0 is a uniformly chosen input unknown to the obfuscator. Further, the $\mathsf{Game}_6(n)$ simply checks whether the output of circuit C' is the correct output value of the obfuscated circuit. Therefore, the correctness of the obfuscator implies that

$$\Pr[\mathsf{Game}_6(n)=1]\ge 1-\epsilon(\ell(n),n).\tag{13}$$

Finally, combining Equation 13 with Equations 6 through 12, we get

$$\begin{aligned}&\Pr[\mathsf{Game}_1(n)=1]\\\ge\,&\Pr[\mathsf{Game}_6(n)=1]-\left|\Pr[\mathsf{Game}_1(n)=1]-\Pr[\mathsf{Game}_6(n)=1]\right|\\\ge\,&1-\epsilon(\ell(n),n)-2\delta(\ell(n),n)-2\mathsf{negl}(n)\end{aligned}$$

thus concluding the proof of Claim 12 and Theorem 4.

Acknowledgments

We are grateful to Andrej Bogdanov, Kai-Min Chung, Siyao Guo, Markulf Kohlweiss, Arno Mittelbach and Vinod Vaikuntanathan for helpful discussions. In particular, Andrej and Vinod pointed out that PAC-learneability implies approximate obfuscation and that thus, CNF formulae are PAC-learneable, which

implies that impossibility results for saiO need to obfuscate more complex functions than CNF formulae. The discussions with Vinod at the Mathematisches Forschungsinstitut Oberwolfach (MFO) inspired the idea of embedding a formula into a PRF. Vinod also suggested that in the absence of one-way functions, there exists a perfectly secure variant of obfuscation where the correctness is on average over the circuit distribution, the input and the obfuscator.

References

1. Boaz Barak, Oded Goldreich, Russell Impagliazzo, Steven Rudich, Amit Sahai, Salil P. Vadhan, and Ke Yang. On the (im)possibility of obfuscating programs. In Joe Kilian, editor, *Advances in Cryptology – CRYPTO 2001*, volume 2139 of *Lecture Notes in Computer Science*, pages 1–18, Santa Barbara, CA, USA, August 19–23, 2001. Springer, Heidelberg, Germany. 1
2. Boaz Barak, Oded Goldreich, Russell Impagliazzo, Steven Rudich, Amit Sahai, Salil P. Vadhan, and Ke Yang. On the (im)possibility of obfuscating programs. *Journal of the ACM*, 59(2):6, 2012. 1
3. Nir Bitansky and Omer Paneth. On the impossibility of approximate obfuscation and applications to resettable cryptography. In Dan Boneh, Tim Roughgarden, and Joan Feigenbaum, editors, *45th Annual ACM Symposium on Theory of Computing*, pages 241–250, Palo Alto, CA, USA, June 1–4, 2013. ACM Press. 1.2
4. Nir Bitansky and Vinod Vaikuntanathan. Indistinguishability obfuscation: From approximate to exact. In Eyal Kushilevitz and Tal Malkin, editors, *TCC 2016-A: 13th Theory of Cryptography Conference, Part I*, volume 9562 of *Lecture Notes in Computer Science*, pages 67–95, Tel Aviv, Israel, January 10–13, 2016. Springer, Heidelberg, Germany. 1.3, B
5. Andrej Bogdanov and Chin Ho Lee. Limits of provable security for homomorphic encryption. In Ran Canetti and Juan A. Garay, editors, *Advances in Cryptology – CRYPTO 2013, Part I*, volume 8042 of *Lecture Notes in Computer Science*, pages 111–128, Santa Barbara, CA, USA, August 18–22, 2013. Springer, Heidelberg, Germany. 2.1, 2.1
6. Dan Boneh and Brent Waters. Constrained pseudorandom functions and their applications. In Kazue Sako and Palash Sarkar, editors, *Advances in Cryptology – ASIACRYPT 2013, Part II*, volume 8270 of *Lecture Notes in Computer Science*, pages 280–300, Bengalore, India, December 1–5, 2013. Springer, Heidelberg, Germany. 1.1, 2.3, B.2
7. Elette Boyle, Shafi Goldwasser, and Ioana Ivan. Functional signatures and pseudorandom functions. In Hugo Krawczyk, editor, *PKC 2014: 17th International Conference on Theory and Practice of Public Key Cryptography*, volume 8383 of *Lecture Notes in Computer Science*, pages 501–519, Buenos Aires, Argentina, March 26–28, 2014. Springer, Heidelberg, Germany. 1.1, 2.3, B.2
8. Ran Canetti, Yael Tauman Kalai, and Omer Paneth. On obfuscation with random oracles. In Yevgeniy Dodis and Jesper Buus Nielsen, editors, *TCC 2015: 12th Theory of Cryptography Conference, Part II*, volume 9015 of *Lecture Notes in Computer Science*, pages 456–467, Warsaw, Poland, March 23–25, 2015. Springer, Heidelberg, Germany. 1.2
9. Whitfield Diffie and Martin E. Hellman. Multiuser cryptographic techniques. In *American Federation of Information Processing Societies: 1976 National Computer Conference*, volume 45 of *AFIPS Conference Proceedings*, pages 109–112, New York, NY, USA, June 7–10, 1976. AFIPS Press. 1

10. Sanjam Garg, Craig Gentry, Shai Halevi, Mariana Raykova, Amit Sahai, and Brent Waters. Candidate indistinguishability obfuscation and functional encryption for all circuits. In *54th Annual Symposium on Foundations of Computer Science*, pages 40–49, Berkeley, CA, USA, October 26–29, 2013. IEEE Computer Society Press. 1
11. Oded Goldreich. *Computational complexity - a conceptual perspective*. Cambridge University Press, 2008. 2.1
12. Oded Goldreich, Shafi Goldwasser, and Silvio Micali. How to construct random functions (extended abstract). In *25th Annual Symposium on Foundations of Computer Science*, pages 464–479, Singer Island, Florida, October 24–26, 1984. IEEE Computer Society Press. 2.3
13. Oded Goldreich, Shafi Goldwasser, and Silvio Micali. How to construct random functions. *Journal of the ACM*, 33(4):792–807, October 1986. 2.3, B.2
14. Shafi Goldwasser and Guy N. Rothblum. On best-possible obfuscation. In Salil P. Vadhan, editor, *TCC 2007: 4th Theory of Cryptography Conference*, volume 4392 of *Lecture Notes in Computer Science*, pages 194–213, Amsterdam, The Netherlands, February 21–24, 2007. Springer, Heidelberg, Germany. 1, 3
15. Shafi Goldwasser and Guy N. Rothblum. On best-possible obfuscation. *Journal of Cryptology*, 27(3):480–505, July 2014. 1, 3
16. Satoshi Hada and Kouichi Sakurai. A note on the (im)possibility of using obfuscators to transform private-key encryption into public-key encryption. In Atsuko Miyaji, Hiroaki Kikuchi, and Kai Rannenberg, editors, *IWSEC 07: 2nd International Workshop on Security, Advances in Information and Computer Security*, volume 4752 of *Lecture Notes in Computer Science*, pages 1–12, Nara, Japan, October 29–31, 2007. Springer, Heidelberg, Germany. 1
17. 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. 2.3, B.2
18. Thomas Holenstein. *Strengthening Key Agreement Using Hard-Core Sets*. PhD thesis, ETH Zurich, 2006. 1.1, B, B.1
19. Russell Impagliazzo and Michael Luby. One-way functions are essential for complexity based cryptography (extended abstract). In *30th Annual Symposium on Foundations of Computer Science*, pages 230–235, Research Triangle Park, North Carolina, October 30 – November 1, 1989. IEEE Computer Society Press. 1.1
20. Russell Impagliazzo and Steven Rudich. Limits on the provable consequences of one-way permutations. In *21st Annual ACM Symposium on Theory of Computing*, pages 44–61, Seattle, Washington, USA, May 15–17, 1989. ACM Press. 1
21. Russell Impagliazzo and Steven Rudich. Limits on the provable consequences of one-way permutations. In Shafi Goldwasser, editor, *Advances in Cryptology – CRYPTO'88*, volume 403 of *Lecture Notes in Computer Science*, pages 8–26, Santa Barbara, CA, USA, August 21–25, 1990. Springer, Heidelberg, Germany. 1
22. Aggelos Kiayias, Stavros Papadopoulos, Nikos Triandopoulos, and Thomas Zacharias. Delegatable pseudorandom functions and applications. In Ahmad-Reza Sadeghi, Virgil D. Gligor, and Moti Yung, editors, *ACM CCS 13: 20th Conference on Computer and Communications Security*, pages 669–684, Berlin, Germany, November 4–8, 2013. ACM Press. 1.1, 2.3, B.2
23. Ilan Komargodski, Tal Moran, Moni Naor, Rafael Pass, Alon Rosen, and Eylon Yogev. One-way functions and (im)perfect obfuscation. In *55th Annual Symposium on Foundations of Computer Science*, pages 374–383, Philadelphia, PA, USA, October 18–21, 2014. IEEE Computer Society Press. 1

24. Huijia Lin, Rafael Pass, Karn Seth, and Sidharth Telang. Output-compressing randomized encodings and applications. Cryptology ePrint Archive, Report 2015/720, 2015. http://eprint.iacr.org/2015/720. 1.3
25. Mohammad Mahmoody, Ameer Mohammed, and Soheil Nematihaji. On the impossibility of virtual black-box obfuscation in idealized models. In Eyal Kushilevitz and Tal Malkin, editors, *TCC 2016-A: 13th Theory of Cryptography Conference, Part I*, volume 9562 of *Lecture Notes in Computer Science*, pages 18–48, Tel Aviv, Israel, January 10–13, 2016. Springer, Heidelberg, Germany. 1.2
26. Mohammad Mahmoody, Ameer Mohammed, Soheil Nematihaji, Rafael Pass, and Abhi Shelat. Lower bounds on assumptions behind indistinguishability obfuscation. In Eyal Kushilevitz and Tal Malkin, editors, *TCC 2016-A: 13th Theory of Cryptography Conference, Part I*, volume 9562 of *Lecture Notes in Computer Science*, pages 49–66, Tel Aviv, Israel, January 10–13, 2016. Springer, Heidelberg, Germany. 1.2, B, B.1
27. Mohammad Mahmoody and David Xiao. On the power of randomized reductions and the checkability of SAT. In *Proceedings of the 25th Annual IEEE Conference on Computational Complexity, CCC 2010, Cambridge, Massachusetts, June 9-12, 2010*, pages 64–75. IEEE Computer Society, 2010. 1, 1.1, 2.1, 5, 2.1
28. Rafael Pass and Abhi Shelat. Impossibility of VBB obfuscation with ideal constant-degree graded encodings. In Eyal Kushilevitz and Tal Malkin, editors, *TCC 2016-A: 13th Theory of Cryptography Conference, Part I*, volume 9562 of *Lecture Notes in Computer Science*, pages 3–17, Tel Aviv, Israel, January 10–13, 2016. Springer, Heidelberg, Germany. 1.2
29. Amit Sahai and Salil P. Vadhan. A complete promise problem for statistical zero-knowledge. In *38th Annual Symposium on Foundations of Computer Science*, pages 448–457, Miami Beach, Florida, October 19–22, 1997. IEEE Computer Society Press. 1.1, 2.1
30. Amit Sahai and Brent Waters. How to use indistinguishability obfuscation: deniable encryption, and more. In David B. Shmoys, editor, *46th Annual ACM Symposium on Theory of Computing*, pages 475–484, New York, NY, USA, May 31 – June 3, 2014. ACM Press. 1, 1.1, B, B.2
31. Leslie G. Valiant. A theory of the learnable. *Commun. ACM*, 27(11):1134–1142, 1984. 1.1
32. Leslie G. Valiant and Vijay V. Vazirani. NP is as easy as detecting unique solutions. In Robert Sedgewick, editor, *17th Annual ACM Symposium on Theory of Computing*, pages 458–463, Providence, Rhode Island, USA, May 6–8, 1985. ACM Press. 1, 1.1, 2.1

A A positive result for Correlation Obfuscation

In this appendix, we instantiate approximately correct correlation obfuscation for a large class of weak parameters. The idea of the construction is fairly simple and is based on two observations. For circuits with only a single bit output, we can efficiently estimate the majority of the outputs by using random sampling. This estimation depends only on the function computed by a circuit and not on the circuit itself. Therefore an obfuscator that simply outputs the estimated majority is fully secure but only correct with probability about 1/2. An obfuscator, that simply outputs the circuit itself, on the other hand, is not secure at all (statistical distance is 1), but is fully correct.

By combining these two obfuscators and outputting the majority with probability 2ϵ and the circuit itself with probability $1-2\epsilon$ we can construct a roughly $(1-\epsilon)$ approximatly correct and $(1-2\epsilon)$ secure obfuscator $\mathsf{O}_{\epsilon,\mu}$ as detailed below. The parameter μ is some inverse polynomial function that describes the amount of approximation error that we allow (and that affects the correctness of $\mathsf{O}_{\epsilon,\mu}$) when the obfuscators samples repeatedly from the output distribution of the circuit to see whether the circuit is closer to the constant 1 or constant 0 function.

For any circuit C, $\mathsf{in}(C)$ denotes the number of input wires. For $b \in \{0,1\}$, Const_b^i is a canonical circuit with input length i and constant output b. The Bernoulli distribution of a parameter $p \in [0,1]$ is defined by $\mathtt{Ber}_p$, i.e., it holds that $\Pr_{b \leftarrow\$ \mathtt{Ber}_p}[b=1] = p$ and $\Pr_{b \leftarrow\$ \mathtt{Ber}_p}[b=0] = 1-p$. Depending on the desired error parameter μ, the obfuscation proceeds as follows.

$\mathsf{O}_{\epsilon,\mu}(C, 1^n)$

$b \leftarrow\$ \mathtt{Ber}_{2\epsilon}$
if $b = 1$:
  $m := \mathsf{EstMaj}(C, \mu, 1^n)$
  $C' := \mathsf{Const}_m^{\mathsf{in}(C)}$
else
  $C' := C$
return C'

$\mathsf{EstMaj}(C, \mu, 1^n)$

for $i := 1, \ldots, \lceil \frac{4n}{\mu^2} \rceil$
  $x_i \leftarrow\$ \{0,1\}^{\mathsf{in}(C)}$
  $y_i := C(x_i)$
return $\mathsf{maj}(y_1, \ldots, y_{\lceil \frac{4n}{\mu^2} \rceil})$

Claim 13. *On input $(C, 1^n)$, the obfuscator $\mathsf{O}_{\epsilon,\mu}$ runs in time linear in $\frac{4n}{\mu^2}|C|$ plus the time needed to sample from $\mathtt{Ber}_{2\epsilon}$ and is an $(1-(\epsilon+\mu))$ approximately correct and $(1-2\epsilon)$ secure correlation obfuscator for circuits with single bit output.*

Proof. Efficiency follow by construction and so does security, because EstMaj only uses the input-output behaviour of the circuit which is the same for two functionally identical circuits. If the function induced by the circuit C is less than $\frac{\mu}{4}$ from being balanced (i.e., 1 with probability $\frac{1}{2}$ on a uniformly random input), then the correctness error is at most $\frac{\mu}{2}$, if $b = 1$, and 0, if $b = 0$ and hence, the overall correctness error is upper bounded by $(1-2\epsilon)\cdot 0 + 2\epsilon \cdot \frac{\mu}{2} = \epsilon\mu \leq \epsilon + \mu$. If the function induced by the circuit C outputs a fixed value, w.l.o.g. 1, with probability at least $\frac{1}{2} + \frac{\mu}{4}$, then via a Chernoff bound, the probability that $\mathsf{EstMaj}(C, \mu, 1^n)$ outputs 1 is at least $1 - \mathsf{negl}(n)$ and in that case, the correctness error is at most $\frac{1}{2} - \frac{\mu}{4}$ and else, the correctness error is at most 1. Hence, for the case that $b = 1$, we obtain an upper bound on the correctness error of $(\frac{1}{2} - \frac{\mu}{4}) \cdot (1 - \mathsf{negl}(n)) + 1 \cdot \mathsf{negl}(n) = \frac{1}{2} - \frac{\mu}{4} + \mathsf{negl}(n)$. As before, when $b = 0$, the correctness error is 0 and hence, we obtain as upper bound on the correctness error $(1-2\epsilon)\cdot 0 + 2\epsilon \cdot (\frac{1}{2} - \frac{\mu}{4} + \mathsf{negl}(n)) = \epsilon - \frac{\mu}{2} \leq \epsilon + \mu$.

B Correctness and Security Parameters for sacO to build a Public-Key Encryption scheme from a One-Way Function

By inspecting the Sahai-Waters [30] construction to transform a one-way function into a public-key encryption scheme (PKE) by using obfuscation, Bitansky and Vaikuntanathan [4] and Mahmoody et al. [26] observe that approximately correct iO suffices for this transformation. Both papers consider approximately correct variants of iO with "full" security, i.e., where the adversary has only negligible advantage in distinguishing obfuscations of two functionally equivalent circuits. As discussed in previous sections, approximately correct correlation obfuscation (sacO) with weaker security might still be useful. We therefore work out the exact correctness and security parameters required of a sacO for the Sahai-Waters transformation to work. Jumping ahead, we note that part of the bounds that we obtain here are ruled out by our impossibility result, but not all of them.

For much weaker parameters, we earlier gave a trivial construction of sacO. We do not deem this construction to be useful. As expected, there is a gap between the parameters that we can construct trivially and the parameters that we can rule out (else, we would have a proof that one-way functions imply the collapse of the polynomial hierarchy). Also, as expected, the trivial bounds do not suffice to instantiate the Sahai-Waters construction (according to our analysis that we have reasons to believe is tight).

On the other hand, our impossibility result does not rule out all useful bounds for sacO. It is an interesting question to (1) show that also for the parameters in this small gap, sacO cannot exist, or (2) show a construction for these parameters, and/or (3) improve the parameters that are needed for meaningful applications. Note that even if it turns out that sacO for these parameters cannot exist, (3) could still be a fruitful research direction, because it might be helpful to weaken the parameters also on variants of acO with *computational* security in order to obtain constructions from weaker assumptions.

We will consider sacO with $(1-\delta)$-security and $(1-\epsilon)$ correctness, and we will also yield a PKE that does not achieve full correctness and that does not achieve full security. In some cases, as observed by Holenstein [18], via amplification, it is possible to achieve full security and correctness with overwhelming probability. However, as we discuss now, amplification is not always possible.

B.1 Amplification

We define $(1-\epsilon_{\mathsf{PKE}})$-correct and $(\frac{1}{2}-\delta_{\mathsf{PKE}})$-secure PKE as follows.

Definition 7 (Approximate Public Key Encryption). *Let* $\mathsf{PKE} = (\mathsf{KGen}, \mathsf{Enc}, \mathsf{Dec})$ *be a public key encryption scheme.*

Correctness We say that PKE *is* $(1-\epsilon_{\mathsf{PKE}})$*-correct, if it holds that*

$$\Pr_{b,\mathsf{KGen},\mathsf{Enc}}[\mathsf{Dec}(\mathsf{sk},\mathsf{Enc}(b,pk)) = b,\ (\mathsf{pk},\mathsf{sk}) \leftarrow_{\$} \mathsf{PKE.KGen}(1^n)] \geq 1-\epsilon_{\mathsf{PKE}}(n).$$

Security We say that PKE *is* $(\frac{1}{2} - \delta_{\mathsf{PKE}})$*-secure, if for all efficient adversaries* $\mathcal{A}$*, there exists a negligible function* $\mathsf{negl}(n)$ *such that*

$$\Pr_{b \leftarrow\!\$ \{0,1\}, \mathsf{KGen}, \mathsf{Enc}}[\mathcal{A}(\mathsf{pk}, \mathsf{Enc}(b, \mathsf{pk})) = b, (\mathsf{pk}, \mathsf{sk}) \leftarrow\!\$ \mathsf{KGen}(1^n)] \leq \frac{1}{2} + \delta_{\mathsf{PKE}}(n) + \mathsf{negl}(n)$$

We would like to amplify such a scheme into "standard" PKE, where ϵ_{PKE} and δ_{PKE} are negligible. We now discuss via a counterexample why such an amplification is not generally possible. Take a bit encryption scheme that outputs the message bit with probability α and a random bit with probability $1-\alpha$ and where decryption is the identity function. This PKE scheme is $(\frac{1}{2} - \frac{\alpha}{2})$-secure and $(\frac{1}{2} + \frac{\alpha}{2})$-correct. Correctness parameters are thus only meaningful if ϵ_{PKE} and δ_{PKE} are bounded away from $\frac{1}{2}$ and if, moreover, there is a meaningful relationship between the security and the correctness parameter. Holenstein [18] shows (and we use the presentation of Mahmoody et al. [26] here) that amplification is possible if there exists a polynomial $\mathsf{poly}(n)$ such that

$$(1 - 2\epsilon_{\mathsf{PKE}}(n))^2 > 2\delta_{\mathsf{PKE}}(n) + \frac{1}{\mathsf{poly}(n)}.$$

Note that Holenstein also shows a tightness result for his amplification technique with respect to restricted black-box reductions.

B.2 The Sahai-Waters Construction

We now present the Sahai-Waters [30] construction of a public-key encryption scheme from a one-way function. We recall that by Håstad et al. [17], Goldreich, Goldwasser and Micali [13], and several independent proofs [6,7,22] that the GGM construction is a puncturable PRF, puncturable PRFs and OWFs are existentially equivalent. The key generation of the Sahai-Waters construction draws a key k for a puncturable PRF as the secret key sk and then outputs an obfuscation of the following circuit $\mathsf{C}_{\mathsf{SW}}[k]$ as a public key pk:

$\underline{\mathsf{C}_{\mathsf{SW}}[k](m, r)}$
$r' := \mathsf{PRG}(r)$
$c := m \oplus \mathsf{PRF}(k, r')$
return (r', c)

The encryption algorithm $\mathsf{Enc}(\mathsf{pk}, m, r)$ interprets the public key pk as a circuit, runs it on (m, r) and returns the result as a ciphertext. Finally, for decryption of a pair (r', c), the decryption algorithm $\mathsf{Dec}(\mathsf{sk}, (r', c))$ outputs $m := c \oplus \mathsf{PRF}(\mathsf{sk}, r')$.

Claim 14 (Sahai-Waters). *The Sahai-Waters construction instantiated with sacO with correctness* $1 - \epsilon$ *and security* $1 - \delta$ *yields a public-key encryption scheme with correctness error* $\epsilon_{\mathsf{PKE}}(n) = \epsilon(|C|, n)$ *and a distinguishing advantage of* $\delta_{\mathsf{PKE}}(n) = \delta(|C|, n) + \epsilon(|C|, n)$

Before we prove this claim, we will first illustrate what this implies for the bounds on parameters allowing for Holenstein amplification. Combining the bound for Holenstein amplification with Claim 14, we get that

$$2\delta_{\mathsf{PKE}}(n) + \frac{1}{\mathsf{poly}(n)} < (1 - 2\epsilon_{\mathsf{PKE}}(n))^2 \tag{14}$$

$$\implies \quad 2\delta(|C|, n) + 2\epsilon(|C|, n) + \frac{1}{\mathsf{poly}(n)} < (1 - 2\epsilon(|C|, n))^2 \tag{15}$$

$$\implies \quad \delta(|C|, n) < \frac{1}{2} - 3\epsilon(|C|, n) + 2\epsilon(|C|, n)^2 - \frac{1}{2\mathsf{poly}(n)}. \tag{16}$$

We thus get the following corollary.

Corollary 15. *Any $(1-\epsilon)$ correct and $(1-\delta)$ secure sacO implies a construction of public key encryption from one-way functions, if there exists some polynomial $\mathsf{poly}(|C|, n)$ such that*

$$\delta(|C|, n) < \frac{1}{2} - 3\epsilon(|C|, n) + 2\epsilon(|C|, n)^2 - \frac{1}{\mathsf{poly}(n)}.$$

Proof (Proof of Claim 14). Note that correctness of the encryption scheme is over a random message, the randomness of the key generation and the randomness of the encryption algorithm. The obfuscated circuit is therefore invoked on a uniformly random input and the probability that it does not output the correct ciphertext can thus be bounded by the correctness error of the obfuscator. Since the decryption of the scheme is perfectly correct, we thus get that $\epsilon_{\mathsf{PKE}}(n) = \epsilon(|C|, n)$.

To prove securtity, we first define the following game

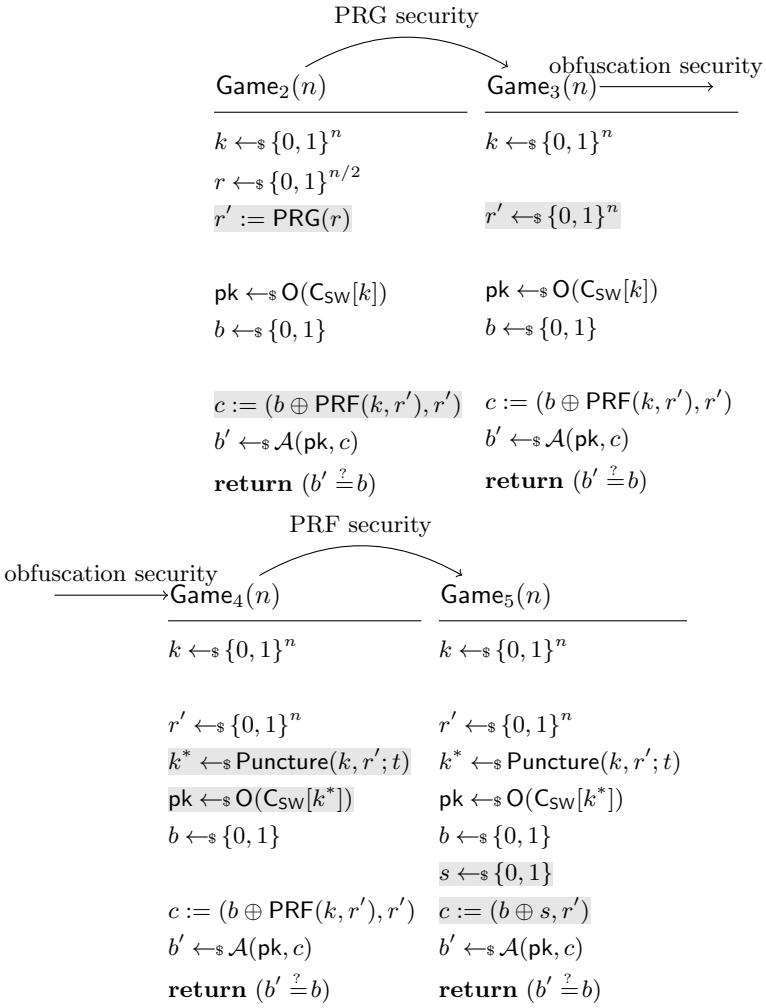
$\mathsf{Game}_1(n)$

$k \leftarrow\!\!\$\ \{0,1\}^n$
$r \leftarrow\!\!\$\ \{0,1\}^{n/2}$
$\mathsf{pk} \leftarrow\!\!\$\ \mathsf{O}(\mathsf{C}_{\mathsf{SW}}[k])$
$b \leftarrow\!\!\$\ \{0,1\}$
$c := \mathsf{pk}(b, r)$
$b' \leftarrow\!\!\$\ \mathcal{A}(\mathsf{pk}, c)$
return $(b' \stackrel{?}{=} b)$

and observe that

$$\begin{aligned}&\Pr_{b \leftarrow\!\!\$\ \{0,1\}, \mathsf{KGen}, \mathsf{Enc}}[\mathcal{A}(\mathsf{pk}, \mathsf{Enc}(b, \mathsf{pk})) = b,\ (\mathsf{pk}, \mathsf{sk}) \leftarrow\!\!\$\ \mathsf{KGen}(1^n)] \\ &= \Pr[\mathsf{Game}_1(n) = 1].\end{aligned}$$

We will now bound this probability using a series of game hops.



$\mathsf{Game}_2(n)$ / $\mathsf{Game}_3(n)$

$\mathsf{Game}_2(n)$	$\mathsf{Game}_3(n)$
$k \leftarrow\!\!\$\ \{0,1\}^n$	$k \leftarrow\!\!\$\ \{0,1\}^n$
$r \leftarrow\!\!\$\ \{0,1\}^{n/2}$	
$r' := \mathsf{PRG}(r)$	$r' \leftarrow\!\!\$\ \{0,1\}^n$
$\mathsf{pk} \leftarrow\!\!\$\ \mathsf{O}(\mathsf{C}_{\mathsf{SW}}[k])$	$\mathsf{pk} \leftarrow\!\!\$\ \mathsf{O}(\mathsf{C}_{\mathsf{SW}}[k])$
$b \leftarrow\!\!\$\ \{0,1\}$	$b \leftarrow\!\!\$\ \{0,1\}$
$c := (b \oplus \mathsf{PRF}(k, r'), r')$	$c := (b \oplus \mathsf{PRF}(k, r'), r')$
$b' \leftarrow\!\!\$\ \mathcal{A}(\mathsf{pk}, c)$	$b' \leftarrow\!\!\$\ \mathcal{A}(\mathsf{pk}, c)$
return $(b' \stackrel{?}{=} b)$	**return** $(b' \stackrel{?}{=} b)$

$\mathsf{Game}_4(n)$ / $\mathsf{Game}_5(n)$

$\mathsf{Game}_4(n)$	$\mathsf{Game}_5(n)$
$k \leftarrow\!\!\$\ \{0,1\}^n$	$k \leftarrow\!\!\$\ \{0,1\}^n$
$r' \leftarrow\!\!\$\ \{0,1\}^n$	$r' \leftarrow\!\!\$\ \{0,1\}^n$
$k^* \leftarrow\!\!\$\ \mathsf{Puncture}(k, r'; t)$	$k^* \leftarrow\!\!\$\ \mathsf{Puncture}(k, r'; t)$
$\mathsf{pk} \leftarrow\!\!\$\ \mathsf{O}(\mathsf{C}_{\mathsf{SW}}[k^*])$	$\mathsf{pk} \leftarrow\!\!\$\ \mathsf{O}(\mathsf{C}_{\mathsf{SW}}[k^*])$
$b \leftarrow\!\!\$\ \{0,1\}$	$b \leftarrow\!\!\$\ \{0,1\}$
	$s \leftarrow\!\!\$\ \{0,1\}$
$c := (b \oplus \mathsf{PRF}(k, r'), r')$	$c := (b \oplus s, r')$
$b' \leftarrow\!\!\$\ \mathcal{A}(\mathsf{pk}, c)$	$b' \leftarrow\!\!\$\ \mathcal{A}(\mathsf{pk}, c)$
return $(b' \stackrel{?}{=} b)$	**return** $(b' \stackrel{?}{=} b)$

We will first bound the differences between each pair of consecutive games and then argue a bound for $\Pr[\mathsf{Game}_5(n) = 1]$.

Hop from Game_1 *to* Game_2. The change between the two games is that the ciphertext is now no longer computed using the obfuscated circuit. Instead, it is computed as specified in the unobfuscated circuit $\mathsf{C}_{\mathsf{SW}}[k]$. Since the input to the circuit is uniformly and independently distributed, we can bound the probability that the two computations differ by the correctness of the sacO. I.e. it holds that

$$|\Pr[\mathsf{Game}_1(n) = 1] - \Pr[\mathsf{Game}_2(n) = 1]| \leq \epsilon(|C|, n). \tag{17}$$

Hop from Game_2 *to* Game_3. The change between the two games is that the bitstring r' is no longer the output of a PRG and instead a uniformly chosen random string. We can thus bound the difference between the two games using the security of the pseudorandom generator. I.e., we can construct a distinguisher $\mathcal{D}$ with advantage $|\Pr[\mathsf{Game}_2(n) = 1] - \Pr[\mathsf{Game}_3(n) = 1]|$ as follows

$\mathcal{D}(r')$

$k \leftarrow\!\!\$\ \{0,1\}^n$
$\mathsf{pk} \leftarrow\!\!\$\ \mathsf{O}(\mathsf{C}_{\mathsf{SW}}[k])$
$b \leftarrow\!\!\$\ \{0,1\}$
$c := (b \oplus \mathsf{PRF}(k, r'), r')$
$b' \leftarrow\!\!\$\ \mathcal{A}(\mathsf{pk}, c)$
return $(b' \stackrel{?}{=} b)$

Observe, that in the case where $\mathcal{D}$ receives the output of the PRG, it holds that

$$\Pr_{r,\mathcal{D}}[\mathcal{D}(\mathsf{PRG}(r)) = 1] = \Pr[\mathsf{Game}_2(n) = 1]\,. \tag{18}$$

If on the other hand, $\mathcal{D}$ receives an r' chosen uniformly at random, then it holds that

$$\Pr_{r',\mathcal{D}}[\mathcal{D}(r') = 1] = \Pr[\mathsf{Game}_3(n) = 1]\,. \tag{19}$$

By definition of a secure PRG, there further exists a negligible function $\mathsf{negl}(n)$, such that

$$|\Pr_{r,\mathcal{D}}[\mathcal{D}(\mathsf{PRG}(r)) = 1] - \Pr_{r',\mathcal{D}}[\mathcal{D}(r') = 1]| \leq \mathsf{negl}(n)\,.$$

Combining this with Equation 18 and Equation 19, we get

$$|\Pr[\mathsf{Game}_2(n) = 1] - \Pr[\mathsf{Game}_3(n) = 1]| \leq \mathsf{negl}(n)\,(n). \tag{20}$$

Hop from Game_3 *to* Game_4. In this hop, the obfuscated circuit is replaced. It is critical to observe, that if r' is *not in the range* of PRG, then the two circuits are functionally equivalent, since the PRF will never be invoked on the point the key is punctured on. In this case, the distance between the two games can therefore be bounded by the security of the sacO. If r' *is in the range* of PRG, then we have no guarantee, but this only occurs with probabilty $2^{-n/2}$. Thus it follows that

$$|\Pr[\mathsf{Game}_3(n) = 1] - \Pr[\mathsf{Game}_4(n) = 1]| \leq \delta(|C|\,, n) + 2^{-n/2}. \tag{21}$$

Hop from Game_4 *to* Game_5. Note that in Game_5, the PRF value is replaced with a uniformly chosen random value. This allows us to bound the difference between the two games by the security of the puncturable pseudorandom function. To bound the difference between games Game_4 and Game_5, we construct a distinguisher $(\mathcal{D}_1, \mathcal{D}_2)$ with advantage

$$|\Pr[\mathsf{Game}_4(n) = 1] - \Pr[\mathsf{Game}_5(n) = 1]|$$

against the puncturable PRF as follows:

$\mathcal{D}_1(1^n; r_1)$	$\mathcal{D}_2(\mathsf{st}, k^*, r', s; r_2)$
$r' \leftarrow\!\!\$\ \{0,1\}^n$	$\mathsf{pk} \leftarrow\!\!\$\ \mathsf{O}(\mathsf{C}_{\mathsf{SW}}[k^*])$
return $(\bot, r')$	$b \leftarrow\!\!\$\ \{0,1\}$
	$c := (b \oplus s)$
	$b' \leftarrow\!\!\$\ \mathcal{A}(\mathsf{pk}, c)$
	return $(C'(x_0) \stackrel{?}{=} b)$

Observe, that in the case where $\mathcal{A}_2$ receives the PRF value, it holds that

$$\Pr_{k,r_1,t,r_2}[\mathcal{D}_2(\mathsf{st}, \mathsf{k}^*, r', \mathsf{PRF}(k, r'); r_2) = 1] = \Pr[\mathsf{Game}_4(n) = 1]. \tag{22}$$

If on the other hand, $\mathcal{D}_2$ receives an s chosen uniformly at random, it holds that

$$\Pr_{k,s,r_1,t,r_2}[\mathcal{D}_2(\mathsf{st}, \mathsf{k}^*, r', s; r_2) = 1] = \Pr[\mathsf{Game}_5(n) = 1] \tag{23}$$

By security of the puncturable PRF, it must hold that there exists a negligible function $\mathsf{negl}(n)$ such that

$$\begin{aligned}|\Pr_{k,r_1,t,r_2}[\mathcal{D}_2(\mathsf{st}, \mathsf{k}^*, r', \mathsf{PRF}(k, r'); r_2) = 1] \\ - \Pr_{k,s,r_1,t,r_2}[\mathcal{D}_2(\mathsf{st}, \mathsf{k}^*, r', s; r_2) = 1]| \leq \mathsf{negl}(n)\end{aligned}$$

Combining this with Equation 22 and Equation 23 yields

$$|\Pr[\mathsf{Game}_5(n) = 1] - \Pr[\mathsf{Game}_4(n) = 1]| \leq \mathsf{negl}(n) \tag{24}$$

It remains to bound the probability $\Pr[\mathsf{Game}_5(n) = 1]$. However, the ciphertext in Game_5 is simply a uniformly distributed random value that does not reveal any information about b. Therefore, it is easy to see that $\Pr[\mathsf{Game}_5(n) = 1] = \frac{1}{2}$. Combining this with Equations 17, 20, 21, and 24, we can conclude that

$$\Pr[\mathsf{Game}_1(n) = 1] \leq \frac{1}{2} + \delta(|C|, n) + \epsilon(|C|, n),$$

thus concluding the proof.