QUANTUM OBFUSCATION

GORJAN ALAGIC AND BILL FEFFERMAN

ABSTRACT. Encryption of data is fundamental to secure communication. Beyond encryption of data lies *obfuscation*, i.e., encryption of functionality. It has been known for some time that the most powerful classical obfuscation, so-called "black-box obfuscation," is impossible. In this work, we initialize the rigorous study of obfuscating programs *via quantum-mechanical means.* We prove quantum analogues of several foundational results in obfuscation, including the aforementioned black-box impossibility result.

In its most powerful "quantum black-box" instantiation, a quantum obfuscator would turn a description of a quantum program f into a quantum state ρ_f, such that anyone in possession of ρ_f can repeatedly evaluate f on inputs of their choice, but never learn *anything else* about the original program. We formalize this notion of obfuscation, and prove an impossibility result: such obfuscation is only possible in a setting where the adversary never has access to more than one obfuscation (of either the same program, or of different programs.) Our proof involves a novel technical idea: chosen-ciphertext-secure encryption for quantum states. In addition, we show that some applications of obfuscation still appear possible in spite of our impossibility result. These include encryption for quantum states, quantum fully-homomorphic encryption, and quantum money.

We also define quantum versions of indistinguishability obfuscation and best-possible obfuscation. We then show that these notions are equivalent, and that their perfect and statistical variants are impossible to achieve. The remaining (i.e., computational) variant would still have an application of interest: witness encryption for QMA.

1. INTRODUCTION

1.1. **Obfuscation.** Obfuscation is *encryption of functionality*, and is arguably the most powerful cryptographic ability that may yet be possible. It implies (with some caveats) almost any other cryptographic construction imaginable. It could also be used to protect intellectual property in software, and provide secure software patching. In all of these applications, an *obfuscator* is an efficient algorithm which rewrites programs so that they satisfy:

(1) *functional equivalence:* input/output functionality is unchanged;
(2) *polynomial slowdown:* efficiency is maintained;
(3) *obfuscation:* the code of the output program is "hard to understand."

The last condition admits several rigorous formulations. The strongest is the so-called "virtual black-box" condition, which says that the obfuscated program is no more useful than a formless, impenetrable box which simply accepts inputs and produces outputs.

1.2. **Classical status.** The first major result in classical obfuscation was the 2001 proof by Barak et al. that virtual black-box obfuscation is impossible, and that many of the applications are impossible to achieve generically [6, 7]. An important step in formulating alternative notions of obfuscation was taken by Goldwasser and Rothblum; they defined *indistinguishability obfuscation* and *best-possible obfuscation* [15]. Indistinguishability requires that functionally-equivalent circuits are mapped to indistinguishable distributions (over circuits); best-possible requires that circuits are mapped to the "least leaky" functionally equivalent circuit. Both definitions have perfect, statistical, and computational variants. In [15] it was shown that indistinguishability and best-possible are equivalent, and that the perfect and statistical versions are impossible (barring PH collapse) [15].

In 2013, in a breakthrough result, Garg et al. proposed a convincing candidate for the one remaining possibility: computational indistinguishability obfuscation. Their proposal was based on the hardness of a certain problem in multilinear maps [12]. It was accompanied by another breakthrough, which showed a range of applications so wide, that indistinguishability obfuscation was proposed as a new "'central hub' for cryptography" [19]. These two breakthroughs were followed by a flurry of new activity in the area, including several new proposals and applications [8, 9, 10, 11, 14, 16]. Unfortunately, the quantum security of the underlying hardness assumptions has recently been put into doubt [18].

1.3. **Quantum status.** Quantum obfuscation is essentially an unexplored topic, and the present work appears to be the first rigorous treatment of the foundational questions. The question of whether quantum obfuscation is possible was posed as one of Scott Aaronson's "semi-grand challenges" for quantum computation [1]. Since so little work on quantum obfuscation has appeared, we briefly discuss some related results. In [2], Aaronson proposed a *complexity-theoretic no-cloning theorem*; this theorem was eventually proved in a paper on quantum money [3]. In related work, Mosca and Stebila showed how to use Aaronson's theorem to give a simple black-box quantum money scheme, and suggested the possibility of using a quantum circuit obfuscator in place of the black box [17]. More recently, Alagic, Jeffery and Jordan proposed obfuscators for both classical (reversible) circuits and quantum circuits, based on ideas from topological quantum computation [4]. The proposed obfuscator satisfies indistinguishability for a restricted set of circuit equivalences; its usefulness is unclear at this time.

2. Our results

We now summarize our contributions; details are explained in the full technical version. In what follows, poly-time algorithms will be called PT, PPT, or QPT; these mean (respectively) classical deterministic, probabilistic, and quantum. The unitary operator implemented by a quantum circuit C is denoted U_C. Functions decaying faster than any inverse-polynomial are denoted $\mathrm{negl}(\cdot)$.

2.1. Quantum black-box obfuscation.

Definition 1. *A **black-box quantum obfuscator** is a quantum algorithm $\mathcal{O}$ and a QPT $\mathcal{J}$ such that for any n-qubit quantum circuit C, $\mathcal{O}(C)$ is a poly(n)-qubit quantum state satisfying:*

(1) *(functional equivalence)* $\left\|\mathcal{J}(\mathcal{O}(C)\otimes\rho) - U_C\rho U_C^\dagger\right\|_{\mathrm{tr}} \leq \mathrm{negl}(n)$ *for all n-qubit states ρ;*

(2) *(virtual black-box)* $\forall$ *QPT* $\mathcal{A}$ $\exists$ *QPT* $\mathcal{S}^{U_C}$ *such that* $|\Pr[\mathcal{A}(\mathcal{O}(C))=1] - \Pr[\mathcal{S}^{U_C}(0^n)=1]| \leq \mathrm{negl}(n)$.

We emphasize that $\mathcal{O}$ need not be poly-time, although its outputs must be poly-sized. A departure from the classical definition is the addition of the "interpreter" $\mathcal{J}$. It is natural since there must be *some efficient way* to run U_C using the state $\mathcal{O}(C)$; whatever it is, we denote it by $\mathcal{J}$. We prove the following impossibility results for these obfuscators.

Theorem 1. *Black-box quantum obfuscation is impossible for pairs: an adversary with access to two outputs*[1] *of the obfuscator can violate the black-box condition.*

Corollary 2. *Black-box obfuscation of quantum circuits into quantum circuits is impossible.*

To prove the theorem, we give an explicit construction of a circuit family $\mathcal{C}$, such that no family of states is an obfuscation of $\mathcal{C}$. Corollary 2 is a quantum generalization of the main result of Barak et al. [6]. We emphasize that, perhaps surprisingly, one cannot conclude Corollary 2 from [6] by appealing to the fact that classical functions are a special case of quantum functions; for one, the

[1] If the two outputs came from the same input circuit, then impossibility only holds if the output states are **reusable**, in the sense that $\mathcal{J}$ outputs another state which again satisfies functional equivalence. This is weaker than cloneability.

adversaries and simulators are now also quantum. Generalizing to the full Theorem 1 is also nontrivial. The main overall technical obstacle is to show that the states output by the obfuscator can be "executed on one another," without also revealing certain secrets to simulators having only black-box access to the original circuits. A crucial ingredient in overcoming this obstacle is a notion of *chosen-ciphertext-secure quantum encryption.* To achieve this, we make use of a new result:

Theorem 3. *If quantum-secure one-way functions (qOWF) exist, then so do IND-CCA1-secure symmetric-key quantum encryption schemes (qSKE).*

The above requires several new definitions (e.g., IND-CCA1 for quantum encryption.) A complete treatment of the subject of quantum encryption under computational assumptions, including the proof of Theorem 3, appears in a recent joint work [5].

In addition, we provide several applications of quantum black-box obfuscation, which still appear feasible (in some form) in spite of Theorem 1. They are briefly outlined as follows.

(1) **Public-key quantum encryption from private-key encryption.** Here the public key is the obfuscation of the encryption circuit. No trapdoor permutations required.
(2) **qOWF imply IND-CPA public-key quantum-fully-homomorphic encryption.** Evaluation keys are obfuscations of a universal decrypt-compute-encrypt circuit.
(3) **Public-key quantum money**. This was proposed by Mosca and Stebila [17], using a result of Aaronson and Christiano [2, 3]. As we show, a certain adaptation survives Theorem 1.

We emphasize that applications 1 and 2 also work for achieving *classical functionality* from a quantum obfuscator; however, they use quantum ciphertexts and quantum keys—another approach not considered before.

2.2. **Quantum indistinguishability obfuscation.** Following the classical approach of Goldwasser and Rothblum [15], we define a version of Definition 1 which guarantees indistinguishability of obfuscator outputs, by replacing the black-box condition (2) with:

(2) *(indistinguishability) if* $\|U_{C_1} - U_{C_2}\| \leq \mathrm{negl}(n)$, *then* $\|\mathcal{O}(C_1) - \mathcal{O}(C_2)\|_* \leq \mathrm{negl}(n)$.

We also define a notion of *quantum best-possible obfuscation*:

(2) *(best-possible) if* $\|U_{C_1} - U_{C_2}\| \leq \mathrm{negl}(n)$, *then for all QPT* $\mathcal{A}$ *there exists a QPT* $\mathcal{S}$ *satisfying* $\|\mathcal{A}(\mathcal{O}(C_1)) - \mathcal{S}(C_2)\|_* \leq \mathrm{negl}(n)$.

Both definitions above have three variants, depending on the nature of the norm $\|\cdot\|_*$: perfect, statistical, and computational (against QPTs). We prove the following equivalence result.

Theorem 4. *A QPT is an indistinguishability obfuscator if and only if it is a best-possible obfuscator.*

We also show that these definitions are only achievable when the distinguishability is guaranteed only against computationally bounded (quantum) adversaries.

Theorem 5. *If efficient quantum statistical-indistinguishability obfuscators exist, then coQMA* $\subseteq$ *QSZK.*[2]

This means that, just as in the classical world, computational indistinguishability is the only surviving variant. We end with an application for such an obfuscator: quantum witness encryption for QMA. Witness encryption for a QMA language L provides for encryption of a plaintext x to a potential instance l. The security guarantee is that (i.) if $l \in L$, then any valid witness allows for decryption of x, and (ii.) if $l \notin L$, then encryptions of different plaintexts are indistinguishable.

Theorem 6. *Quantum computational-indistinguishability obfuscation implies witness encryption for QMA.*

Classical witness encryption is known to have numerous applications [13]. We conjecture that many of the other recently discovered classical applications of computational indistinguishability obfuscation (see, e.g., [19]) also have interesting quantum analogues or extensions.

[2] In fact, we can show that obfuscation of CPTP circuits implies PSPACE $\subseteq$ QSZK!

REFERENCES

[1] Scott Aaronson. Ten semi-grand challenges for quantum computing theory, July 2005. URL http://www.scottaaronson.com/writings/qchallenge.html. Retrieved 02/16.

[2] Scott Aaronson. Quantum copy-protection and quantum money. In *Computational Complexity, 2009. CCC'09. 24th Annual IEEE Conference on*, pages 229–242. IEEE, 2009.

[3] Scott Aaronson and Paul Christiano. Quantum money from hidden subspaces. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 41–60. ACM, 2012.

[4] Gorjan Alagic, Stacey Jeffery, and Stephen Jordan. Circuit obfuscation using braids. In *Proceedings of TQC 2014*, volume 27, pages 141–160, 2014.

[5] Gorjan Alagic, Anne Broadbent, Bill Fefferman, Tommaso Gagliardoni, Christian Schaffner, and Michael St. Jules. Computational security of quantum encryption. 2016. URL http://arxiv.org/abs/1602.01441.

[6] Boaz Barak, Oded Goldreich, Russell Impagliazzo, Steven Rudich, Amit Sahai, Salil P. Vadhan, and Ke Yang. On the (im)possibility of obfuscating programs. In *Proceedings of the 21st Annual International Cryptology Conference on Advances in Cryptology*, CRYPTO '01, pages 1–18, London, UK, UK, 2001. Springer-Verlag. ISBN 3-540-42456-3. URL http://dl.acm.org/citation.cfm?id=646766.704152.

[7] Boaz Barak, Oded Goldreich, Russell Impagliazzo, Steven Rudich, Amit Sahai, Salil Vadhan, and Ke Yang. On the (im)possibility of obfuscating programs. *J. ACM*, 59(2):6:1–6:48, May 2012. ISSN 0004-5411. doi:10.1145/2160158.2160159. URL http://doi.acm.org/10.1145/2160158.2160159.

[8] Boaz Barak, Sanjam Garg, Yael Tauman Kalai, Omer Paneth, and Amit Sahai. Protecting obfuscation against algebraic attacks. In PhongQ. Nguyen and Elisabeth Oswald, editors, *Advances in Cryptology EUROCRYPT 2014*, volume 8441 of *Lecture Notes in Computer Science*, pages 221–238. Springer Berlin Heidelberg, 2014. ISBN 978-3-642-55219-9. doi:10.1007/978-3-642-55220-5_13. URL http://dx.doi.org/10.1007/978-3-642-55220-5_13.

[9] Nir Bitansky, Ran Canetti, Henry Cohn, Shafi Goldwasser, Yael Tauman Kalai, Omer Paneth, and Alon Rosen. The impossibility of obfuscation with auxiliary input or a universal simulator. In Juan A. Garay and Rosario Gennaro, editors, *Advances in Cryptology CRYPTO 2014*, volume 8617 of *Lecture Notes in Computer Science*, pages 71–89. Springer Berlin Heidelberg, 2014. ISBN 978-3-662-44380-4. doi:10.1007/978-3-662-44381-1_5. URL http://dx.doi.org/10.1007/978-3-662-44381-1_5.

[10] Dan Boneh and Mark Zhandry. Multiparty key exchange, efficient traitor tracing, and more from indistinguishability obfuscation. In JuanA. Garay and Rosario Gennaro, editors, *Advances in Cryptology CRYPTO 2014*, volume 8616 of *Lecture Notes in Computer Science*, pages 480–499. Springer Berlin Heidelberg, 2014. ISBN 978-3-662-44370-5. doi:10.1007/978-3-662-44371-2_27. URL http://dx.doi.org/10.1007/978-3-662-44371-2_27.

[11] Zvika Brakerski and Guy N. Rothblum. Virtual black-box obfuscation for all circuits via generic graded encoding. In Yehuda Lindell, editor, *Theory of Cryptography*, volume 8349 of *Lecture Notes in Computer Science*, pages 1–25. Springer Berlin Heidelberg, 2014. ISBN 978-3-642-54241-1. doi:10.1007/978-3-642-54242-8_1. URL http://dx.doi.org/10.1007/978-3-642-54242-8_1.

[12] S. Garg, C. Gentry, S. Halevi, M. Raykova, A. Sahai, and B. Waters. Candidate indistinguishability obfuscation and functional encryption for all circuits. In *Foundations of Computer Science (FOCS), 2013 IEEE 54th Annual Symposium on*, pages 40–49, Oct 2013. doi:10.1109/FOCS.2013.13.

[13] Sanjam Garg, Craig Gentry, Amit Sahai, and Brent Waters. Witness encryption and its applications. In *Proceedings of the Forty-fifth Annual ACM Symposium on Theory of Computing*, STOC '13, pages 467–476, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2029-0. doi:10.1145/2488608.2488667. URL http://doi.acm.org/10.1145/2488608.2488667.

[14] Sanjam Garg, Craig Gentry, Shai Halevi, and Daniel Wichs. On the implausibility of differing-inputs obfuscation and extractable witness encryption with auxiliary input. In JuanA. Garay and Rosario Gennaro, editors, *Advances in Cryptology CRYPTO 2014*, volume 8616 of *Lecture Notes in Computer Science*, pages 518–535. Springer Berlin Heidelberg, 2014. ISBN 978-3-662-44370-5. doi:10.1007/978-3-662-44371-2_29. URL http://dx.doi.org/10.1007/978-3-662-44371-2_29.

[15] Shafi Goldwasser and Guy N. Rothblum. On best-possible obfuscation. In SalilP. Vadhan, editor, *Theory of Cryptography*, volume 4392 of *Lecture Notes in Computer Science*, pages 194–213. Springer Berlin Heidelberg, 2007. ISBN 978-3-540-70935-0. doi:10.1007/978-3-540-70936-7_11. URL http://dx.doi.org/10.1007/978-3-540-70936-7_11.

[16] Susan Hohenberger, Amit Sahai, and Brent Waters. Replacing a random oracle: Full domain hash from indistinguishability obfuscation. In PhongQ. Nguyen and Elisabeth Oswald, editors, *Advances in Cryptology EUROCRYPT 2014*, volume 8441 of *Lecture Notes in Computer Science*, pages 201–220. Springer Berlin Heidelberg, 2014. ISBN 978-3-642-55219-9. doi:10.1007/978-3-642-55220-5_12. URL http://dx.doi.org/10.1007/978-3-642-55220-5_12.

[17] Michele Mosca and Douglas Stebila. Quantum coins. *Error-Correcting Codes, Finite Geometries and Cryptography*, 523:35–47, 2010.

[18] Chris Peikert. What does gchq "cautionary tale" mean for lattice cryptography? http://web.eecs.umich.edu/ cpeikert/soliloquy.html, June 2015. Retrieved 09/2015.

[19] 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*, STOC '14, pages 475–484, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2710-7. doi:10.1145/2591796.2591825. URL http://doi.acm.org/10.1145/2591796.2591825.