Zero-knowledge proof systems for QMA

Extended Abstract for QCrypt 2016. Full version available at: arXiv:1604.02804.

Anne Broadbent, Zhengfeng Ji, Fang Song and John Watrous

Prior work has established that all problems in NP admit classical zero-knowledge proof systems, and under reasonable hardness assumptions for quantum computations, these proof systems can be made secure against quantum attacks. We prove a result representing a further quantum generalization of this fact, which is that every problem in the complexity class QMA has a quantum zero-knowledge proof system. More specifically, assuming the existence of an unconditionally binding and quantum computationally concealing commitment scheme, we prove that every problem in the complexity class QMA has a quantum interactive proof system that is zero-knowledge with respect to efficient quantum computations.

Our QMA proof system is sound against arbitrary quantum provers, but only requires an honest prover to perform polynomial-time quantum computations, provided that it holds a quantum witness for a given instance of the QMA problem under consideration. The proof system relies on a new variant of the QMA-complete local Hamiltonian problem in which the local terms are described by Clifford operations and standard basis measurements. We believe that the QMA-completeness of this problem may have other uses in quantum complexity.

Zero-knowledge proof systems, first introduced by Goldwasser, Micali and Rackoff [10], are interactive protocols that allow a prover to convince a verifier of the validity of a statement while revealing no additional information beyond the statement's validity. Although paradoxical as it appears, several problems that are not known to be efficiently computable, such as the Quadratic Non-Residuosity, Graph Isomorphism, and Graph Non-Isomorphism problems, were shown to admit zero-knowledge proof systems [9,10]. Under reasonable intractability assumptions, Goldreich, Micali and Wigderson [9] gave a zero-knowledge protocol for the Graph 3-Coloring problem and, because of its NP-completeness, for all NP problems. This line of work was further extended in [4], which showed that all problems in IP have zero-knowledge proof systems.

Since the invention of this concept, zero-knowledge proof systems have become a cornerstone of modern theoretical cryptography. In addition to the conceptual innovation of formulating a complexity-theoretic notion of knowledge, zero-knowledge proof systems are essential building blocks in a host of cryptographic constructions. One notable example is the design of secure two-party and multi-party computation protocols [8].

The extensive works on zero-knowledge largely reside in a classical world. The development of quantum information science and technology has urged another look at the landscape of zero-knowledge proof systems in a *quantum* world. Namely, both honest users and adversaries may potentially possess the capability to exchange and process quantum information. There are, of course, zero-knowledge protocols that immediately become insecure in the presence of quantum attacks due to efficient quantum algorithms that break the intractability assumptions upon which these protocols rely. For instance, Shor's quantum algorithms for factoring and computing discrete logarithms [12] invalidate the use of these problems, generally conjectured to be classically hard, as a basis for the security of zero-knowledge protocols against quantum attacks. Even with computational assumptions against quantum adversaries, however, it is still highly nontrivial to establish the security of classical zero-knowledge proof systems in the presence of malicious *quantum* verifiers because of a technical reason that we now briefly explain.

The zero-knowledge property of a proof system for a fixed input string is concerned with the computations that may be realized through an interaction between a (possibly malicious) verifier and the prover. That is, the malicious verifier may take an arbitrary input (usually called the *auxiliary input* to distinguish it from the input string to the proof system under consideration), interact

with the prover in any way it sees fit, and produce an output that is representative of what it has learned through the interaction. Roughly speaking, the prover is said to be *zero-knowledge* on the fixed input string if any computation of the sort just described can be efficiently approximated[1] by a *simulator* operating entirely on its own—meaning that it does not interact with the prover, and in the case of an NP problem it does not possess a witness for the fixed problem instance being considered. The proof system is then said to be zero-knowledge when this zero-knowledge property holds for all yes-instances of the problem under consideration.

Classically speaking, the zero-knowledge property is typically established through a technique known as *rewinding*. In essence, the simulator can store a copy of its auxiliary input, and it can make guesses and store intermediate states representing a hypothetical prover/verifier interaction—and if it makes a bad guess or otherwise experiences bad luck when simulating this hypothetical interaction, it simply reverts to an earlier stage (or possibly back to the beginning) of the simulation and tries again. Indeed, it is generally the simulator's freedom to disregard the temporal restrictions of the actual prover/verifier interaction in a way such as this that makes it possible to succeed.

However, rewinding a quantum simulation is more problematic; the *no-cloning theorem* [16] forbids one from copying quantum information, making it impossible to store a copy of the input or of an intermediate state, and measurements generally have an irreversible effect [7] that may partially destroy quantum information. Such difficulties were first observed by van de Graaf [13] and further studied in [6,14]. Later, a *quantum rewinding* technique was found [15] to establish that several interactive proof systems, including the Goldreich-Micali-Wigderson Graph 3-Coloring proof system [9], remain zero-knowledge against malicious quantum verifiers (under appropriate quantum intractability assumptions in some cases). It follows that all NP problems have zero-knowledge proof systems even against quantum malicious verifiers, provided that a quantum analogue of the intractability assumption required by the Goldreich-Micali-Wigderson Graph 3-Coloring proof system are in place.

This work studies the quantum analogue of NP, known as QMA, in the context of zero-knowledge. These are problems with a succinct *quantum* witness satisfying similar completeness and soundness to NP (or its randomized variant MA). Quantum witnesses and verification are conjectured to be more powerful than their classical counterparts: there are problems that admit short quantum witnesses, whereas there is no known method for verification using a polynomial-sized classical witness. In other words, NP $\subseteq$ QMA holds trivially, and the containment is typically conjectured to be proper. The question we address in this paper is: *Does every problem in* QMA *have a zero-knowledge quantum interactive proof system?* In more philosophical terms, viewing quantum witnesses as precious sources of knowledge: *Can one always devise a proof system that reveals nothing about a quantum witness beyond its validity?*

Our contributions We answer the above question positively by constructing a quantum interactive proof system for any problem in QMA that is zero-knowledge against any polynomial-time quantum adversary, under a reasonable quantum intractability assumption.

Theorem 1. *Assuming the existence of an unconditionally binding and quantum computationally concealing bit commitment scheme, every problem in* QMA *has a quantum computational zero-knowledge proof system.*

[1] Different notions of approximations are considered, including *statistical* approximations and *computational* approximations, which require that the simulator's computation is either statistically (or information-theoretically) indistinguishable or computationally indistinguishable from the malicious verifier's computation. This paper is primarily concerned with the computational variant.

A few of the desirable features of our proof system are as follows:

1. Our proof system has a simple structure, similar to the classical Goldreich-Micali-Wigderson Graph 3-Coloring proof system (and to the so-called Σ-protocols more generally). It can be viewed as a three-phase process: the prover commits to a quantum witness, the verifier makes a random challenge, and finally the prover responds to the challenge by partial opening of the committed information that suffices to certify the validity.
2. All communications in our proof system are classical except for the first commitment message, and the verifier can measure the quantum message immediately upon its arrival (which has a strong technological appeal).
3. Our protocol is based on mild computational assumptions. The sort of bit commitment scheme it requires can be implemented, for instance, under the existence of injective one-way functions that are hard to invert in quantum polynomial time.
4. Our protocol is prover-efficient. It is sound against general quantum provers, but given a valid quantum witness, an honest prover only needs to perform efficient quantum computations. As has already been suggested, aside from the preparation of the first quantum message, all of the remaining computations performed by the honest prover are classical polynomial-time computations.

As a key ingredient of our zero-knowledge proof system, we introduce a new variant of the k-local Hamiltonian problem and prove that it remains QMA-complete (with respect to Karp reductions). The k-local Hamiltonian problem asks if the minimum eigenvalue (or ground state energy in physics parlance) of an n-qubit Hamiltonian $H = \sum_j H_j$, where each H_j is k-local (i.e., acts trivially on all but k of the n qubits), is below a particular threshold value. This problem was introduced and proved to be QMA-complete (for the case $k = 5$) by Kitaev [11]. We show that each H_j can be restricted to be realized by a Clifford operation, followed by a standard basis measurement, and the QMA-completeness is preserved. Beyond its use in this paper, this fact has the potential to provide other insights into the study of quantum Hamiltonian complexity. For an arbitrary problem $A \in$ QMA, we can reduce an instance of A efficiently to an instance of the k-local Clifford Hamiltonian problem, and a valid witness for A can also be transformed into a witness for the corresponding k-local Clifford Hamiltonian problem instance by an efficient quantum procedure. As a result, A has a zero-knowledge proof system by composing this reduction with our zero-knowledge proof system for the k-local Clifford Hamiltonian problem.

Our proof system also employs a new encoding scheme for quantum states, which we construct by extending the *trap scheme* proposed in [5]. While our new scheme can be seen as a *quantum authentication scheme* (cf. [1–3]), it in addition allows performing arbitrary constant-qubit Clifford circuits and measuring in the computational basis directly on authenticated data without the need for auxiliary states. The only previously known scheme supporting this feature requires high-dimensional quantum systems (i.e., qudits rather than qubits) [3], which make it inconvenient in our setting where all quantum operations are on qubits.

References

[1] AHARONOV, D., BEN-OR, M., AND EBAN, E. Interactive proofs for quantum computations. In *Innovations in Computer Science* (2010), pp. 453–469.

[2] BARNUM, H., CRÉPEAU, C., GOTTESMAN, D., SMITH, A., AND TAPP, A. Authentication of quantum messages. In *Proceedings of the 43th Annual IEEE Symposium on Foundations of Computer Science* (2002), pp. 449–458.

[3] BEN-OR, M., CRÉPEAU, C., GOTTESMAN, D., HASSIDIM, A., AND SMITH, A. Secure multiparty quantum computation with (only) a strict honest majority. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science* (2006), pp. 249–260.

[4] BEN-OR, M., GOLDREICH, O., GOLDWASSER, S., HÅSTAD, J., KILIAN, J., MICALI, S., AND ROGAWAY, P. Everything provable is provable in zero-knowledge. In *Advances in Cryptology – CRYPTO 1988* (1990), vol. 403 of *Lecture Notes in Computer Science*, Springer-Verlag, pp. 37–56.

[5] BROADBENT, A., GUTOSKI, G., AND STEBILA, D. Quantum one-time programs. In *Advances in Cryptology – CRYPTO 2013* (2013), vol. 8043 of *Lecture Notes in Computer Science*, Springer, pp. 344–360.

[6] DAMGÅRD, I., FEHR, S., AND SALVAIL, L. Zero-knowledge proofs and string commitments withstanding quantum attacks. In *Advances in Cryptology – CRYPTO 2004* (2004), vol. 3152 of *Lecture Notes in Computer Science*, Springer, pp. 254–272.

[7] FUCHS, C. A., AND PERES, A. Quantum-state disturbance versus information gain: Uncertainty relations for quantum information. *Physical Review A 53*, 4 (1996), 2038.

[8] GOLDREICH, O., MICALI, S., AND WIGDERSON, A. How to play ANY mental game. In *Proceedings of the 19th Annual ACM Symposium on Theory of Computing* (1987), pp. 218–229.

[9] GOLDREICH, O., MICALI, S., AND WIGDERSON, A. Proofs that yield nothing but their validity or all languages in NP have zero-knowledge proof systems. *Journal of the ACM 38*, 3 (1991), 690–728.

[10] GOLDWASSER, S., MICALI, S., AND RACKOFF, C. The knowledge complexity of interactive proof systems. *SIAM Journal on Computing 18*, 1 (1989), 186–208.

[11] KITAEV, A. Y., SHEN, A. H., AND VYALYI, M. N. *Classical and Quantum Computation*, vol. 47 of *Graduate Studies in Mathematics*. American Mathematical Society, 2002.

[12] SHOR, P. W. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Journal on Computing 26*, 5 (1997), 1484–1509.

[13] VAN DE GRAAF, J. *Towards a Formal Definition of Security for Quantum Protocols*. PhD thesis, Université de Montréal, 1997.

[14] WATROUS, J. Limits on the power of quantum statistical zero-knowledge. In *Proceedings of the 43rd Annual IEEE Symposium on Foundations of Computer Science* (2002), pp. 459–468.

[15] WATROUS, J. Zero-knowledge against quantum attacks. *SIAM Journal on Computing 39*, 1 (2009), 25–58.

[16] WOOTTERS, W. K., AND ZUREK, W. H. A single quantum cannot be cloned. *Nature 299* (1982), 802–803.