Quantum Proofs of Knowledge

Dominique Unruh

University of Tartu, Estonia

Abstract. We motivate, define and construct quantum proofs of knowledge, proofs of knowledge secure against quantum adversaries. Our constructions are based on a new quantum rewinding technique that allows us to extract witnesses in many classical proofs of knowledge. We give criteria under which a classical proof of knowledge is a quantum proof of knowledge. Combining our results with Watrous' results on quantum zero-knowledge, we show that there are zero-knowledge quantum proofs of knowledge for all languages in NP (assuming quantum 1-1 one-way functions).

1 Introduction

Cryptographic protocols, with few exceptions, are based on the assumption that certain problems are computationally hard. Typical examples include specific number-theoretic problems such as the difficulty of finding discrete logarithms, and general problems such as inverting one-way functions. It is well-known, however, that many such problems would become easy in the advent of quantum computers. Shor's algorithm [16], e.g., efficiently solves the discrete logarithm problem and allows to factor large integers. While quantum computers do not exist today, it is not unreasonable to expect quantum computers to be available in the future. To meet this threat, we need cryptographic protocols that are secure even in the presence of an adversary with a quantum computer. We stress that this does not necessarily imply that the protocol itself should make use of quantum technology; instead, it is preferable that the protocol itself can be easily implemented on today's readily-available classical computers.

Finding such quantum-secure protocols, however, is not trivial. Even when we have found suitable complexity-theoretic assumptions such as the hardness of certain lattice problems, a classical protocol based on these assumptions may fail to be secure against quantum computers. The reason for this is that many cryptographic proofs use a technique called rewinding. This technique requires that it is possible, when simulating some machine, to make snapshots of the state of that machine and then later to go back to that snapshot. As first observed by van de Graaf [9], classical rewinding-based proofs do not carry over to the quantum case. Two features unique to the quantum setting prohibit (naive) rewinding: The no-cloning theorem [21] states that quantum-information cannot be copied, so we cannot make snapshots. Furthermore, measurements destroy information, so interacting with a simulated machine may destroy information that would be needed later.

This leads to the following observation: Even if a classical protocol is proven secure based on the hardness of some problem, and even if that problem is hard even for quantum computers, we have no guarantee that the protocol is secure against quantum computers. The reduction of the protocol's security to the problem's hardness may be based on inherently classical features such as the possibility of rewinding.

An example of a protocol construction that suffers from this difficulty is zero-knowledge proofs. Zero-knowledge proofs are interactive proofs with the special property that the verifier does not learn anything except the validity of the proven statement. Zero-knowledge proofs are inherently based on rewinding (at least as long as we do not assume additional trusted setup such as so-called common-reference strings). Yet, zero-knowledge proofs are one of the most powerful tools available to the cryptographer; a multitude of protocol constructions use zero-knowledge proofs. These protocol constructions cannot be proven secure without using rewinding. To resolve this issue, Watrous [19] introduced a quantum rewinding technique. This technique allows to prove the quantum security of many common zero-knowledge proofs. One should note, however, that Watrous' technique is restricted to a specific type of rewinding: If we use Watrous' technique, whenever some machine rewinds another machine to an earlier point, the rewinding machine forgets everything it learned after that point (we call this oblivious rewinding). That is, we can only use Watrous' technique to backtrack if the rewinding machine made a mistake that should be corrected, but it cannot be used to collect and combine information from different branches of an execution.

Constructing quantum zero-knowledge proofs solves, however, only half of the problem. In many, if not most, applications of zero-knowledge proofs one needs zero-knowledge *proofs of knowledge*. A proof of knowledge [7,3] is a proof system which does not only show the truth of a certain statement, but also that the prover knows a witness for that statement. This is made clearer by an example: Assume that Alice wishes to convince Bob that she (the prover) is in possession of a signature issued by some certification authority. For privacy reasons, Alice does not wish to reveal the signature itself. If Alice uses a zero-knowledge *proof*, she can only show the statement "there exists a signature with respect to the CA's public key". This does not, however, achieve anything: A signature always exists in a mathematical sense, even if it has never been computed. What Alice wishes to say is: "I *know* a signature with respect to the CA's public key." To prove such a statement, Alice needs a zero-knowledge *proof of knowledge*; a proof of knowledge would convince Bob that Alice indeed knows a witness, i.e., a signature. Very roughly, the definition of a proof of knowledge is the following: Whenever the prover can convince the verifier, one can extract the witness from the prover given oracle access to the prover. Here oracle access means that one can interact with the prover and *rewind* him. Thus, we have the same problem as in the case of quantum zero-knowledge proofs: To get proofs of knowledge that are secure against quantum adversaries, we need to use quantum rewinding. Unfortunately, Watrous' *oblivious* rewinding does not work here; proofs of

knowledge use rewinding to produce two (or more) different protocol traces and compute the witness by combining the information from both traces. Thus, we are back to where we started: to make classical cryptographic protocols work in a quantum setting, we need (in many cases) quantum zero-knowledge *proofs of knowledge*, but we only have constructions for quantum zero-knowledge *proofs*.

Our contribution. We define and construct quantum proofs of knowledge. Our protocols are classical (i.e., honest parties do not use quantum computation or communication) but secure against quantum adversaries. Our constructions are based on a new quantum rewinding technique (different from Watrous' technique) that allows us to extract witnesses in many classical proofs of knowledge. We give criteria under which a classical proof of knowledge is a quantum proof of knowledge. Combining our results with Watrous' results on zero-knowledge, we can show that there are zero-knowledge quantum proofs of knowledge for all languages in NP (assuming quantum 1-1 one-way functions). (We leave it as an open question whether unconditionally secure protocols exist for more restricted languages related, e.g., to lattice-problems.)

Also, we believe that the use of our rewinding technique is not limited to QPoKs. For example, we encourage the reader to try to prove the following without using our technique: Given a quantum computationally binding commitment scheme, first let the adversary commit, and then give a random value v to the adversary. Then the probability that the adversary opens the commitment to v is negligible.[1]

Follow-up work. In subsequent work, Lunemann and Nielsen [14] and Hallgren, Smith, and Song [12] developed zero-knowledge QPoKs with the additional advantage of allowing to simultaneously simulate an interaction with the malicious prover and extract the witness; this property is necessary in some multi-party computations. (In contrast, in our setting the initial state of the prover could be lost after extracting.) We stress, however, that this powerful feature comes at a cost: They need considerably stronger assumptions, namely quantum mixed commitments (while we only need quantum 1-1 one-way functions). Both their zero-knowledge property and their extractability hold only against polynomial-time adversaries. In contrast, we get unconditional extractability and computational zero-knowledge; and by adapting our construction to unconditionally hiding commitments, we could instead make the zero-knowledge property unconditional – this would be necessary, e.g., for constructions that achieve everlasting security. Finally, note that the protocols from [14,12] are much more involved than their classical counterparts while we only slightly modify existing classical protocols. Thus, [14,12] give valuable alternatives to our protocols but do not supersede them.

---

[1] The definition of a computationally binding commitment only guarantees that the adversary cannot simultaneously produce opening information for two different values. Thus, to get a contradiction, we need to rewind the adversary to extract two values. If the commitment is strictly binding (Definition 9), our rewinding technique can be used.

Organization. In Section 1.1, we give an overview over the techniques underlying our results. In Section 2 we present and discuss the definition of quantum proofs of knowledge (QPoKs). In Section 3, we give criteria under which a proof system is a QPoK. In Section 4, we show that zero-knowledge QPoKs exist for all languages in NP. Omitted proofs and definitions are presented in the full version [18].

1.1 Our techniques

Defining proofs of knowledge. In the classical setting, proofs of knowledge are defined as follows:[2] A proof system consisting of a prover P and a verifier V is a proof of knowledge (PoK) with knowledge error κ if there is a polynomial-time machine K (the extractor) such that the following holds: For any prover P^*, if P^* convinces V with probability $\mathrm{Pr}_\mathsf{V} \geq \kappa$, then $\mathsf{K}^{\mathsf{P}^*}$ (the extractor K with rewinding black-box access to P^*) outputs a witness with probability $\mathrm{Pr}_\mathsf{K} \geq \frac{1}{p}(\mathrm{Pr}_\mathsf{V} - \kappa)^d$ for some polynomial p and constant $d > 0$. In order to transfer this definition to the quantum setting, we need to specify what it means that K has quantum rewinding black-box access to P^*. We choose the following definition: Let U denote the unitary transformation describing one activation of P^* (if P^* is not unitary, this needs to work for *all* purifications of P^*). K may invoke U (this corresponds to running P^*), he may invoke the inverse $U^\dagger$ of U (this corresponds to rewinding P^* by one activation), and he may read/write a shared register N for exchanging messages with P^*. But K may not make snapshots of the state of P^*. Allowing K to invoke $U^\dagger$ is justified by the fact that all quantum circuits are reversible; given a circuit for U, we can efficiently apply $U^\dagger$. Note that previous black-box constructions such as Watrous' rewinding technique and Grover's algorithm [10] make use of this fact. We can now define quantum proofs of knowledge: (P, V) is a quantum proof of knowledge (QPoK) with knowledge error κ iff there is a polynomial-time quantum algorithm K such that for all malicious provers P^*, $\mathsf{K}^{\mathsf{P}^*}$ (the extractor K with quantum rewinding black-box access to P^*) outputs a witness with probability $\mathrm{Pr}_\mathsf{K} \geq \frac{1}{p}(\mathrm{Pr}_\mathsf{V} - \kappa)^d$ for some polynomial p and constant $d > 0$.

We illustrate that QPoKs according to this definition are indeed useful for analyzing cryptographic protocols. Assume the following toy protocol: In phase 1, a certification authority (CA) signs the pair $(\texttt{Alice}, a)$ where a is Alice's age. In phase 2, Alice uses a zero-knowledge QPoK with negligible knowledge error κ to prove to Bob that she possesses a signature σ on $(\texttt{Alice}, a')$ for some $a' \geq 21$. That is, a witness in this QPoK would consist of an integer $a' \geq 21$ and a signature σ on $(\texttt{Alice}, a')$ with respect to the CA's public key. We can now show that, if Alice is underage, i.e., if $a < 21$, Bob accepts the QPoK only with negligible probability: Assume that Bob accepts with non-negligible probability ν. Then, by the definition of QPoKs, $\mathsf{K}^{\texttt{Alice}}$ will, with probability $\frac{1}{p}(\nu - \kappa)^d$,

---

[2] This is one of different possible definitions, loosely following [11]. It permits us to avoid the use of expected polynomial-time. We discuss alternatives in Section 2.2 "On the success probability of the extractor".

output an integer $a' \geq 21$ and a (forged) signature σ on $(\texttt{Alice}, a')$ with respect to the CA's public key (given the information learned in phase 1 as auxiliary input). Notice that $\frac{1}{p}(\nu - \kappa)^d$ is non-negligible. However, the CA only signed $(\texttt{Alice}, a)$ with $a < 21$. This implies that $\mathsf{K}^{\texttt{Alice}}$ can produce with non-negligible probability a valid signature of a message that has never been signed by the CA. This contradicts the security of the signature scheme (assuming, e.g., existential unforgeability [8]). This shows the security of our toy protocol.

Relation to classical proofs of knowledge. Notice that a quantum proof of knowledge according to our definition is not necessarily a classical PoK because the quantum extractor might have more computational power. (E.g., in a proof system where the witness is a factorization, a quantum extractor could just compute this witness himself.) We stress that this "paradox" is not particular to our definition, it occurs with all simulation-based definitions (e.g., zero-knowledge [19], universal composability [17]). If needed, one can avoid this "paradox" by requiring the extractor/simulator to be classical if the malicious prover/verifier is. (This would actually be equivalent to requiring that the scheme is both a classical ZK PoK and a quantum one.)

Amplification. Our toy example shows that QPoKs with negligible knowledge error can be used to show the security of protocols. But what about QPoKs with non-negligible knowledge error? In the classical case, we know that the knowledge error of a PoK can be made exponentially small by sequential repetition. Fortunately, this result carries over to the quantum case; its proof follows the same lines.

Elementary constructions. In order to understand our constructions of QPoKs, let us first revisit a common method for constructing classical PoKs. Assume a protocol that consists of three messages: the commitment (sent by the prover), the challenge (picked from a set C and sent by the verifier), and the response (sent by prover). Assume that there is an efficient algorithm K_0 that computes a witness given two conversations with the same commitment but different challenges; this property is called special soundness. Then we can construct the following (classical) extractor K: $\mathsf{K}^{\mathsf{P}^*}$ runs P^* using a random challenge ch. Then $\mathsf{K}^{\mathsf{P}^*}$ rewinds P^* to the point after it produced the commitment, and then $\mathsf{K}^{\mathsf{P}^*}$ runs P^* with a random challenge ch'. If both executions lead to an accepting conversation, and $ch \neq ch'$, K_0 can compute a witness. The probability of getting two accepting conversations can be shown to be $\mathrm{Pr}_{\mathsf{V}}^2$, where Pr_{V} is the probability of the verifier accepting P^*'s proof. From this, a simple calculation shows that the knowledge error of the protocol is $1/\#C$.

If we directly translate this approach to the quantum setting, we end up with the following extractor: K runs one step of P^*, measures the commitment com, provides a random challenge ch, runs the second step of P^*, measures the response, runs the inverse of the second step of P^*, provides a random challenge ch', runs the second step of P^*, and measures the response $resp'$. If $ch \neq ch'$, and both $(com, ch, resp)$ and $(com, ch', resp')$ are accepting conversations, then we get a witness using K_0. We call this extractor the canonical extractor. The

problem is to bound the probability F of getting two accepting conversations. In the classical setting, one uses that the two conversations are essentially independent (given a fixed commitment), and each of them is, from the point of view of P^*, the same as an interaction with the honest verifier V. In the quantum setting, this is not the case. Measuring *resp* disturbs the state of P^*; we hence cannot make any statement about the probability that the second conversation is accepting.

How can we solve this problem? Note that we cannot use Watrous' oblivious rewinding since we need to remember both responses *resp* and *resp′* from two different execution paths of P^*. Instead, we observe that, the more information we measure in the first conversation (i.e., the longer *resp* is), the more we destroy the state of P^* used in the second conversation. Conversely, if would measure only one bit, the disturbance of P^*'s state would be small enough to still get a sufficiently high success probability. But if *resp* would contain only one bit, it would clearly be too short to be of any use for K_0. Yet, it turns out that this conflict can be resolved: In order not to disturb P^*'s state, we only need that the *resp* information-theoretically contains little information. For K_0, however, even an information-theoretically determined *resp* is still useful; it might, for example, reveal a value which P^* was already committed to. To make use of this observation, we introduce an additional condition on our proof systems, strict soundness. A proof system has strict soundness if for any commitment and challenge, there is at most one response that makes the conversation accepting. Given a proof system with special and strict soundness, we can show that measuring *resp* does not disturb P^*'s state too much; the canonical extractor is successful with probability approximately $\Pr_{\mathsf{V}}^3$. A precise calculation shows that a proof system with special and strict soundness has knowledge error $1/\sqrt{\#C}$.

QPoKs for all languages in NP. Blum [4] presents a classical zero-knowledge PoK for showing the knowledge of a Hamiltonian cycle. Using a suitable commitment scheme (it should have the property that the opening information is uniquely determined by the commitment), the proof system is easily seen to have special and strict soundness, thus it is a QPoK. By sequential repetition, we get a QPoK for Hamiltonian cycles. Using the Watrous' results, we get that the QPoK is also zero-knowledge. Using the fact that the Hamiltonian cycle problem is NP-complete, we get zero-knowledge QPoKs for all languages in NP (assuming quantum 1-1 one-way functions).

1.2 Preliminaries

General. A non-negative function μ is called negligible if for all $c > 0$ and all sufficiently large k, $\mu(k) < k^{-c}$. $\oplus$ denotes the XOR operation on bitstrings. $\#C$ is the cardinality of the set C.

Quantum systems. We can only give a terse overview over the formalism used in quantum computing. For a thorough introduction, we recommend the textbook by Nielsen and Chuang [15, Chap. 1–2]. A (pure) state in a quantum system

is described by a unit vector $|\Phi\rangle$ in some Hilbert space $\mathcal{H}$. We always assume a designated orthonormal basis for each Hilbert space, called the computational basis. The tensor product of several states (describing a joint system) is written $|\Phi\rangle \otimes |\Psi\rangle$. We write $\langle\Psi|$ for the linear transformation mapping $|\Phi\rangle$ to the scalar product $\langle\Psi|\Phi\rangle$. The norm $\||\Phi\rangle\|$ is defined as $\sqrt{\langle\Phi|\Phi\rangle}$. A unit vector is a vector with $\||\Phi\rangle\| = 1$. The Hermitean transpose of a linear operator A is written $A^\dagger$.

2 Quantum Proofs of Knowledge

2.1 Definitions

Interactive machines. A *quantum interactive machine* M (machine, for short) is a machine that gets two inputs, a classical input x and a quantum input $|\Phi\rangle$. M operates on two quantum registers; a network register N and a register S_M for the state. S_M is initialized with $|\Phi\rangle$. The operation of M is described by a unitary transformation M_x (depending on the classical input x). In each activation of M, M_x is applied to N, S_M. We write $\langle\mathsf{M}(x, |\Phi\rangle), \mathsf{M}'(x', |\Phi'\rangle)\rangle$ for the classical output of M' in an interaction where M is activated first (and where M and M' share the register N). Often, we will omit the quantum input $|\Phi\rangle$ or $|\Phi'\rangle$. In this case, we assume the input $|0\rangle$.

Oracles algorithms with rewinding. A *quantum oracle algorithm* A is an algorithm that has oracle access to a machine M. In an execution $\mathsf{A}^{\mathsf{M}(x', |\Phi\rangle)}(x)$, two registers N, S_M are used for the communication with and the state of M. A's behavior is described by a quantum circuit; A has access to two special gates $\square$ and $\square^\dagger$ that invoke the unitary transformations $\mathsf{M}_{x'}$ and $\mathsf{M}^\dagger_{x'}$, respectively. This corresponds to running and rewinding M. A is not allowed to access S_M directly, and he is allowed to apply $\square$ and $\square^\dagger$ only to N, S_M. (I.e., A has no access to the internal state and the quantum input of the prover. Any access to this information is done by communicating with M.) Details on the definitions of interactive quantum machines and quantum oracle algorithms are given in the full version [18].

Proof systems. A *quantum proof system* for a relation R is a pair of two machines (P, V). We call P the prover and V the verifier. The prover expects a classical input (x, w) with $(x, w) \in R$, the verifier expects only the input x. We call (P, V) *complete* if there is a negligible function μ such that for all $(x, w) \in R$, we have that $\Pr[\langle\mathsf{P}(x, w), \mathsf{V}(x)\rangle = 1] \geq 1 - \mu(|x|)$. (Remember that, if we do not explicitly specify a quantum input, we assume the quantum input $|0\rangle$.) Although we allow P and V to be quantum machines, and in particular to send and receive quantum messages, we will not need this property in the following; all protocols constructed in this paper will consist of classical machines. We call a (P, V) *sound* with soundness error s iff for all malicious prover P^*, all auxiliary inputs $|\Phi\rangle$, and all x with $\nexists w : (x, w) \in R$, we have $\Pr[\langle\mathsf{P}^*(x, |\Phi\rangle), \mathsf{V}(x)\rangle = 1] \leq s(|x|)$. A proof system is computational zero-knowledge iff for all polynomial-time verifiers V^* there is a polynomial-time machine S (the simulator) such that for all auxiliary

inputs $|\Phi\rangle$, and all $(x, w) \in R$, we have that the quantum state of V^* after an interaction $\langle \mathsf{P}(x, w), \mathsf{V}^*(x, |\Phi\rangle)\rangle$ is computationally indistinguishable from the output of $\mathsf{S}(x, |\Phi\rangle)$; we refer to [19] for details.

Quantum Proofs of Knowledge. We can now define quantum proofs of knowledge (QPoKs). Roughly, a quantum proof system (P, V) is a QPoK if there is a quantum oracle algorithm K (the extractor) that achieves the following: Whenever some malicious prover P^* convinces V that a certain statement holds, the extractor $\mathsf{K}^{\mathsf{P}^*}$ with oracle access to P^* is able to return a witness. Here, we allow a certain knowledge error κ; if P^* convinces V with a probability smaller than κ, we do not require anything. Furthermore, we also do not require that the success probability of $\mathsf{K}^{\mathsf{P}^*}$ is as high as the success probability of P^*; instead, we only require that it is polynomially related. Finally, to facilitate the use of QPoKs as subprotocols, we give the malicious prover an auxiliary input $|\Phi\rangle$. We get the following definition:

Definition 1 (Quantum Proofs of Knowledge). *We call a proof system* (P, V) *for a relation R quantum extractable with knowledge error κ if there exists a constant $d > 0$, a polynomially-bounded function $p > 0$, and a polynomial-time quantum oracle machine K such that for any interactive quantum machine P^*, any state $|\psi\rangle$, and any $x \in \{0, 1\}^*$, we have that*

$$\Pr[\langle \mathsf{P}^*(x, |\psi\rangle), \mathsf{V}(x)\rangle = 1] \geq \kappa(|x|) \implies$$

$$\Pr[(x, w) \in R : w \leftarrow \mathsf{K}^{\mathsf{P}^*(x, |\psi\rangle)}(x)] \geq \frac{1}{p(|x|)}\Big(\Pr\big[\langle \mathsf{P}^*(x, |\psi\rangle), \mathsf{V}(x)\rangle = 1\big] - \kappa(|x|)\Big)^d.$$

A quantum proof of knowledge *for R with knowledge error κ (QPoK, for short) is a complete[3] quantum extractable proof system for R with knowledge error κ.*

Note that by quantifying over all unitary provers P^*, we implicitly quantify over *all* purifications of *all* possible non-unitary provers. Note that extractability with knowledge error κ implies soundness with soundness error κ. We thus do not need to explicitly require soundness in Definition 1. The knowledge error κ can be made exponentially small by sequential repetition:

Theorem 2. *Let n be a polynomially bounded and efficiently computable function. Let (P, V) be extractable with knowledge error κ. Let $(\mathsf{P}', \mathsf{V}')$ be the proof system consisting of n-sequential executions of (P, V). Then $(\mathsf{P}', \mathsf{V}')$ is extractable with knowledge error κ^n.*

2.2 Discussion

In this section, we motivate various design choices made in the definition of QPoKs.

Access to the black-box prover's state and input. The extractor has no access to the prover's state nor to its quantum input. (This is modeled by the fact

---

[3] I.e., for honest prover and verifier, the proof succeeds with overwhelming probability.

that an oracle algorithm may not apply any gates except for $\Box, \Box^\dagger$ to the register containing the oracle's state and quantum input.) In this, we follow [3] who argue in Section 4.3 that a proof of knowledge is supposed to "capture the knowledge of the prover *demonstrated by the interaction*" and that thus the extractor is not supposed to see the internal state of the prover. We stress, however, that our results are independent of this issue; they also hold if we allow the extractor to access the prover's state directly.

Unitary & invertible provers – technical view. Probably the most important design choice in our definition is to require the prover to be a unitary operation, and to allow the extractor to also execute the inverse of this operation. We begin with a discussion of this design choice from a technical point of view. First, we stress that seems that these assumptions are necessary: Since in a quantum world, making a snapshot/copy of a state is not possible or even well-defined, we have to allow the extractor to run the prover "backwards". But the inverse of a non-unitary quantum operation does not, in general, exist. Thus rewinding seems only possible with respect to unitary provers. Second, the probably most important question is: Does the definition, from an operational point of view, make sense? That is, does our definition behave well in cryptographic, reduction-based proofs? A final answer to this question can only be given when more protocols using QPoKs have been analyzed. However, the toy protocol discussed on page 4 gives a first indication that our definition can be used in a similar fashion to classical proofs of knowledge. Third, we would like to remind the reader that any non-unitary prover can be transformed into a unitary one by purification before applying the definition of QPoKs. Thus allowing only unitary malicious provers does not seem to be a restriction in practice.

Unitary & invertible provers – philosophical view. Intuitively, a QPoK should guarantee that a prover that convinces the verifier "knows" the witness.[4] The basic idea is that if an extractor can extract the witness using only what is available to the prover, then the prover "knew" the witness (or could have computed it). In particular, we may allow the extractor to run a purified (unitary) version of the prover because the prover himself could have done so. Similarly for the inverse of that operation. Of course, this leaves the question why we give these two capabilities to the extractor but not others (e.g., access to the circuit of the prover)? We would like to stress that analogous questions are still open (from a philosophical point) even in the classical case: Why is it natural to allow an extractor to rewind the prover? Why is it natural to give a trapdoor for a common reference string to the extractor? We would like to point out one justification for the assumption that the prover is unitary, though: [3] suggests that we "capture the knowledge of the prover *demonstrated by the interaction*". A prover that performs non-unitary operations is identical in terms of its interaction to one that is purified. Thus, by restricting to unitary provers, we come closer to only capturing the interaction but not the inner workings of the prover.

---

[4] We believe, though, that this issue is secondary to the technical suitability; it is much more important that a QPoK is useful as a cryptographic subprotocol.

On the success probability of the extractor. We require the extractor to run in polynomial-time and to succeed with probability $\frac{1}{p}(\mathrm{Pr_V} - \kappa)^d$ where $\mathrm{Pr_V}$ is the probability that the prover convinces the verifier. (We call this an A-style definition.) In classical PoKs, a more common definition is to require the extractor to have expected runtime $\frac{p}{\mathrm{Pr_V} - \kappa}$ and to succeed with probability 1. (We call this a B-style definition.) This definition is known to be equivalent to the definition in which the extractor runs in expected polynomial-time and succeeds with probability $\frac{1}{p}(\mathrm{Pr_V} - \kappa)$. (We call this a C-style definition.) The advantage of an A-style definition (which follows [11]) is that we can consider polynomial-time extractors (instead of expected polynomial-time extractors). To get extractors for B-style and C-style definitions, one has to increase the success probability of an extractor by repeatedly invoking it until it outputs a correct witness. In the quantum case, however, this does not work directly: If the invoked extractor fails once, the auxiliary input of the prover is destroyed. The oblivious rewinding technique by Watrous' would seem to help here, but when trying to apply that technique one gets the requirement that the invoked extractors' success probability must be independent of the auxiliary input. This condition is not necessarily fulfilled. To summarize, all three styles of definitions have their advantages, but it is not clear how one could fulfil B- and C-style definitions in the quantum case. This is why we chose an A-style definition. There are, however, applications that would benefit from a proof system fulfilling a C-style definition. For example, general multi-party computation protocols such as [5] use extractors as part of the construction of the simulator for the multi-party computation; these extractors must then succeed with probability close to 1. We leave the construction of C-style QPoKs as an open problem.

3   Elementary constructions

In this section, we show that under certain conditions, a classical PoK is also a QPoK (i.e., secure against malicious quantum provers). The first condition refers to the outer form of the protocol; we require that the proof systems is a protocol with three messages (commitment, challenge, and response) with a public-coin verifier. Such protocols are called Σ-protocols. Furthermore, we require that the proof system has special soundness. This means that given two accepting conversations between prover and verifier that have the same commitment but different challenges, we can efficiently compute a witness. Σ-protocols with special soundness are well-studied in the classical case; many efficient classical protocols with these properties exist. The third condition (strict soundness) is non-standard. We require that given the commitment and the challenge of a conversation, there is at most one response that would make the verifier accept. We require strict soundness to ensure that the response given by the prover does not contain too much information; measuring it will then not disturb the state of the prover too much. Not all known protocols have strict soundness (the proof for graph isomorphism [6] is an example). Fortunately, many protocols do satisfy

strict soundness; a slight variation of the proof for Hamiltonian cycles [4] is an example (see Section 4).

Definition 3 (Σ-protocol). *A proof system* (P, V) *is called a Σ-protocol if P and V are classical, the interaction consists of three messages $com, ch, resp$ (sent by P, V, and P, respectively, and called* commitment, challenge, *and* response*), and ch is uniformly chosen from some set C_x (the* challenge space*) that may only depend on the statement x. Furthermore, the verifier decides whether to accept or not by a deterministic polynomial-time computation on $x, com, ch, resp$. (We call $(com, ch, resp)$ an* accepting conversation *for x if the verifier would accept it.) We also require that it is possible in polynomial time to sample uniformly from C_x, and that membership in C_x should be decidable in polynomial time.*

Definition 4 (Special soundness). *We say a Σ-protocol (P, V) for a relation R has* special soundness *if there is a deterministic polynomial-time algorithm K_0 (the* special extractor*) such that the following holds: For any two accepting conversations $(com, ch, resp)$ and $(com, ch', resp')$ for x such that $ch \neq ch'$ and $ch, ch' \in C_x$, we have that $w := \mathsf{K}_0(x, com, ch, resp, ch', resp')$ satisfies $(x, w) \in R$.*

Definition 5 (Strict soundness). *We say a Σ-protocol (P, V) has* strict soundness *if for any two accepting conversations $(com, ch, resp)$ and $(com, ch, resp')$ for x, we have that $resp = resp'$.*

Canonical extractor. Let (P, V) be a Σ-protocol with special soundness and strict soundness. Let K_0 be the special extractor for that protocol. We define the *canonical extractor* K for (P, V). K will use measurements, even though our definition of quantum oracle algorithms only allows for unitary operations. This is only for the sake of presentation; by purifying K one can derive a unitary algorithm with the same properties. Given a malicious prover P^*, $\mathsf{K}^{\mathsf{P}^*(x, |\Phi\rangle)}(x)$ operates on two quantum registers N, S_{P^*}. N is used for communication with P^*, and S_{P^*} is used for the state of P^*. The registers N, S_{P^*} are initialized with $|0\rangle, |\Phi\rangle$. Let P_x^* denote the unitary transformation describing a single activation of P. First, K applies P_x^* to N, S_{P^*}. (This can be done using the special gate $\square$.) This corresponds to running the first step of P^*; in particular, N should now contain the commitment. Then K measures N in the computational basis; call the result com. Then K initializes N with $|0\rangle$. Then K chooses uniformly random values $ch, ch' \in C_x$. Let U_{ch} denote the unitary transformation operating on N such that $U_{ch}|x\rangle = |x \oplus ch\rangle$. Then K applies $\mathsf{P}_x^* U_{ch}$. (Now N is expected to contain the response for challenge ch.) Then K measures N in the computational basis; call the result $resp$. Then K applies $(\mathsf{P}_x^* U_{ch})^\dagger$ (we rewind the prover). Then $\mathsf{P}_x^* U_{ch'}$ is applied. (Now N is expected to contain the response for challenge ch'.) Then N is measured in the computational basis; call the result $resp'$. Then $(\mathsf{P}_x^* U_{ch'})^\dagger$ is applied. Finally, K outputs $w := \mathsf{K}_0(x, com, ch, resp, ch', resp')$.

Analysis of the canonical extractor. In order to analyze the canonical extractor (Theorem 8 below), we first need a lemma that bounds the probability that two consecutive binary measurements P_{ch} and $P_{ch'}$ with random $ch \neq ch'$

succeed in terms of the probability that a single such measurement succeeds. In a classical setting (or in the case of commuting measurements), the answer is simple: the outcomes of the measurements are independent; thus the probability that two measurements succeed is the square of the probability that a single measurement succeeds. In the quantum case, however, the first measurement may disturb the state; this makes the analysis considerably more involved. We first prove some inequalities needed in the proof:

Lemma 6. *Let C be a set with $\#C = c$. Let $(P_i)_{i \in C}$ be orthogonal projectors on a Hilbert space $\mathcal{H}$. Let $|\Phi\rangle \in \mathcal{H}$ be a unit vector. Let $V := \sum_{i \in C} \frac{1}{c} \||P_i|\Phi\rangle\|^2$ and $F := \sum_{i,j \in C} \frac{1}{c^2} \||P_i P_j |\Phi\rangle\|^2$. Then $F \geq V^3$.*

Proof. To prove the lemma, we first show two simple facts:

Claim. For any positive operator A on $\mathcal{H}$ and any unit vector $|\Phi\rangle \in \mathcal{H}$, we have that $(\langle \Phi | A | \Phi \rangle)^3 \leq \langle \Phi | A^3 | \Phi \rangle$.

Since A is positive, it is diagonalizable. Thus we can assume without loss of generality that A is diagonal (by applying a suitable basis transform to A and $|\Phi\rangle$). Let a_i be the i-th diagonal element of A, and let f_i be the i-th component of $|\Phi\rangle$. Then

$$(\langle \Phi | A | \Phi \rangle)^3 = \left(\sum_i |f_i|^2 a_i \right)^3 \overset{(*)}{\leq} \sum_i |f_i|^2 a_i^3 = \langle \Phi | A^3 | \Phi \rangle.$$

Here $(*)$ uses Jensen's inequality [13] and the facts that $a_i \geq 0$, that $a_i \mapsto a_i^3$ is a convex function on nonnegative numbers, and that $\sum_i |f_i|^2 = 1$. This concludes the proof of Lemma 3.

Claim. For vectors $|\Psi_1\rangle, \ldots, |\Psi_c\rangle \in \mathcal{H}$, it holds that $\|\frac{1}{c} \sum_i |\Psi_i\rangle\|^2 \leq \frac{1}{c} \sum_i \||\Psi_i\rangle\|^2$.

To show the claim, let $|\bar{\Psi}\rangle := \sum_i \frac{1}{c} |\Psi_i\rangle$. Then

$$\sum_i \left(\||\Psi_i\rangle\|^2 - \||\bar{\Psi}\rangle\|^2 \right) = \sum_i \left(\||\Psi_i\rangle\| - \||\bar{\Psi}\rangle\| \right) \left(\||\Psi_i\rangle\| - \||\bar{\Psi}\rangle\| + 2\||\bar{\Psi}\rangle\| \right)$$

$$= \sum_i \left(\||\Psi_i\rangle\| - \||\bar{\Psi}\rangle\| \right)^2 + 2\||\bar{\Psi}\rangle\| \sum_i \left(\||\Psi_i\rangle\| - \||\bar{\Psi}\rangle\| \right)$$

$$\geq 2\||\bar{\Psi}\rangle\| \sum_i \left(\||\Psi_i\rangle\| - \||\bar{\Psi}\rangle\| \right) = 2\||\bar{\Psi}\rangle\| \left(\sum_i \||\Psi_i\rangle\| - \|n|\bar{\Psi}\rangle\| \right) \qquad (1)$$

$$= 2\||\bar{\Psi}\rangle\| \left(\sum_i \||\Psi_i\rangle\| - \left\| \sum_i |\Psi_i\rangle \right\| \right) \qquad (2)$$

From the triangle inequality, it follows that $\sum_i \||\Psi_i\rangle\| \geq \|\sum_i |\Psi_i\rangle\|$, hence with (2), we have $\sum_i \left(\||\Psi_i\rangle\|^2 - \||\bar{\Psi}\rangle\|^2 \right) \geq 0$. Since $\frac{1}{c} \sum_i \||\Psi_i\rangle\|^2 - \|\frac{1}{c} \sum_i |\Psi_i\rangle\|^2 = \frac{1}{c} \sum_i \left(\||\Psi_i\rangle\|^2 - \||\bar{\Psi}\rangle\|^2 \right) \geq 0$, Lemma 3 follows.

We proceed to prove Lemma 6. Let $A := \sum_i \frac{1}{c} P_i$, let $|\Psi_{ij}\rangle := P_j P_i |\Phi\rangle$. Then A is positive. Furthermore,

$$
V^3 = \left(\sum_i \frac{1}{c} \langle \Phi | P_i | \Phi \rangle \right)^3 = \left(\langle \Phi | A | \Phi \rangle \right)^3 \overset{(*)}{\leq} \langle \Phi | A^3 | \Phi \rangle = \sum_{i,j,k} \frac{1}{c^3} \langle \Phi | P_i P_j P_k | \Phi \rangle
$$

$$
= \sum_{i,j,k} \frac{1}{c^3} \langle \Psi_{ij} | \Psi_{kj} \rangle = \sum_j \frac{1}{c} \left(\sum_i \frac{1}{c} \langle \Psi_{ij} | \right) \left(\sum_k \frac{1}{c} | \Psi_{kj} \rangle \right) = \sum_j \frac{1}{c} \left\| \sum_i \frac{1}{c} | \Psi_{ij} \rangle \right\|^2
$$

$$
\overset{(**)}{\leq} \sum_j \frac{1}{c} \sum_i \frac{1}{c} \| |\Psi_{ij}\rangle \|^2 = F.
$$

Here $(*)$ uses Lemma 3 and $(**)$ uses Lemma 3. Thus we have $F \geq V^3$ and Lemma 6 follows.

Lemma 7. *Let C be a set with $\#C = c$. Let $(P_i)_{i \in C}$ be orthogonal projectors on a Hilbert space $\mathcal{H}$. Let $|\Phi\rangle \in \mathcal{H}$ be a unit vector. Let $V := \sum_{i \in C} \frac{1}{c} \| P_i |\Phi\rangle \|^2$ and $E := \sum_{i,j \in C, i \neq j} \frac{1}{c^2} \| P_i P_j |\Phi\rangle \|^2$. Then, if $V \geq \frac{1}{\sqrt{c}}$, $E \geq V(V^2 - \frac{1}{c})$.*

Proof. Let F be as in Lemma 6. Then

$$
E = \sum_{\substack{i,j \in C \\ i \neq j}} \frac{1}{c^2} \| P_i P_j |\Phi\rangle \|^2 = \sum_{i,j \in C} \frac{1}{c^2} \| P_i P_j |\Phi\rangle \|^2 - \sum_{i \in C} \frac{1}{c^2} \| P_i P_i |\Phi\rangle \|^2
$$

$$
\overset{(*)}{=} \sum_{i,j \in C} \frac{1}{c^2} \| P_i P_j |\Phi\rangle \|^2 - \sum_{i \in C} \frac{1}{c^2} \| P_i |\Phi\rangle \|^2 = F - \frac{V}{c} \overset{(**)}{\geq} V^3 - \frac{V}{c} = V(V^2 - \frac{1}{c})
$$

Here $(*)$ uses that $P_i = P_i P_i$ since P_i is a projection, and $(**)$ uses Lemma 6. $\square$

Theorem 8. *A Σ-protocol (P, V) for a relation R with special and strict soundness and challenge space C_x is extractable with knowledge error $\frac{1}{\sqrt{\#C_x}}$.*

Proof. To show that (P, V) is extractable, we will use the canonical extractor K. Fix a malicious prover P^*, a statement x, and an auxiliary input $|\Phi\rangle$. Let Pr_V denote the probability that the verifier accepts when interacting with P^*. Let Pr_K denote the probability that $\mathsf{K}^{\mathsf{P}^*(x,|\Phi\rangle)}(x)$ outputs some w with $(x,w) \in R$. We will show that $\mathrm{Pr}_\mathsf{K} \geq \mathrm{Pr}_\mathsf{V} \cdot (\mathrm{Pr}_\mathsf{V}^2 - \frac{1}{\#C_x})$. For $\mathrm{Pr}_\mathsf{V} \geq \frac{1}{\sqrt{\#C_x}}$, we have that $\mathrm{Pr}_\mathsf{V}(\mathrm{Pr}_\mathsf{V}^2 - \frac{1}{\#C_x}) \geq (\mathrm{Pr}_\mathsf{V} - \frac{1}{\sqrt{\#C_x}})^3$. Since furthermore K is polynomial-time, this implies that (P, V) is extractable with knowledge error $\frac{1}{\sqrt{\#C_x}}$.

In order to show $\mathrm{Pr}_\mathsf{K} \geq \mathrm{Pr}_\mathsf{V} \cdot (\mathrm{Pr}_\mathsf{V}^2 - \frac{1}{\#C_x})$, we will use a short sequence of games. Each game will contain an event Succ, and in the first game, we will have $\Pr[\mathsf{Succ} : \text{Game } 1] = \mathrm{Pr}_\mathsf{K}$. For any two consecutive games, we will have $\Pr[\mathsf{Succ} : \text{Game } i] \geq \Pr[\mathsf{Succ} : \text{Game } i + 1]$, and for the final game, we will have $\Pr[\mathsf{Succ} : \text{Game } 7] \geq \mathrm{Pr}_\mathsf{V} \cdot (\mathrm{Pr}_\mathsf{V}^2 - \frac{1}{\#C_x})$. This will then conclude the proof. The description of each game will only contain the changes with respect to the preceding game.

Game 1. An execution of $\mathsf{K}^{\mathsf{P}^*(x,|\Phi\rangle)}(x)$. Succ denotes the event that K outputs a witness for x. By definition, $\Pr_\mathsf{K} = \Pr[\mathsf{Succ} : \text{Game 1}]$.

Game 2. Succ denotes the event that $(com, ch, resp)$ and $(com, ch', resp')$ are accepting conversations for x and $ch \neq ch'$. (The variables $(com, ch, resp)$ and $(com, ch', resp')$ are as in the definition of the canonical extractor.) Since (P, V) has special soundness, if Succ occurs, K outputs a witness. Thus $\Pr[\mathsf{Succ} :$ Game 1$] \geq \Pr[\mathsf{Succ} :$ Game 2$]$.

Game 3. Before K measures $resp$, it first measures whether measuring $resp$ would yield an accepting conversation. More precisely, it measures N with the orthogonal projector P_{ch} projecting onto $V_{ch} := \text{span}\{|resp\rangle :$ $(com, ch, resp)$ is accepting$\}$. Analogously for the measurement of $resp'$ (using the projector $P_{ch'}$.) Since a complete measurement (of $resp$ and $resp'$, respectively) is performed on N after applying the measurement P_{ch} and $P_{ch'}$, introducing the additional measurements does not change the outcomes $resp$ and $resp'$ of these complete measurements, nor their post-measurement state. Thus $\Pr[\mathsf{Succ} :$ Game 2$] = \Pr[\mathsf{Succ} :$ Game 3$]$.

Game 4. Succ denotes the event that $ch \neq ch'$ and both measurements P_{ch} and $P_{ch'}$ succeed. By definition of these measurements, this happens iff $(com, ch, resp)$ and $(com, ch', resp')$ are accepting conversations. Thus $\Pr[\mathsf{Succ} :$ Game 3$] = \Pr[\mathsf{Succ} :$ Game 4$]$.

Game 5. We do not execute K_0, i.e., we stop after applying $(\mathsf{P}_x^* U_{ch'})^\dagger$. Since at that point, Succ has already been determined, $\Pr[\mathsf{Succ} :$ Game 4$] = \Pr[\mathsf{Succ} :$ Game 5$]$.

Game 6. We remove the measurements of $resp$ and $resp'$. Note that the outcomes of these measurements are not used any more. Since (P, V) has strict soundness, $V_{ch} = \text{span}\{|resp_0\rangle\}$ for a single value $resp_0$ (depending on com and ch, of course). Thus if the measurement P_{ch} succeeds, the post-measurement state in N is $|resp_0\rangle$. That is, the state in N is classical at this point. Thus, measuring N in the computational basis does not change the state. Hence, the measurement of $resp$ does not change the state. Analogously for the measurement of $resp'$. It follows that $\Pr[\mathsf{Succ} :$ Game 5$] = \Pr[\mathsf{Succ} :$ Game 6$]$.

Game 7. First, N and S_{P^*} are initialized with $|0\rangle$ and $|\Phi\rangle$. Then the unitary transformation P_x^* is applied. Then com is measured (complete measurement on N), and N is initialized to $|0\rangle$. Random $ch, ch' \in C_x$ are chosen. Then $\mathsf{P}_x^* U_{ch}$ is applied. Then the measurement P_{ch} is performed. Then $(\mathsf{P}_x^* U_{ch})^\dagger$ is applied. Then $\mathsf{P}_x^* U_{ch'}$ is applied. Then the measurement $P_{ch'}$ is performed. Then $(\mathsf{P}_x^* U_{ch'})^\dagger$ is applied. The event Succ holds if $ch \neq ch'$ and both measurements succeed. Games 6 and 7 are identical; we have just recapitulated the game for clarity. Thus, $\Pr[\mathsf{Succ} :$ Game 6$] = \Pr[\mathsf{Succ} :$ Game 7$]$.

In Game 7, for some value d, let p_d denote the probability that $com = d$ is measured. Let $|\Phi_d\rangle$ denote the state of N, S_{P^*} after measuring $com = d$ and initializing N with $|0\rangle$. (I.e., the state directly before applying $\mathsf{P}_x^* U_{ch}$.) Let K_d denote the probability that starting from state $|\Phi_d\rangle$, both measurements P_{ch} and

$P_{ch'}$ succeed. Let $c := \#C_x$. Then we have that $\Pr[\mathsf{Succ} : \mathsf{Game}\ 7] = \sum_d p_d K_d$
and

$$K_d = \sum_{\substack{ch, ch' \in C_x \\ ch \neq ch'}} \frac{1}{c^2} \|(\mathsf{P}_x^* U_{ch'})^\dagger P_{ch'} (\mathsf{P}_x^* U_{ch'})(\mathsf{P}_x^* U_{ch})^\dagger P_{ch}(\mathsf{P}_x^* U_{ch})|\varPhi_d\rangle\|^2$$
$$= \sum_{\substack{ch, ch' \in C_x \\ ch \neq ch'}} \frac{1}{c^2} \|P_{ch'}^* P_{ch}^* |\varPhi_d\rangle\|^2$$

where $P_{ch}^* := (\mathsf{P}_x^* U_{ch})^\dagger P_{ch}(\mathsf{P}_x^* U_{ch})$. Since P_{ch} is an orthogonal projector and $\mathsf{P}_x^* U_{ch}$ is unitary, P_{ch}^* is an orthogonal projector. Let $\varphi(v) := v(v^2 - \frac{1}{c})$ for $v \in [\frac{1}{\sqrt{c}}, 1]$ and $\varphi(v) := 0$ for $v \in [0, \frac{1}{\sqrt{c}}]$. Then, by Lemma 7, $K_d \geq \varphi(V_d)$ for $V_d := \sum_{ch \in C_x} \frac{1}{c} \|P_{ch}^*|\varPhi_d\rangle\|^2$.

Furthermore, by construction of the honest verifier V, we have that

$$\Pr_\mathsf{V} = \sum_d p_d \sum_{ch \in C_x} \tfrac{1}{c}\|P_{ch}\mathsf{P}_x^* U_{ch}|\varPhi_d\rangle\|^2$$
$$\overset{(*)}{=} \sum_d p_d \sum_{ch \in C_x} \tfrac{1}{c}\|(\mathsf{P}_x^* U_{ch})^\dagger P_{ch}(\mathsf{P}_x^* U_{ch})|\varPhi_d\rangle\|^2 = \sum_d p_d V_d$$

where $(*)$ uses that $(\mathsf{P}_x^* U_{ch})^\dagger$ is unitary. Finally, we have

$$\Pr_\mathsf{K} = \Pr[\mathsf{Succ} : \mathsf{Game}\ 1] \geq \Pr[\mathsf{Succ} : \mathsf{Game}\ 7]$$
$$= \sum_d p_d K_d \geq \sum_d p_d \varphi(V_d) \overset{(*)}{\geq} \varphi(\Pr_\mathsf{V}).$$

Here $(*)$ uses Jensen's inequality [13] and the fact that φ is convex on $[0, 1]$. As discussed in the beginning of the proof, $\Pr_\mathsf{K} \geq \varphi(\Pr_\mathsf{V}) = \Pr_\mathsf{V} \cdot (\Pr_\mathsf{V}^2 - \frac{1}{c})$ for $\Pr_\mathsf{V} \geq \frac{1}{\sqrt{c}}$ implies that (P, V) is a QPoK with knowledge error $1/\sqrt{\#C_x}$.

4 QPoKs for all languages in NP

In the preceding section, we have seen that complete proof systems with strict and special soundness are QPoKs. The question that remains to be asked is: do such proof systems, with the additional property of being zero-knowledge, exist for interesting languages? In this section, we will show that for any language in NP (more precisely, for any NP-relation), there is a zero-knowledge QPoK. (Assuming the existence of quantum 1-1 one-way functions.) Here and in the following, by zero-knowledge we mean quantum computational zero-knowledge.

The starting point for our construction will be the Blum's zero-knowledge PoK for Hamiltonian cycles [4]. In this Σ-protocol, the prover's commits to the vertices of a graph using a perfectly binding commitment scheme. In the prover's response, some of these commitments are opened. That is, the response contains the opening information for some of the commitments. The problem is that standard definitions of commitment schemes do not guarantee that the opening

information is unique; only the actual content of the commitment has to be determined by the commitment. This means that the prover's response is not unique. Thus, with a standard commitment scheme we do not get strict soundness. Instead we need a commitment scheme such that the sender of the commitment scheme is committed not only to the actual content of the commitment, but also to the opening information.

Definition 9 (Strict binding). *A commitment scheme* COM *is a deterministic polynomial-time function taking two arguments a, y, the opening information a and the message y. We say* COM *is* strictly binding *if for all a, y, a', y' with $(a, y) \neq (a', y')$, we have that* $\mathrm{COM}(a, y) \neq \mathrm{COM}(a', y')$.

Furthermore, in order to get the zero-knowledge property, we will need that our commitment schemes are quantum computationally concealing. We refer to [19] for a precise definition of this property. In [2], an unconditionally binding, quantum computationally concealing commitment scheme based on quantum 1-1 one-way function is presented.[5] Unfortunately, to the best of our knowledge, no candidates for quantum 1-1 functions are known. Their definitions differ somewhat from those of [19], but as mentioned in [19], their proof carries over to the definitions from [19]. Furthermore, in the scheme from [2], the commitment contains the image of the opening information under a quantum 1-1 one-way function. Thus the strict binding property is trivially fulfilled. Thus strictly binding, quantum computationally concealing commitment schemes exist under the assumption that quantum 1-1 one-way functions exist.

Given such a commitment scheme COM, we can construct the proof system (P, V). This proof system differs from the original proof system for Hamiltonian cycles [4] only in the following aspect: The prover does not only commit to the vertices in the graph $\pi(x)$, but also to the permutation π and the cycle H. Without these additional commitments, we would not get strict soundness; there might be several permutations leading to the same graph, or the graph might contain several Hamiltonian cycles. The full description of the protocol is given in Figure 1.

Theorem 10. *Let $(x, w) \in R$ iff w is a Hamiltonian cycle of the graph x. Assume that* COM *is a strictly binding, quantum computationally concealing commitment scheme. Then the proof system (P, V) is a zero-knowledge QPoK for R with knowledge error $\frac{1}{\sqrt{2}}$.*

The zero-knowledge property is proven using the techniques from [19]. Extractability is shown by proving special and strict soundness. The strict soundness follows from the fact that the prover is committed to all the information sent in his response using a strictly binding commitment.

---

[5] In [2], the result is stated for quantum one-way permutations $f : \{0,1\}^n \to \{0,1\}^n$. (To the best of our knowledge, no candidates for quantum one-way permutations are known.) Inspection of their proof reveals, however, that the result also holds for families of quantum 1-1 one-way functions $f_i : \{0,1\}^n \to D$ for arbitrary domain D and efficiently samplable indices i, assuming that given an index i, it can be efficiently verified that f_i is injective.

> **Inputs:** A directed graph x (the statement) with vertices W, and a Hamiltonian cycle w in x (the witness).
> **Protocol:**
> 1. P picks a random permutation π on W. Let A be the adjacency matrix of the graph $\pi(x)$. Let $H := \{(\pi(i), \pi(j)) : (i,j) \in w\}$. Using COM, P commits to π, H, and to each entry A_{ij} of A. P sends the resulting commitments to V.
> 2. V picks $ch \in \{0,1\}$ and sends ch to P.
> 3. If $ch = 0$, P opens the commitments to π and A. If $ch = 1$, P opens the commitments to H and to all A_{ij} with $(i,j) \in H$.
> 4. If $ch = 0$, V checks that the commitments are opened correctly, that π is a permutation, and that A is the adjacency matrix of $\pi(x)$. If $ch = 1$, V checks that the commitments are opened correctly, that H is a cycle, that exactly the A_{ij} with $(i,j) \in H$ are opened, and that $A_{ij} = 1$ for all $(i,j) \in H$. If all checks succeed, V outputs 1.

Fig. 1. A QPoK (P, V) for Hamiltonian cycles.

Corollary 11 (QPoKs for all languages in NP). *Let R be an NP-relation.*[6] *Then there is a zero-knowledge QPoK for R with negligible knowledge error.*

References

1. Aaronson, S.: Limitations of quantum advice and one-way communication. Theory of Computing 1(1), 1–28 (2005), online available at http://www.theoryofcomputing.org/articles/v001a001
2. Adcock, M., Cleve, R.: A quantum goldreich-levin theorem with cryptographic applications. In: STACS 2002. LNCS, vol. 2285, pp. 323–334. Springer, Berlin, Heidelberg (2002)
3. Bellare, M., Goldreich, O.: On defining proofs of knowledge. In: Brickell, E.F. (ed.) CRYPTO '92. LNCS, vol. 740, pp. 390–420. Springer-Verlag (1993), extended version online available at http://www-cse.ucsd.edu/users/mihir/papers/pok.ps

---

[6] An NP-relation is a relation R such that $(x,w) \in R$ is decidable in deterministic polynomial time, and there is a polynomial p such that for all $(x,w) \in R$, $|w| \leq p(|x|)$.

4. Blum, M.: How to prove a theorem so no one else can claim it. In: Proceedings of the International Congress of Mathematicians. pp. 1444–1451. Berkeley (1986)
5. Goldreich, O., Micali, S., Wigderson, A.: How to play any mental game – or – a completeness theorem for protocols with honest majority. In: STOC 87. pp. 218–229 (1987)
6. Goldreich, O., Micali, S., 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), 690–728 (1991), online available at `http://www.wisdom.weizmann.ac.il/~oded/X/gmw1j.pdf`
7. Goldwasser, S., Micali, S., Rackoff, C.: The knowledge complexity of interactive proof-systems. In: Proceedings of the seventeenth annual ACM symposium on Theory of computing. pp. 291–304. ACM Press (1985)
8. Goldwasser, S., Micali, S., Rivest, R.L.: A digital signature scheme secure against adaptive chosen-message attacks. SIAM Journal on Computing 17(2), 281–308 (1988), online available at `http://theory.lcs.mit.edu/~rivest/GoldwasserMicaliRivest-ADigitalSignatureSchemeSecureAgainstAdaptiveChosenMessageAttacks.ps`
9. van de Graaf, J.: Towards a formal definition of security for quantum protocols. Ph.D. thesis, Départment d'informatique et de r.o., Université de Montréal (1998), online available at `http://www.cs.mcgill.ca/~crepeau/PS/these-jeroen.ps`
10. Grover, L.K.: A fast quantum mechanical algorithm for database search. In: STOC. pp. 212–219 (1996)
11. Halevi, S., Micali, S.: More on proofs of knowledge. IACR ePrint 1998/015 (1998)
12. Hallgren, S., Smith, A., Song, F.: Classical cryptographic protocols in a quantum world. In: Crypto 2011. LNCS, vol. 6841, pp. 411–428. Springer (2011)
13. Jensen, J.L.W.V.: Sur les fonctions convexes et les inégalités entre les valeurs moyennes. Acta Mathematica 30(1), 175–193 (1906), in French
14. Lunemann, C., Nielsen, J.B.: Fully simulatable quantum-secure coin-flipping and applications. In: Africacrypt 2011. LNCS, vol. 6737, pp. 21–40. Springer (2011)
15. Nielsen, M.A., Chuang, I.L.: Quantum Computation and Quantum Information. Cambridge University Press (2000)
16. Shor, P.W.: Algorithms for quantum computation: Discrete logarithms and factoring. In: 35th Annual Symposium on Foundations of Computer Science, Proceedings of FOCS 1994. pp. 124–134. IEEE Computer Society (1994)
17. Unruh, D.: Universally composable quantum multi-party computation. In: Eurocrypt 2010. pp. 486–505. LNCS, Springer (2010), preprint on arXiv:0910.2912 [quant-ph]
18. Unruh, D.: Quantum proofs of knowledge. IACR ePrint 2010/212 (2012), full version
19. Watrous, J.: Zero-knowledge against quantum attacks. SIAM J. Comput. 39(1), 25–58 (2009)
20. Winter, A.: Coding Theorems of Quantum Information Theory. Ph.D. thesis, Universität Bielefeld (1999), arXiv:quant-ph/9907077v1
21. Wootters, W.K., Zurek, W.H.: A single quantum cannot be cloned. Nature 299, 802–803 (1982)