2014 IEEE Annual Symposium on Foundations of Computer Science

Quantum Attacks on Classical Proof Systems

The Hardness of Quantum Rewinding

Andris Ambainis
University of Latvia and
Institute for Advanced Study
Princeton

Ansis Rosmanis
Institute for Quantum Computing
School of Computer Science
University of Waterloo

Dominique Unruh
University of Tartu

***Abstract*—Quantum zero-knowledge proofs and quantum proofs of knowledge are inherently difficult to analyze because their security analysis uses rewinding. Certain cases of quantum rewinding are handled by the results by Watrous (SIAM J Comput, 2009) and Unruh (Eurocrypt 2012), yet in general the problem remains elusive. We show that this is not only due to a lack of proof techniques: relative to an oracle, we show that classically secure proofs and proofs of knowledge are insecure in the quantum setting.**

More specifically, sigma-protocols, the Fiat-Shamir construction, and Fischlin's proof system are quantum insecure under assumptions that are sufficient for classical security. Additionally, we show that for similar reasons, computationally binding commitments provide almost no security guarantees in a quantum setting.

To show these results, we develop the "pick-one trick", a general technique that allows an adversary to find one value satisfying a given predicate, but not two.

***Keywords*-quantum cryptography; quantum query complexity; rewinding; random oracles**

I. Introduction

Quantum computers threaten classical cryptography. With a quantum computer, an attacker would be able to break all schemes based on the hardness of factoring, or on the hardness of discrete logarithms [1], this would affect most public key encryption and signature schemes is use today. For symmetric ciphers and hash functions, longer key and output lengths will be required due to considerable improvements in brute force attacks [2], [3]. These threats lead to the question: how can classical cryptography be made secure against quantum attacks? Much research has been done towards cryptographic schemes based on hardness assumptions not known to be vulnerable to quantum computers, e.g., lattice-based cryptography. (This is called *post-quantum cryptography*; see [4] for a somewhat dated survey.) Yet, identifying useful quantum-hard assumptions is only half of the problem. Even if the underlying assumption holds against quantum attackers, for many classically secure protocols it is not clear if they also resist quantum attacks: the proof techniques used in the classical setting often cannot be applied in the quantum world. This raises the question whether it is just our proof techniques that are insufficient, or whether the protocols themselves are quantum insecure. The most prominent example are zero-knowledge proofs. To show the security of a zero-knowledge proof system, one typically uses rewinding. That is, in a hypothetical execution, the adversary's state is saved, and the adversary is executed several times starting from that state. In the quantum setting, we cannot do that: saving a quantum state means cloning it, violating the no-cloning theorem [5]. Watrous [6] showed that for many zero-knowledge proofs, security can be shown using a quantum version of the rewinding technique. (Yet this technique is not as versatile as classical rewinding. For example, the quantum security of the graph non-isomorphism proof system [7] is an open problem.) Unruh [8] noticed that Watrous' rewinding cannot be used to show the security of proofs of knowledge; he developed a new rewinding technique to show that so-called sigma-protocols are proofs of knowledge. Yet, in [8] an unexpected condition was needed: their technique only applies to proofs of knowledge with *strict soundness* (which roughly means that the last message in the interaction is determined by the earlier ones); this condition is not needed in the classical case. The security of sigma-protocols without strict soundness (e.g., graph isomorphism [7]) was left open. The problem also applies to arguments as well (i.e., computationally-sound proof systems, without "of knowledge"), as these are often shown secure by proving that they are actually arguments of knowledge. Further cases where new proof techniques are needed in the quantum setting are schemes involving random oracles. Various proof techniques were developed [9]–[13], but all are restricted to specific cases, none of them matches the power of the classical proof techniques.

To summarize: For many constructions that are easy to prove secure classically, proofs in the quantum setting are much harder and come with additional conditions limiting their applicability. The question is: does this only reflect our lack of understanding of the quantum setting, or are those additional conditions indeed necessary? Or could it be that those classically secure constructions are actually insecure quantumly?

Our contribution. We show, relative to an oracle, that the answer is indeed **yes**:

- Sigma-protocols are not necessarily quantum proofs of knowledge, even if they are classical proofs of knowledge. In particular, the strict soundness condition from [8] is necessary. (Theorem 15)
- In the computational setting, sigma-protocols are not necessarily quantum arguments, even if they are classical arguments. (Theorem 19)
- The Fiat-Shamir construction [14] for non-interactive

0272-5428/14 $31.00 © 2014 IEEE
DOI 10.1109/FOCS.2014.57

proofs of knowledge in the random oracle model does not give rise to quantum proofs of knowledge. And in the computational setting, not even to quantum arguments. (Theorems 24 and 25)
- Fischlin's non-interactive proof of knowledge in the random oracle model [15] is not a quantum proof of knowledge. (This is remarkable because in contrast to Fiat-Shamir, the classical security proof of Fischlin's scheme does not use rewinding.) And in the computational setting, it is not even an argument. (Theorems 27 and 28)
- Besides proof systems, we also have negative results for commitment schemes. The usual classical definition of computationally binding commitments is that the adversary cannot provide openings to two different values for the same commitment. Surprisingly, relative to an oracle, there are computationally binding commitments where a quantum adversary can open the commitment to any value he chooses (just not to two values *simultaneously*). (Theorem 11)
- The results on commitments in turn allow us to strengthen the above results for proof systems. While it is known that even in the quantum case, sigma-protocols with so-called "strict soundness" (the third message is uniquely determined by the other two) are proofs and proofs of knowledge [8], using the computational variant of this property leads to schemes that are not even computationally secure. (Theorems 15, 19, 24, 25, 27, and 28.)

Figure 1 gives an overview of the results relating to proofs of knowledge.

Our main result are the separations listed in the bullet points above. Towards that goal, we additionally develop two tools that may be of independent interest in quantum cryptographic proofs:

- Section III: We develop the "pick-one" trick, a technique for providing the adversary with the ability to compute a value with a certain property, but not two of them. (See "our technique" below.) This technique and the matching lower bound on the adversary's query complexity may be useful for developing further oracle separations between quantum and classical security. (At least it gives rise to all the separations listed above.)
- We show (in the full version) how to create an oracle that allows us to create arbitrarily many copies of a given state $|\Psi\rangle$, but that is not more powerful than having many copies of $|\Psi\rangle$, even if queried in superposition. Again, this might be useful for other oracle separations, too. (The construction of $\mathcal{O}_\Psi$ in Section III is an example for this.)

Related work. Van der Graaf [16] first noticed that security definitions based on rewinding might be problematic in the quantum setting. Watrous [6] showed how the problems with quantum rewinding can be solved for a large class of zero-knowledge proofs. Unruh [8] gave similar results for proofs of knowledge; however he introduced the additional condition "strict soundness" and they did not cover the computational case (arguments and arguments of knowledge). Our work (the results on sigma-protocols, Section V) shows that these restrictions are not accidental: both strict soundness and statistical security are required for the result from [8] to hold. Protocols that are secure classically but insecure in the quantum setting were constructed before: [17] presented classically secure pseudorandom functions that become insecure when the adversary is not only quantum, but can also *query the pseudorandom function in superposition*. Similarly for secret sharing schemes [18] and one-time MACs [19]. But, in all of these cases, the negative results are shown for the case when the adversary is allowed to interact with the honest parties in superposition. Thus, the cryptographic protocol is different in the classical case and the quantum case. In contrast, we keep the protocols the same, with only classical communication and only change adversary's internal power (by allowing it to be a polynomial-time quantum computer which may access quantum oracles). We believe that this is the first such separation. Boneh, Dagdelen, Fischlin, Lehmann, Schaffner, and Zhandry [9] first showed how to correctly define the random oracle in the quantum setting (namely, the adversary has to have superposition access to it). For the Fiat-Shamir construction (using random oracles as modeled by [9]), an impossibility result was given by Dagdelen, Fischlin, and Gagliardoni [20]. However, their impossibility only shows that security of Fiat-Shamir cannot be shown using extractors that do not perform quantum rewinding;[1] but such quantum rewinding is possible and used in the existing positive results from [6], [8] which would also not work in a model without quantum rewinding. A *variant of* Fiat-Shamir has been shown to be a quantum secure signature scheme [20]. Probably their scheme can also be shown to be a quantum zero-knowledge proof of knowledge.[2] However, their construction assumes sigma-protocols with "oblivious commitments". These are a much stronger assumption that usual sigma-protocols: as shown in [21, Appendix A], sigma-protocols with oblivious commitments are by themselves already non-interactive zero-knowledge proofs in the CRS model (albeit single-theorem, non-adaptive ones). [21] presents a non-interactive quantum zero-knowledge proof of knowledge in the random oracle model, based on arbitrary sigma-protocols (it does not even need strict soundness). That protocol uses ideas different from both Fiat-Shamir and Fischlin's scheme to avoid rewinding.

It was known for a long time that it is difficult to use classical definitions for computational binding in the quantum setting ([22] is the first reference we are aware of), but none showed so far that the computational definition was truly insufficient.

Our technique. The schemes we analyze are all based on sigma-protocols which have the *special soundness* property: In a proof of a statement s, given two accepting conversations $(com, ch, resp)$ and $(com, ch', resp')$, one can efficiently ex-

[1] They do allow extractors that restart the adversary with the same classical randomness from the very beginning. But due to the randomness inherent in quantum measurements, the adversary will then not necessarily reach the same state again. They also do not allow the extractor to use a purified (i.e., unitary) adversary to avoid measurements that introduce randomness.

[2] The unforgeability proof from [20] is already almost a proof of the proof of knowledge property. And the techniques from [21] can probably be applied to show that the protocol form [20] is zero-knowledge.

Underlying sigma-protocol			Sig.-pr. used directly		Fiat-Shamir		Fischlin	
zero-knowledge	special soundness	strict soundness	PoK	proof	PoK	proof	PoK	proof
stat	perf	comp	attack[15]	stat [6]	attack[24]	?	attack[27]	?
stat	comp	comp	attack[19]	attack[19]	attack[25]	attack[25]	attack[28]	attack[28]
stat	perf	perf	stat [8]	stat [6]	?	?	?	?

Fig. 1: Taxonomy of proofs of knowledge. For different combinations of security properties of the underlying sigma-protocol (statistical (stat)/perfect (perf)/computational (comp)), is there an attack in the quantum setting (relative to an oracle)? Or do we get a statistically/computationally secure proof/proof of knowledge (PoK)? The superscripts refer to theorem numbers in this paper or to literature references. Note that in all cases, classically we have at least computational security.

tract a witness for s. (The *commitment* com and the *response* $resp$ are sent by the prover, and the *challenge* ch by the verifier.) In the classical case, we can ensure that the prover cannot produce one accepting conversation without having enough information to produce two. This is typically proven by rewinding the prover to get two conversations. So in order to break the schemes in the quantum case, we need to give the prover some information that allows him to succeed in one interaction, but not in two.

To do so, we use the following trick (we call it the *pick-one trick*): Let S be a set of values (e.g., accepting conversations). Give the quantum state $|\Psi\rangle := \frac{1}{\sqrt{|S|}} \sum_{x\in S} |x\rangle$ to the adversary. Now the adversary can get a random $x \in S$ by measuring $|\Psi\rangle$. However, on its own that is not more useful than just providing a random $x \in S$. So in addition, we provide an oracle that applies the unitary $\mathcal{O}_F$ with $\mathcal{O}_F|\Psi\rangle = -|\Psi\rangle$ and $\mathcal{O}_F|\Psi^\perp\rangle = |\Psi^\perp\rangle$ for all $|\Psi^\perp\rangle$ orthogonal to $|\Psi\rangle$. Now the adversary can use (a variant of) Grover's search starting with state $|\Psi\rangle$ to find some $x \in S$ that satisfies a predicate $P(x)$ of his choosing, as long as $|S|/|\{x \in S : P(x)\}|$ is polynomially bounded. Note however: once the adversary did this, $|\Psi\rangle$ is gone, he cannot get a second $x \in S$.

How do we use that to break proofs of knowledge? The simplest case is attacking the sigma-protocol itself. Assume the challenge space is polynomial. (I.e., $|ch|$ is logarithmic.) Fix a commitment com, and let S be the set of all $(ch, resp)$ that form an accepting conversation with com. Give com and $|\Psi\rangle$ to the malicious prover. (Actually, in the full proof we provide an oracle $\mathcal{O}_\Psi$ that allows us to get $|\Psi\rangle$ for a random com.) He sends com and receives a challenge ch'. And using the pick-one trick, he gets $(ch, resp) \in S$ such that $ch = ch'$. Thus sending $resp$ will make the verifier accept.

This in itself does not constitute a break of the protocol. A malicious prover is allowed to make the verifier accept, as long as he knows a witness. Thus we need to show that even given $|\Psi\rangle$ and $\mathcal{O}_F$, it is hard to compute a witness. Given two accepting conversations $(com, ch, resp)$ and $(com, ch', resp')$ we can compute a witness. So we need that given $|\Psi\rangle$ and $\mathcal{O}_F$, it is hard to find two different $x, x' \in S$. We show this below (under certain assumptions on the size of S, see Theorem 4, Corollary 7). Thus the sigma-protocol is indeed broken: the malicious prover can make the verifier accept using information that does not allow him to compute a witness. (The full counterexample will need additional oracles, e.g., for membership test in S etc.) Counterexamples for the other constructions (Fiat-Shamir, Fischlin, etc.) are constructed similarly. We stress that this does not contradict the security of sigma-protocols with strict soundness [8]. Strict soundness implies that there is only one response per challenge. Then $|S|$ is polynomial and it becomes possible to extract two accepting conversations from $|\Psi\rangle$ and $\mathcal{O}_F$.

The main technical challenge is to prove that given $|\Psi\rangle$ and $\mathcal{O}_F$, it is hard to find two different $x, x' \in S$. This is done using the representation-theoretic form of "quantum adversary" lower bound method for quantum algorithms [23], [24]. The method is based on viewing a quantum algorithm as a sequence of transformations on a bipartite quantum system that consists of two registers: one register $\mathcal{H}_A$ that contains the algorithm's quantum state and another register $\mathcal{H}_I$ that contains the information which triples $(com, ch, resp)$ belong to S. The algorithm's purpose is to obtain two elements $x_1, x_2 \in S$ using only a limited type of interactions betweeen $\mathcal{H}_A$ and $\mathcal{H}_I$. (From a practical perspective, a quantum register $\mathcal{H}_I$ holding the membership information about S would be huge. However, we do not propose to implement such a register. Rather, we use it as a tool to prove a lower bound which then implies a corresponding lower bound in the usual model where S is accessed via oracles.)

We then partition the state-space of $\mathcal{H}_I$ into subspaces corresponding to group representations of the symmetry group of $\mathcal{H}_I$ (the set of all permutations of triples $(com, ch, resp)$ that satisfy some natural requirements). Informally, these subspaces correspond to possible states of algorithm's knowledge about the input data: having no information about any $s \in S$, knowing one value $x \in S$, knowing two values $x_1, x_2 \in S$ and so on.

The initial state in which the algorithm has $|\Psi\rangle$ corresponds to $\mathcal{H}_I$ being in the state "the algorithm knows one $x \in S$". (This is very natural because measuring $|\Psi\rangle$ gives one value $x \in S$ and there is no way to obtain two values $x \in S$ from this state with a non-negligible probability.) We then show that each application of the available oracles (such as O_F and the membership test for S) can only move a tiny part of the state in $\mathcal{H}_I$ from the "the algorithm knows one $x \in S$" subspace of $\mathcal{H}_I$ to the "the algorithm knows two $x \in S$" subspace. Therefore, to obtain two values $x_1, x_2 \in S$, we need to apply the available oracles a large number of times.

While the main idea is quite simple, implementing it requires a sophisticated analysis of the representations of the symmetry group of $\mathcal{H}_I$ and how they evolves when the oracles are applied.

Actually, below we prove an even stronger result: We do not wish to give the state $|\Psi\rangle$ as input to the adversary. (Because that would mean that the attack only works with an input that is not efficiently computable, even in our relativized model.) Thus, instead, we provide an oracle $\mathcal{O}_\Psi$ for efficiently constructing this state. But then, since the oracle can be invoked arbitrarily many times, the adversary could create two copies of $|\Psi\rangle$, thus easily obtaining two $x, x' \in S$! Instead, we provide an oracle $\mathcal{O}_\Psi$ that provides a state $|\Sigma\Psi\rangle$ which is a superposition of many $|\Psi\rangle = |\Psi(y)\rangle$ for independently chosen sets S_y. Now the adversary can produce $|\Sigma\Psi\rangle$ and using a measurement of y, get many states $|\Psi(y)\rangle$ for random y's, but no two states $|\Psi(y)\rangle$ for the same y. Taking these additional capabilities into account complicates the proof further, as does the presence of additional oracles that are needed, e.g., to construct the prover (who does need to be able to get several $x \in S$).

On the meaning of oracle separations. At this point, we should say a few words about what it implies that our impossibility results are relative to a certain oracle. Certainly, our results do not necessarily imply that the investigated schemes are insecure or unprovable in the "real world", i.e., without oracles. However, our results give a number of valuable insights. Foremost, they tell us which proof techniques cannot be used for showing security of those schemes: only non-relativizing proofs can work. This cuts down the search space for proofs considerable. Also, it shows that security proofs would need new techniques; the proof techniques from [6], [8] at least are relativizing. And even non-relativizing proof techniques such as (in the classical setting) [25] tend to use specially designed (and more complicated) protocols than their relativizing counterparts, so our results might give some evidence that the specific protocols we investigate here have no proofs at all, whether relativizing or non-relativizing. Furthermore, oracle-based impossibilities can give ideas for non-oracle-based impossibilities. If we can find computational problems that exhibit similar properties as our oracles, we might get analogous impossibilities without resorting to oracles (using computational assumptions instead).[3] However, we should stress that even if we get rid of the oracles, our results do not state that *all* sigma-protocols lead to insecure schemes. It would not be excluded that, e.g., the graph-isomorphism sigma-protocol [7] is still a proof of knowledge. What our approach aims to show is the impossibility of *general* constructions that are secure for *all* sigma-protocols.

Finally, we mention one point that is important in general when designing oracle separations in the quantum world: even relative to an oracle, the structural properties of quantum circuits should not change. For example, any quantum algorithm (even one that involves intermediate measurements or other non-unitary operations) can be replaced by a unitary quantum circuit, and that unitary circuit can be reversed. If we choose oracles that are not reversible, then we lose this property. (E.g., oracles that perform measurements or that perform random choices are non-reversible.) So an impossibility result based on such oracles would only apply in a world where quantum circuits are not reversible. Thus for meaningful oracle separations, we need to ensure that: (a) all oracles are unitary, and (b) all oracles have inverses. This makes some of the definitions of oracles in our work (Definition 6) more involved than would be necessary if we had used non-unitary oracles.

Organization. Section II introduces security definitions. Section III develops the pick-one trick. Section IV shows the insecurity of computationally binding commitments, Section V that of sigma-protocols, Section VI that of the Fiat-Shamir construction, and Section VII that of Fischlin's construction. Additional details and full proofs are given in the full version [27].

II. Security definitions

A *sigma-protocol* for a relation R is a three message proof system. It is described by the lengths $\ell_{com}, \ell_{ch}, \ell_{resp}$ of the messages, a polynomial-time prover (P_1, P_2) and a polynomial-time verifier V. The first message from the prover is $com \leftarrow P_1(s, w)$ with $(s, w) \in R$ and is called *commitment*, the uniformly random reply from the verifier is $ch \xleftarrow{\$} \{0,1\}^{\ell_{ch}}$ (called *challenge*), and the prover answers with $resp \leftarrow P_2(ch)$ (the *response*). We assume P_1, P_2 to share state. Finally $V(s, com, ch, resp)$ outputs whether the verifier accepts.

We will make use of the following standard properties of sigma-protocols. Note that we have chosen to make the definition stronger by requiring honest entities (simulator, extractor) to be classical while we allow the adversary to be quantum.

Definition 1 (Properties of sigma-protocols): Let $(\ell_{com}, \ell_{ch}, \ell_{resp}, P_1, P_2, V, R)$ be a sigma-protocol. We define:

- **Completeness:** For all $(s, w) \in R$, $\Pr[ok = 0 : com \leftarrow P_1(s, w), ch \xleftarrow{\$} \{0,1\}^{\ell_{ch}}, resp \leftarrow P_2(ch), ok \leftarrow V(s, com, ch, resp)]$ is negligible.
- **Perfect special soundness:** There is a polynomial-time classical algorithm E_Σ (the extractor) such that for any $(s, com, ch, resp, ch', resp')$ with $ch \neq ch'$, we have that $\Pr[(s, w) \notin R \wedge ok = ok' = 1 : ok \leftarrow V(s, com, ch, resp), ok' \leftarrow V(s, com, ch', resp'), w \leftarrow E_\Sigma(s, com, ch, resp, ch', resp')] = 0$.
- **Computational special soundness:** There is a polynomial-time classical algorithm E_Σ (the extractor) such that for any polynomial-time quantum algorithm A (the adversary), we have that $\Pr[(s, w) \notin R \wedge ch \neq ch' \wedge ok = ok' = 1 : (s, com, ch, resp, ch', resp') \leftarrow A, ok \leftarrow V(s, com, ch, resp), ok' \leftarrow V(s, com, ch', resp'), w \leftarrow E_\Sigma(s, com, ch, resp, ch', resp')]$ is negligible.
- **Statistical honest-verifier zero-knowledge (HVZK):**[4] There is a polynomial-time classical algorithm S_Σ (the

[3] For example, [26] presents a construction that might allow to implement an analogue to the oracle $\mathcal{O}_F$. Essentially, if the set S (called A in [26]) is a linear code, then they give a candidate for how to obfuscate $\mathcal{O}_F$ (called V_A in [26]) such that one can apply $\mathcal{O}_F$ but does not learn A. Of course, this does not give us a candidate for how to construct the other oracles needed in this work, but it shows that the idea of actually replacing our custom made oracles by computational assumptions may not be far fetched.

[4] In the context of this paper, HVZK is equivalent to zero-knowledge because our protocols have logarithmic challenge length ℓ_{ch} [6].

simulator) such that for any (possibly unlimited) quantum algorithm A and all $(s,w) \in R$, the following is negligible:

$$\big|\Pr[b=1 : com \leftarrow P_1(s,w), ch \xleftarrow{\$} \{0,1\}^{\ell_{ch}}, resp \leftarrow P_2(ch), b \leftarrow A(com,ch,resp)] - \Pr[b=1 : (com,ch,resp) \leftarrow S(s), b \leftarrow A(com,ch,resp)]\big|$$

- **Strict soundness:** For any (s, com, ch) and any $resp \neq resp'$ we have $\Pr[ok = ok' = 1 : ok \leftarrow V(s,com,ch,resp), ok' \leftarrow V(s,com,ch,resp')] = 0$.
- **Computational strict soundness:**[5] For any polynomial-time quantum algorithm A (the adversary), we have that $\Pr[ok = ok' = 1 \wedge resp \neq resp' : (s,com,ch,resp,resp') \leftarrow A, ok \leftarrow V(s,com,ch,resp), ok' \leftarrow V(s,com,ch,resp')]$ is negligible.
- **Commitment entropy:** For all $(s,w) \in R$ and $com \leftarrow P_1(s,w)$, the min-entropy of com is superlogarithmic.

In a relativized setting, all quantum algorithms additionally get access to all oracles, and all classical algorithms additionally get access to all classical oracles. ◇

In this paper, we will mainly be concerned with proving that certain schemes are *not* proofs of knowledge. Therefore, we will not need to have precise definitions of these concepts; we only need to know what it means to break them.

Definition 2 (Total breaks): Consider an interactive or non-interactive proof system (P,V) for a relation R. Let $L_R := \{s : \exists w.(s,w) \in R\}$ be the language defined by R. A *total break* is a polynomial-time quantum algorithm A such that the following probability is overwhelming:

$$\Pr[ok = 1 \ \wedge \ s \notin L_R : s \leftarrow A, ok \leftarrow \langle A, V(s)\rangle]$$

Here $\langle A, V(s)\rangle$ denotes the output of V in an interaction between A and $V(s)$. (Intuitively, the adversary performs a total break if the adversary manages with overwhelming probability to convince the verifier V of a statement s that is not in the language L_R.)

A *total knowledge break* is a polynomial-time quantum algorithm A such that for all polynomial-time quantum algorithms E we have that:

- Adversary success: $\Pr[ok = 1 : s \leftarrow A, ok \leftarrow \langle A, V(s)\rangle]$ is overwhelming.
- Extractor failure: $\Pr[(s,w) \in R : s \leftarrow A, w \leftarrow E(s)]$ is negligible.

Here E has access to the final state of A. (Intuitively, the adversary performs a total knowledge break if the adversary manages with overwhelming probability to convince the verifier V of a statement s, but the extractor E cannot extract a witness w for that statement.)

When applied to a proof system relative to an oracle $\mathcal{O}$, both A and E get access to $\mathcal{O}$. In settings where R and $\mathcal{O}$ are probabilistic, the probabilities are averaged over all values of R and $\mathcal{O}$. ◇

Note that these definitions of attacks are quite strong. In particular, A does not get any auxiliary state. And A needs to succeed with overwhelming probability and make the extraction fail with overwhelming probability. (Usually, proofs / proofs of knowledge are considered broken already when the adversary has non-negligible success probability.) Furthermore, we require A to be polynomial-time.

In particular, a total break implies that a proof system is neither a proof nor an argument. And total knowledge break implies that it is neither a proof of knowledge nor an argument of knowledge, with respect to all definitions the authors are aware of.[6]

III. The pick-one trick

In this section, we first show a basic case of the pick-one trick which focusses on the core query complexity aspects. In Section III-A, we extend this by a number of additional oracles that will be needed in the rest of the paper.

Definition 3 (Two values problem): Let X, Y be finite sets and let $k \leq |X|$ be a positive integer. For each $y \in Y$, let S_y be a uniformly random subset of X of cardinality k, let $|\Psi(y)\rangle := \sum_{x\in S_y} |x\rangle/\sqrt{k}$. Let $|\Sigma\Psi\rangle = \sum_{y\in Y} |y\rangle|\Psi(y)\rangle/\sqrt{|Y|}$ and $|\Sigma\Phi\rangle = \sum_{y\in Y, x\in X} |y\rangle|x\rangle/\sqrt{|Y|\cdot|X|}$. The *Two Values* problem is to find $y \in Y$ and $x_1, x_2 \in S_y$ such that $x_1 \neq x_2$ given the following resources:

- one instance of the state $\bigotimes_{\ell=1}^{h}(\alpha_{\ell,0}|\Sigma\Psi\rangle + \alpha_{\ell,1}|\Sigma\Phi\rangle)$, where h and the coefficients α are independent of the S_y's and are such that this state has unit norm;
- an oracle $\mathcal{O}_V$ such that for all $y \in Y$, $x \in X$, $\mathcal{O}_V(y,x) = 0$ if $x \notin S_y$ and $\mathcal{O}_V(y,x) = 1$ if $x \in S_y$.
- on oracle $\mathcal{O}_F$ that, for all $y \in Y$, maps $|y,\Psi(y)\rangle$ to $-|y,\Psi(y)\rangle$ and, for any $|\Psi^\perp\rangle$ orthogonal to $|\Psi(y)\rangle$, maps $|y,\Psi^\perp\rangle$ to itself. ◇

The two values problem is at the core of the *pick-one trick*: if we give an adversary access to the resources described in Definition 3, he will be able to search for one $x \in S_y$ satisfying a predicate P (shown in Theorem 5 below). But he will not be able to find two different $x, x' \in S_y$ (Theorem 4 below); we will use this to foil any attempts at extracting by rewinding.

Theorem 4 (Hardness of the two values problem): Let $\mathcal{A}$ be an algorithm for the Two Values problem that makes q_V and q_F queries to oracles $\mathcal{O}_V$ and $\mathcal{O}_F$, respectively. The success probability for $\mathcal{A}$ to find $y \in Y$ and $x_1, x_2 \in S_y$ such that $x_1 \neq x_2$ is at most

$$O\left(\frac{h}{|Y|^{1/2}} + \frac{(q_V+q_F)^{1/2}k^{1/4}}{|X|^{1/4}} + \frac{(q_V+q_F)^{1/2}}{k^{1/4}}\right). \quad \diamond$$

The proof uses the adversary-method from [23], [24] as described in the introduction. In Section III-A we extend this hardness result to cover additional oracles.

[5] Also known as *unique responses* in [15].

[6] Definitions that would not be covered would be such where the extractor gets additional auxiliary input not available to the adversary. We are, however, not aware of such in the literature.

Theorem 5 (Searching one value): Let $S_y \subseteq X$ and $\mathcal{O}_F, \mathcal{O}_V$ be as in Definition 3.

There is a polynomial-time oracle algorithm E_1 that on input $|\Sigma\Psi\rangle$ returns a uniformly random $y \in Y$ and $|\Psi(y)\rangle$. There is a polynomial-time oracle algorithm E_2 such that: For any $\delta_{\min} > 0$, for any $y \in Y$, for any predicate P on X with $|\{x \in S_y : P(x) = 1\}|/|S_y| \geq \delta_{\min}$, and for any $n \geq 0$ we have

$$\Pr[x \in S_y \ \wedge \ P(x) = 1 : x \leftarrow E_2^{\mathcal{O}_V, \mathcal{O}_F, P}(n, \delta_{\min}, y, |\Psi(y)\rangle)] \geq 1 - 2^{-n}.$$

(The running time of E_2 is polynomial-time in n, $1/\delta_{\min}$, $|y|$.) ◇

This theorem is proven with a variant of Grover's algorithm [2]: Using Grover's algorithm, we search for an x with $P(x) = 1$. However, we do not search over all $x \in \{0,1\}^\ell$ for some ℓ, but instead over all $x \in S_y$. When searching over S_y, the initial state of Grover's algorithm needs to be $\sum_x \frac{1}{\sqrt{|S_y|}}|x\rangle = |\Psi(y)\rangle$ instead of $\sum_x 2^{-\ell/2}|x\rangle =: |\Phi\rangle$. And the diffusion operator $I - 2|\Phi\rangle\langle\Phi|$ needs to be replaced by $I - 2|\Psi(y)\rangle\langle\Psi(y)|$. Fortunately, we have access both to $|\Psi(y)\rangle$ (given as input), and to $I - 2|\Psi(y)\rangle\langle\Psi(y)|$ (through the oracle $\mathcal{O}_F$). To get an overwhelming success probability, Grover's algorithm is usually repeated until it succeeds. (In particular, when the number of solutions is not precisely known [28].) We cannot do that: we have only one copy of the initial state. Fortunately, by being more careful in how we measure the final result, we can make sure that the final state in case of failure is also a suitable initial state for Grover's algorithm.

A. Additional oracles

In this section, we extend the hardness of the two values problem to cover additional oracles that we will need in various parts of the paper.

Definition 6 (Oracle distribution): Fix integers $\ell_{com}, \ell_{ch}, \ell_{resp}$ (that may depend on the security parameter) such that ℓ_{com}, ℓ_{resp} are superlogarithmic and ℓ_{ch} is logarithmic. Let $\ell_{rand} := \ell_{com} + \ell_{resp}$.

Let $\mathcal{O}_{all} = (\mathcal{O}_E, \mathcal{O}_P, \mathcal{O}_R, \mathcal{O}_S, \mathcal{O}_F, \mathcal{O}_\Psi, \mathcal{O}_V)$ be chosen according to the following distribution:

- Let s_0 be arbitrary but fixed (e.g., $s_0 := 0$). Pick $w_0 \xleftarrow{\$} \{0,1\}^{\ell_{rand}}$.
- Choose S_y, $\mathcal{O}_V$, $\mathcal{O}_F$ as in Definition 3 with $Y := \{0,1\}^{\ell_{com}}$ and $X := \{0,1\}^{\ell_{ch}} \times \{0,1\}^{\ell_{resp}}$ and $k := 2^{\ell_{ch} + \lfloor \ell_{resp}/3 \rfloor}$.
- For each $z \in \{0,1\}^{\ell_{rand}}$, pick $y \xleftarrow{\$} Y$ and $x \xleftarrow{\$} S_y$, and set $\mathcal{O}_S(z) := (y, x)$.
- Let $|\perp\rangle$ be a quantum state orthogonal to all $|com, ch, resp\rangle$ (i.e., we extend the dimension of the space in which $|\Sigma\Psi\rangle$ lives by one). $\mathcal{O}_\Psi|\perp\rangle := |\Sigma\Psi\rangle$, $\mathcal{O}_\Psi|\Sigma\Psi\rangle := |\perp\rangle$, and $\mathcal{O}_\Psi|\Phi\rangle := |\Phi\rangle$ for $|\Phi\rangle$ orthogonal to $|\Sigma\Psi\rangle$ and $|\perp\rangle$.
- Let $\mathcal{O}_E(com, ch, resp, ch', resp') := w_0$ iff $(ch, resp), (ch', resp') \in S_{com} \wedge (ch, resp) \neq (ch', resp')$ and $\mathcal{O}_E := 0$ everywhere else.
- Let $\mathcal{O}_R(s_0, w_0) := 1$ and $\mathcal{O}_R := 0$ everywhere else.
- For each $com \in \{0,1\}^{\ell_{com}}$, $ch \in \{0,1\}^{\ell_{ch}}$, $z \in \{0,1\}^{\ell_{rand}}$, let $\mathcal{O}_P(w_0, com, ch, z)$ be assigned a uniformly random $resp$ with $(ch, resp) \in S_{com}$. (Or $\perp$ if no such $resp$ exists.) Let $\mathcal{O}_P(w, \cdot, \cdot, \cdot) := 0$ for $w \neq w_0$. ◇

The following corollary is a strengthening of Theorem 4 to the oracle distribution from Definition 6. For later convenience, we express the soundness additionally in terms of guessing w_0.

Corollary 7 (Hardness of two values 2): Let $\mathcal{O}_{all} = (\mathcal{O}_E, \mathcal{O}_P, \mathcal{O}_R, \mathcal{O}_S, \mathcal{O}_F, \mathcal{O}_\Psi, \mathcal{O}_V), w_0$ be as in Definition 6. Let A be an oracle algorithm making at most $q_E, q_P, q_R, q_S, q_F, q_\Psi, q_V$ queries to $\mathcal{O}_E, \mathcal{O}_P, \mathcal{O}_R, \mathcal{O}_S, \mathcal{O}_F, \mathcal{O}_\Psi, \mathcal{O}_V$, respectively. Assume that $q_E, q_P, q_R, q_S, q_F, q_V$ are polynomially-bounded (and ℓ_{com}, ℓ_{resp} are superlogarithmic by Definition 6). Then:

(i) $\Pr[(ch, resp) \neq (ch', resp') \wedge (ch, resp), (ch', resp') \in S_{com} : (com, ch, resp, ch', resp') \leftarrow A^{\mathcal{O}_{all}}]$ is negligible.

(ii) $\Pr[w = w_0 : w \leftarrow A^{\mathcal{O}_{all}}]$ is negligible. ◇

This corollary is shown by reduction to Theorem 4 (Hardness of the two values problem). Given an adversary that violates (ii), we remove step by step the oracles that are not covered by Theorem 4. First, we remove the oracles $\mathcal{O}_P, \mathcal{O}_R$. Those do not help the adversary (much) to find w_0 because $\mathcal{O}_P$ and $\mathcal{O}_R$ only give non-zero output if their input already contains w_0. Next we change A to output a collision $(ch, resp) \neq (ch', resp') \wedge (ch, resp), (ch', resp') \in S_{com}$ instead of the witness w_0; since w_0 can only be found by querying $\mathcal{O}_E$ with such a collision, this adversary succeeds with non-negligible probability, too. Furthermore, A then does not need access to $\mathcal{O}_E$ any more since $\mathcal{O}_E$ only helps in finding w_0. Next we get rid of $\mathcal{O}_\Psi$: $\mathcal{O}_\Psi$ can be emulated (up to an inversely polynomial error) using (suitable superpositions on) copies of the state $|\Sigma\Psi\rangle$. Finally we remove $\mathcal{O}_S$: Using the "small range distribution" theorem from [17], $\mathcal{O}_S$ can be replaced by an oracle that provides only a polynomial number of triples $(com, ch, resp)$. Those triples the adversary can produce himself by measuring polynomially-many copies of $|\Sigma\Psi\rangle$ in the computational basis. Thus we have shown that without loss of generality, we can assume an adversary that only uses the oracles $\mathcal{O}_F, \mathcal{O}_V$ and (suitable superpositions of) polynomially-many copies of $|\Sigma\Psi\rangle$, and that tries to find a collision. But that such an adversary cannot find a collision was shown in Theorem 4.

And (i) is shown by observing that an adversary violating (ii) leads to one violating (i) using one extra $\mathcal{O}_E$-query.

IV. Attacking commitments

In the classical setting, a non-interactive commitment scheme is usually called computationally binding if it is hard to output a commitment and two different openings (Definition 8 below). We now show that in the quantum setting, this definition is extremely weak. Namely, it may still be possible to commit to a value and then to open the commitment to an arbitrary value (just not to two values *at the same time*).

Security definitions. To state this more formally, we define the security of commitments: A *non-interactive commitment*

scheme consists of algorithms $\mathrm{COM}, \mathrm{COM}_{verify}$, such that $(c,u) \leftarrow \mathrm{COM}(m)$ returns a commitment c on the message m, and an opening information u. The sender then sends c to the recipient, who is not supposed to learn anything about m. Only when the sender later sends m, u, the recipients learns m. But, intuitively speaking, the sender should not be able to "change his mind" about m after sending c (binding property). We require *perfect completeness*, i.e., for any m and $(c,u) \leftarrow \mathrm{COM}(m)$, $\mathrm{COM}_{verify}(c,m,u) = 1$ with probability 1. In our setting, c, m, u are all classical.

Definition 8 (Computationally binding): A commitment scheme $\mathrm{COM}, \mathrm{COM}_{verify}$ is *computationally binding* iff for any quantum polynomial-time algorithm A the following probability is negligible:

$$\Pr[ok = ok' = 1 \ \wedge \ m \neq m' : (c,m,u,m',u') \leftarrow A, \\ ok \leftarrow \mathrm{COM}_{verify}(c,m,u), \ ok \leftarrow \mathrm{COM}_{verify}(c,m',u')] \quad \diamond$$

We will show below that this definition is *not* the right one in the quantum setting.

[8] also introduces a stronger variant of the binding property, called strict binding, which requires that also the opening information u is unique (not only the message). The results from [8] show that strict binding commitments can behave better under rewinding, so perhaps strict binding commitments can avoid the problems that merely binding commitments have? We define a computational variant of this property here:

Definition 9 (Computationally strict binding): A commitment scheme $\mathrm{COM}, \mathrm{COM}_{verify}$ is *computationally strict binding* iff for any quantum polynomial-time algorithm A the following probability is negligible:

$$\Pr[ok = ok' = 1 \wedge (m,u) \neq (m',u') : (c,m,u,m',u') \leftarrow A, \\ ok \leftarrow \mathrm{COM}_{verify}(c,m,u), \ ok \leftarrow \mathrm{COM}_{verify}(c,m',u')] \quad \diamond$$

We will show below that this stronger definition is also not sufficient.

Definition 10 (Statistically hiding): A commitment scheme $\mathrm{COM}, \mathrm{COM}_{verify}$ is *statistically hiding* iff for all m_1, m_2 with $|m_1| = |m_2|$ and $c_i \leftarrow \mathrm{COM}(m_i)$ for $i = 1, 2$, c_1 and c_2 are statistically indistinguishable.

The attack. We now state the insecurity of computationally binding commitments. The remainder of this section will prove the following theorem.

Theorem 11 (Insecurity of binding commitments): There is an oracle $\mathcal{O}$ and a non-interactive commitment scheme $\mathrm{COM}, \mathrm{COM}_{verify}$ such that:

- The scheme is perfectly complete, computationally binding, computationally strict binding, and statistically hiding.
- There is a quantum polynomial-time adversary B_1, B_2 such that for all m,

$$\Pr[ok = 1 : c \leftarrow B_1(|m|), u \leftarrow B_2(m), \\ ok \leftarrow \mathrm{COM}_{verify}(c,m,u)]$$

is overwhelming. (In other words, the adversary can open to a value m that he did not know while committing.) $\diamond$

In the rest of this section, when referring to the sets S_{com} from Definition 6, we will call them S_y and we refer to their members as $x \in S_y$. (Not $(ch, resp) \in S_{com}$.) In particular, oracles such as $\mathcal{O}_S$ will returns pairs (y,x), not triples $(com, ch, resp)$, etc.

We construct a commitment scheme relative to the oracle $\mathcal{O}_{all}$ from Definition 6. (Note: that oracle distribution contains more oracles than we need for Theorem 11. However, we will need in later sections that our commitment scheme is defined relative to the same oracles as the proof systems there.)

Definition 12 (Bad commitment scheme): Let $\mathsf{bit}_i(x)$ denote the i-th bit of x. We define $\mathrm{COM}, \mathrm{COM}_{verify}$ as follows:

- $\mathrm{COM}(m)$: For $i = 1, \ldots, |m|$, pick $z_i \xleftarrow{\$} \{0,1\}^{\ell_{rand}}$ and let $(y_i, x_i) := \mathcal{O}_S(z_i)$. Let $p_i \xleftarrow{\$} \{1, \ldots, \ell_{ch} + \ell_{resp}\}$. Let $b_i := m_i \oplus \mathsf{bit}_{p_i}(x_i)$. Let $c := (p_1, \ldots, p_{|m|}, y_1, \ldots, y_{|m|}, b_1, \ldots, b_{|m|})$ and $u := (x_1, \ldots, x_{|m|})$. Output (c,u).
- $\mathrm{COM}_{verify}(c,m,u)$ with $c = (p_1, \ldots, p_n, y_1, \ldots, y_n, b_1, \ldots, b_n)$ and $u = (x_1, \ldots, x_n)$: Check whether $|m| = n$. Check whether $\mathcal{O}_V(y_i, x_i) = 1$ for $i = 1, \ldots, n$. Check whether $b_i = m_i \oplus \mathsf{bit}_{p_i}(x_i)$ for $i = 1, \ldots, n$. Return 1 if all checks succeed. $\diamond$

For the results of the current section, there is actually no need for the values p_i which select which bit of x_i is used for masking the committed bit m_i. (E.g., we could always use the least significant bit of x_i.) But in Section VII (attack on Fischlin's scheme) we will need commitments of this particular form to enable a specific attack where we need to open commitments to certain values while *simultaneously* searching for these values in the first place.

Lemma 13 (Properties of COM*):* The scheme from Definition 12 is perfectly complete, computationally binding, computationally strict binding, and statistically hiding. (Relative to $\mathcal{O}_{all}$.) $\diamond$

The computational binding and computational strict binding property are a consequence of Corollary 7 (Hardness of two values 2): to open a commitment to two different values, the adversary would need to find one y_i (part of the commitment) and two $x_i \in S_{y_i}$ (part of the two openings). Corollary 7 states that this only happens with negligible probability. Statistical hiding follows from the fact that for each y_i, there are superpolynomially many $x_i \in S_{y_i}$, hence $\mathsf{bit}_{p_i}(x_i)$ is almost independent of y_i.

Lemma 14 (Attack on COM*):* There is a quantum polynomial-time adversary B_1, B_2 such that for all m, $\Pr[ok = 1 : c \leftarrow B_1(|m|), u \leftarrow B_2(m), ok \leftarrow \mathrm{COM}_{verify}(c,m,u)]$ is overwhelming. $\diamond$

Basically, the adversary B_1, B_2 commits to a random commitment. And to unveil to a message m, he needs to find values $x_i \in S_{y_i}$ with $\mathsf{bit}_{p_i}(x_i) = m_i \oplus b_i$. Since half of all x_i have this property, such x_i can be found using Theorem 5 (Searching one value).

Theorem 11 then follows immediately from

Lemmas 13 and 14.

V. Attacking sigma-protocols

We will now show that in general, sigma-protocols with special soundness are not necessarily proofs of knowledge. [8] showed that if a sigma-protocol additionally has strict soundness, it is a proof of knowledge. It was left as an open problem whether that additional condition is necessary. The following theorem resolves that open question by showing that the results from [8] do not hold without strict soundness (not even with computational strict soundness), relative to an oracle.

Theorem 15 (Insecurity of sigma-protocols): There is an oracle $\mathcal{O}_{all}$ and a relation R and a sigma-protocol relative to $\mathcal{O}_{all}$ with logarithmic ℓ_{ch} (challenge length), completeness, perfect special soundness, computational strict soundness, and statistical honest-verifier zero-knowledge for which there exists a total knowledge break.

In contrast, a sigma-protocol relative to $\mathcal{O}_{all}$ with completeness, perfect special soundness, and statistical honest-verifier zero-knowledge is a classical proof of knowledge.

Note that a corresponding theorem with polynomially bounded ℓ_{ch} follows immediately by parallel repetition of the sigma-protocol.

The remainder of this section will prove Theorem 15. As a first step, we construct the sigma-protocol.

Definition 16 (Sigma-protocol): Let $\mathrm{COM}, \mathrm{COM}_{verify}$ be the commitment scheme from Definition 12.[7]

Relative to the oracle distribution from Definition 6, we define the following sigma-protocol $(\ell_{com}, \ell_{ch}, \ell_{resp}, P_1, P_2, V, R)$ for the relation $R := \{(s_0, w_0)\}$:

- $P_1(s, w)$ picks $com \xleftarrow{\$} \{0,1\}^{\ell_{com}}$. For each $ch \in \{0,1\}^{\ell_{ch}}$, he picks $z_{ch} \xleftarrow{\$} \{0,1\}^{\ell_{rand}}$ and computes $resp_{ch} := \mathcal{O}_P(w, com, ch, z_{ch})$ and $(c_{ch}, u_{ch}) \leftarrow \mathrm{COM}(resp_{ch})$. Then P_1 outputs $com^* := (com, (c_{ch})_{ch \in \{0,1\}^{\ell_{ch}}})$.
- $P_2(ch)$ outputs $resp^* := (resp_{ch}, u_{ch})$.
- For $com^* = (com, (c_{ch})_{ch \in \{0,1\}^{\ell_{ch}}})$ and $resp^* = (resp, u)$, let $V(s, com^*, ch, resp^*) := 1$ iff $\mathcal{O}_V(com, ch, resp) = 1$ and $s = s_0$ and $\mathrm{COM}_{verify}(c_{ch}, resp, u) = 1$. ◇

The commitments c_{ch} are only needed to get computational strict soundness. A slightly weaker Theorem 15 without computational strict soundness can be achieved using the sigma-protocol from Definition 16 without the commitments c_{ch}; the proofs stay the same, except that the steps relating to the commitments are omitted.

Lemma 17 (Security of the sigma-protocol): The sigma-protocol from Definition 16 has: completeness, perfect special soundness, computational strict soundness, statistical honest-verifier zero-knowledge, commitment entropy. ◇

[7]The commitment described there has the property that it is computationally binding, but still it is possible for the adversary to open the commitment to any value, only not to several values at the same time. The commitment is defined relative to the same oracle distribution as the sigma-protocol here, which is why we can use it.

Perfect special soundness follows from the existence of the oracle $\mathcal{O}_E$. That oracle provides the witness w_0 given two accepting conversations, as required by perfect special soundness. Computational strict soundness stems from the fact that the message com^* contains commitments c_{ch} to all possible answers. Thus to break computational strict soundness (i.e., to find two different accepting $resp^*$), the adversary would need to open one of the commitments c_{ch} in two ways. This happens with negligible probability since COM is computationally strict binding. Statistical honest-verifier zero-knowledge follows from the existence of the oracle $\mathcal{O}_S$ which provides simulations. (And the commitment c_{ch} that are not opened can be filled with arbitrary values due to the statistical hiding property of COM.)

Lemma 18 (Attack on the sigma-protocol): Assume that ℓ_{ch} is logarithmically bounded. Then there exists a total knowledge break (Definition 2) against the sigma-protocol from Definition 16. ◇

To attack the sigma protocol, the malicious prover uses Theorem 5 (Searching one value) to get a com and a corresponding state $|\Psi(com)\rangle$. Then, when receiving ch, he needs to find $(ch', resp) \in S_{com}$ with $ch' = ch$. Since an inversely polynomial fraction of $(ch', resp)$ satisfy $ch' = ch$ (ℓ_{ch} is logarithmic), this can be done with Theorem 5. This allows the prover to succeed in the proof with overwhelming probability. (He additionally needs to open the commitments c_{ch} to suitably. This can be done using Lemma 14 (Attack on COM).) However, an extractor that has the same information as the prover (namely, access to the oracle $\mathcal{O}_{all}$) will fail to find w_0 by Corollary 7 (Hardness of two values 2).

Now Theorem 15 follows from Lemmas 17 and 18. (The fact that the sigma-protocol is a classical proof of knowledge is shown in [29].)

Note that we cannot expect to get a total break (as opposed to a total knowledge break): Since the sigma-protocol is a classical proof of knowledge, it is also a classical proof. But a classical proof is also a quantum proof, because an unlimited classical adversary can simulate a quantum adversary. However, this argument does not apply when we consider computationally limited provers, see Section V-A below.

A. The computational case

We now consider the variant of the impossibility result from the previous section. Namely, we consider sigma-protocols that have only computational security (more precisely, for which the special soundness property holds only computationally) and show that these are not even arguments in general (the results from the previous section only say that they are not arguments of knowledge).

Theorem 19 (Insecurity of sigma-protocols, computational): There is an oracle $\mathcal{O}_{all}$ and a relation R' and a sigma-protocol relative to $\mathcal{O}_{all}$ with logarithmic ℓ_{ch} (challenge length), completeness, *computational* special soundness, and statistical honest-verifier zero-knowledge for which there exists a *total break*.

In contrast, a sigma-protocol relative to $\mathcal{O}_{all}$ with completeness, computational special soundness, and statistical honest-verifier zero-knowledge is a classical argument. $\diamond$

Note that a corresponding theorem with polynomially bounded ℓ_{ch} follows immediately by parallel repetition of the sigma-protocol. The remainder of this section is dedicated to proving Theorem 19.

Definition 20 (Sigma-protocol, computational): We define a sigma-protocol $(\ell_{com}, \ell_{ch}, \ell_{resp}, P_1, P_2, V, R')$ as in Definition 16, except that the relation is $R' := \varnothing$. $\diamond$

Lemma 21 (Security of the sigma-protocol, computational): The sigma-protocol from Definition 20 has: completeness. *computational* special soundness. computational strict soundness. statistical honest-verifier zero-knowledge. commitment entropy. $\diamond$

Most properties are either immediate or shown as in Lemma 17 (Security of the sigma-protocol). However, perfect special soundness does not hold for the sigma-protocol from Definition 20: There exist pairs of accepting conversations $(ch, resp), (ch', resp') \in S_{com}$. But these do not allow us to extract a valid witness for s_0 (because $R' = \varnothing$, so no witnesses exist). However, we have computational special soundness: by Corollary 7 (Hardness of two values 2), it is computationally infeasible to find those pairs of conversations.

Lemma 22 (Attack on the sigma-protocol, computational): Assume that ℓ_{ch} is logarithmically bounded. Then there exists a total break (Definition 2) against the sigma-protocol from Definition 20. $\diamond$

In this lemma, we use the same malicious prover as in Lemma 18 (Attack on the sigma-protocol). That adversary proves the statement s_0. Since $R' = \varnothing$, that statement is not in the language, thus this prover performs a total break.

Now Theorem 19 follows from Lemmas 21 and 22. (And sigma-protocols with computational special soundness are arguments of knowledge and thus arguments; we are not aware of an explicit write-up in the literature, but the proof from [29] for sigma-protocols with special soundness applies to this case, too.)

VI. Attacking Fiat-Shamir

Definition 23 (Fiat-Shamir): Fix a sigma-protocol $(\ell_{com}, \ell_{ch}, \ell_{resp}, P_1, P_2, V, R)$ and an integer $r > 0$. Let $H : \{0,1\}^* \to \{0,1\}^{r \cdot \ell_{ch}}$ be a random oracle. The *Fiat-Shamir construction* (P_{FS}, V_{FS}) is the following non-interactive proof system:

- Prover $P_{FS}(s, w)$: For $(s, w) \in R$, invoke $com_i \leftarrow P_1(s, w)$ for $i = 1, \ldots, r$. Let $ch_1 \| \ldots \| ch_r := H(s, com_1, \ldots, com_r)$. Invoke $resp_i \leftarrow P_2(ch_i)$. Return $\pi := (com_1, \ldots, com_r, resp_1, \ldots, resp_r)$.
- Verifier $V_{FS}(s, (com_1, \ldots, com_r, resp_1, \ldots, resp_r))$: Let $ch_1 \| \ldots \| ch_r := H(s, com_1, \ldots, com_r)$. Check whether $V(s, com_i, ch_i, resp_i) = 1$ for all $i = 1, \ldots, r$. If so, return 1. $\diamond$

Theorem 24 (Insecurity of Fiat-Shamir): There is an oracle $\mathcal{O}_{all}$ and a relation R and a sigma-protocol relative to $\mathcal{O}_{all}$ with logarithmic ℓ_{ch} (challenge length), completeness, perfect special soundness, computational strict soundness, statistical honest-verifier zero-knowledge, and commitment entropy, such that there is total knowledge break on the Fiat-Shamir construction.

In contrast, the Fiat-Shamir construction based on a sigma-protocol with the same properties is a classical argument of knowledge (assuming that $r\ell_{ch}$ is superlogarithmic). $\diamond$

As the underlying sigma-protocol, we use the one from Definition 16. The attack on Fiat-Shamir is analogous to that on the sigma-protocol itself. The only difference is that the challenge ch now comes from H and not from the verifier; this does not change the attack strategy.

Again, we get even stronger attacks if the special soundness holds only computationally.

Theorem 25 (Insecurity of Fiat-Shamir, computational): There is an oracle $\mathcal{O}_{all}$ and a relation R and a sigma-protocol relative to $\mathcal{O}_{all}$ with logarithmic ℓ_{ch} (challenge length), completeness, *computational* special soundness, computational strict soundness, statistical honest-verifier zero-knowledge, and commitment entropy, such that there is a *total break* on the Fiat-Shamir construction.

In contrast, the Fiat-Shamir construction based on a sigma-protocol with the same properties is a classical argument of knowledge (assuming that $r\ell_{ch}$ is superlogarithmic). $\diamond$

The proof is along the lines of those of Theorem 24 and Lemma 22.

VII. Attacking Fischlin's scheme

In the preceding sections we have used the pick-one trick to give negative results for the (knowledge) soundness of sigma protocols and of the Fiat-Shamir construction. Classically, both protocols are shown sound using rewinding. This leads to the conjecture that the pick-one trick is mainly useful for getting impossibilities for protocols with rewinding-based security proofs. Yet, in this section we show that this is not the case; we use the pick-one trick to give an impossibility result for Fischlin's proof system with online-extractors [15]. The crucial point of that construction is that in the classical security proof, no rewinding is necessary. Instead, a witness is extracted by passively inspecting the list of queries performed by the adversary.

Definition 26 (Fischlin's scheme): Fix a sigma-protocol $(\ell_{com}, \ell_{ch}, \ell_{resp}, P_1, P_2, V, R)$. Fix integers $b, r, \mathbf{S}, t$ such that br and 2^{t-b} are superlogarithmic, b, r, t are logarithmic, $\mathbf{S} \in O(r)$ ($\mathbf{S} = 0$ is permitted), and $b \leq t \leq \ell_{ch}$.

Let $H : \{0,1\}^* \to \{0,1\}^b$ be a random oracle. *Fischlin's construction* (P_{Fis}, V_{Fis}) is the non-interactive proof system is defined as follows:

- $P_{Fis}(s, w)$: See [15]. (Omitted here since we only need to analyze V_{Fis} for our results.)
- $V_{Fis}(s, \pi)$ with $\pi = (com_i, ch_i, resp_i)_{i=1,\ldots,r}$: Check if $V(com_i, ch_i, resp_i) = 0$ for all $i = 1, \ldots, r$. Check if $\sum_{i=1}^{r} H(x, (com_i)_i, i, ch_i, resp_i) \leq \mathbf{S}$ (where $H(\ldots)$ is interpreted as a binary unsigned integer). If all checks succeed, return 1. $\diamond$

The idea (in the classical case) is that, in order to produce triples $(com_i, ch_i, resp_i)$ that make $H(x, (com_i)_i, i, ch_i, resp_i)$ sufficiently small, the prover needs try out several accepting $ch_i, resp_i$ for each com_i. So with overwhelming probability, the queries made to H will contain at least two $ch_i, resp_i$ for the same com_i. This then allows extraction by just inspecting the queries.

In the quantum setting, this approach towards extraction does not work: the "list of random oracle queries" is not a well-defined notion, because the argument of H is not measured when a query is performed. In fact, we show that Fischlin's scheme is in fact not an argument of knowledge in the quantum setting (relative to an oracle):

Theorem 27 (Insecurity of Fischlin's construction): There is an oracle $\mathcal{O}_{all}$ and a relation R and a sigma-protocol relative to $\mathcal{O}_{all}$ with logarithmic ℓ_{ch} (challenge length), completeness, perfect special soundness, computational strict soundness, statistical honest-verifier zero-knowledge, and commitment entropy, such that there is a total knowledge break of Fischlin's construction.

Yet, Fischlin's construction based on a sigma-protocol with the same properties is a classical argument of knowledge. $\diamond$

As the underlying sigma-protocol, we use the one from Definition 16. The basic idea is that the malicious prover finds conversations $(com_i^*, ch_i, resp_i^*)$ by first fixing the values com_i^*, and then using Theorem 5 to find $ch, resp^*$ where $resp_i^*$ contains $resp_i$ such that $(ch_i, resp_i) \in S_{com_i}$ and $H(x, (com_i^*)_i, i, ch_i, resp_i^*) = 0$. If $resp_i^*$ would not additionally contain commitments c_{ch} (see Definition 16), this would already suffice to break Fischlin's scheme. To additionally make sure we can open the commitments to the right value, we use a specific fixpoint property of COM.

Theorem 28 (Insecurity of Fischlin, computational): There is an oracle $\mathcal{O}_{all}$ and a relation R and a sigma-protocol relative to $\mathcal{O}_{all}$ with logarithmic ℓ_{ch} (challenge length), completeness, computational special soundness, computational strict soundness, statistical honest-verifier zero-knowledge, and commitment entropy, such that there is a *total break* on Fischlin's construction.

Yet, Fischlin's construction based on a sigma-protocol with the same properties is a classical argument of knowledge. $\diamond$

Acknowledgments. We thank Marc Fischlin and Tommaso Gagliardoni for valuable discussions and the initial motivation for this work. Andris Ambainis was supported by FP7 FET project QALGO and ERC Advanced Grant MQC (at the University of Latvia) and by National Science Foundation under agreement No. DMS-1128155 (at IAS, Princeton). Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. Ansis Rosmanis was supported by the Mike and Ophelia Lazaridis Fellowship, the David R. Cheriton Graduate Scholarship, and the US ARO. Dominique Unruh was supported by the Estonian ICT program 2011-2015 (3.2.1201.13-0022), the European Union through the European Regional Development Fund through the sub-measure "Supporting the development of R&D of info and communication technology", by the European Social Fund's Doctoral Studies and Internationalisation Programme DoRa, by the Estonian Centre of Excellence in Computer Science, EXCS.

REFERENCES

[1] P. W. Shor, "Algorithms for quantum computation: Discrete logarithms and factoring," in *FOCS 1994*. IEEE, 1994, pp. 124–134.
[2] L. K. Grover, "A fast quantum mechanical algorithm for database search," in *STOC 1996*. IEEE, 1996, pp. 212–219.
[3] G. Brassard, P. Høyer, and A. Tapp, "Quantum algorithm for the collision problem," *ACM SIGACT News*, vol. 28, pp. 14–19, 1997.
[4] D. J. Bernstein, J. Buchmann, and E. Dahmen, Eds., *Post-Quantum Cryptography*. Springer, 2009.
[5] W. K. Wootters and W. H. Zurek, "A single quantum cannot be cloned," *Nature*, vol. 299, pp. 802–803, 1982.
[6] J. Watrous, "Zero-knowledge against quantum attacks," *SIAM J. Comput.*, vol. 39, no. 1, pp. 25–58, 2009.
[7] O. Goldreich, S. Micali, and A. Wigderson, "Proofs that yield nothing but their validity or all languages in NP have zero-knowledge proof systems," *Journal of the ACM*, vol. 38, no. 3, pp. 690–728, 1991.
[8] D. Unruh, "Quantum proofs of knowledge," in *Eurocrypt 2012*, ser. LNCS, vol. 7237. Springer, 2012, pp. 135–152.
[9] D. Boneh, O. Dagdelen, M. Fischlin, A. Lehmann, C. Schaffner, and M. Zhandry, "Random oracles in a quantum world," in *Asiacrypt 2011*. Springer, 2011, pp. 41–69.
[10] M. Zhandry, "Secure identity-based encryption in the quantum random oracle model," in *Crypto 2012*, ser. LNCS, vol. 7417. Springer, 2012, pp. 758–775.
[11] D. Unruh, "Revocable quantum timed-release encryption," in *Eurocrypt 2014*, ser. LNCS, vol. 8441. Springer, 2014, pp. 129–146.
[12] D. Boneh and M. Zhandry, "Secure signatures and chosen ciphertext security in a quantum computing world," in *Crypto 2013*, ser. LNCS. Springer, 2013, pp. 361–379.
[13] D. Unruh, "Quantum position verification in the random oracle model," in *Crypto 2014*, ser. LNCS. Springer, 2014, to appear.
[14] A. Fiat and A. Shamir, "How to prove yourself: Practical solutions to identification and signature problems," in *Crypto '86*, ser. LNCS, no. 263. Springer, 1987, pp. 186–194.
[15] M. Fischlin, "Communication-efficient non-interactive proofs of knowledge with online extractors," in *Crypto 2005*, ser. LNCS, vol. 3621. Springer, 2005, pp. 152–168.
[16] J. van de Graaf, "Towards a formal definition of security for quantum protocols," Ph.D. dissertation, Université de Montréal, 1998.
[17] M. Zhandry, "How to construct quantum random functions," in *FOCS 2013*. IEEE Computer Society, 2012, pp. 679–687.
[18] I. Damgård, J. Funder, J. Buus Nielsen, and L. Salvail, "Superposition attacks on cryptographic protocols," in *ICITS 2013*, ser. LNCS, vol. 8317. Springer, 2014, pp. 142–161.
[19] D. Boneh and M. Zhandry, "Quantum-secure message authentication codes," in *Eurocrypt 2013*, ser. LNCS, vol. 7881. Springer, 2013, pp. 592–608.
[20] Ö. Dagdelen, M. Fischlin, and T. Gagliardoni, "The Fiat-Shamir transformation in a quantum world," in *Asiacrypt 2013*, ser. LNCS, vol. 8270. Springer, 2013, pp. 62–81.
[21] D. Unruh, "Non-interactive zero-knowledge proofs in the quantum random oracle model," IACR ePrint 2014/587, 2014.
[22] P. Dumais, D. Mayers, and L. Salvail, "Perfectly concealing quantum bit commitment from any quantum one-way permutation," in *Eurocrypt '00*, ser. LNCS, vol. 1807. Springer, 2000, pp. 300–315.
[23] A. Ambainis, "A new quantum lower bound method, with an application to a strong direct product theorem for quantum search," *Theory of Computing*, vol. 6, no. 1, pp. 1–25, 2010.
[24] A. Ambainis, L. Magnin, M. Roetteler, and J. Roland, "Symmetry-assisted adversaries for quantum state generation," in *IEEE Conference on Computational Complexity*. IEEE, 2011, pp. 167–177.
[25] B. Barak, "How to go beyond the black-box simulation barrier," in *FOCS 2001*. IEEE, 2001, pp. 106–115.
[26] S. Aaronson and P. Christiano, "Quantum money from hidden subspaces," in *STOC '12*. ACM, 2012, pp. 41–60.
[27] A. Ambainis, A. Rosmanis, and D. Unruh, "Quantum attacks on classical proof systems – the hardness of quantum rewinding," arXiv:1404.6898 [quant-ph], 2014, full version of this paper.
[28] M. Boyer, G. Brassard, P. Høyer, and A. Tapp, "Tight bounds on quantum searching," *Fortschritte der Physik*, vol. 46, no. 4-5, pp. 493–505, 1998.
[29] I. Damgård, "On σ-protocols," Course notes for "Cryptologic Protocol Theory", http://www.cs.au.dk/~ivan/Sigma.pdf, 2010, retrieved 2014-03-17. Archived at http://www.webcitation.org/6O9USFecZ.