Unprovable Security of Two-Message Zero Knowledge

Kai-Min Chung[*] Edward Lui[*]
Mohammad Mahmoody[*] Rafael Pass[*]

December 19, 2012

Abstract

Goldreich and Oren (JoC'94) show that only trivial languages have 2-message zero-knowledge arguments. In this note we consider weaker, *super-polynomial-time* simulation (SPS), notions of zero-knowledge. We present barriers to using black-box reductions for demonstrating soundness of 2-message protocols with efficient prover strategies satisfying SPS zero-knowledge. More precisely, we show that assuming the existence of poly($T(n)$)-hard one-way functions, the following holds:

- For sub-exponential (or smaller) $T(\cdot)$, *polynomial-time* black-box reductions cannot be used to prove soundness of 2-message $T(\cdot)$-simulatable arguments based on any polynomial-time intractability assumption. This matches known 2-message quasi-polynomial-time simulatable arguments using a quasi-polynomial-time reduction (Pass'03), and 2-message exponential-time simulatable proofs using a polynomial-time reduction (Dwork-Naor'00, Pass'03).
- poly($T(\cdot)$)-time black-box reductions cannot be used to prove soundness of 2-message *strong* $T(\cdot)$-simulatable (efficient prover) arguments based on any poly($T(\cdot)$)-time intractability assumption; strong $T(\cdot)$-simulatability means that the output of the simulator is indistinguishable also for poly($T(\cdot)$)-size circuits. This matches known 3-message strong quasi-polynomial-time simulatable proofs (Blum'86, Canetti et al' 00).

[*] Cornell University. {chung,luied,mohammad,rafael}@cs.cornell.edu
Chung is supported in part by a Simon's Foundation postdoctoral fellowship.
Pass is supported in part by a Alfred P. Sloan Fellowship, Microsoft New Faculty Fellowship, NSF Award CNS-1217821, NSF CAREER Award CCF-0746990, NSF Award CCF-1214844, AFOSR YIP Award FA9550-10-1-0093, and DARPA and AFRL under contract FA8750-11-2- 0211. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the US Government.

1 Introduction

The notion of *zero-knowledge*, and the *simulation-paradigm* used to define it, is of fundamental importance in modern cryptography—most definitions of protocol security rely on it. In a zero-knowledge protocol, a prover P can convince a verifier V of the validity of some mathematical statement $x \in L$, while revealing "zero (additional) knowledge" to V. This zero-knowledge property is formalized by requiring that for every potentially malicious efficient verifier V^*, there exists an efficient simulator S that, without talking to P, is able to "indistinguishably reconstruct" the view of V^* in a true interaction with P. Namely, the output of S cannot be distinguished (with more than negligible probability) from the true view of V^* by any efficient distinguisher D.

Assuming standard cryptographic hardness assumptions, 3-message zero-knowledge proofs with constant soundness [Blu86], 4-message zero-knowledge *arguments* (where the soundness is guaranteed to hold only against *efficient* provers) with negligible soundness [FS90], and 5-message zero-knowledge proofs with negligible soundness [GK96] are known for all languages in $\mathcal{NP}$; additionally these interactive proofs/arguments have efficient prover strategies. On the other hand, by the results of Goldreich and Oren [GO94], 2-message zero-knowledge arguments only exist for languages in $\mathcal{BPP}$. In the rest of this note, we focus on interactive proofs/arguments with negligible soundness error and efficient prover strategies

Super-Polynomial-Simulation (SPS) Zero-Knowledge. The usual notion of zero-knowledge requires the simulator to be efficient (i.e., it runs in polynomial time). However, the notion of *super-polynomial-simulation (SPS) zero-knowledge* [Pas03] allows the simulator to run in super-polynomial time. More specifically, the notion of *SPS zero-knowledge* is defined similarly to zero-knowledge except that the simulator is allowed to run in super-polynomial time $T(\cdot)$; such protocols are referred to as *$T(\cdot)$-simulatable.* [Pas03] also defined the (stronger) notion of *strong* SPS zero-knowledge with the additional requirement that any $\mathrm{poly}(T(\cdot))$-time distinguisher cannot distinguish the simulated transcript from a true transcript with better than $\mathrm{negl}(T(\cdot))$ advantage; such protocols are referred to as *strong $T(\cdot)$-simulatable.*

It is known that under *sub-exponential* hardness assumptions 2-message quasi-polynomial-time (i.e., $T(n) = n^{\mathrm{poly}\log n}$) simulatable *arguments* for $\mathcal{NP}$ exist, but 2-message $T(\cdot)$-simulatable *proofs* only exist for languages in $\mathcal{BPTIME}(\mathrm{poly}(T(\cdot)))$ [Pas03]. On the other hand, for 3-message protocols, strong quasi-polynomial-time simulatable proofs for $\mathcal{NP}$ exist [Blu86, CGGM00] (based on sub-exponential hardness assumptions).

This leaves open the following questions regarding 2-message SPS zero-knowledge:

1. *Do 2-message SPS zero-knowledge arguments for $\mathcal{NP}$ exist based on standard polynomial-time hardness assumptions?*
2. *Do 2-message strong SPS zero-knowledge arguments for $\mathcal{NP}$ exist (even under sub-exponential hardness assumptions)?*

In this note, we present barriers to using black-box reductions for providing affirmative answers to the above two questions. In particular, we show the following:

Theorem 1 (Informally Stated)**.** *Assuming the existence of* $\mathrm{poly}(T(n))$*-hard one-way functions, the following holds:*

1. *For sub-exponential (or smaller) $T(\cdot)$, polynomial-time black-box reductions cannot be used to prove soundness of 2-message $T(\cdot)$-simulatable (efficient prover) arguments based on any intractability assumption that can be modeled as a security game with a polynomial-time challenger.*

2. poly$(T(\cdot))$*-time black-box reductions cannot be used to prove soundness of 2-message* strong $T(\cdot)$*-simulatable (efficient prover) arguments based on any intractability assumption that can be modeled as a security game with a poly$(T(\cdot))$-time challenger.*

The first part of our theorem matches known 2-message quasi-polynomial-time simulatable arguments using a quasi-polynomial-time reduction [Pas03], and 2-message exponential-time simulatable proofs using a polynomial-time reduction [DN00, Pas03]. The second part of our theorem matches (in terms of the round-complexity) the 3-message strong quasi-polynomial-time simulatable proofs of [Blu86, CGGM00].

On the Fiat-Shamir Heuristic (added on December 19th, 2012). We were recently made aware of two e-print reports [DSJKLA12, BGW12] (independent of our work) demonstrating barriers to provable security of the Fiat-Shamir heuristic when applied to *proof* systems. Let us briefly point out that a direct corollary of our Theorem 1 yields an even stronger provability barrier.[1] As we mentioned above, [CGGM00] shows (assuming one-way permutations with subexponential hardness), the existence of a 3-message strong quasi-polynomial-time simulatable proof (with negligible soundness error); additionally, this protocol is public coin. Assuming the soundness of the Fiat-Shamir heuristic (when applied only to proof systems), this 3-message proof system can be collapsed to a 2-message strong quasi-polynomial-time simulatable proof system (the "collapsed" protocol is still strongly quasi-polynomial-time simulatable since the hash-function used in the Fiat-Shamir heuristic can just be viewed as a particular malicious verifier. Our Theorem 1 shows that this 2-message proof system can not be proven sound through a black-box reduction to any "standard" assumption.

2 Intractability Assumptions and Black-Box Reductions

Our definition of an intractability assumption closely follows [Pas11]. Following Naor [Nao03] (see also [DOP05, HH09, RV10, GW11]), we model an intractability assumption as an interactive game between a probabilistic machine C—called the challenger—and an attacker A. Both parties get as input 1^n where n is the security parameter. For any $t(n) \in [0,1]$ and any "adversary" A, if $\Pr[\langle A, C\rangle(1^n) = 1] \geq t(n) + p(n)$, then we say that A *breaks* C *with advantage* $p(n)$ over the "threshold" $t(n)$. When this happens, we might also say that A *breaks* $(C, t(\cdot))$ *with advantage* $p(n)$. Any pair $(C, t(\cdot))$ intuitively corresponds to the following assumption:

> **Assumption** $(C, t(\cdot))$**:** *For every polynomial-time adversary A, there exists a negligible function $\nu(\cdot)$ such that for every $n \in \mathbb{N}$, A breaks C with advantage at most $\nu(n)$ over the threshold $t(n)$.*

[1] Our result rules out also *nonuniform security reductions*, as well as reductions that only need to work for deterministic attackers, two techniques that are commonly used in cryptographic proofs. In constrast, as far as we can tell, the results of [DSJKLA12, BGW12] only rule out uniform reductions that need to work for randomized attackers.

If the challenger C of the assumption $(C, t(\cdot))$ is polynomial-time in the security parameter n and the total length of the messages it receives, then we say that the assumption is *efficient challenger*; such assumptions are referred to as *falsifiable* assumptions by Naor [Nao03] and Gentry and Wichs [GW11]. More generally, we refer to an assumption $(C, t(\cdot))$ as having a $T(\cdot, \cdot)$-time (resp. size) challenger if C can be implemented in time (resp. size) $T(n, \ell)$ on input the security parameter 1^n, and when receiving messages of total length ℓ. $(C, t(\cdot))$ is an efficient challenger assumption if and only if $(C, t(\cdot))$ has a $T(\cdot, \cdot)$-time (or size) challenger where $T(n, \ell)$ is polynomial in both n and ℓ. For simplicity, we here consider either poly(n, ℓ)-time (or size) challengers, or $T(n, \ell) = T(n)$-time (or size) challengers, where the running-time of the challenger is bounded only as a function of the security parameter.

Note that we can capture super-polynomial hardness of an assumption by allowing for super-polynomial-time reductions to the assumption.

Black-Box Reductions. We consider probabilistic polynomial-time Turing reductions—i.e., *black-box reductions*. A black-box reduction refers to a probabilistic polynomial-time oracle algorithm. Roughly speaking, a black-box reduction for basing the security of a primitive P on the hardness of an assumption $(C, t(\cdot))$, is a probabilistic polynomial-time oracle machine R such that whenever the oracle O "breaks" P with respect to the security parameter n, then R^O "breaks" $(C, t(\cdot))$ with respect to a polynomially related security parameter n' such that n' can be efficiently computed given n. We restrict ourselves to the case where $n' = n$, since without loss of generality we can always redefine the challenger C so that it acts as if its input was actually n' (since n' can be efficiently computed given n). To formalize this notion, we thus restrict ourselves to oracle machines R that on input 1^n always query the oracle on inputs of the form $(1^n, \cdot)$.

Definition 1. *We say that R is a* valid *black-box reduction if R is an oracle machine such that $R(1^n)$ only queries its oracle with inputs of the form $(1^n, y)$, where $y \in \{0, 1\}^*$.*

The reason to restrict R to only query its oracle on a single "input length" n (which is the case also in all known security reductions in the literature), is that standard cryptographic definitions require ruling out the existence of attackers that break some primitive even for *any* infinite sequence of input lengths; as these input lengths can be very sparse, a black-box reduction might only get to access the adversary over a single "good" input length (and that input length could as well be equal to the length n' over which they win the challenge). Therefore, it must successfully use the adversary even if it has access to an attacker that only succeeds on a single input length.

3 Barriers to Proving Soundness of 2-Message SPS-ZK

We recall the definition of interactive proofs/arguments and SPS-ZK.

Definition 2 (Interactive Proofs and Arguments [GMR89, BCC88])**.** *A pair of probabilistic interactive algorithms (P, V) is said to be an* interactive proof system *for an* $\mathcal{NP}$-*language* L *with* witness relation R_L *if V is probabilistic polynomial-time and the following two conditions hold:*

- *Completeness: There exists a negligible function $\nu(\cdot)$ such that for every $x \in L$ and every $y \in R_L(x)$, it holds that*

$$\Pr[\langle P(y), V\rangle(x) = 1] \geq 1 - \nu(|x|).$$

- *Soundness: For every (computationally unbounded) interactive algorithm P^*, $x \notin L$, and $y \in \{0,1\}^*$, it holds that*
$$\Pr\left[\langle P^*(y), V\rangle(x) = 0\right] \geq 1/2.$$

In case the soundness condition holds only with respect to polynomial-time provers P^*, the pair (P, V) is called an interactive argument *system.*

We now give the definition of $T(\cdot)$-simulatability.

Definition 3 ($T(\cdot)$-Simulatability [Pas03])**.** *Let (P, V) be an interactive proof/argument system for an $\mathcal{NP}$-language L with witness relation R_L. We say that (P, V) is* $T(\cdot)$-simulatable *if for every probabilistic polynomial-time adversary V^*, there exists a $T(\cdot)$-time simulator S such that for every probabilistic polynomial-time distinguisher D, there exists a negligible function $\nu(\cdot)$ such that for every $x \in L$, $y \in R_L(x)$, and $z, z' \in \{0,1\}^*$, it holds that*

$$\left|\Pr\left[D(x, z', \langle P(y), V^*(z)\rangle(x)) = 1\right] - \Pr\left[D(x, z', S(x, z)) = 1\right]\right| \leq \nu(|x|).$$

We now give the definition of strong $T(\cdot)$-simulatability.

Definition 4 (Strong $T(\cdot)$-Simulatability [Pas03])**.** *Let (P, V) be an interactive proof/argument system for an $\mathcal{NP}$-language L with witness relation R_L. We say that (P, V) is strong* $T(\cdot)$-simulatable *if for every probabilistic polynomial-time adversary V^*, there exists a $T(\cdot)$-time simulator S such that for every probabilistic* $\mathrm{poly}(T(\cdot))$*-time distinguisher D, there exists a negligible function $\nu(\cdot)$ such that for every $x \in L$, $y \in R_L(x)$, and $z, z' \in \{0,1\}^*$,*

$$\left|\Pr\left[D(x, z', \langle P(y), V^*(z)\rangle(x)) = 1\right] - \Pr\left[D(x, z', S(x, z)) = 1\right]\right| \leq \nu(T(|x|)).$$

The notions of *SPS zero-knowledge* and *strong SPS zero-knowledge* correspond, respectively, to $T(\cdot)$-simulatability and strong $T(\cdot)$-simulatability for a super-polynomial function $T(\cdot)$. It is shown in [Pas03] that both plain and strong $\mathrm{poly}(T(\cdot))$-simulatability is closed under sequential composition; we will rely on the proof of this result.

Barriers to 2-message SPS-ZK. We aim to prove limitations of basing soundness for 2-message SPS-ZK on intractability assumptions. Let us first explicitly define what it means to break soundness.

Definition 5 (Breaking Soundness)**.** *We say that A* breaks soundness *of (P, V) w.r.t. L with probability $\mu(\cdot)$ if for every $n \in \mathbb{N}$,*

$$\Pr\left[\ (x, z) \leftarrow A(1^n)\ :\ \langle A(1^n, x, z), V(x)\rangle = 1 \wedge x \notin L\ \right] \geq \mu(n).$$

Let us turn to defining what it means to base soundness on an intractability assumption $(C, t(\cdot))$.

Definition 6 (Basing Soundness on the Hardness of $(C, t(\cdot))$)**.** *We say that R is a* black-box reduction for basing soundness of (P, V) w.r.t. L on the hardness of $(C, t(\cdot))$ *if R is a valid black-box reduction and there exists a positive polynomial $p(\cdot, \cdot)$, such that for every deterministic (computationally unbounded) adversary A that breaks soundness of (P, V) w.r.t. L with probability $\mu(\cdot)$, for every $n \in \mathbb{N}$, R^A breaks $(C, t(\cdot))$ with advantage $p(\mu(n), 1/n)$ on input 1^n.*

Note that we here require that R^A breaks the assumption $(C, t(\cdot))$ on the security parameter n by querying A on the *same* security parameter n. As previously mentioned, a seemingly more general definition would allow R^A to break $(C, t(\cdot))$ on a polynomially-related security parameter n' (which can be efficiently computed given n), but this extra generality does not buy us anything as we can always redefine C so that on input n it acts as if its input was n'.

Additionally, note that we only consider *deterministic* attackers; this only makes our result stronger (and will be useful to us, as we shall see later). We now state and prove our results:

Theorem 2. *Suppose one-way functions secure against* $\mathrm{poly}(T(n))$*-size circuits exist. Then, there exists an* $\mathcal{NP}$*-language* L *such that if* (P, V) *is a 2-message* $T(\cdot)$*-simulatable protocol for* L*, and* P *runs in polynomial time (given a witness), then for any efficient-challenger assumption* $(C, t(\cdot))$*, if there exists a probabilistic polynomial-time black-box reduction* R *for basing soundness of* (P, V) *w.r.t.* L *on the hardness of* $(C, t(\cdot))$*, then there exists a probabilistic polynomial-time machine* B *and a positive polynomial* $p'(\cdot)$ *such that for sufficiently* $n \in \mathbb{N}$*,* B *breaks* $(C, t(\cdot))$ *with advantage* $1/p'(n)$ *on input* 1^n*. Furthermore, if* (P, V) *is strong* $T(\cdot)$*-simulatable, then the above holds even if we allow* C *and* R *to run in* $\mathrm{poly}(T(n))$ *time, where in this case our algorithm* B *runs in time* $\mathrm{poly}(T(n))$ *as well.*

Before proving Theorem 2, let us remark that since our lower bound rules out reductions that only need to work for deterministic attackers, by using techniques from [CLMP13] one can directly extend the proof of Theorem 2 to handle *non-uniform* reductions as well. A non-uniform reduction R also gets a function $z(A)$ of the adversary's (perhaps exponential-sized) description as advice before interacting with A; we refer the reader to [CLMP13] for further details.

Proof of Theorem 2. We first prove the theorem for the "plain simulatability" case; we next extend this proof to cover the "strong simulatability" case as well.

By the result of [HILL99], the existence of one-way functions secure against $\mathrm{poly}(T(n))$-size circuits implies the existence of PRGs secure against $\mathrm{poly}(T(n))$-size circuits.[2] Let $g : \{0,1\}^* \to \{0,1\}^*$ be a length-doubling PRG secure against $\mathrm{poly}(T(n))$-size circuits. Consider the language $L = \{g(s) \mid s \in \{0,1\}^*\}$ with witness relation $R_L(x) = \{s \in \{0,1\}^* \mid g(s) = x\}$.

Suppose (P, V) is a 2-message $T(\cdot)$-simulatable protocol for L, and P runs in polynomial time given any witness $w \in R_L(x)$. Suppose further that there exists a polynomial-time black-box reduction R and a polynomial $p(\cdot, \cdot)$ such that R^A breaks the assumption $(C, t(\cdot))$ with advantage $p(\mu(n), 1/n)$ on input 1^n, whenever A is a deterministic (computationally unbounded) adversary that breaks soundness of (P, V) with probability $\mu(\cdot)$. Following the "meta-reduction" paradigm by Boneh and Venkatesan [BV98] (which is also used in [Pas11, GW11, Pas13]), we will use R to directly break $(C, t(\cdot))$ with non-negligible probability. More formally, we exhibit a particular (inefficient) attacker A that breaks soundness of (P, V) with overwhelming probability, and we next show how to "emulate" this attacker for R efficiently without disturbing R's interaction with C.

We first describe our attacker A, and next explain how to emulate it efficiently. More precisely (as in [Pas11]), we define a class of *deterministic* attackers A^f, parametrized by a function $f\colon \{0,1\}^* \to \{0,1\}^\infty$. Given that A^f is deterministic, we may assume without loss of generality that R never asks its oracle the same query twice. Let $S = S(x, z)$ be the $T(\cdot)$-time simulator for the verifier $V^*(x, z) = z$, i.e., V^* sends z to the prover P to get a response a, and then simply

[2] Even though [HILL99] proved their result for $T(n) = \mathrm{poly}(n)$, since it is black-box, it immediately "scales up" to handle larger $T(\cdot)$ as well.

outputs a. On input 1^n, A^f samples $x \leftarrow \{0,1\}^n$ using $f(1^n)$ as randomness, and then outputs x. Next, on input a "first message" q, $A^f(1^n)$ computes $a = S(x,q)$ using $f(1^n,q)$ as randomness, and responds with the message a.

Let $\mathbf{RO} : \{0,1\}^* \to \{0,1\}^\infty$ be a uniformly distributed random oracle. Our first claim is that $A^{\mathbf{RO}}$ breaks soundness of (P,V) with overwhelming probability. First note that except with negligible probability (over the choice of $\mathbf{RO}$), $A^{\mathbf{RO}}$ selects a false statement $x \notin L$. Now, consider an alternative attacker $\widehat{A}^f$ that selects $s \in \{0,1\}^{n/2}$ (again using $f(1^n)$ as the randomness), lets $x = g(s)$, and then proceeds just as A^f does. It follows from the indistinguishability property of the simulator S and the completeness of (P,V) that with overwhelming probability $\widehat{A}^{\mathbf{RO}}$ convinces the honest verifier. Because of this fact and the poly$(T(n))$-indistinguishability of $g(U_{n/2})$ and U_n, it holds that $A^{\mathbf{RO}}$ convinces the honest verifier with overwhelming probability. By the union bound, we thus have that except with negligible probability, $A^{\mathbf{RO}}$ selects a false statement and yet convinces the honest verifier; that is, $A^{\mathbf{RO}}$ breaks soundness of (P,V) with probability $\mu(\cdot) = 1 - \nu(\cdot)$, where $\nu(\cdot)$ is a negligible function.

By an averaging argument, with probability at least $1-10\nu(n)$ over the choice of a random oracle $f \leftarrow \mathbf{RO}$, A^f breaks soundness of (P,V) with probability at least 0.9, and for each such "good" choice of f we have that $R^{A^f}(1^n)$ breaks $(C, t(\cdot))$ with non-negligible advantage $p(0.9, 1/n)$; let $\alpha(n) = p(0.9, 1/n)$. By the union bound, it follows that $R^{A^{\mathbf{RO}}}(1^n)$ breaks $(C, t(\cdot))$ with advantage $\alpha(n)/2$ for sufficiently large n.

We now construct a probabilistic *polynomial-time* attacker $\widetilde{A}$ that emulates $A^{\mathbf{RO}}$. $\widetilde{A}(1^n)$ uniformly samples $s \in \{0,1\}^{n/2}$ and outputs $x = g(s)$; next, on input a first message q, $\widetilde{A}$ runs the honest prover strategy $P(x,s)$ on input the message q and outputs whatever P outputs. We now show the following claim, which concludes the proof of the first part of Theorem 2 by letting $B = R^{\widetilde{A}}$.

Claim 1. *For sufficiently large n, $R^{\widetilde{A}}$ breaks $(C, t(\cdot))$ with advantage at least $\alpha(n)/6$ on common input 1^n.*

Proof. From above, we have that $R^{A^{\mathbf{RO}}}(1^n)$ breaks $(C, t(\cdot))$ with advantage $\alpha(n)/2$ for sufficiently large n. Recall the alternative attacker $\widehat{A}$ defined above. The only difference between $A^{\mathbf{RO}}$ and $\widehat{A}^{\mathbf{RO}}$ is that the former samples a statement from U_n while the latter samples a statement from $g(U_{n/2})$. Since $R(1^n)$ only queries its oracle on the security parameter n, by the poly$(T(n))$-indistinguishability of U_n and $g(U_{n/2})$, it follows that $R^{\widehat{A}^{\mathbf{RO}}}(1^n)$ breaks $(C, t(\cdot))$ with advantage $\alpha(n)/3$ for sufficiently large n.

Now, we note that (since R never asks the same query twice) the only difference between $\widehat{A}^{\mathbf{RO}}$ and $\widetilde{A}$ is that the former uses simulated proofs (of true statements) whereas the latter uses honestly generated proofs. Thus, intuitively, the claim should directly follow by the indistinguishability property of the simulation (and the fact that C and R are polynomial-size). There is a small catch: note that R can query its oracle on several first messages q which is like the execution of a verifier V^* in a sequential composition of (P,V) (on the same fixed statement x). Indeed, by the same argument behind the sequential composition theorem for SPS simulation [Pas03], we will show that indistinguishability still holds. More precisely, let $m(n)$ be an upper-bound on the running-time of R (in this case, $m(n) = \text{poly}(n)$), and define a sequence of $m(n)$ hybrids $H_0, \ldots, H_{m(n)}$ as follows. The hybrid H_i is the output of C when interacting with $R^{(\cdot)}$ where the first i oracle responses (apart from the returned x) are simulated (i.e., answered by $\widehat{A}^{\mathbf{RO}}$), and the remaining queries are answered by running the honest prover strategy (i.e., answered by $\widetilde{A}$). Note that H_0 is the output of C after interacting with $R^{\widetilde{A}}$, and $H_{m(n)}$ is the output of C after interacting with $R^{\widehat{A}^{\mathbf{RO}}}$.

Indistinguishability of any two consecutive hybrids H_i and H_{i+1} follows by the indistinguishability of the simulation and the fact that oracle responses for all $j > i+1$ can be generated in polynomial-time (given the witness to the selected statement). More formally, if the outputs of hybrids H_i and H_{i+1} are $\frac{\alpha(n)}{6m(n)}$-distinguishable, we can always fix the first $i+1$ queries and the first i oracle responses so that the same $\frac{\alpha(n)}{6m(n)}$-distinguishability holds, and then use this fact to distinguish between an honest proof and a simulated proof (i.e., the answers to the $(i+1)^{th}$ query) with advantage $\frac{\alpha(n)}{6m(n)}$ (by answering the subsequent oracle queries efficiently using a hard-wired witness), which contradicts the (non-uniform) indistinguishability of the simulation from the honest proof. Thus, the statistical distance between the output bit of the challenger C in hybrids H_0 and $H_{m(n)}$ is at most $\frac{\alpha(n)}{6}$ for sufficiently large n. Since $R^{\widehat{A}^{\mathbf{RO}}}(1^n)$ breaks $(C, t(\cdot))$ with advantage $\frac{\alpha(n)}{3}$ for sufficiently large n, the claim follows. □

Second part of Theorem 2. We finally note that if (P, V) is *strong* $T(\cdot)$-simulatable, then the very same argument works even if C and R run in time $\mathrm{poly}(T(n))$ (as opposed to $\mathrm{poly}(n)$). The only difference is that now we shall use $m(n) = \mathrm{poly}(T(n))$ hybrids in the proof of Claim 1 (because the reduction R can call its oracle $\mathrm{poly}(T(n))$ times). Now, for every pair of consecutive hybrids H_i and H_{i+1} the distinguishability gap that could be obtained by any $\mathrm{poly}(T(n))$-time distinguisher is at most negligible in $T(n)$ due to the strong $T(\cdot)$-simulatable property. Therefore, the statistical distance between the output of the challenger in hybrids H_0 and $H_{m(n)}$ is at most negligible in $T(n)$ which is indeed at most $\mathrm{negl}(n)$. Therefore the statement of Claim 1 still holds the same as before. □

References

[AGGM06] Adi Akavia, Oded Goldreich, Shafi Goldwasser, and Dana Moshkovitz. On basing one-way functions on NP-hardness. In *STOC '06*, pages 701–710, 2006. 10

[BCC88] Gilles Brassard, David Chaum, and Claude Crépeau. Minimum disclosure proofs of knowledge. *J. Comput. Syst. Sci.*, 37(2):156–189, 1988. 3

[BGW12] Nir Bitansky, Sanjam Garg, and Daniel Wichs. Why fiat-shamir for proofs lacks a proof. *IACR Cryptology ePrint Archive*, 2012, 2012. 2

[Blu86] M. Blum. How to prove a theorem so no one else can claim it. *Proc. of the International Congress of Mathematicians*, pages 1444–1451, 1986. 1, 2

[BMG07] Boaz Barak and Mohammad Mahmoody-Ghidary. Lower bounds on signatures from symmetric primitives. In *IEEE Symposium on Foundations of Computer Science (FOCS)*, 2007. 9

[Bra83] Gilles Brassard. Relativized cryptography. *IEEE Transactions on Information Theory*, 29(6):877–893, 1983. 10

[BT03] Andrej Bogdanov and Luca Trevisan. On worst-case to average-case reductions for np problems. In *FOCS*, pages 308–317, 2003. 10

[BV98] Dan Boneh and Ramarathnam Venkatesan. Breaking rsa may not be equivalent to factoring. In *EUROCRYPT*, pages 59–71, 1998. 5

[CGGM00] Ran Canetti, Oded Goldreich, Shafi Goldwasser, and Silvio Micali. Resettable zero-knowledge (extended abstract). In *STOC '00*, pages 235–244, 2000. 1, 2

[CLMP13] Kai-min Chung, Huijia Lin, Mohammad Mahmoody, and Rafael Pass. On the power of non-uniform proof of security. In *ITCS'13*, 2013. 5

[DN00] Cynthia Dwork and Moni Naor. Zaps and their applications. In *In 41st FOCS*, pages 283–293. IEEE, 2000. 2

[DOP05] Yevgeniy Dodis, Roberto Oliveira, and Krzysztof Pietrzak. On the generic insecurity of the full domain hash. In *CRYPTO*, pages 449–466, 2005. 2

[DSJKLA12] Dana Dachman-Soled, Abhishek Jain, Yael Tauman Kalai, and Adriana Lopez-Alt. On the (in)security of the fiat-shamir paradigm, revisited. *IACR Cryptology ePrint Archive*, 2012, 2012. 2

[FF93] Joan Feigenbaum and Lance Fortnow. Random-self-reducibility of complete sets. *SIAM Journal on Computing*, 22(5):994–1005, 1993. 10

[FS90] Uriel Feige and Adi Shamir. Witness indistinguishable and witness hiding protocols. In *STOC '90*, pages 416–426, 1990. 1

[GK96] Oded Goldreich and Ariel Kahan. How to construct constant-round zero-knowledge proof systems for NP. *Journal of Cryptology*, 9(3):167–190, 1996. 1

[GKM$^+$00] Yael Gertner, Sampath Kannan, Tal Malkin, Omer Reingold, and Mahesh Viswanathan. The Relationship between Public Key Encryption and Oblivious transfer. In *FOCS*, pages 325–335, 2000. 9

[GMR89] Shafi Goldwasser, Silvio Micali, and Charles Rackoff. The knowledge complexity of interactive proof systems. *SIAM Journal on Computing*, 18(1):186–208, 1989. 3

[GO94] Oded Goldreich and Yair Oren. Definitions and properties of zero-knowledge proof systems. *Journal of Cryptology*, 7:1–32, 1994. 1

[GW11] Craig Gentry and Daniel Wichs. Separating succinct non-interactive arguments from all falsifiable assumptions. In *STOC*, pages 99–108, 2011. 2, 3, 5, 10

[HH09] Iftach Haitner and Thomas Holenstein. On the (im)possibility of key dependent encryption. In *TCC*, pages 202–219, 2009. 2

[HHRS07] Iftach Haitner, Jonathan J. Hoch, Omer Reingold, and Gil Segev. Finding collisions in interactive protocols – A tight lower bound on the round complexity of statistically-hiding commitments. In *Proceedings of the 47th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE Computer Society, 2007. 9

[HILL99] Johan Håstad, Russell Impagliazzo, Leonid Levin, and Michael Luby. A pseudorandom generator from any one-way function. *SIAM Journal on Computing*, 28:12–24, 1999. 5

[IR88] Russell Impagliazzo and Steven Rudich. Limits on the provable consequences of one-way permutations. In *CRYPTO '88*, pages 8–26, 1988. 9

[Nao03] Moni Naor. On cryptographic assumptions and challenges. In *CRYPTO*, pages 96–109, 2003. 2, 3

[Pas03] Rafael Pass. Simulation in quasi-polynomial time, and its application to protocol composition. In *EUROCRYPT*, pages 160–176, 2003. 1, 2, 4, 6

[Pas06] Rafael Pass. Parallel repetition of zero-knowledge proofs and the possibility of basing cryptography on np-hardness. In *IEEE Conference on Computational Complexity*, pages 96–110, 2006. 10

[Pas11] Rafael Pass. Limits of provable security from standard assumptions. In *STOC*, pages 109–118, 2011. 2, 5, 10

[Pas13] Rafael Pass. Unprovable security of statistical nizk and non-interactive non-malleable commitments. In *TCC*, 2013. 5, 10

[PTV11] Rafael Pass, Wei-Lung Dustin Tseng, and Muthuramakrishnan Venkitasubramaniam. Towards non-black-box lower bounds in cryptography. In *TCC*, pages 579–596, 2011. 10

[RTV04] Omer Reingold, Luca Trevisan, and Salil P. Vadhan. Notions of reducibility between cryptographic primitives. In *TCC*, pages 1–20, 2004. 9

[RV10] Guy N. Rothblum and Salil P. Vadhan. Are pcps inherent in efficient arguments? *Computational Complexity*, 19(2):265–304, 2010. 2

[Sim98] Daniel R. Simon. Finding Collisions on a One-Way Street: Can Secure Hash Functions Be Based on General Assumptions? In *EUROCRYPT*, pages 334–345, 1998. 9

A Related Separation Results

There is a large literature on separation results between cryptographic primitives/assumptions. We distinguish between two types of results:

Separations for fully black-box constructions. The seminal work of Impagliazzo and Rudich [IR88] provides a framework for proving black-box separations between cryptographic primitives. We highlight that this framework refutes the possibility of so-called "fully-black-box constructions" (see [RTV04] for a taxonomy of various black-box separations); that is, this framework considers both black-box *constructions* (i.e., the higher-level primitive only uses the underlying primitive as a black-box), and black-box *proofs of security* (i.e., the security reduction only uses the adversary against the constructed scheme as a black-box). Most black-box separation results fall into this framework (e.g., [Sim98, GKM$^+$00, BMG07, HHRS07] to name a few). As it was shown by [RTV04], some of these separations extend to the setting where the security reduction is "semi" or even "weakly" black-box, but we emphasize that the construction is always black-box.

Separations for black-box security reductions. In recent years, new types of separations between cryptographic primitives/assumptions have emerged. These separations apply even to non-black-box constructions as long as the proof of security is black-box: Pass [Pas06] and Pass, Tseng and Venkitasubramaniam [PTV11] demonstrate that under certain (new) complexity theoretic assumptions, various cryptographic tasks cannot be based on *one-way functions* using a black-box security reduction, even if the protocol uses the one-way function in a non-black-box way. (These results follow techniques used by Brassard [Bra83] and Akavia et al [AGGM06] to demonstrate limitations of "NP-hard cryptography".)[3]

Recently, two independent works demonstrate similar types of separation results, but this time ruling out security reductions to a *general* set of intractability assumptions: Pass [Pas11] demonstrates impossibility of using black-box reductions to prove the security of several primitives (e.g., Schnorr's identification scheme, commitment schemes secure under weak notions of selective opening, Chaum blind signatures, etc.) based on any "bounded-round" intractability assumption (where the challenger uses an a-priori bounded number of messages, but is otherwise unbounded). Gentry and Wichs [GW11] (assuming the existence of strong pseudorandom generators) demonstrate impossibility of using black-box security reductions to prove soundness of "succinct non-interactive arguments" based on any falsifiable assumption (where the challenger is computationally bounded). An even more recent work by Pass [Pas13], developed in parallel with the current note, rules out constructions of statistical NIZK with adaptive soundness and non-interactive non-malleable commitments, based on falsifiable assumptions.

Our results in this work fall into this second category of results and rule out black-box security reductions for proving the soundness of various forms of SPS zero-knowledge protocols even if the construction is arbitrarily non-black-box.

[3] See also the results of Feigenbaum and Fortnow [FF93] and the result of Bogdanov and Trevisan [BT03] that demonstrate limitations of NP-hard cryptography for *restricted* types of reductions.