— A merged version of this work and the work of [DJKL12] appears in the proceedings of the *Theory of Cryptography Conference* - TCC 2013 —

Why "Fiat-Shamir for Proofs" Lacks a Proof

Nir Bitansky* Tel Aviv University

Sanjam Garg† UCLA

Daniel Wichs IBM Research, T.J. Watson

December 19, 2012

Abstract

The Fiat-Shamir heuristic (CRYPTO '86) is used to convert any 3-message public-coin proof or argument system into a non-interactive argument, by hashing the prover's first message to select the verifier's challenge. It is known that this heuristic is sound when the hash function is modeled as a random oracle. On the other hand, the surprising result of Goldwasser and Kalai (FOCS '03) shows that there exists a computationally sound *argument* on which the Fiat-Shamir heuristic is *never* sound, when instantiated with any *actual* efficient hash function.

This leaves us with the following interesting possibility: perhaps there exists a hash function that securely instantiates the Fiat-Shamir heuristic for all 3-message public-coin *statistically sound proofs*, even if it can fail for some computationally sound arguments. Indeed, the existence of such hash functions has been conjectured by Barak, Lindell and Vadhan (FOCS '03), who also gave a seemingly reasonable and sufficient condition under which such hash functions exist. However, we do not have any provably secure construction of such hash functions, under any standard assumption such as the hardness of DDH, RSA, QR, LWE, etc.

In this work we give a broad black-box separation result, showing that the security of such hash functions *cannot* be proved under virtually any standard cryptographic assumption via a *black-box reduction*.

*Research was done while visiting IBM T.J., Watson Research Center. Supported by the Check Point Institute for Information Security, an ISF grant 20006317, and the Fulbright program.

†Research conducted while at the IBM Research, T.J.Watson funded by NSF Grant No.1017660.

1 Introduction

The Fiat-Shamir heuristic. The Fiat-Shamir (FS) heuristic [FS87] allows us to convert an interactive *public-coin* protocol between a *prover* P and a *verifier* V into a one-message (non-interactive) protocol. Recall that, in a public-coin protocol, the verifier sends a uniformly random *challenge* to the prover in each round. According to the FS heuristic, the prover executes the original interactive protocol "in his head", computing the verifier's challenge in each round by applying some *public hash function* to the transcript of the protocol so far. The prover then only sends the final protocol transcript to the actual verifier, who verifies its validity. The hash function can be initialized with some randomly chosen public seed, which we think of as a "common random string (CRS)", and therefore the compiled protocol is non-interactive in the CRS model. Alternatively, the seed can also be chosen by the verifier in an additional initial message, in which case the compiled protocol consists of two messages. This heuristic has numerous remarkable applications in cryptography, such as constructing practical *signature schemes* [Sch91, GQ90, Oka93], *non-interactive zero knowledge (NIZK)* [BR93], and non-interactive succinct arguments [Mic00].

Soundness of FS. Although the FS heuristic seems to produce secure cryptographic schemes in practice, its formal security properties remain elusive. Perhaps the most basic question is to understand the *soundness* of the heuristic when applied to a statistically sound *proof* or computationally sound *argument* for some NP language. We say that an instance of the FS-heuristic is sound if the resulting non-interactive protocol is a computationally sound argument, for the same language. We can ask what kind of protocols do we need to start with, and what kind of hash functions should we use, to make the FS-heuristic sound. Since we are interested in a negative result, we restrict our attention to *3-message public-coin (3PC)* protocols.

Applying FS to arguments. On the positive side, if the FS heuristic uses a *random oracle* as its hash function, then it is known to be sound when applied to *any* 3PC argument [BR93, PS00]. On the other hand, the work of Goldwasser and Kalai [GK03] shows a surprising negative result: the FS heuristic *cannot* be securely instantiated with any *actual* efficient hash function that would achieve the same result. In particular, there exists *some* 3PC argument on which the FS heuristic is *never* sound, no matter which efficient hash function we try to instantiate it with.

Applying FS to proofs. The above negative result only applies to computationally sound arguments, and therefore we are still left with the following interesting possibility: perhaps the FS heuristic could be instantiated with some hash function that makes it sound for *all* 3PC *proofs*, even if it can fail for some arguments. We call such a hash function *FS-universal*. When instantiated with an FS-universal hash function, the FS heuristic should successfully compile any 3PC (statistically sound) proof into a non-interactive (computationally sound) argument.

Barak, Lindell, and Vadhan [BLV03] conjecture that such FS-universal hash functions should indeed exist, and define a plausible hash-function property called *entropy-preservation*, which they show to be sufficient. Variants of this entropy-preservation property were further studied by Dodis, Ristenpart and Vadhan [DRV12], who also showed that it is *necessary*. Nevertheless, despite the amazing possibility that such hash functions may exist, we do not have any candidate construction that is provably secure under some "standard" cryptographic hardness assumption.

Relating soundness to ZK. Dwork et al. [DNRS99] show an interesting connection between the soundness of the FS heuristic and zero-knowledge: if the starting protocol is a zero-knowledge proof or argument, for a non-trivial language, then applying the FS heuristic to this protocol is *never*

sound, no matter which hash function is used. The reason is intuitive: we can always just run the original zero-knowledge simulator for the "Fiat-Shamir verifier" that chooses its challenges by hashing the protocol transcript so far. Since this efficient simulator cannot decided the language, it is likely to output an accepting transcript even for some false statement, and therefore breaks the soundness of the FS heuristic.

The celebrated work of Barak [Bar01] constructs a constant-round public-coin zero-knowledge argument; in particular, the FS heuristic is never sound on this argument. The result of [GK03] cleverly extends this technique to also show the existence of a 3PC argument for which the FS heuristic always fails. On the other hand, we do not know of any 3PC (or even constant-round) zero-knowledge *proofs* for a hard language, and hence we do not have any such negative results for proofs. Indeed, if true, the prevalent conjecture that there exist FS-universal hash functions implies that 3PC proofs can *never* be zero-knowledge.

An alternative to FS. We mention that the work of Kalai and Raz [KR09] offers a provably secure method for converting any interactive proof into a *two-message* argument using private information retrieval (PIR). We can think of this as an alternative to the Fiat-Shamir heuristic. However, the FS heuristic has a crucial advantage in that it gives us a publicly-verifiable *non-interactive* argument in the "common random string model", where the random string is the seed of the hash function. In other words, once the seed of the hash function is fixed, anybody can non-interactively compute and verify proofs (which is needed in central applications, such as digital signatures). This is not the case for the Kalai-Raz compiler, which gives a two-message protocol where the verifier needs to keep some secret state associated with the first message (private coins) to later decide if the prover's second message is accepting.

1.1 Our Results

In this work, we re-examine the possibility of having an *FS-universal* hash function, which securely instantiates the FS heuristic for all 3PC statistically sound proofs. As our main result, we show that the existence of such FS-universal hash functions cannot be proved under virtually any *standard assumption* via a *black-box reduction*. We elaborate on these concepts in the next two paragraphs. We wish to emphasize that this result does *not* refute the highly believable conjecture that FS-universal hash functions exist. However, it shows that we will need to rely on new "non-standard" assumptions or develop new "non-black box" proof techniques if we ever hope to prove this conjecture.

Cryptographic-game assumptions. To capture all "standard assumptions", we consider a general class assumptions that have the syntactic format of an interactive *cryptographic game* between an attacker and a (possibly inefficient) challenger. For a given game, the associated assumption states that every efficient attacker has at most negligible probability in winning this game. This notion, due to [DOP05, HH09], captures, essentially, all of the concrete assumptions we use in cryptography, such as the hardness of factoring, the RSA problem, the discrete logarithm problem, the computational/decissional Diffie-Hellman problem (CDH/DDH), learning with errors (LWE), etc. Note that this notion of cryptographic games refers to *concrete* assumptions (e.g., the RSA function is one-way) rather than just *general* assumptions (e.g., one-way function exist). Of course, this also means that we get a black-box separation from the corresponding general assumption, such as existence of one-way functions, collision-resistant hashing, trapdoor permutations, oblivious transfer, fully homomorphic encryption etc., as long as some concrete instantiation of the assumption exists.

We stress that the notion of cryptographic games is defined as liberally as possible so as to include essentially everything that could be considered a "standard assumption", and to make our negative result as strong as possible. Of course, it may also capture many non-standard (and false) assumptions, as well as trivially true and uninteresting assumptions.

Black-box reduction for FS. Let's say that we want to prove the FS-universal security of some hash-function family $\mathcal{H}$, under some cryptographic-game assumption. We let $\mathcal{FS}_{\mathcal{H}}^{\langle P,V\rangle}$ denote the non-interactive protocol that we get by applying the FS-heuristic with the hash function $\mathcal{H}$ to some proof system $\langle P, V\rangle$. Then, the hash function $\mathcal{H}$ *fails* to be FS-universal if there exists some 3PC proof system $\langle P, V\rangle$, and an efficient attacker $\mathcal{A}$, that breaks the computational soundness of the non-interactive protocol $\mathcal{FS}_{\mathcal{H}}^{\langle P,V\rangle}$. Therefore, proving the FS-universal security of $\mathcal{H}$ is equivalent to showing that the existence of any such triple $(P, V, \mathcal{A})$ implies the existence of an efficient attack against the assumption. A natural and constructive approach for proving this would be via a *black-box reduction*. This is an efficient algorithm $\mathcal{B}^{P,V,\mathcal{A}}$ such that, given black-box access to *any* (possibly inefficient) oracles $(P, V, \mathcal{A})$ satisfying the above conditions, the reduction $\mathcal{B}^{P,V,\mathcal{A}}$ manages to break the assumption. Notice that we are restricting ourselves to reductions which treat the attacker $\mathcal{A}$ *as well as* the 3PC proof system $\langle P, V\rangle$ as a black box. This is a natural notion of a "black-box reduction" for FS-universal hash functions, since the definition of FS-universality treats all of the components $P, V, \mathcal{A}$ as "worst-case" and therefore we treat them as adversarial objects. We note that, for example, the result of Kalai and Raz [KR09] offering a provably secure alternative to FS, *does* use a black-box reduction of the type outlined above in its proof of security.

FS-universal vs. FS for specific proofs. We note that our black-box separation result does *not* rule out the possibility of hash functions that provably (under standard assumptions via a black-box reduction) instantiate the FS heuristic for some *specific* 3PC proof. In fact, the *assumption* that a hash function makes the FS heuristic sound for a *particular* 3PC proof, *is* already a cryptographic-game assumption according to our definition.[1] Our result applies to proving *FS-universal* security, which is a stronger condition that requires the hash function to securely instantiate the FS heuristic for *all* 3PC proofs. The assumption that some hash function is FS-universal no longer has the format of a cryptographic game assumption, making our result possible.

1.2 Our Techniques

We now give a high-level overview of the techniques that we use to prove the above separation. We first recall the notion of *entropy-preserving* hashing introduced by Barak, Lindell and Vadhan [BLV03], which is tightly connected to FS-universality.

Entropy-preserving hashing. We can think of entropy-preserving hashing as a simple and compelling but "non-standard" assumption under which FS-universal hash functions exist. A hash function $h_s(\cdot)$ with a public seed s is entropy-preserving if for any efficient attacker $\mathcal{A}$ that sees a uniformly random seed s and adaptively chooses some correlated value $x = \mathcal{A}(s)$, the conditional Shannon entropy $\mathbf{H}(h_s(x) \mid x)$ is sufficiently large. In other words, given the adversarial choice of $x = \mathcal{A}(s)$, but no other information about s, the value $y = h_s(x)$ should *not* be uniquely determined. This property only makes sense when the seed size $|s|$ is bigger than the input size $|x|$ so that the value $x = \mathcal{A}(s)$ loses information about the seed s. Notice that this property is defined in terms

[1] Indeed, the notion of a cryptographic game will allow the challenger to be inefficient and, in particular, identify that indeed $x \notin \mathcal{L}$, even for a hard language $\mathcal{L}$.

of entropy and does not have the usual format of a cryptographic-game assumption. The works of [BLV03, DRV12] show that any entropy-preserving hash function (with appropriate parameters) is also FS-universal. The work of [DRV12] also shows the reverse direction, that any FS-universal hash function must also be entropy-preserving (at least in some weak sense).

Separation for entropy preserving hashing. As our starting point, we give a black-box separation showing that one cannot prove the *entropy-preserving* security of any hash-function family $\mathcal{H} = \{h_s(\cdot)\}$ from any cryptographic-game assumption via a black-box reduction. We prove this via the "*simulatable attacker*" paradigm (also known as the "meta-reduction" paradigm) which has been used in several prior black-box separation results [BV98, Cor02, Bro05, PV05, GBL08, DOP05, HH09, GW11, Pas11, Seu12, DHT12, Wic12].

The main idea of this paradigm is to construct a special *inefficient* attacker $\mathcal{A}$ that breaks the security of the target primitive (in our case, the entropy-preserving security of $\mathcal{H}$), but for which there is an *efficient* simulator Sim such that no distinguisher can tell the difference between "black-box" interaction with Sim and $\mathcal{A}$. This means that any efficient black-box reduction which can win some cryptographic game, given oracle access to the inefficient attacker $\mathcal{A}$, can also win the cryptographic game, given oracle access to the efficient simulator Sim. Hence, if we have a black-box reduction showing the entropy-preserving security of $\mathcal{H}$ under some cryptographic-game assumption, it implies that the reduction, together with the efficient simulator Sim, give us an efficient stand-alone attack against the assumption, and so it cannot be secure to begin with!

We show that, for any hash function family $\mathcal{H}$, there is a simulatable attack against $\mathcal{H}$. Our inefficient attacker $\mathcal{A}$ breaks the entropy-preserving security of $\mathcal{H}$ by outputting values $x = \mathcal{A}(s)$ in a very careful manner so that $h_s(x)$ is uniquely determined by x (without any additional knowledge of s). However, we show that getting polynomially many queries to $\mathcal{A}$ on various inputs s, can be (statistically) simulated by an efficient simulator Sim that just outputs uniformly random values! Of course, this naive simulator is completely innocuous and unlikely to break the entropy-preserving security of $\mathcal{H}$, but nobody can tell the difference. Showing the above is the main technical result of the paper.

Separation for FS-universal hashing. Next, we show that any black-box reduction proving the FS-universal security of some hash function $\mathcal{H}$ under a cryptographic game assumption, would also give us a black-box reduction showing the entropy-preserving security of $\mathcal{H}$ under the same assumption. Therefore, our separation for entropy-preserving hash functions also gives us a black-box separation for FS-universal hash functions. Here, we adapt the result of Dodis, Ristenpart and Vadhan [DRV12], showing that FS-universal hash functions must also be entropy-preserving, to the setting of black-box reductions.

1.3 Concurrent Work

In a concurrent and independent work, Dachman-Soled, Jain, Kalai and López-Alt [DJKL12] show a very similar result to ours, namely that the soundness of the Fiat-Shamir heuristic cannot be proved via black-box reductions from a large class of "standard" assumptions. Interestingly, the two works rely on very different techniques and the the end results are technically incomparable. We now provide a brief high-level overview of the main differences between the two results. A merged version of the two works [BDG+13] will appear at TCC 2013.

Our work focuses on first giving a separation for *entropy-preserving hash functions*, and then shows that it implies a similar separation for FS-universal hash functions. On the other hand,

[DJKL12] take a more direct approach by showing a separation whenever applying the FS heuristic to any zero-knowledge (ZK) protocol with a super-polynomial-time simulator (cleverly extending a prior-known result that the FS heuristic is simply insecure when applied to any ZK protocol with a poly-time simulator). By showing that every honest-verifier zero-knowledge (HVZK) proof is also ZK with a super-polynomial simulator, [DJKL12] gets a black-box separation for *any* such 3PC proof. The different techniques lead to two main differences in the end results. Our negative result only applies to a *universal* Fiat-Shamir compiler that must preserve soundness when applied to *any* 3PC proof. In contrast, the negative result of [DJKL12] is stronger in that it even applies to having a Fiat-Shamir compiler for many *specific* 3PC proof systems. On the other hand, we show a separation from a somewhat larger class of assumptions, consisting of "cryptographic games" with a possibly unbounded challenger. In contrast, [DJKL12] show a separation from a smaller class of "falsifiable assumptions", where the challenger is required to be efficient.

1.4 Organization

In Section 2, we define the basic concepts discussed in this paper, including FS-universality and entropy-preserving hashing. In Section 3, we define the concepts of cryptographic games, black-box reductions, and simulatable attacks. In Section 4 we show that the entropy-preserving security of any hash function cannot be proved under a black-box reduction from any cryptographic game. In Section 5, we relate the concept of a black-box reduction for FS-universality and the concept of a black-box reduction for entropy-preserving hashing, concluding that FS-universality also cannot be proved via black-box reduction from any cryptographic games.

2 Preliminaries and Definitions

Let n denote the security parameter. We say that a function is *negligible* in the security parameter, and denote it by $\mathsf{negl}(n)$, if it is asymptotically smaller than the inverse of any polynomial, i.e. $1/n^{\omega(1)}$. We consider the class of efficient schemes to be ones that can be implemented by a probabilistic polynomial-time Turing machine, denoted by PPT. In contrast, we consider the class of efficient adversaries $\mathcal{A} = \{\mathcal{A}_n\}$ to be non-uniform families of polynomial-size circuits, denoted by $\mathsf{polysize}$.

Interactive proofs. In an interactive proof system [GMR89] a *prover* P interacts with a PPT *verifier* V to convince him of accepting some common input x. Typically, we require that the honest prover is also PPT, when given some additional auxiliary input (e.g., a witness). For common input x and prover auxiliary prover input w, we denote by $\langle P(w), V\rangle(1^n, x)$ the random variable representing the output of V at the end of the protocol.

Definition 2.1 (An interactive proof system)**.** *A pair of* PPT *machines* $\langle P, V\rangle$ *is called an* interactive proof system *for a relation* $\mathcal{R}$ *with an associated language* $\mathcal{L}(\mathcal{R}) \stackrel{\text{def}}{=} \{\ x\ |(x,w) \in \mathcal{R}\}$ *if the following two conditions hold:*

- Completeness: *For every* $(x,w) \in \mathcal{R}$*, we have*

$$\Pr[\langle P(w), V\rangle(1^n, x) = 1] \geq 1 - \mathsf{negl}(n)\ .$$

(We also say that the protocol has perfect completeness if the above probability is exactly 1).

- Statistical Soundness: *For every $x \notin \mathcal{L}(\mathcal{R})$, and every unbounded P^*,*

$$\Pr[\langle P^*, V\rangle(1^n, x) = 1] \leq \mathsf{negl}(n) \ .$$

(We also say that the protocol is ε-sound if the above probability is upper bounded by $\varepsilon(n)$.)

The proof system is said to be public-coin *if all the messages sent by the verifier are random strings, and the verifier's decision is made solely based on the messages exchanged during the protocol.*

The Fiat-Shamir heuristic. Throughout the paper, we focus on the special case of applying the FS heuristic to a *3-message public-coin (3PC)* interactive proof system $\langle P, V\rangle$ for an NP relation $\mathcal{R}$.[2] Denote the first message of the prover by α, the verifier's challenge by β, and the final message of the prover by γ. Also, let $\pi = (\alpha, \beta, \gamma)$ denote the transcript of the execution.

For security parameter n, let $m(n)$ and $k(n)$ denote the lengths of α and β, respectively. Let $\mathcal{H} = \{h_s : \{0,1\}^{m(n)} \to \{0,1\}^{k(n)}\}_{n\in\mathbb{N}, s\in\{0,1\}^{\ell(n)}}$ be a family of hash functions mapping m bits to k bits. The Fiat-Shamir collapse (or FS-collapse in short) of protocol $\langle P, V\rangle$ using $\mathcal{H}$ is a two-message protocol $\mathcal{FS}_{\mathcal{H}}^{\langle P,V\rangle} = \langle P_{FS}, V_{FS}\rangle$ defined as follows:

- In the first message, the FS verifier $V_{FS}(1^n, x)$ selects a random seed $s \leftarrow \{0,1\}^{\ell(n)}$ for the hash function. (We can also skip this step by thinking of s as a common reference string).
- In the second message, the FS prover $P_{FS}(1^n, x, w)$ runs $P(1^n, x, w)$ to derive its first message α. It then computes the challenge $\beta := h_s(\alpha)$ by hashing α, and passes β to P to get its third message γ. Finally, P_{FS} outputs the tuple (α, β, γ).
- The FS verifier $V_{FS}(1^n, x)$ accepts the proof if $\beta = h_s(\alpha)$ *and* the original verifier $V(1^n, x)$ accepts the protocol (α, β, γ) when executed with random-coins β.

We say that the *FS-collapse is sound* if the resulting protocol $\mathcal{FS}_{\mathcal{H}}^{\langle P,V\rangle}$ is a *computationally-sound argument system* as specified below.

Definition 2.2 (Fiat-Shamir soundness)**.** *We say that $\mathcal{FS}_{\mathcal{H}}^{\langle P,V\rangle}$ is computationally sound if, for any* polysize *prover $P^* = \{P_n^*\}$ and $x \notin \mathcal{L}(\mathcal{R})$*

$$\Pr_{s\xleftarrow{\$}\{0,1\}^{\ell(n)}}\left[V(1^n, x, \pi) = 1 \;\middle|\; \begin{array}{c} \pi \leftarrow P_n^*(x, s) \\ \pi = (\alpha, \beta, \gamma) \\ h_s(\alpha) = \beta \end{array}\right] \leq \mathsf{negl}(n) \ .$$

We call the above probability the advantage *of P^* in breaking computational soundness.*

We could also consider a stronger security definition, according to which P^* *adaptively* chooses the false theorem statement, depending on the hash function h_s. Since we are interested in a negative result, we focus on the above weaker definition.

As explained in the introduction, we will be interested in the existence of *universal Fiat-Shamir hash function* (or a FS-universal hash for short); namely, families of hash functions with respect to which the FS-collapse of any statistically sound 3PC protocol is a computationally sound argument.

[2] Indeed, this is the most common but also minimal case for which Fiat-Shamir is expected to work, and therefore restricting ourselves to this case gives us the strongest negative result.

Definition 2.3 ((m,k)-FS-universal hash function)**.** *We say that a hash-function family*

$$\mathcal{H} = \{h_s : \{0,1\}^{m(n)} \to \{0,1\}^{k(n)}\}_{s\in\{0,1\}^{\ell(n)}}$$

is $(m(n),k(n))$-FS-universal if for every 3PC (statistically sound) proof system $\langle P,V\rangle$ with first and second messages of respective lengths $m = m(n)$ and $k = k(n)$, the FS-collapse $\mathcal{FS}_{\mathcal{H}}^{\langle P,V\rangle}$ is a (computational sound) argument.

It is not hard to show that the seed-length $\ell(n)$ of an FS-universal hash function must depend on (and exceed) the input length m. Therefore, we cannot have a single fixed-seed-length hash function family that is FS-universal for all 3PC proofs. Instead, we only desire a (m,k)-FS-universal hash function which works for all 3-PC proofs whose messages α,β are of lengths $m(n),k(n)$ respectively. In particular, the seed length $\ell(n)$ may be any polynomial, which may arbitrarily depend on (m, k).

Entropy preserving hash functions. Barak et al. [BLV03] formulated a relatively simple entropy preservation property for hash functions, and showed that it is sufficient for FS-universality. Recall that the *(Shannon) entropy* of a random variable $\mathbf{x}$ is $\mathbf{H}(\mathbf{x}) = \mathbf{E}_{x\xleftarrow{\$}\mathbf{x}}[-\log(\Pr[\mathbf{x}=x])]$. For jointly distributed random variables $(\mathbf{x},\mathbf{y})$, the *conditional entropy* of $\mathbf{x}$ given $\mathbf{y}$ is defined by $\mathbf{H}(\mathbf{x}\mid\mathbf{y}) = \mathbf{E}_{y\xleftarrow{\$}\mathbf{y}}[\mathbf{H}(\mathbf{x}\mid\mathbf{y}=y)]$, where $\mathbf{x}|_{\mathbf{y}=y}$ is a random variable distributed according to $\mathbf{x}$ conditioned on $\mathbf{y}=y$.

Definition 2.4 (Definition 9.2 [BLV03])**.** *We say that a hash function family $\mathcal{H} = \{h_s : \{0,1\}^{m(n)} \to \{0,1\}^{k(n)}\}_{s\in\{0,1\}^{\ell(n)}}$* preserves $u(n)$-entropy*, if for any* polysize $\mathcal{A}$*, and all large enough values of the security parameter $n\in\mathbb{N}$ we have*

$$\mathbf{H}(h_{\mathbf{s}}(\mathbf{x})\mid\mathbf{x}) > u(n)\ ,$$

where $\mathbf{s},\mathbf{x}$ are correlated random variables defined by choosing $\mathbf{s}$ uniformly at random over $\{0,1\}^{\ell(n)}$, and setting $\mathbf{x}$ to be the first $m(n)$ bits of the output of $\mathcal{A}(1^n,\mathbf{s})$. We say that the hash function (just plain) preserves entropy *if it preserves $u(n)$-entropy for $u(n)=0$.*

The work of [BLV03] shows that any hash function family that preserves $u(n) = k(n) - O(\log n)$ entropy, is also (m,k)-FS-universal. An alternative take on the notion of "entropy preserving" hash functions and a detailed exploration of the parameters is also given in [DRV12]. The work of [DRV12] also shows an implication in the reverse direction, that any (m,k)-FS-universal hash function family must also preserve entropy, at least in the weak sense for $u(n)=0$. Note that even this weak notion of preserving entropy is already an interesting and highly non-trivial. In Section 4 we show that it cannot be proved under any standard assumptions via black box reductions.

3 Games, Black-Box Separations, and Simulatable Attacks

Cryptographic games. Cryptographic games present a general framework for defining cryptographic assumptions and security properties. A game is given by a protocol specified via a *challenger* who interacts with an arbitrary *attacker* – security mandates that no efficient attacker should be able to win the game with better than negligible probability.

Definition 3.1 (Cryptographic game [HH09])**.** *A* cryptographic game $\mathcal{G} = (\Gamma, c)$ *is defined by a (possibly inefficient) random system* Γ*, called the* challenger*, and a constant* $c \in [0, 1)$*. On security parameter* n*, the challenger* $\Gamma(1^n)$ *interacts with some attacker* $\mathcal{A}_n$ *and outputs a bit* b*. We denote the output of this interaction by* $b = (\mathcal{A}_n \leftrightarrows \Gamma(1^n))$*. The* advantage *of an attacker* $\mathcal{A}_n$ *in the game* $\mathcal{G}$ *is defined as*

$$\mathbf{Adv}^{\mathcal{A}}_{\mathcal{G}}(n) \stackrel{def}{=} \Pr[\ (\mathcal{A}_n \leftrightarrows \Gamma(1^n)) = 1\ ] - c\ .$$

A cryptographic game $\mathcal{G}$ *is* secure *if for all* polysize*attackers* $\mathcal{A} = \{\mathcal{A}_n\}$*, the advantage* $\mathbf{Adv}^{\mathcal{A}}_{\mathcal{G}}(n)$ *is negligible.*

When $c = 0$, the above definition of cryptographic games captures *search problems* such as factoring, the discrete logarithm problem, signature security etc. When $c = \frac{1}{2}$, it captures *decisional problems* such as DDH, encryption security etc. Note that cryptographic games may be highly interactive and may not even have any a-priori bound on the number of rounds of interaction between $\mathcal{A}$ and Γ. The work of [GW11] defined a more restricted notion of cryptographic games (called "efficiently falsifiable assumptions") where the challenger is also required to be efficient. We do *not* rely on this requirement in the current work.

Remark 3.2 (δ-exponential security)**.** We can also define a cryptographic game $\mathcal{G}$ to be δ-exponentially secure for some constant $\delta > 0$ if for all $\mathcal{A}_n$ of size $2^{O(n^\delta)}$ the advantage $\mathbf{Adv}^{\mathcal{A}}_{\mathcal{G}}(n) = 2^{-\Omega(n^\delta)}$.

Black-box reductions. We define what it means to have a black-box reduction showing that some hash function family $\mathcal{H}$ is (m, k)-FS-universal. Recall that an FS-universal hash function should be able to compile *all* efficient 3-PC proofs $\langle P, V\rangle$ with first message of length $m(n)$ and challenge length $k(n)$ into a corresponding 2-message argument system. Therefore, we think of the proof system $\langle P, V\rangle$ itself as an adversarial entity given to the reduction, and hence it is natural to require that the reduction is black box in the attacker $\mathcal{A}$ as well as the proof system $\langle P, V\rangle$.

Since we are proving a negative result, we want to place as few requirements on the reduction as possible. Therefore, to simplify things, we will only require that the reduction works if the proof system $\langle P, V\rangle$ has perfect completeness and near-perfect $2^{-}k$-soundness (where k is the challenge size), meaning that for any choice of a false statement and first message of the prover, there is at most a single challenge on which the prover can succeed. We also assume the reduction only works for attackers $\mathcal{A}$ that *always* break the soundness of the FS-collapse argument $\mathcal{FS}^{\langle P,V\rangle}_{\mathcal{H}}$ with advantage 1.

Definition 3.3 (BB reduction for Fiat-Shamir)**.** *Let* $\mathcal{G} = (\Gamma, c)$ *be a cryptographic game assumption and let* $\mathcal{H}$ *be an* (m, k)*-hash function family for some polynomials* $m = m(n), k = k(n)$*. A* black-box reduction showing the (m, k)-FS-universality of $\mathcal{H}$ under the assumption $\mathcal{G}$ *is a* PPT *oracle-access machine* $\mathcal{B}^{(\cdot,\cdot,\cdot)}$*, for which there exists some polynomial* $p(n)$ *such that the following holds. Let* $P = \{P_n\}, V = \{V_n\}$ *describe an interactive/stateful protocol and* $\mathcal{A} = \{\mathcal{A}_n\}$ *be any stateless attacker, all given as (possibly inefficient) circuit-families, such that:*

1. *The protocol* $\langle P, V\rangle$ *is a 3PC interactive proof for some relation* $\mathcal{R}$*, with first-message length* m *and challenge-length* k*. Furthermore, it has perfect completeness and* $2^{-k(n)}$*-soundness. (See Definition 2.1)*

2. *The attacker $\mathcal{A}$ breaks the computational soundness of $\mathcal{FS}_{\mathcal{H}}^{\langle P,V\rangle}$ with advantage 1. (See Definition 2.2)*

The reduction $\mathcal{B}$ gets black-box oracle access to $\mathcal{A}_n$, along with black-box "rewinding access" to the stateful entities P_n, V_n, meaning that it can interleave arbitrarily many concurrent executions and arbitrarily rewind them to any prior state. The statement is also chosen adversarially by the attacker (and provided to the reduction); however, this choice is independent of the choice of the hash seed (i.e., it is non-adaptive).

Then the advantage of $\mathcal{B}^{P_n,V_n,\mathcal{A}_n}(1^n)$ in the game $\mathcal{G}$ must be at least $1/p(n)$.

We note that, since we are proving a negative result, requiring that the reductions only work for proofs with strong soundness $2^{-k(n)}$ and attackers $\mathcal{A}$ with high-advantage 1, and a non-adpative choice of statement, only makes the result stronger (compared to insisting that the reduction works for any $\mathsf{negl}(n)$-soundness and $1/\mathsf{poly}(n)$ advantage). On the other hand, we do insist that the reduction itself has some *noticeable* advantage $1/p(n)$ rather than the standard requirement that its advantage is simply *non-negligible*. Furthermore, we also insist that the reduction is *security-parameter preserving* meaning that when it is called with security parameter 1^n it only accesses the oracles $P_n, V_n, \mathcal{A}_n$ on the *same* security parameter n. The above two requirements come with some loss of generality, but they hold for all of the natural reductions in cryptography.

We note that, although in general we need to consider the issue of the reduction running many executions and rewinding P_n, V_n, in our eventual result this will not play an important role. In particular, our counterexample will construct a proof system $P_n^{\mathcal{A}_n}, V_n^{\mathcal{A}_n}$ which is efficiently defined in terms of a single inefficient but *stateless* adversary $\mathcal{A}_n$. Therefore, everything the reduction sees can just be derived from simple oracle access to the stateless $\mathcal{A}_n$.

We now define an analogous notion of a black-box reduction for entropy-preserving hashing. Since, without loss of generality, the attacker against entropy-preserving hashing is stateless, we do not have to worry about issues of rewinding.

Definition 3.4 (BB Reduction for Entropy Preserving Hash)**.** *Let $\mathcal{G} = (\Gamma, c)$ be a cryptographic game and let $\mathcal{H}$ be an (m,k)-hash function family for some polynomials $m = m(n), k = k(n)$. A* black-box reduction showing that $\mathcal{H}$ is entropy-preserving from the security of the game $\mathcal{G}$ *is an oracle-access* PPT *machine $\mathcal{B}^{(\cdot)}$ for which there exists some polynomial p such that the following holds. Let $\mathcal{A} = \{\mathcal{A}_n\}$ be any (possibly inefficnet) attacker such that $\mathbf{H}\left(h_{\mathbf{s}}(\mathbf{x}) \mid \mathbf{x}\right) = 0$, where the random variable $\mathbf{s}, \mathbf{x}$ are defined the same way as in Definition 2.4, i.e., $\mathbf{s} \xleftarrow{\$} \{0,1\}^{\ell(n)}$, and $\mathbf{x} \leftarrow \mathcal{A}_n(\mathbf{s})$. Then, the advantage of $\mathcal{B}^{\mathcal{A}_n}(1^n)$ in the game $\mathcal{G}$ is at least $1/p(n)$.*

Remark 3.5 (Reductions from δ-exponential security assumptions)**.** For both of Definition 3.3 and Definition 3.4, we can also consider a variant, where the black-box reduction is from the δ-exponential security of the cryptographic game $\mathcal{G}$. In this case we allow the reduction $\mathcal{B}^{(\cdot)}$ to run in time $2^{O(n^\delta)}$ and only insist that its advantage is $\geq 2^{-o(n^\delta)}$.

BB Separation via Simulatable Attack. We now outline a general strategy for proving a black-box separations via a technique called a *simulatable attack*. This strategy has been used in several prior works [DOP05, HH09, GW11, Pas11, DHT12, Wic12], and was recently formalized as a general technique in [Wic12]. Here we will rely on the notation and the results from that work. However, for concreteness, we only restrict ourselves to describing this strategy for the specific case of *entropy preserving hash functions*.

Definition 3.6 (Simulatable Attack for Entropy-Preserving Hashing)**.** *Let $\mathcal{H}$ be some (m,k)-hash function family. A $\varepsilon(n)$-*simulatable attack *on the entropy-preserving security of $\mathcal{H}$ consists of: (1) an ensemble of (possibly inefficient) stateless non-uniform attackers $\{\mathcal{A}_{n,f}\}_{n\in\mathbb{N},f\in\mathcal{F}_n}$ where $\{\mathcal{F}_n\}$ is some ensemble of finite sets, and (2) a stateful* PPT *simulator* Sim*. We require that the following two properties hold:*

- *For each $n \in \mathbb{N}, f \in \mathcal{F}_n$, the (inefficient) attacker $\mathcal{A}_{n,f}$ successfully breaks the entropy-preserving security of $\mathcal{H}$.*
- *For every (possibly inefficient) oracle access machine $\mathcal{M}^{(\cdot)}$, making at most $q = q(n)$ queries to its oracle:*

$$\left| \Pr_{f\xleftarrow{\$}\mathcal{F}_n,\mathcal{M}}[\mathcal{M}^{\mathcal{A}_{n,r}}(1^n)=1] \;-\; \Pr_{(\mathcal{M},\mathsf{Sim})}[\mathcal{M}^{\mathsf{Sim}(1^n)}(1^n)=1] \right| \leq \mathsf{poly}(q(n))\cdot\varepsilon(n) \;\;.$$

  namely, oracle access to $\mathcal{A}_{n,f}$ for a random $f \xleftarrow{\$} \mathcal{F}_n$ is indistinguishable from that to Sim*.*

We omit the $\varepsilon(n)$ and just say "simulatable attack" as shorthand for an $\varepsilon(n)$-simulatable attack with some negligible $\varepsilon(n) = \mathsf{negl}(n)$.

As discussed in the introduction, the existence of a simulatable attack against some scheme $\mathcal{H}$ ensures that one cannot prove the security of $\mathcal{H}$ using black-box reduction from cryptographic game assumption, unless the assumption is false. This is because a reduction must be able to use the simulatable attacker $\mathcal{A}$ against $\mathcal{H}$ to break the underlying assumption, but then this means that the reduction and the simulator together would give us an efficient stand-alone attack against the assumption to begin with. A general version of this theorem was given in [Wic12] and therefore we get the following as a special case.

Theorem 3.7 (Special case of [Wic12])**.** *If there exists a simulatable attack against the entropy preserving security of $\mathcal{H}$, and there is a black-box reduction showing the entropy preserving security of $\mathcal{H}$ from the security of some cryptographic game $\mathcal{G}$, then $\mathcal{G}$ is* not *secure.*

Furthermore, for any constant $\delta > 0$, if there exists an $(\varepsilon(n) = 2^{-\omega(n^\delta)})$-simulatable attack against $\mathcal{H}$ and there is a black-box reduction from the δ-exponential security of $\mathcal{G}$, then $\mathcal{G}$ is not *δ-exponentially secure.*

4 Constructing a Simulatable Attack

We now show that, for any family of hash functions $\mathcal{H}$, there is a simulatable attack against the entropy preserving security of $\mathcal{H}$.

Theorem 4.1. *Let $\mathcal{H} = \{h_s : \{0,1\}^{m(n)} \to \{0,1\}^{k(n)}\}_{n\in\mathbb{N},s\in\{0,1\}^{\ell(n)}}$ be any family of hash functions. Then there is a $2^{-\Omega(m-k)}$-simulatable attack against the entropy preserving security of $\mathcal{H}$.*

Proof of Theorem 4.1. Let $\mathcal{F}_n$ be the set of functions $f : \{0,1\}^{m(n)} \to \{0,1\}^{k(n)}$, and let $\mathcal{F}_n^* \subseteq \mathcal{F}_n$ be a subset consisting of all the functions f such that for every $s \in \{0,1\}^{\ell(n)}$, there is some $x \in \{0,1\}^m$ on which $h_s(x) = f(x)$. We will define a family of inefficient attackers $\{\mathsf{Break}_f\}$, indexed by functions $f \in \mathcal{F}_n^*$, that break the entropy preserving security of $\mathcal{H}$. Before we do so, we first show that $\mathcal{F}_n^*$ is non-empty, and in fact forms a very dense subset of $\mathcal{F}_n$.

Claim 4.2. *The subset $\mathcal{F}_n^*$ is dense in $\mathcal{F}_n$ with $\frac{|\mathcal{F}_n^*|}{|\mathcal{F}_n|} = (1 - 2^{-\Omega(2^{m-k})})$-fraction of $\mathcal{F}_n$.*

Proof of Claim 4.2. If we choose $f \leftarrow \mathcal{F}_n$ uniformly at random then

$$\begin{aligned} 1 - \frac{|\mathcal{F}_n^*|}{|\mathcal{F}_n|} &= \Pr_{f \leftarrow \mathcal{F}_n}[f \notin \mathcal{F}_n^*] = \Pr_{f \leftarrow \mathcal{F}_n}[\ \exists s \in \{0,1\}^{\ell(n)}\ \forall x \in \{0,1\}^m\ : h_s(x) \neq f(x)] \\ &\leq \sum_{s \in \{0,1\}^{\ell(n)}} \Pr_{f \leftarrow \mathcal{F}_n}[\forall x \in \{0,1\}^m\ : h_s(x) \neq f(x)] \leq 2^{\ell} \left(1 - 2^{-k}\right)^{2^m} \\ &\leq 2^{2^{O(\log n)}} 2^{-2^{m-k}} \leq 2^{-\Omega(2^{m-k})} \end{aligned}$$

□

Constructing an attack. Now we are ready to define a family of inefficient attackers $\{\mathsf{Break}_f\}$, indexed by functions $f \in \mathcal{F}_n^*$, that break the entropy preserving security of $\mathcal{H}$ as follows:

Break_f **:** $f \in \mathcal{F}_n^*$
Given input $s \in \{0,1\}^{\ell(n)}$, output a random x from the set of all values satisfying $h_s(x) = f(x)$. (By the definition of $\mathcal{F}_n^*$, at least one such x always exists.)

Figure 1

The attack is successful. For any fixed $f \in \mathcal{F}_n^*$, it is easy to see that the attacker Break_f breaks the entropy preserving security of $\mathcal{H}$. This is because, conditioned on seeing any output $x \leftarrow \mathsf{Break}_f(s)$, we can completely determine the value $h_s(x)$ without knowing the seed s, via the relation $h_s(x) = f(x)$. Therefore, defining the random variables $\mathbf{s}$ to be uniform over $\{0,1\}^{\ell(n)}$ and $\mathbf{x} \leftarrow \mathsf{Break}_f(\mathbf{s})$, we have $\mathbf{H}(h_{\mathbf{s}}(\mathbf{x}) \mid \mathbf{x}) = 0$ as desired.

The simulator for the attack. The more interesting part of the proof is showing that for random $f \leftarrow \mathcal{F}_n^*$, the attacker Break_f can be simulated very efficiently, with a small statistical error. Our (stateful) simulator is incredibly simple and, on each invocation, just outputs a fresh random value (which wasn't output previously). It is easy to see that the simulator satisfies the efficiency

$\mathsf{Sim}(1^n)$
Initialize the set $X := \emptyset$. On input $s \in \{0,1\}^{\ell(n)}$: Sample $x \leftarrow \{0,1\}^m \setminus X$, add x to the set X, and output x.

Figure 2

requirements of the definition of a simulatable attack.

Indistinguishability of simulator. Our next step is to show that a random attacker from the class $\{\mathsf{Break}_f\}$ and the above simulator are statistically indistinguishable. In particular, for any (computationally unbounded) q-query distinguisher $\mathcal{M}$,

$$\left| \Pr_{f \xleftarrow{\$} \mathcal{F}_n^*} \left[\mathcal{M}^{\mathsf{Break}_f}(1^n) = 1\right] - \Pr_{\mathsf{Sim}} \left[\mathcal{M}^{\mathsf{Sim}(1^n)}(1^n) = 1\right] \right| \leq q^2 \cdot 2^{-\Omega(m-k)} \ . \tag{1}$$

Sim_i

- Initialize an empty set $X := \emptyset$ and an (initially empty) partial mapping $f|_X : X \to \{0,1\}^k$.
- Answer the first i queries the same way as Sim: on input s, sample uniformly at random $x \leftarrow \{0,1\}^m \setminus X$ and output x. Add x to X and set $f|_X(x) := h_s(x)$.
- After answering the first i queries, choose the rest of the function $f : \{0,1\}^m \to \{0,1\}^k$ extending $f|_X$ uniformly at random.
- Answer the rest of the queries the same way as Break_f. That is, on query s, output a random x such that $h_s(x) = f(x)$. If none exists, output $\perp$.

Figure 3

To prove the above, we define a series of (inefficient) "hybrid" simulators $\mathsf{Sim}_0, \ldots, \mathsf{Sim}_q$, defined as follows:

We can write Sim_0 as equivalent to the oracle Break_f where $f \leftarrow \mathcal{F}_n$ is chosen as a uniformly random function. By Claim 4.2, the statistical distance between the uniform distributions over $\mathcal{F}_n^*$ and $\mathcal{F}_n$ is $2^{-\Omega(m-k)}$ and therefore this bounds the distance between the oracle $\{\mathsf{Break}_f : f \leftarrow \mathcal{F}_n^*\}$ and Sim_0. On the other hand Sim_q is exactly the same as the original simulator Sim. Therefore, to prove equation 1, we only need to prove that the oracles Sim_i and Sim_{i+1} are statistically close; concretely, we show:

$$\left|\Pr\left[\mathcal{M}^{\mathsf{Sim}_i(1^n)}(1^n) = 1\right] - \Pr\left[\mathcal{M}^{\mathsf{Sim}_{i+1}(1^n)}(1^n) = 1\right]\right| \leq \sqrt{\frac{2^k}{2^m - i}} + 2i \cdot \frac{2^k}{2^m - i} \ , \tag{2}$$

Notice that Sim_i and Sim_{i+1} behave exactly the same on the first i queries. Let us fix *any* (worst-case) sequence of the first i queries and their possible answers. This defines some set X and some partial mapping $f|_X : X \to \{0,1\}^k$ at the end of this sequence. Let us also fix the $(i+1)$ query s. We can bound the statistical distance between Sim_i and Sim_{i+1} by how they answer the query $i+1$ and the rest of the queries going forward. The oracle Sim_i chooses the function f extending $f|_X$ uniformly at random and answers the query $i+1$ via a random x such that $h_s(x) = f(x)$, and then answers the rest of the queries after that similarly according to Break_f. Let $\mathbf{x}, \mathbf{f}$ be random variables for the above choice of x, f. The oracle Sim_{i+1} answers query $i+1$ with a random $x \leftarrow \{0,1\}^m \setminus X$, then chooses a function f uniformly at random subject to extending $f|_X$ and $f(x) = h_s(x)$, and answers the rest of the queries according to Break_f. Let $\mathbf{x}', \mathbf{f}'$ be random variables for the above choice of x, f. We can now bound the distance in equation (2) by $\Delta := \mathbf{SD}(\ (\mathbf{x}, \mathbf{f})\ ,\ (\mathbf{x}', \mathbf{f}')\ )$. Therefore, we are left to bound Δ.

Let $\mathcal{F}_{\text{ext}}$ be the set of all functions $\{0,1\}^m \to \{0,1\}^k$ extending $f|_X$. Then we can write:

$$2\Delta \ = \sum_{\substack{f \in \mathcal{F}_{\text{ext}} \\ x \in \{0,1\}^m}} \Pr\left[(\mathbf{x}, \mathbf{f}) = (x, f)\right] - \Pr\left[(\mathbf{x}', \mathbf{f}') = (x, f)\right]$$

Next, for a function $f\colon \{0,1\}^m \to \{0,1\}^k$, we denote by

$$C_f = \{x \in \{0,1\}^m \colon f(x) = h_s(x)\}$$
$$D_f = C_f \setminus X$$

the set of elements on which f coincides with h_s, and the set of such elements outside of X, respectively.

Then denoting $M_i = |\{0,1\}^m \setminus X| = 2^m - i$ and $K = \left|\{0,1\}^k\right| = 2^k$. We have for each (x,f) such that $x \in C_f$:

$$\Pr\left[(\mathbf{x},\mathbf{f}) = (x,f)\right] = \frac{1}{|C_f|} \cdot \left(\frac{1}{K}\right)^{M_i}$$

whereas for any other (x,f) this probability is zero. Also, for any (x,f) such that $x \in D_f$:

$$\Pr\left[(\mathbf{x}',\mathbf{f}') = (x,f)\right] = \frac{1}{M_i} \cdot \left(\frac{1}{K}\right)^{M_i-1} = \frac{K}{M_i} \cdot \left(\frac{1}{K}\right)^{M_i} \quad ,$$

while for any other (x,f) this probability is zero. Thus, we have

$$2\Delta = \sum_{\substack{f \in \mathcal{F}_{\text{ext}} \\ x \in D_f}} \left(\frac{1}{K}\right)^{M_i} \left|\frac{1}{|C_{\mathbf{f}}|} - \frac{K}{M_i}\right| + \sum_{\substack{f \in \mathcal{F}_{\text{ext}} \\ x \in X \cap C_f}} \left(\frac{1}{K}\right)^{M_i} \cdot \frac{1}{|C_f|}$$

Now, recalling that $\mathbf{f} : \{0,1\}^m \to \{0,1\}^k$ is a random function extending $f|_X$, we can rewrite the last term as:

$$\begin{aligned}
&\mathop{\mathbf{E}}_{\mathbf{f}}\left[|D_{\mathbf{f}}| \cdot \left|\frac{1}{|C_f|} - \frac{K}{M_i}\right|\right] + \mathop{\mathbf{E}}_{\mathbf{f}}\left[|X \cap C_{\mathbf{f}}| \cdot \frac{1}{|C_{\mathbf{f}}|}\right] = \\
&\mathop{\mathbf{E}}_{\mathbf{f}}\left[(|C_{\mathbf{f}}| - |X \cap C_{\mathbf{f}}|) \cdot \left|\frac{1}{|C_f|} - \frac{K}{M_i}\right|\right] + \mathop{\mathbf{E}}_{\mathbf{f}}\left[|X \cap C_{\mathbf{f}}| \cdot \frac{1}{|C_{\mathbf{f}}|}\right] = \\
&\mathop{\mathbf{E}}_{\mathbf{f}}\left[|C_{\mathbf{f}}| \cdot \left|\frac{1}{|C_{\mathbf{f}}|} - \frac{K}{M_i}\right|\right] + \mathop{\mathbf{E}}_{\mathbf{f}}\left[|X \cap C_{\mathbf{f}}| \cdot \left(\frac{1}{|C_{\mathbf{f}}|} - \left|\frac{1}{|C_{\mathbf{f}}|} - \frac{K}{M_i}\right|\right)\right] \leq \\
&\mathop{\mathbf{E}}_{\mathbf{f}}\left[\left|1 - |C_{\mathbf{f}}| \cdot \frac{K}{M_i}\right|\right] + \mathop{\mathbf{E}}_{\mathbf{f}}\left[|X \cap C_{\mathbf{f}}| \cdot \frac{K}{M_i}\right] \leq \\
&\mathop{\mathbf{E}}_{\mathbf{f}}\left[\left|1 - |D_{\mathbf{f}}| \cdot \frac{K}{M_i}\right| + |X \cap C_{\mathbf{f}}| \cdot \frac{K}{M_i}\right] + \mathop{\mathbf{E}}_{\mathbf{f}}\left[|X \cap C_{\mathbf{f}}| \cdot \frac{K}{M_i}\right] \leq \\
&\mathop{\mathbf{E}}_{\mathbf{f}}\left[\left|1 - |D_{\mathbf{f}}| \cdot \frac{K}{M_i}\right|\right] + 2 \cdot |X| \cdot \frac{K}{M_i}
\end{aligned}$$

We now note that $|D_{\mathbf{f}}|$ describes the number of collisions of a random function $\mathbf{f}|_{\{0,1\}^m \setminus X}$ with $h_{s_{i+1}}$, and is hence the sum of M_i independent indicators, each with probability $\frac{1}{K}$, implying that:

$$\mathop{\mathbf{E}}_{\mathbf{f}}\left[|D_{\mathbf{f}}|\right] = \frac{M_i}{K} \text{ and } \mathop{\mathbf{Var}}_{\mathbf{f}}\left[|D_{\mathbf{f}}|\right] = \frac{M_i}{K}\left(1 - \frac{1}{K}\right) \quad .$$

Using the above, Jensen's inequality, and the fact that $|X| < i$:

$$\begin{aligned}
&\mathop{\mathbf{E}}_{\mathbf{f}}\left[\left|1-|D_{\mathbf{f}}|\frac{K}{M_i}\right|\right]+2\cdot|X|\cdot\frac{K}{M_i}\leq\\
&\sqrt{\mathop{\mathbf{E}}_{\mathbf{f}}\left[\left|1-|D_{\mathbf{f}}|\cdot\frac{K}{M_i}\right|^2\right]}+2i\cdot\frac{K}{M_i}=\\
&\frac{K}{M_i}\sqrt{\mathop{\mathbf{E}}_{\mathbf{f}}\left[\left|\mathop{\mathbf{E}}_{\mathbf{f}}\left[|D_{\mathbf{f}}|\right]-|D_{\mathbf{f}}|\right|^2\right]}+2i\cdot\frac{K}{M_i}=\\
&\frac{K}{M_i}\cdot\sqrt{\mathop{\mathbf{Var}}_{\mathbf{f}}\left[|D_{\mathbf{f}}|\right]}+2i\cdot\frac{K}{M_i}\leq\\
&\sqrt{\frac{K}{M_i}}+2i\cdot\frac{K}{M_i}\ .
\end{aligned}$$

The above proves the hybrid step in equation (2). By combining q such hybrid steps, we get the indistinguishability claimed in equation 1, which proves the theorem. □

Theorem 3.7 and Theorem 4.1 allow us to conclude the following.

Corollary 4.3. *Let $\mathcal{G} = (\Gamma, c)$ be a cryptographic game assumption and let $\mathcal{H}$ be an (m, k)-hash function family for some polynomials $m = m(n), k = k(n)$ such that $m(n) - k(n) = \omega(\log(n))$. If there is a black-box reduction showing that $\mathcal{H}$ is entropy-preserving from the security of the game $\mathcal{G}$, then $\mathcal{G}$ is not secure. Furthermore, if $m(n) - k(n) = n^\delta$ for some constant $\delta > 0$, and there is a black-box reduction showing that $\mathcal{H}$ is entropy preserving from the δ-exponential security of $\mathcal{G}$, then $\mathcal{G}$ is not δ-exponentially secure.*

5 Black-Box Impossibility for Fiat-Shamir

As mentioned in the introduction, the work of Dodis, Ristenpart and Vadhan [DRV12], shows that any FS-universal hash function family $\mathcal{H}$ must also be entropy-preserving. Intuitively, this should imply that our negative result for entropy-preserving hashing from the previous section should yield a similar negative result for FS-universal hashing. Indeed, we do show a theorem along these lines. However, formalizing the above intuition requires some care. For example, it becomes important that our notion of black-box reductions for FS-universal hashing treats the 3PC proof-system as a black box. Intuitively, this is because the result of [DRV12] uses the attacker $\mathcal{A}$ against the entropy-preserving security of a hash family $\mathcal{H}$ to *construct* a 3PC proof system $\Pi^{\mathcal{A}} = \langle P^{\mathcal{A}}, V^{\mathcal{A}} \rangle$ as well as to *attacker* $\mathcal{D}^{\mathcal{A}}$ that breaks the soundness of the FS-collapse of $\Pi^{\mathcal{A}}$. Therefore, any black-box reduction that shows the FS-univeraslity of $\mathcal{H}$ under some game assumption by treating the proof system $\Pi^{\mathcal{A}} = \langle P^{\mathcal{A}}, V^{\mathcal{A}} \rangle$ and the attacker $\mathcal{D}^{\mathcal{A}}$ as a black box, can also be used as a reduction showing the entropy-preserving security of $\mathcal{H}$ under the same assumption by treating the attacker $\mathcal{A}$ as a black box. We give the full proof for completeness, adapting the result of [DRV12] to the setting of black-box reductions.

Theorem 5.1. *Let $\mathcal{G} = (\Gamma, c)$ be a cryptographic game and let $\mathcal{H}$ be an (m, k)-hash function family for some polynomials $m = m(n), k = k(n)$. Let $\mathcal{B}^{(\cdot,\cdot,\cdot)}$ be a black-box reduction showing the (m, k)-FS-universality of $\mathcal{H}$ from the security of the game $\mathcal{G}$ (Definition 3.3). Then there exists a black-box reduction $\mathcal{C}^{(\cdot)}$ showing that $\mathcal{H}$ is entropy-preserving from the security of the game $\mathcal{G}$ (Definition 3.4).*

The reduction $\mathcal{C}$ has the same running time as $\mathcal{B}$ up to a polynomial blowup; in particular, the above also holds for the case of δ-exponential security.

Proof. We first show that, given an attacker $\mathcal{A}$ that breaks the entropy-preserving security of $\mathcal{H}$, we can construct a 3-message protocol $\Pi^{\mathcal{A}}$, where the honest verifier V has oracle access $\mathcal{A}$, such that the protocol is a proof, but its FS-collapse is not sound. Indeed, let $\mathcal{A} = \{\mathcal{A}_n\}$ be any deterministic (and possibly inefficient) attacker against the entropy-preserving security of $\mathcal{H}$. In other words, we have $\mathbf{H}(h_{\mathbf{s}}(\mathbf{x})|\mathbf{x}) = 0$, where $\mathbf{s}, \mathbf{x}$ are correlated random variables defined by $\mathbf{s} \xleftarrow{\$} \{0,1\}^{\ell(n)}, \mathbf{x} \xleftarrow{\$} \mathcal{A}_n(\mathbf{s})$. (For simplicity of notation, we will skip the subscript n for the oracles). We construct a 3PC protocol $\Pi^{\mathcal{A}} = \langle P, V^{\mathcal{A}} \rangle$, where the verifier $V^{\mathcal{A}}$ is efficient up to having oracle access to the (possibly inefficient) oracle $\mathcal{A}$. The protocol will be a proof system for the empty relation $\mathcal{R} = \emptyset$, meaning that the verifier should always reject (with all but negligible probability) when interacting with any prover. We call a protocol of this type an unwinnable protocol. In this case, we do not need to specify an honest prover P (i.e., the vacuous prover that doesn't do anything already meets the definition), and the completeness property just holds trivially. (See further discussion on the use of unwinnable games below.) The protocol is shown in Figure 4.

$\langle P, V^{\mathcal{A}} \rangle$

- The honest prover P does not do anything at all.
- The verifier $V^{\mathcal{A}}$ interacts with a (possibly adversarial) prover P^* as follows. In the first round, it receives some string $\alpha \in \{0,1\}^m$. In the second round, it responds with a random challenge $\beta \in \{0,1\}^k$. In the third round, it receives some string $\gamma \in \{0,1\}^\ell$. Finally, it outputs 1 if $\beta = h_\gamma(\alpha)$ and $\alpha = \mathcal{A}(\gamma)$.

Figure 4

We now show that the above protocol $\Pi^{\mathcal{A}}$ is an 3PC unwinnable protocol (proof for the empty relation) with soundness 2^{-k}. This follows by the following simple argument. Recall that the oracle $\mathcal{A}$ is such that $\mathbf{H}(h_{\mathbf{s}}(\mathbf{x})|\mathbf{x}) = 0$ where $\mathbf{s}, \mathbf{x}$ are correlated random variables defined by $\mathbf{s} \xleftarrow{\$} \{0,1\}^{\ell(n)}, \mathbf{x} \xleftarrow{\$} \mathcal{A}(\mathbf{s})$. This implies that given α, there is a unique value β for which there exists some γ satisfying: $\mathcal{A}(\gamma) = \alpha$ and $\beta = h_\gamma(\alpha)$. However the verifier $V^{\mathcal{A}}$ in the above protocol chooses β randomly in $\{0,1\}^k$. This implies that no (even unbounded) prover can make the verifier output 1 except with probability 2^{-k}.

Next, let $\Pi^{\mathcal{A}}_{FS} = \mathcal{FS}^{\Pi^{\mathcal{A}}}_{\mathcal{H}}$ be the FS-collapse of $\Pi^{\mathcal{A}}$ with respect to $\mathcal{H}$. We construct an oracle-aided (and otherwise efficient) attacker $\mathcal{D}^{\mathcal{A}}$ that breaks the soundness of $\Pi^{\mathcal{A}}_{FS}$ with probability 1. Define $\mathcal{D}^{\mathcal{A}}$ so that, on input s, it outputs (α, β, γ) where $\alpha := \mathcal{A}(s)$, $\beta = h_s(\alpha)$ and $\gamma := s$. Note that the verifier $V^{\mathcal{A}}$ on input (α, β, γ) will always output 1 and $\beta = h_\gamma(\alpha)$. Hence, $\mathcal{D}^{\mathcal{A}}$ always breaks the computational soundness of $\Pi^{\mathcal{A}}_{FS}$.

To complete the proof, we observe that a black-box reduction showing the FS-universal security of $\mathcal{H}$ based on a cryptographic game $\mathcal{G}$ can be converted to a black-box reduction showing the entropy-preserving security of $\mathcal{H}$ from the same game $\mathcal{G}$. Indeed, let $\mathcal{B}^{(\cdot,\cdot,\cdot)}$ be a black-box reduction from the FS-universality of $\mathcal{H}$ to a game $\mathcal{G}$, the new reduction $\mathcal{C}^{(\cdot)}$, given oracle access to an attacker $\mathcal{A}$ against the entropy-preserving security of $\mathcal{H}$, simply emulates $\mathcal{B}^{P, V^{\mathcal{A}}, \mathcal{D}^{\mathcal{A}}}(1^n)$, where $P, V^{\mathcal{A}}$ are the prover and verifier of $\Pi^{\mathcal{A}}$ as defined above, and $\mathcal{D}^{\mathcal{A}}$ is the attacker on FS-universality defined above; indeed, all of these algorithms can be efficiently emulated, given oracle-access to $\mathcal{A}$. If $\mathcal{A}$

breaks the entropy-preserving security of $\mathcal{H}$, then $P, V^{\mathcal{A}}, \mathcal{D}^{\mathcal{A}}$ break its FS-universality, implying that the reduction $\mathcal{B}^{P,V^{\mathcal{A}},\mathcal{D}^{\mathcal{A}}}(1^n)$, and hence also $\mathcal{C}^{\mathcal{A}}(1^n)$, has a noticeable advantage $1/p(n)$ in the game $\mathcal{G}$.

The analogous result for δ-exponentially security follows similarly. □

Theorem 5.1 along with Corollary 4.3 directly implies the following corollary.

Corollary 5.2. *Let $\mathcal{G} = (\Gamma, c)$ be a cryptographic game assumption and let $\mathcal{H}$ be an (m, k)-hash function family for some polynomials $m = m(n), k = k(n)$ such that $m(n) - k(n) = \omega(\log(n))$. If there is a black-box reduction showing the (m, k)-FS-universality of $\mathcal{H}$ from the security of the game $\mathcal{G}$, then $\mathcal{G}$ is not secure. Furthermore, if $m(n) - k(n) = n^{\delta}$ for some constant $\delta > 0$, and there is a black-box reduction showing the (m, k)-FS-universality of $\mathcal{H}$ from the δ-exponential security of $\mathcal{G}$, then $\mathcal{G}$ is not δ-exponentially secure.*

FS-Universality for non-trivial languages and special classes of 3PC proofs. Our negative result, as stated in the above theorem and corollary, applies to FS-*universal* hash functions that are sound for *all* 3PC proofs and for all languages. In particular, the proof of impossibility discusses a protocol for the trivial relation $\mathcal{R} = \emptyset$. Such protocols may indeed seem unnatural, and one may hope to prove the security of FS-universal hash functions for restricted classes of "interesting languages," e.g. some NP-complete language. However, it is not hard to see that the result directly extends to rule out such FS-universal functions as well. Concretely, looking into the above proof, we can modify the prover P and verifier $V^{\mathcal{A}}$, by taking any 3PC proof system $\Pi_{\mathcal{L}}$ (e.g., for some NP-complete language $\mathcal{L}$) and welding it together with the above unwinnable game; namely, the prover will act as in the original system $\Pi_{\mathcal{L}}$, and the verifier $V^{\mathcal{A}}$ will accept if the transcript is accepting with respect to either $\Pi_{\mathcal{L}}$, or the unwinnable game (assume the message lengths are the same in both $\Pi^{\mathcal{A}}$ and $\Pi_{\mathcal{L}}$).

Another possible approach towards obtaining a positive result is to restrict the class of protocols for which we require the FS heuristic to work (while still having a rich and interesting enough class of protocols). Here, a very natural type of restriction, for NP languages, concerns prover privacy. Indeed, we would, typically, try to apply the FS heuristic to protocols that satisfy some notion of privacy such as zero-knowledge, witness-hiding, or witness indistinguishability. Intuitively, adding a privacy guarantee for the prover may be an obstruction to achieving protocols with "stronger soundness properties", and in particular, an obstruction to the security of the FS heuristic. For example, as mentioned in the introduction, we already know that for zero-knowledge protocols the FS heuristic cannot be sound (however, we also don't have any public-coin zero-knowledge proofs candidates to apply it to). It may be interesting to ask whether we can get FS-universality for all protocols that are not zero-knowledge, but do meet some weaker notion of privacy. For example, all protocols that are *not* zero-knowledge but *are* witness hiding, or witness indistinguishable. However, our negative result also extends to cover this type of restrictions. As above, this is done by welding a protocol $\Pi_{\mathcal{L}}$ that has given privacy properties, e.g. it is witness-hiding but not zero-knowledge, with an unwinnable game as the above. The new protocol inherits the privacy properties of the original protocol as it does not effect the honest prover behavior, but rather only the verifier's decision algorithm.

References

[Bar01] Boaz Barak. How to go beyond the black-box simulation barrier. In *42nd Annual Symposium on Foundations of Computer Science*, pages 106–115. IEEE Computer Society Press, October 2001.

[BDG+13] Nir Bitansky, Dana Dachman-Soled, Sanjam Garg, Abhishek Jain, Yael Tauman Kalai, Adriana López-Alt, and Daniel Wichs. Why "fiat-shamir for proofs" lacks a proof. In *TCC*, 2013.

[BLV03] Boaz Barak, Yehuda Lindell, and Salil P. Vadhan. Lower bounds for non-black-box zero knowledge. In *44th Annual Symposium on Foundations of Computer Science*, pages 384–393. IEEE Computer Society Press, October 2003.

[BR93] Mihir Bellare and Phillip Rogaway. Random oracles are practical: A paradigm for designing efficient protocols. In V. Ashby, editor, *ACM CCS 93: 1st Conference on Computer and Communications Security*, pages 62–73. ACM Press, November 1993.

[Bro05] Daniel R. L. Brown. Breaking rsa may be as difficult as factoring. Cryptology ePrint Archive, Report 2005/380, 2005. http://eprint.iacr.org/.

[BV98] Dan Boneh and Ramarathnam Venkatesan. Breaking rsa may not be equivalent to factoring. In Kaisa Nyberg, editor, *EUROCRYPT*, volume 1403 of *Lecture Notes in Computer Science*, pages 59–71. Springer, 1998.

[Cor02] Jean-Sébastien Coron. Security proof for partial-domain hash signature schemes. In Moti Yung, editor, *Advances in Cryptology – CRYPTO 2002*, volume 2442 of *Lecture Notes in Computer Science*, pages 613–626. Springer, August 2002.

[Cra12] Ronald Cramer, editor. *Theory of Cryptography - 9th Theory of Cryptography Conference, TCC 2012, Taormina, Sicily, Italy, March 19-21, 2012. Proceedings*, volume 7194 of *Lecture Notes in Computer Science*. Springer, 2012.

[DHT12] Yevgeniy Dodis, Iftach Haitner, and Aris Tentes. On the instantiability of hash-and-sign rsa signatures. In Cramer [Cra12], pages 112–132.

[DJKL12] Dana Dachman-Soled, Abhishek Jain, Yael Tauman Kalai, and Adriana Lopez-Alt. On the (in)security of the fiat-shamir paradigm, revisited. Cryptology ePrint Archive, Report 2012/706, 2012. http://eprint.iacr.org/.

[DNRS99] Cynthia Dwork, Moni Naor, Omer Reingold, and Larry J. Stockmeyer. Magic functions. In *40th Annual Symposium on Foundations of Computer Science*, pages 523–534. IEEE Computer Society Press, October 1999.

[DOP05] Yevgeniy Dodis, Roberto Oliveira, and Krzysztof Pietrzak. On the generic insecurity of the full domain hash. In Victor Shoup, editor, *Advances in Cryptology – CRYPTO 2005*, volume 3621 of *Lecture Notes in Computer Science*, pages 449–466. Springer, August 2005.

[DRV12] Yevgeniy Dodis, Thomas Ristenpart, and Salil P. Vadhan. Randomness condensers for efficiently samplable, seed-dependent sources. In Cramer [Cra12], pages 618–635.

[FS87] Amos Fiat and Adi Shamir. How to prove yourself: practical solutions to identification and signature problems. In *Proceedings of the 6th Annual International Cryptology Conference*, CRYPTO '87, pages 186–194, 1987.

[GBL08] Sanjam Garg, Raghav Bhaskar, and Satyanarayana V. Lokam. Improved bounds on security reductions for discrete log based signatures. In David Wagner, editor, *Advances in Cryptology – CRYPTO 2008*, volume 5157 of *Lecture Notes in Computer Science*, pages 93–107. Springer, August 2008.

[GK03] Shafi Goldwasser and Yael Tauman Kalai. On the (in)security of the Fiat-Shamir paradigm. In *44th Annual Symposium on Foundations of Computer Science*, pages 102–115. IEEE Computer Society Press, October 2003.

[GMR89] Shafi Goldwasser, Silvio Micali, and Charles Rackoff. The knowledge complexity of interactive proof systems. *SIAM Journal on Computing*, 18(1):186–208, 1989. Preliminary version appeared in STOC '85.

[GQ90] Louis C. Guillou and Jean-Jacques Quisquater. A "paradoxical" indentity-based signature scheme resulting from zero-knowledge. In Shafi Goldwasser, editor, *Advances in Cryptology – CRYPTO'88*, volume 403 of *Lecture Notes in Computer Science*, pages 216–231. Springer, August 1990.

[GW11] Craig Gentry and Daniel Wichs. Separating succinct non-interactive arguments from all falsifiable assumptions. In Lance Fortnow and Salil P. Vadhan, editors, *43rd Annual ACM Symposium on Theory of Computing*, pages 99–108. ACM Press, June 2011.

[HH09] Iftach Haitner and Thomas Holenstein. On the (im)possibility of key dependent encryption. In Omer Reingold, editor, *TCC 2009: 6th Theory of Cryptography Conference*, volume 5444 of *Lecture Notes in Computer Science*, pages 202–219. Springer, March 2009.

[KR09] Yael Tauman Kalai and Ran Raz. Probabilistically checkable arguments. In Shai Halevi, editor, *Advances in Cryptology – CRYPTO 2009*, volume 5677 of *Lecture Notes in Computer Science*, pages 143–159. Springer, August 2009.

[Mic00] Silvio Micali. Computationally sound proofs. *SIAM Journal on Computing*, 30(4):1253–1298, 2000. Preliminary version appeared in FOCS '94.

[Oka93] Tatsuaki Okamoto. Provably secure and practical identification schemes and corresponding signature schemes. In Ernest F. Brickell, editor, *Advances in Cryptology – CRYPTO'92*, volume 740 of *Lecture Notes in Computer Science*, pages 31–53. Springer, August 1993.

[Pas11] Rafael Pass. Limits of provable security from standard assumptions. In Lance Fortnow and Salil P. Vadhan, editors, *43rd Annual ACM Symposium on Theory of Computing*, pages 109–118. ACM Press, June 2011.

[PS00] David Pointcheval and Jacques Stern. Security arguments for digital signatures and blind signatures. *Journal of Cryptology*, 13(3):361–396, 2000.

[PV05] Pascal Paillier and Damien Vergnaud. Discrete-log-based signatures may not be equivalent to discrete log. In Bimal K. Roy, editor, *Advances in Cryptology – ASIACRYPT 2005*, volume 3788 of *Lecture Notes in Computer Science*, pages 1–20. Springer, December 2005.

[Sch91] Claus-Peter Schnorr. Efficient signature generation by smart cards. *Journal of Cryptology*, 4(3):161–174, 1991.

[Seu12] Yannick Seurin. On the exact security of schnorr-type signatures in the random oracle model. In David Pointcheval and Thomas Johansson, editors, *EUROCRYPT*, volume 7237 of *Lecture Notes in Computer Science*, pages 554–571. Springer, 2012.

[Wic12] Daniel Wichs. Barriers in cryptography with weak, correlated and leaky sources. Cryptology ePrint Archive, Report 2012/459, 2012. http://eprint.iacr.org/.