On the (In)security of the Fiat-Shamir Paradigm

Shafi Goldwasser* Yael Tauman Kalai*

Abstract

In 1986, Fiat and Shamir proposed a general method for transforming secure 3*-round public-coin identification schemes into digital signature schemes. The idea of the transformation was to replace the random message of the verifier in the identification scheme, with the value of some deterministic"hash" function evaluated on various quantities in the protocol and on the message to be signed.*

The Fiat-Shamir methodology for producing digital signature schemes quickly gained popularity as it yields efficient and easy to implement digital signature schemes. The most important question however remained open: are the digital signatures produced by the Fiat-Shamir methodology secure?

In this paper, we answer this question negatively. We show that there exist secure 3*-round public-coin identification schemes for which the Fiat-Shamir transformation yields insecure digital signature schemes for any "hash" function used by the transformation. This is in contrast to the work of Pointcheval and Stern which proved that the Fiat-Shamir methodology always produces digital signatures secure against chosen message attack in the "Random Oracle Model" – when the hash function is modelled by a random oracle.*

Among other things, we make new usage of Barak's technique for taking advantage of non black-box access to a program, this time in the context of digital signatures.

1 Introduction

In their famous paper laying the foundations for modern cryptography, Diffie and Hellman [DH76] introduced the notion of *digital signatures* and proposed a general method for designing them. Their method uses trapdoor functions as its basic primitive and is known as the *trapdoor-function signature method*. Several drawbacks of the trapdoor function approach have surfaced. In terms of security, by its very definition, it is prone to *existential forgery* as defined in [GMR88]. In terms of efficiency, the time to sign and verify are proportional to the time to invert and compute the underlying trapdoor function – a cost, which for some trapdoor functions, is prohibitive for certain applications.

Addressing the security concerns inherent in the trapdoor function model several other digital signature schemes were proposed and proved existentially unforgeable against chosen message attacks under standard intractability assumptions [GMR88, BM84, NY89, GHR99, CS99]. Most notably, [NY89] and [Rom90] showed that the existence of secure digital signature schemes is equivalent to the existence of one-way functions. These schemes, however, are rarely used in applications as they are often considered too inefficient.

A general paradigm for designing digital signature schemes was proposed by Fiat and Shamir [FS86]. Their starting observation was that designing secure interactive identification protocols (in which a sender merely identifies himself to a receiver) can be done with greater ease and efficiency than seems to be the case for secure digital signature schemes (in which a signer produces digital signatures for messages to be verified valid by a verifier). Building on this observation, they proposed a two-step approach for how to design secure digital signatures.

- First, design a secure 3-round public-coin identification scheme. Namely, a secure 3-round identification scheme $(\alpha,;\beta;\gamma)$ where α is the prover's first message, β is a random message sent by the verifier, and γ is the prover's response.
- Second, obtain a digital signature scheme as follows. Let the signer publish a "hash" function h as part of his public-key. To sign a message M, the legal signer produces an accepting transcript of the interactive identification protocol $(\alpha;\beta;\gamma)$, where $\beta = h(\alpha, M)$. The legal signer who knows the secret key can produce accepting transcripts for any M. The intuition for why this signature scheme is secure is that when h is a sufficiently complicated function chosen by the the real signer it should be hard for a forger to find any message M and a transcript $(\alpha;\beta;\gamma)$ for which it is true both that $\beta = h(\alpha, M)$ and that γ is an answer which

*Department of Computer Science and Applied Math, The Weizmann Institute of Science at Rehovot, ISRAEL, and The Department of Computer Science and Electrical Engineering at MIT. {shafi,yael}@theory.lcs.mit.edu

Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science (FOCS'03)
0272-5428/03 $17.00 © 2003 **IEEE**

makes $(\alpha; \beta; \gamma)$ an accepting transcript of the identification protocol.

The complexity of a digital signature scheme resulting from the above paradigm is equivalent to the complexity of the starting identification scheme and the cost of evaluating the public function h. Current proposals for a public (keyless) function h are very efficient [MD5].

Due to the efficiency and the ease of design, the Fiat-Shamir paradigm quickly gained much popularity both in theory and in practice. Several digital signature schemes, including [Sch91, GQ88, Ok92], were designed following this paradigm. The paradigm has also been applied in other domains so as to achieve forward secure digital signature schemes [AABN02] and to achieve better exact security [MR02]. Both of the above applications actually use a variation of the Fiat-Shamir paradigm. Still, they share the same basic structure: start with some secure 3-round identification scheme and transform it into a digital signature scheme, eliminating the random move of the verifier by an application of a fixed function h to different quantities determined by the protocol and to the message to be signed.

The main question regarding any of these proposals is what can be proven about the security of the resulting signature schemes.

Pointcheval and Stern [PS96] made a first step towards answering this question. They proved that for every 3-round public-coin identification protocol, which is zero-knowledge with respect to an honest verifier, the signature scheme, obtained by applying the Fiat-Shamir transformation, is secure in the *Random Oracle Model.* This work was extended by Abdalla et. al. [AABN02] to show necessary and sufficient conditions on 3-round identification protocols for which the signature scheme, obtained by applying the Fiat-Shamir transformation, is secure in the Random Oracle Model.[1]

The Random Oracle Model is an *idealization* which assumes that all parties (including the adversary) have oracle access to a truly random function. The so called *random oracle methodology* is a popular methodology that uses the Random Oracle Model for designing cryptographic schemes. It consists of two steps. First, design a secure scheme in the Random Oracle Model. Then, replace the random oracle with a function, chosen at random from some function ensemble, and provide all parties (including the adversary) with a succinct description of this function. This gives an *implementation* of the idealized scheme in the real world. This methodology, introduced implicitly by [FS86], was formalized by Bellare and Rogaway [BR93].

[1] The conditions are that the identification scheme is secure against impersonation under passive attacks, and that the first message sent by the prover is drawn at random from a large space. [AABN02] show that the latter can be removed for a randomized version of the Fiat-Shamir transformation.

As attractive as the methodology is for obtaining security "proofs", the obvious question was whether it is indeed always possible to replace the random oracle with a real world implementation. This question was answered negatively by Canetti, Goldreich and Halevi [CGH98]. They showed that there exists a signature scheme and an encryption scheme which are secure in the Random Oracle Model but are insecure with respect to any implementation of the random oracle by a function ensemble, thus showing that the random oracle methodology fails "in principle."

The work of [CGH98] left open the possibility that for particular "natural" cryptographic practices, such as the Fiat-Shamir paradigm, the random oracle methodology does work.

In this paper we show that this is not the case.

1.1 Our Results

We prove that the Fiat-Shamir paradigm for designing digital signatures can lead to universally forgeable digital signatures. We do so by demonstrating the existence of a secure 3-round public-coin identification scheme for which the corresponding signature scheme, obtained by applying the Fiat-Shamir paradigm, is insecure with respect to any function ensemble implementing the public function.

Our result relies on the existence of one-way functions. Note, however, that if one-way functions do not exist then secure signature schemes do not exist and thus the Fiat-Shamir paradigm always fails to produce secure signature schemes, as none exist. In this sense, our result is unconditional. Moreover, the problems we demonstrate for the Fiat-Shamir paradigm apply to all other variations of the Fiat-Shamir paradigm proposed in the literature [MR02, AABN02].

We stress that our result does not imply that particular ID schemes such as [FS86, Sch91] cannot be proven to yield secure signature schemes, with respect to some tailor-made function $\mathcal{H}$, under the Fiat-Shamir paradigm. What it does imply is that any proof of security would have to involve the particulars of the ID scheme and the $\mathcal{H}$ in question.

Our first idea is to make use of Barak's technique [Bar01] of taking advantage of non black-box access to the program of the verifier. Intuitively, the idea is to take any secure 3-round public-coin identification scheme (which is not necessarily zero-knowledge) and extend its verdict function so that the verifier also accepts views which convince him that the prover knows the verifier's next message. Since the verifier chooses the next message at random, there is no way that the prover can guess the verifier's next message during a real interaction, except with negligible probability, and therefore the scheme remains secure. However, when the identification scheme is converted into a signature scheme, by applying the Fiat-Shamir paradigm, the "verifier's next

message" is computed by a public function which is chosen at random from some function ensemble and is known in advance to everyone. A forger, who will now know in advance the "verifier's next message" on any input, will be able to generate an accepting view for the verifier. This makes the signature scheme insecure regardless of which function ensemble is used to compute the "verifier's next message" in the identification scheme.

The main technical challenge with implementing this approach is the following: How can the prover convince the verifier that he knows the verifier's next message using a 3-round protocol?

We make strong use of the non-interactive CS-proofs of Micali [Mi94] to overcome this challenge. However, non-interactive CS-proofs themselves are only known to hold in the Random Oracle Model, and thus we *first* get the (somewhat odd-looking) conditional result that if CS-proofs are realizable in the real world by some function ensemble, then there exists a secure identification scheme for which the Fiat-Shamir paradigm always fails in the real world for all hash-function ensembles. Next, we show that even if CS-proofs are not realizable in the real world by any function ensemble, then again the Fiat-Shamir paradigm fails. This part of the proof contains the bulk of difficulty and technical complication. It entails showing different extensions of secure 3-round public-coin identification schemes, which become insecure as digital signature schemes when the Fiat-Shamir paradigm is applied to them. All in all, we construct three ID schemes ID^1, ID^2 and ID^3, and prove that at least one of them demonstrates the failure of the Fiat-Shamir paradigm.

1.2 Related Work: Fiat-Shamir Paradigm and Zero Knowledge

Following the work of [CGH98], Dwork, Naor, Reingold and Stockmeyer [DNRS99] investigated the security of the Fiat-Shamir paradigm, and showed that it is closely related to two previously studied problems: *the selective decommitment problem*[2], and *the existence of* 3*-round public-coin weak zero-knowledge arguments for non BPP languages*. We note that our negative results, regarding the security of the Fiat-Shamir paradigm, have implications on these related problems.

In particular, the result of [DNRS99], that the existence of 3-round public-coin zero-knowledge protocols for non BPP languages implies the insecurity of the Fiat-Shamir paradigm, is worth elaborating on. It follows from the following simple observation. Suppose there exists a 3-round public-coin zero-knowledge argument for some hard language. View this zero-knowledge argument as a secure identification protocol.[3] The fact that the identification protocol is zero-knowledge (and not only honest verifier zero-knowledge) means that for *every verifier* there exists a simulator that can generate identical views to the ones produced during the run of the identification protocol. As the Fiat-Shamir paradigm (applied to this identification protocol) essentially fixes a public program for the verifier of the zero-knowledge argument, any forger can now simply run the simulator for this fixed verifier to produce a view of the identification protocol, i.e. a valid digital signature.

This simple argument extends to any k-round public-coin zero-knowledge argument. Namely, if such a k-round public-coin zero-knowledge argument exists, it can be viewed as an identification protocol. Now, extend the original Fiat-Shamir paradigm to an *Extended-Fiat-Shamir* paradigm which replaces each message of the verifier (one round at a time) by applying a fixed public function to previous messages in the protocol. Then the same argument as above says, that the simulator for the k-round zero-knowledge protocol can be used to produce forgeries in the signature scheme resulting from the Extended-Fiat-Shamir paradigm, and thus the Extended-Fiat-Shamir paradigm fails.

Barak [Bar01] has shown that under the assumption that collision resistant function ensembles exist, every language in NP has a k-round (for some constant $k > 3$) public-coin zero-knowledge argument. Thus, it follows from [DNRS99] and [Bar01] that the k-round Extended-Fiat-Shamir paradigm is insecure.

However, the Fiat-Shamir paradigm was defined, and has always been used, only for 3-round identification schemes. Barak's work does not apply to this case. Moreover, whereas all that can be deduced from [DNRS99, Bar01] is that the Fiat-Shamir paradigm (extended or otherwise) fails on zero-knowledge identification schemes (indeed it is the simulator for the zero-knowledge system which will produce forgeries), it leaves open the possibility that the (extended and ordinary) Fiat-Shamir paradigm works when the starting identification schemes are secure with respect to a less strict security requirement and are not zero-knowledge.

2 The Fiat-Shamir Paradigm

We use standard definitions for identification schemes (ID schemes) and signature schemes (see [GMR88, FFS88, Gol01]).

An ID scheme is identified with a triplet of algorithms (G, S, R): G is the key generation algorithm which takes

[2] In the selective decommitment problem, an adversary is given commitments to a collection of messages, and the adversary can ask for some subset of the commitments to be opened. The question is whether seeing the decommitments to these open plaintexts allows the adversary to learn something unexpected about the plaintexts that are still hidden.

[3] It is not necessarily a proof of knowledge but it is certainly a proof of ability of proving membership in L, which is hard for polynomial-time impersonating algorithms.

as input a security parameter 1^n and outputs a pair of keys (SK, PK), known as the secret key and the public key; S is the algorithm to be used by the sender who knows SK and wishes to prove his identity to the receiver; and R is the algorithm to be used by receiver.

Formally, (S, R) is a pair of probabilistic-polynomial-time interactive Turing machines that take a public key PK as a common input. We require that a legal sender S, on a pair (SK, PK) generated by G, can always convince the receiver to accept, and that for sufficiently large values of the security parameter, no impersonator (on input a public key PK generated by G) can convince the receiver to accept with non negligible probability even after interacting with the real sender (in the role of the receiver) polynomially many times. If the latter requirement holds we say that an ID scheme is *secure*.

In this paper we are interested in a particular type of ID schemes, known as *canonical ID schemes*.

Definition 1 (Canonical ID Scheme): *A canonical ID scheme is a* 3*-round ID scheme* $(\alpha; \beta; \gamma)$*, in which* α *is sent by the sender* S*,* β *is sent by the receiver* R *and consists of* R*'s random coins, and* γ *is sent by the sender* S*.*

We identify a signature scheme SIG with a triplet of algorithms $(GEN, SIGN, VERIFY)$. GEN is the key generation algorithm which takes as input a security parameter 1^n and outputs a pair of keys (SK, VK), known as the signing key and the verification key. $SIGN$ is the signing algorithm which takes as input a pair of keys (SK, VK) and a message M to be signed and outputs a signature of M (with respect to (SK, VK)). $VERIFY$ is the verification algorithm which takes as input a verification key VK, a message M, and a string c (supposedly a signature of M). It outputs 1 if c is a valid signature of M with respect to VK and it outputs 0 otherwise. We say that a signature scheme is *secure* if for a sufficiently large security parameter, no forger, on input a randomly chosen verification key VK where $(SK, VK) \leftarrow GEN(1^n)$ and who can ask for signatures of polynomially many messages of his choice, can generate a valid signature to any new message with non-negligible probability (with respect to VK).

For completeness, before presenting the formal definition of the Fiat-Shamir paradigm, let us define the notion of a *hash-function ensemble*.

Definition 2 *(*Hash-Function Ensemble): $\mathcal{H} = \{\mathcal{H}_n\}_{n\in\mathbb{N}}$ *is a hash-function ensemble if for every* n *and for every* $h_n \in \mathcal{H}_n$*,* $h_n : \{0,1\}^* \to \{0,1\}^n$ *is polynomial-time computable.*

Definition 3 (The Fiat-Shamir Paradigm): *Given any canonical ID scheme* $ID = (G, S, R)$ *and any hash-function ensemble* $\mathcal{H} = \{\mathcal{H}_n\}_{n\in\mathbb{N}}$*, the Fiat-Shamir paradigm transforms* ID *and* $\mathcal{H}$ *into a signature scheme* $SIG_{\mathcal{H}} = (GEN_{\mathcal{H}}, SIGN_{\mathcal{H}}, VERIFY_{\mathcal{H}})$*, defined as follows.*

- *The key generation algorithm* $GEN_{\mathcal{H}}$*, on input* 1^n*, emulates algorithm* $G(1^n)$ *to generate a pair* (SK, PK) *of secret key and public key. It then chooses at random a function* $h \in \mathcal{H}_n$*, and outputs* SK *as the signing key and* $VK = (PK, h)$ *as the verification key.*
- *The signing algorithm* $SIGN_{\mathcal{H}}$*, on input a signing key* SK*, a corresponding verification key* $VK = (PK, h)$*, and a message* M*, emulates the sender* S*, with respect to* (SK, PK)*, to produce* (α, β, γ) *where* $\beta = h(\alpha, M)$*.*
- *The verification algorithm* $VERIFY_{\mathcal{H}}$*, on input a verification key* $VK = (PK, h)$*, a message* M *and a triplet* (α, β, γ) *(which is supposedly a signature of* M*), accepts if and only if* $\beta = h(\alpha, M)$ *and* $(\alpha; \beta; \gamma)$ *is an accepted transcript by* R*, with respect to the public key* PK*.*

Throughout this paper, the Fiat-Shamir paradigm is referred to as the FS paradigm. We denote by $FS_{\mathcal{H}}(ID)$ the signature scheme obtained by applying the FS paradigm to ID and $\mathcal{H}$.

We say that the the FS paradigm is *secure* if for every secure canonical ID scheme ID, there exists a hash-function ensemble $\mathcal{H}$ such that $FS_{\mathcal{H}}(ID)$ is secure. Otherwise, we say that the FS paradigm *fails*.

We note that the FS paradigm, taken outside of the context of ID schemes and digital signature schemes, provides a general way of eliminating interaction from protocols by replacing the verifier with a function ensemble. As such, it has also been applied in other contexts, such as in the context of CS proofs [Mi94].

3 Main Result: FS Paradigm Fails

Let us begin by defining a few notions that will be used throughout this paper.

Definition 4 (negligible): *We say that a function* $g(\cdot)$ *is negligible if for every polynomial* $p(\cdot)$ *there exists* $n_0 \in \mathbb{N}$ *such that for every* $n \geq n_0$*,* $g(n) < \frac{1}{p(n)}$*.*

Definition 5 (one-way function): *We say that a polynomial-time computable function* $f : \{0,1\}^* \to \{0,1\}^*$ *is one-way if for every polynomial-size circuit* $C = \{C_n\}_{n\in\mathbb{N}}$*,*

$$Pr[C_{|y|}(y) = x \text{ s.t. } f(x) = y] = negl(n)$$

(where the probability is over uniformly chosen $y \in f(U_n)$*).*

Definition 6 (collision resistant hash-function ensemble): *We say that a hash-function ensemble $\mathcal{F} = \{\mathcal{F}_n\}_{n\in\mathbb{N}}$ is collision resistant if for every polynomial-size circuit $C = \{C_n\}_{n\in\mathbb{N}}$,*

$$Pr[C_n(f_n) = (x_1, x_2) \text{ s.t. } f_n(x_1) = f_n(x_2)] = negl(n)$$

(where the probability is over uniformly chosen $f_n \in \mathcal{F}_n$).

We prove the following two theorems.

Theorem 1 *If collision resistant hash-function ensembles do not exist and one-way functions do exist then the FS paradigm fails.*

Theorem 2 *If collision resistant hash-function ensembles exist then the FS paradigm fails.*

Corollary 1 *If one-way functions exist then the FS paradigm fails.*

It is well known that if one-way functions do not exist then neither do secure digital signature schemes. Thus, in a sense our result is unconditional since we get that the FS paradigm either fails or is useless (i.e., never produces secure digital signatures, as none exist).

We note that the proof of the first theorem is relatively simple and that the main contribution of this paper is in proving the second theorem. In what follows we give the main ideas in the proofs of the above two theorems.

3.1 Outline of Proof of Theorem 1

In this subsection, we assume that collision resistant hash-function ensembles do not exist and that one-way functions do exist. That is, we assume that for every hash-function ensemble $\mathcal{H} = \{\mathcal{H}_n\}$, for infinitely many n's, given a random $h \in \mathcal{H}_n$, it is easy to find $m_1 \neq m_2$ such that $h(m_1) = h(m_2)$. For every $\mathcal{H}$, we denote the set of all such n's by $S_{\mathcal{H}}$. Our goal is to construct a secure canonical ID scheme ID such that for every $\mathcal{H}$, the corresponding signature scheme $FS_{\mathcal{H}}(ID)$ will be insecure. More specifically, we will demonstrate the insecurity of $FS_{\mathcal{H}}(ID)$ by constructing a forger that for every $n \in S_{\mathcal{H}}$ will succeed in forging signatures, with respect to $VK = (PK, h)$ generated by $GEN(1^n)$, with non-negligible probability.

Let $SIG = (GEN, SIGN, VERIFY)$ be any secure signature scheme (the existence of secure signature schemes follows from the existence of one-way functions [Rom90, NY89]). Consider the ID scheme which generates keys according to GEN and which operates as follows: in the first round the sender sends the empty-string $\alpha = \epsilon$ to the receiver; upon receiving a random message β from the receiver, the sender replies with γ which is a signature of β with respect to his public key. In other words, the verdict function of the ID scheme accepts the transcript $(\alpha; \beta; \gamma)$ if and only if $\alpha = \epsilon$ and γ is a signature of β with respect to the sender's public key. It can be easily verified that this ID scheme is secure assuming the underlying signature scheme is secure.

When ID is converted into a signature scheme, by applying the FS paradigm with respect to some hash-function ensemble $\mathcal{H}$, a forger, given a pair $(PK, h) \leftarrow GEN(1^n)$ where $n \in S_{\mathcal{H}}$, can easily find two messages $m_1 \neq m_2$ such that $h(m_1) = h(m_2)$. Since any valid signature of m_1 is also a valid signature of m_2, the forger can query its signing oracle with message m_1 to obtain a signature of m_2.

Thus, ID is secure whereas the corresponding signature scheme $FS_{\mathcal{H}}(ID)$ is insecure for any $\mathcal{H}$, demonstrating the failure of the FS paradigm.

3.2 Outline of Proof of Theorem 2

In this subsection we assume the existence of a collision resistant hash-function ensemble, which we denote by $\mathcal{F}$. Actually, we restrict our attention to collision-resistant function ensembles from $\{0,1\}^{2n}$ to $\{0,1\}^n$.

Our goal is again to construct a secure canonical ID scheme ID such that for any hash-function ensemble $\mathcal{H}$, $FS_{\mathcal{H}}(ID)$ will be an insecure digital signature scheme. In fact we cannot point to one explicit construction of such an ID scheme. Instead we show three explicit constructions of ID schemes: ID^1, ID^2, ID^3, and prove that there exists an $1 \leq i \leq 3$ for which ID^i is a secure ID scheme, whereas for every $\mathcal{H}$, $FS_{\mathcal{H}}(ID^i)$ is an insecure digital signature scheme. Namely, we show that the FS paradigm must fail with respect to one of the three.

The idea is the following. Take any secure canonical ID scheme and extend its verdict function so as to also accept transcripts which convince the receiver that the sender knows in advance the receiver's next message. Since the receiver chooses the next message at random (by definition of a canonical ID scheme), there is no way that a sender can guess in advance the receiver's next message, except with negligible probability, and therefore the scheme remains secure. However, when the ID scheme is converted into a signature scheme via the FS paradigm, the receiver is replaced with a succinct public function, and thus everyone knows in advance the "receiver's next message" on any input, and so can generate an accepting transcript, which corresponds to a legitimate signature. Hence, the corresponding signature scheme will be insecure with respect to any hash-function ensemble.

The main problem with this approach is the following: How can the sender convince the receiver that he knows the receiver's next message? One idea is for the sender to send the receiver a polynomial-size circuit which computes the receiver's next message. The problem is that we must first

fix the ID scheme (in particular, fix a polynomial bound on the size of its messages) and only then show that for *any* hash-function ensemble $\mathcal{H}$ replacing the receiver in the signature scheme, $FS_{\mathcal{H}}(ID)$ is insecure. In other words, we need to find a protocol of a-priori bounded size, in which the sender will be able to convince the receiver of knowledge of *any* polynomial-size circuit corresponding to *any* $\mathcal{H}$.

To achieve this goal, the sender, instead of sending his circuit in hand (which may be too big), will send a size-reducing commitment to his circuit. The type of commitment we use is a tree-commitment, which allows a fixed polynomial-size commitment for any polynomial-size string. The notion of *tree-commitment* was introduced by Merkle [Mer90] and is defined as follows.

Definition 7 (Tree-Commitment): *A tree-commitment to $x \in \{0,1\}^*$, with respect to the function $f : \{0,1\}^{2n} \to \{0,1\}^n$, is defined as follows. Consider a complete binary tree of depth $\lg(|x|/n)$, where each node has a label in $\{0,1\}^n$. The leaves are labeled by the bits of x (n bits per leaf). Each internal node is labeled by applying f to the label of its children. The tree-commitment to x, with respect to f, is denoted by $TC_f(x)$, and consists of the label of the root and the depth of the tree.*

We note that if f is chosen at random from a collision resistant hash-function ensemble then the tree-commitment, with respect to f, is computationally binding.

Thus, start with a secure ID scheme ID and extend its verdict function so as to also accept views in which the sender first sends message a (supposedly a tree-commitment to a circuit C), the receiver replies with b, and only then the sender proves to the receiver that he knows a circuit C, such that both $TC_f(C) = a$ and $C(a) = b$ (where f is pre-specified and chosen at random from a collision resistant hash-function ensemble $\mathcal{F}$). More precisely, the sender proves that he knows a circuit C, which is a witness to (f, a, b) in the following relation:

$$\mathcal{R}_{\mathcal{F}} = \{((f,a,b),C) \ : \ C(a) = b \quad \wedge \quad TC_f(C) = a \quad \wedge \quad |C| < n^{\lg n}\}$$

We slightly abuse notation and let C here stand for an encoding of the circuit. Moreover, we need to be careful of which encoding is used, as is elaborated on in the full version.

Note that as we cannot bound the size of the witness C by any fixed polynomial, $\mathcal{R}_{\mathcal{F}}$ is not an NP relation. We restrict the witness C to be of size at most $n^{\lg n}$, so that the language corresponding to $\mathcal{R}_{\mathcal{F}}$ will be in $NTIME(n^{O(\lg n)})$. We remark that the exact same relation was used by Barak and Goldreich in [BG02].

In order to carry out the above idea towards establishing the insecurity of the FS paradigm, we need a proof-of-knowledge system for $\mathcal{R}_{\mathcal{F}}$. Moreover, since canonical ID schemes are confined to 3-rounds, we need a proof-of-knowledge system for $\mathcal{R}_{\mathcal{F}}$ which is either one round, or two rounds in which the first round consists of the verifier's random coin tosses. Note that this is not an easy task as $\mathcal{R}_{\mathcal{F}}$ is not an NP-relation.

If there somehow existed a 2-round public-coin proof-of-knowledge system for $\mathcal{R}_{\mathcal{F}}$ then we would be done, since we could take the secure canonical ID scheme ID, extend its public-key by appending a random $f \in_R \mathcal{F}$ to it, and extend its verdict function so as to also accept transcripts of the form

$$\overline{\quad a \quad} \longrightarrow$$

$$\longleftarrow \overline{\quad b, q \quad}$$

$$\overline{\quad ans \quad} \longrightarrow$$

where $(q; ans)$ is a 2-round public-coin proof-of-knowledge of C such that $((f,a,b),C) \in \mathcal{R}_{\mathcal{F}}$.

Unfortunately, we do not know whether a 2-round proof-of-knowledge system for $\mathcal{R}_{\mathcal{F}}$ exists.

However, a 4-round public-coin proof-of-knowledge system for $NEXP$ is implicit in the work of [Ki92, Mi94][4]. In [BG02] this system was explicitly formulated for $NEXP$ (and in particular for $\mathcal{R}_{\mathcal{F}}$) and it was proven that, under the existence of a collision resistant hash-function ensemble, this system satisfies that each bit of the witness can be extracted in polynomial time. They referred to this system by *Universal Argument*.

Our idea is to reduce interaction in Universal Arguments by applying a Fiat-Shamir type step to them (in order to eliminate the verifier's second message) and to use this new (2-round) reduced interaction argument for $\mathcal{R}_{\mathcal{F}}$ in our ID scheme. This seems like a strange idea, since our goal is to prove the failure of the FS paradigm, but it will take us one step further in the proof. This idea is carried out as follows.

Denote by (P^0, V^0) the (4-round) Universal Argument for $\mathcal{R}_{\mathcal{F}}$ and denote the transcript of (P^0, V^0) by $(\alpha; \beta; \gamma; \delta)$. Fix any hash-function ensemble $\mathcal{G}$. Denote by $(P^{\mathcal{G}}, V^{\mathcal{G}})$ the 2-round (reduced interaction) argument for $\mathcal{R}_{\mathcal{F}}$, obtained by applying a Fiat-Shamir type step to (P^0, V^0) with respect to $\mathcal{G}$. Formally, define $(P^{\mathcal{G}}, V^{\mathcal{G}})$ as follows: In the first round $V^{\mathcal{G}}$ sends to $P^{\mathcal{G}}$ a message of the form (α, g), where α is the first message that V^0 sends in the protocol (P^0, V^0) and g is chosen at random from $\mathcal{G}$. In the second round $P^{\mathcal{G}}$ sends to $V^{\mathcal{G}}$ a message of the form (β, γ, δ), where $\gamma = g(\beta)$ and the transcript $(\alpha; \beta; \gamma; \delta)$ is accepted by V^0. To be consistent with previous notations, we denote the first message sent by $V^{\mathcal{G}}$ by $q = (\alpha, g)$ and the second message sent by $P^{\mathcal{G}}$ by $ans = (\beta, \gamma, \delta)$.

[4] Explicitly, Kilian constructs an argument for NP and Micali constructs a non-interactive proof for $NEXP$.

Thus, we extend the secure ID scheme ID to a new ID scheme $ID^1_{\mathcal{G}}$, by appending a uniformly chosen $f \in \mathcal{F}$ to the public-key and by extending the verdict function so as to also accept views of the form

$$\overline{\quad a \quad}\longrightarrow$$

$$\longleftarrow \overline{b, q}$$

$$\overline{\quad ans \quad}\longrightarrow$$

where $(q; ans)$ is an accepted transcript by $V^{\mathcal{G}}(f, a, b)$.

To establish the failure of the FS paradigm, we need to show that there exists a $\mathcal{G}$ such that:

- $ID^1_{\mathcal{G}}$ is secure.
- For all $\mathcal{H}$, $SIG^1_{\mathcal{G},\mathcal{H}} = FS_{\mathcal{H}}(ID^1_{\mathcal{G}})$ is insecure.

Proving the latter is straightforward, since any forger, given a pair (PK, h) where $h \in \mathcal{H}$, knows that the "verifier's next message" is computed by h, and thus can easily generate an accepting transcript for the extended verdict function.

It remains to show that there exists a hash-function ensemble $\mathcal{G}$, such that $ID^1_{\mathcal{G}}$ is secure. This question is closely related to questions arising in the study of CS proofs.

CS Proofs and the FS Paradigm: CS proofs, defined by Micali [Mi94], are non-interactive (one-round) proof-of-knowledge systems for any language in $NEXP$ (and in particular for $\mathcal{R}_{\mathcal{F}}$). They have been shown to exist in the Random Oracle Model, but only hypothesized to preserve the proof-of-knowledge property when the random oracle is replaced by a succinct hash-function ensemble $\mathcal{G}$. Thus, an immediate partial result would be that under this hypothesis the FS paradigm fails. Simply, replace the above $(q; ans)$ by replacing q with a uniformly chosen hash-function $g \in \mathcal{G}$ (instantiating the random oracle) and by replacing ans with a one-round CS proof for $\mathcal{R}_{\mathcal{F}}$ with respect to g.

Unfortunately, we do not know whether CS proofs exist in the "real world", and in particular, we do not know whether $ID^1_{\mathcal{G}}$ is secure. Instead, we consider two cases.

In case 1, we simply assume that there exists a hash-function ensemble $\mathcal{G}$, such that $ID^1_{\mathcal{G}}$ is secure, and thus immediately reach the conclusion that the FS paradigm fails.

In case 2, we assume the opposite. Namely, we assume that for every hash-function ensemble $\mathcal{G}$, there exists an impersonator for $ID^1_{\mathcal{G}}$ that for infinitely many n's succeeds in fooling the receiver to believe that he is the "real sender". We denote the set of all such n's by $S^1_{\mathcal{G}}$. Since the original identification scheme ID was assumed to be secure, it must be the case that the impersonation is with respect to the extended verdict function. Thus, in effect we assume that for every $n \in S^1_{\mathcal{G}}$ and for a random $f \in \mathcal{F}_n$ it is easy to find an a such that for random b, the impersonator can convince $V^{\mathcal{G}}(f, a, b)$ to accept, with non-negligible probability (over random $f \in \mathcal{F}$, random b and the random coin tosses of $V^{\mathcal{G}}(f, a, b)$).

We refer to this case by $(\forall \mathcal{G}\ \exists$ IMPERSONATOR$)$. It remains to prove the following lemma.

Lemma 3.1 *If $(\forall \mathcal{G}\ \exists$ IMPERSONATOR$)$, then the FS paradigm fails.*

To prove this Lemma we construct yet two more ID schemes ID^2 and ID^3, such that one of them demonstrates the failure of the FS paradigm. This proof contains most of the technical difficulties. In what follows, we try to explain the main ideas of the proof without diving too deep into the technical issues. Nevertheless, understanding the next subsection will probably require an extra effort from the reader.

3.2.1 Outline of Proof of Lemma 3.1

The assumption $(\forall \mathcal{G}\ \exists$ IMPERSONATOR$)$ implies in particular, that for every $n \in S^1_{\mathcal{G}}$, given a random $f \in \mathcal{F}_n$, it is easy to find a and $b_1 \neq b_2$, and to convince both $V^{\mathcal{G}}(f, a, b_1)$ and $V^{\mathcal{G}}(f, a, b_2)$, with non-negligible probability.

In contrast, it is hard to convince both $V^0(f, a, b_1)$ and $V^0(f, a, b_2)$, since (P^0, V^0) is a proof of knowledge, and anyone who knows a witness to both (f, a, b_1) and (f, a, b_2) can be used to find collisions to f.

This contrast between V^0 and $V^{\mathcal{G}}$ suggests constructing a new ID scheme, ID^2, whose security will follow from the proof-of-knowledge property of (P^0, V^0) on one hand, and on the other hand the insecurity of the corresponding digital signature scheme (obtained from the Fiat-Shamir paradigm) will follow from the assumption $(\forall \mathcal{G}\ \exists$ IMPERSONATOR$)$.

Take any secure identification scheme ID, and extend it by appending a uniformly chosen $f \in \mathcal{F}$ and a uniformly chosen α_1, α_2 to the public-key, and by extending the verdict function so as to also accept views of the form

$$\overline{\quad a, b_1, b_2, \beta_1, \beta_2 \quad}\longrightarrow$$

$$\longleftarrow \overline{\gamma_1, \gamma_2}$$

$$\overline{\quad \delta_1, \delta_2 \quad}\longrightarrow$$

where $b_1 \neq b_2$, $(\alpha_1; \beta_1; \gamma_1; \delta_1)$ is an accepted transcript by $V^0(f, a, b_1)$ and $(\alpha_2; \beta_2; \gamma_2; \delta_2)$ is an accepted transcript by $V^0(f, a, b_2)$. Denote this extended identification scheme by ID^2.

First we claim that ID^2 is secure. Suppose not, then there would exist a cheating prover, which given a random f, could generate a and $b_1 \neq b_2$ for which it

could convince both $V^0(f,a,b_1)$ and $V^0(f,a,b_2)$ to accept. This implies that the cheating prover can be used to extract two different circuits C_1, C_2 such that $a = TC_f(C_1) = TC_f(C_2)$.[5] This contradicts the fact that the tree-commitment is computationally-binding, which in turn contradicts the assumption that $\mathcal{F}$ is a collision resistant hash-function ensemble. A subtle point to address is that the size of the witness here (i.e. the circuit) may be of size up to $n^{\lg n}$, and yet we need to find a collision to the tree-commitment in polynomial-time in order to contradict the existence of collision resistant functions. The same issue comes up in the work of Barak and Goldreich [BG02]. It is resolved by showing that there is no need to fully extract the witness, as a collision to the tree-commitment can be found after extracting only $2n$ bits of the witness, where each bit can be extracted in polynomial time.

Proving the insecurity of $SIG^2_{\mathcal{H}} = FS_{\mathcal{H}}(ID^2)$ is tricky. Intuitively, we would like to use an impersonator for $\mathcal{H}$ to forge signatures, as follows. Fix $h \in \mathcal{H}$. Given a random verification key $VK = (PK, f, \alpha_1, \alpha_2, h)$, use the impersonator to find an a such that for random $b_1 \neq b_2$, the impersonator can find (1) $ans_1 = (\beta_1, \gamma_1, \delta_1)$ such that $((\alpha_1, h); ans_1)$ is an accepted transcript by $V^{\mathcal{H}}(f,a,b_1)$ (i.e., such that $\gamma_1 = h(\beta_1)$ and $(\alpha_1; \beta_1; \gamma_1; \delta_1)$ is an accepted transcript by $V^0(f,a,b_1)$); and independently finds (2) $ans_2 = (\beta_2, \gamma_2, \delta_2)$ such that $((\alpha_2, h); ans_2)$ is an accepted transcript by $V^{\mathcal{H}}(f,a,b_2)$ (i.e., such that $\gamma_2 = h(\beta_2)$ and $(\alpha_2; \beta_2; \gamma_2; \delta_2)$ is an accepted transcript by $V^0(f,a,b_2)$). However, in this approach γ_1 depends only on β_1 and γ_2 depends only on β_2, whereas in valid signatures γ_1 and γ_2 are functions of both β_1 and β_2. Thus, to obtain a valid signature, we cannot simply run $\tilde{P}_1^n$ twice independently, since the value of β_2 affects the value of γ_1 and vise versa.

To get around this problem we distinguish between the following two cases:

- *Case 2a:* ($\forall \mathcal{G}\ \exists$ *strong*-IMPERSONATOR)
- *Case 2b:* $\neg$($\forall \mathcal{G}\ \exists$ *strong*-IMPERSONATOR)

Where ($\forall \mathcal{G}\ \exists$ *strong*-IMPERSONATOR) refers to the case that for every function ensemble $\mathcal{G}$ there exists a "*strong*"-impersonator, that for infinitely many n's, on input a random $f \in \mathcal{F}_n$, finds a and b_1 such that he can both convince $V^0(f,a,b_1)$ to accept and convince $V^{\mathcal{G}}(f,a,b_2)$ to accept for a random b_2. We denote the set of all such n's by $S^2_{\mathcal{G}}$. Formally speaking, ($\forall \mathcal{G}\ \exists$ *strong*-IMPERSONATOR) refers to the case that for every function ensemble $\mathcal{G}$ there exists a polynomial-size circuit family $F_2 = \{F_2^n\}$, a polynomial-size circuit family $\tilde{P}_2 = \{\tilde{P}_2^n\}$ and a polynomial $p(\cdot)$ such that for every $n \in S^2_{\mathcal{G}}$,

$$Pr[(\tilde{P}_2^n, V^0)(f,a,b_1) = 1 \quad \wedge \quad (\tilde{P}_2^n, V^{\mathcal{G}})(f,a,b_2) = 1 : (a,b_1) = F_2^n(f)] \geq \frac{1}{p(n)}$$

(where the probability is over $f \in_R \mathcal{F}_n$, over $b_2 \in_R \{0,1\}^n$ and over the random coin tosses of $V^{\mathcal{G}}$ and V^0).

We proceed by proving the failure of the FS paradigm is case 2a and in case 2b.

The Failure of the FS Paradigm in Case 2a:

In this case, we proceed with ID^2 and show that $SIG^2_{\mathcal{H}}$ is insecure for every $\mathcal{H}$, and for every $n \in S^2_{\mathcal{G}}$. Fix any message M to be forged. Define $\mathcal{H}^M = \{h^M\}_{h \in \mathcal{H}}$ where $h^M(x) = h(x, M)$. We want to use the fact ($\forall \mathcal{G}\ \exists$ *strong*-IMPERSONATOR), and in particular we want to use a strong impersonator for $\mathcal{H}^{\mathcal{M}}$. A problem in doing so is that $\mathcal{H}^{\mathcal{M}}$ takes the 5-tuple $(a, b_1, b_2, \beta_1, \beta_2)$ to a pair (γ_1, γ_2), whereas strong impersonators are defined only with respect to hash functions $\mathcal{G}$ which take β to γ. Dealing with the incompatible size of the output of $\mathcal{H}^M$ is easy, by redefining $\mathcal{H}^M$ to truncate its output to be restricted to γ_2. To address the incompatible size of the input, we actually modify ID^2 so that the first message of the the sender is a length reducing (computationally binding) commitment to $(a, b_1, b_2, \beta_1, \beta_2)$ (which is decommitted in the senders last message) instead of $(a, b_1, b_2, \beta_1, \beta_2)$ itself. However, for the sake of this extended abstract, we shall ignore this problem and the modification and pretend that there exists a strong-impersonator $(F_2, \tilde{P}_2)$ for $\mathcal{H}^M$. We proceed to describe the high level ideas necessary to prove the insecurity of $SIG^2_{\mathcal{H}}$ (see full version of the paper for full treatment).

Fix a message M. A forger for M takes as input a random verification key $VK = (PK, f, \alpha_1, \alpha_2, h)$ and works as follows: Set $(a, b_1) = F_2^n(f)$; choose at random b_2; simulate $(\tilde{P}_2^n, V^0|_{\alpha_1})(f,a,b_1)$ to compute β_1 (the notation $V^0|_{\alpha_1}$ means that V^0's first message is restricted to α_1); simulate $(\tilde{P}_2^n, V^{\mathcal{H}^M}|_q)(f,a,b_2)$, where $q = (\alpha_2, h^M)$, to compute $ans_2 = (\beta_2, \gamma_2, \delta_2)$; compute $(\gamma_1, \cdot) = h^M(a, b_1, b_2, \beta_1, \beta_2)$; complete the simulation of $(\tilde{P}_2^n, V^0|_{\alpha_1, \gamma_1})(f,a,b_1)$ to obtain δ_1 (the notation $V^0|_{\alpha_1,\gamma_1}$ means that V^0's first message is restricted to α_1 and V^0's second message is restricted to γ_1). Finally, output the transcript $((a, b_1, b_2, \beta_1, \beta_2), (\gamma_1, \gamma_2), (\delta_1, \delta_2))$ as a forgery of M.

By definition of a strong-impersonator for $\mathcal{H}^M$, the above forgery will succeed with non-negligible probability[6]

[5] We use here the that fact that (P^0, V^0) remains a proof-of-knowledge even if the prover chooses the instance (f,a,b) after receiving α. This fact was proven in [BG02].

[6] Actually, since γ_1 is not uniformly distributed as required for the non-negligible success probability of $(\tilde{P}_2, V^0)$, in the full version we append uniformly chosen γ' to the public-key and xor it with γ. This xor is used instead of γ both in definition of the verdict function of ID^2 and in the forgers algorithm.

(for $n \in S^2_{\mathcal{H}^{\mathcal{M}}}$), and thus the insecurity of $SIG^2_{\mathcal{H}}$ is established.

This completes the proof the failure of the FS paradigm in case $2a$.

The Failure of the FS Paradigm in Case 2b:

In this case, we need to prove the failure of the FS paradigm assuming

1. $(\forall \mathcal{G}\ \exists$ IMPERSONATOR$)$
2. $\neg(\forall \mathcal{G}\ \exists\ strong$-IMPERSONATOR$)$

We construct yet another and final ID scheme, ID^3, which will be secure assuming $\neg(\forall \mathcal{G}\ \exists\ strong$-IMPERSONATOR$)$, and for which it will be easy to forge signatures in $SIG^3_{\mathcal{H}} = FS_{\mathcal{H}}(ID^3)$ (for $n \in S^1_{\mathcal{H}}$) assuming $(\forall \mathcal{G}\ \exists$ IMPERSONATOR$)$.

Fix a hash-function ensemble $\mathcal{G}$ that does not have a $strong$-IMPERSONATOR (one exists by assumption). Take any secure ID scheme ID and extend it by appending a uniformly chosen $f \in \mathcal{F}$ and a uniformly chosen α to the public-key, and by extending the verdict function so as to also accept views of the form

$$\xrightarrow[a, b_1, \beta_1]{}$$

$$\leftarrow \underset{\gamma_1, b_2, q}{\_\_\_\_\_}$$

$$\xrightarrow[\delta_1, ans]{}$$

where $(\alpha; \beta_1; \gamma_1; \delta_1)$ is accepted by $V^0(f, a, b_1)$ and $(q; ans)$ is accepted by $V^{\mathcal{G}}(f, a, b_2)$. Denote this extended ID scheme by $ID^3_{\mathcal{G}}$ (or in short by ID^3).

First, we claim that $ID^3_{\mathcal{G}}$ is secure. This is so, since by assumption no $strong$-IMPERSONATOR for $\mathcal{G}$ exists, which amounts to the security of $ID^3_{\mathcal{G}}$.

Second, we claim that signature scheme $SIG^3_{\mathcal{G},\mathcal{H}} = FS_{\mathcal{H}}(ID^3_{\mathcal{G}})$ is insecure for every $\mathcal{H}$. Fix any message M to be forged. Define $\mathcal{H}^M = \{h^M\}_{h \in \mathcal{H}}$ where $h^M(x) = h(x, M)$. The crux of the idea is that we would like to take advantage of the existence of an impersonator which succeeds against both $\mathcal{H}^{\mathcal{M}}$ and $\mathcal{G}$, in order to produce a forgery. One problem in doing so is that $\mathcal{H}^{\mathcal{M}}$ takes the triple (a, b_1, β_1) to a triple (γ_1, b_2, q) whereas impersonators are defined with respect to hash functions which take β to γ. Dealing with the incompatible size of the output of $\mathcal{H}^M$ is easy, by redefining $\mathcal{H}^M$ to truncate its output to be restricted to the γ_1 part. To address the incompatible size of the input, we again (as in case $2a$) actually modify ID^3 so that the first message of the the sender is a length reducing (computationally binding) commitment to (a, b_1, β_1) (decommitted in the senders last message) instead of (a, b_1, β_1) itself. As before, we ignore this problem and the modification in this extended abstract (for details see full version) and proceed as if there exists an impersonator for $\mathcal{H}^M$.

We define the combined ensemble $\mathcal{H}' = \mathcal{G} \cup \mathcal{H}^{\mathcal{M}}$ and we use an impersonator for $\mathcal{H}'$.

Formally, an impersonator for $\mathcal{H}'$ is formulated as two polynomial-size circuit families: $F_1 = \{F_1^n\}$ and $\tilde{P}_1 = \{\tilde{P}_1^n\}$, where F_1 is used to generate the first message a and $\tilde{P}_1$ is a cheating prover used to interact with $V^{\mathcal{H}'}(f, a, b)$, for random b. It is required that there exists a polynomial $p(\cdot)$ such that for for every $n \in S^1_{\mathcal{H}'}$,

$$Pr[(\tilde{P}_1^n, V^{\mathcal{H}'})(f, a, b) = 1\ :\ a = F_1^n(f)] \geq \frac{1}{p(n)}$$

(where the probability is over $f \in_R \mathcal{F}_n$, over $b \in_R \{0,1\}^n$ and over the random coin tosses of $V^{\mathcal{H}'}$).

We are finally ready to describe a forger for a fixed message M. On input a random verification key $VK = (PK, f, \alpha, h)$, where $h \in_R \mathcal{H}$, the forger operates as follows: compute $a = F_1^n(f)$; choose b_1 at random; simulate $(\tilde{P}_1^n, V^{\mathcal{H}^M}|_{q_1})(f, a, b_1)$ where $q_1 = (\alpha, h^M)$ to obtain $ans_1 = (\beta_1, \gamma_1, \delta_1)$ such that $(\alpha; \beta_1; \gamma_1; \delta_1)$ is an accepted transcript by $V^0(f, a, b_1)$ and such that $\gamma_1 = h^M(\beta_1)$; compute $(\cdot, b_2, q) = h^M(a, b_1, \beta_1)$; Finally, simulate $(\tilde{P}_1^n, V^{\mathcal{G}}|_q)(f, a, b_2)$ to obtain an ans such that $(q; ans)$ is an accepting transcript by $V^{\mathcal{G}}(f, a, b_2)$. Output the transcript $((a, b_1, \beta_1), (\gamma_1, b_2, q), (\delta_1, ans))$ as forgery for M.

By definition of an impersonator for $\mathcal{H}'$, the above forgery will succeed with non-negligible probability[7] (for $n \in S^1_{\mathcal{H}'}$), and thus the insecurity of $SIG^2_{\mathcal{H}}$ is established. This proves the failure of the FS paradigm in case 2b.

Figure 1 summarizes the outline of the proof of Theorem 2.

4 On the Insecurity of FS Modifications

Two modifications of the FS paradigm were considered in the literature: One due to Micali and Reyzin [MR02] and the other due to Abdalla, An, Bellare and Nampremre [AABN02].

Micali in Reyzin presented a method for constructing FS-like signature schemes that yield better "exact security" than the original FS method. In their method, they convert

[7] Actually, we encounter a few technical complications. First, to ensure that our choice of $\mathcal{H}'$ implies there exists an $(F_1^n, \tilde{P}_1^n)$ which succeeds with non-negligible probability in fooling both $V^{\mathcal{G}}$ and $V^{\mathcal{H}^{\mathcal{M}}}$, requires reducing the soundness error in the original $(P^{\mathcal{H}}, V^{\mathcal{H}})$. Second, by assumption, when b_2 and q are randomly chosen $(\tilde{P}_1^n, V^{\mathcal{G}}|_q)(f, a, b_2)$ will produce an accepting transcript $(q; ans)$ with non-negligible probability. However, b_2, q are not random in the above forger sketch. We resolve this, by appending a uniformly chosen (b_2', q') to the public-key, and running $(\tilde{P}_1^n, V^{\mathcal{G}}|_{q \oplus q'})(f, a, b_2 \oplus b_2')$ instead both in the definition of the verdict function of ID^3 and in the forgers algorithm. For details we refer the reader to the full version.

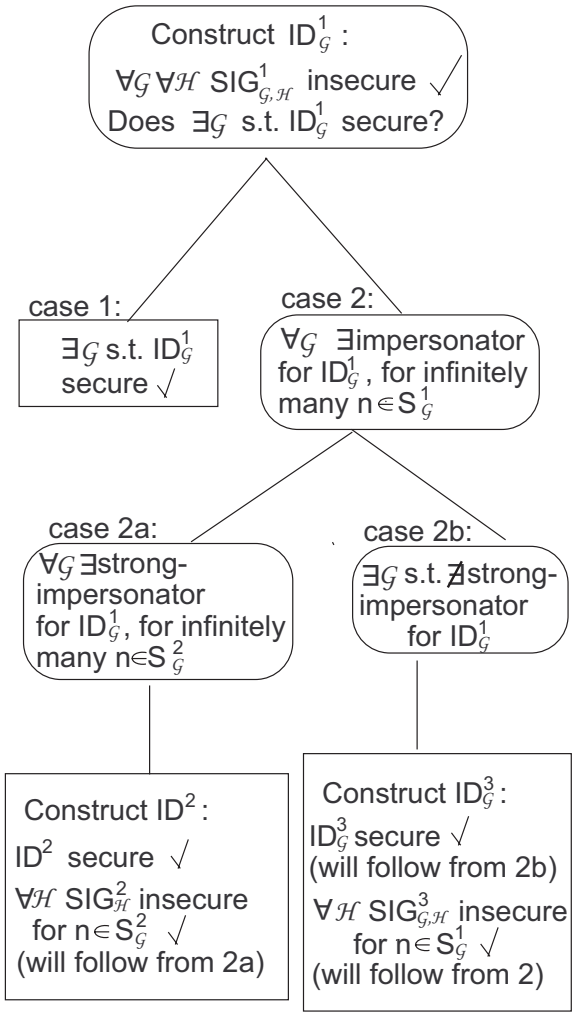

Figure 1. Proof of Theorem 2

any ID scheme $(\alpha; \beta; \gamma)$ into a signature scheme, in which the signer first chooses β and only then produces α by computing $\alpha = h(\beta, M)$, where M is the message to be signed and h is the function used to reduce interaction.[8]

Abdalla et. al. defined a randomized generalization of the FS paradigm, and showed that signature schemes, obtained from the generalized FS paradigm, are secure (resp. forward secure) in the Random Oracle Model if and only if the underlying ID scheme is secure (resp. forward secure) against impersonation under passive attacks. Their randomized method transforms any canonical ID scheme $(\alpha; \beta; \gamma)$ into a signature scheme by replacing the random β with $h(\alpha, M, R)$, where M is the message to be signed, h is the function used to reduce interaction, and R is randomness chosen by the signer.

Using similar ideas to the ones presented in this paper, one can show the failure of these two FS modifications.

[8] Note that this method can be applied only to ID schemes in which the sender can compute γ only given (SK, PK, α, β), and does not need any additional information on α.

5 Future Directions

We have shown examples of digital signature schemes, that are obtained from secure identification schemes by applying the Fiat-Shamir Paradigm, and are insecure regardless of which "hash" function is used. Several related questions arise.

1. Our proof does not imply that the ID schemes used in practice such as [FFS88] or [Sch91] combined with some particular hash function ensemble H necessarily yield insecure digital signature schemes. It does imply that a proof of security would have to involve the particulars of the ID scheme and the H in question. Can one exhibit a proof of security (based on standard intractability assumptions) of $FS_{\mathcal{H}}(ID)$ for *any* practiced ID scheme and *any* H.

2. We showed that the FS paradigm and its known modifications [MR02, AABN02] fail. But, perhaps there exists another general efficient transformation from secure interactive ID schemes to digital signature schemes which can be proven secure?

3. Do there exist other "natural" cryptographic practices which are secure in the Random Oracle Model, and become insecure when the random oracle is replaced with any public function (chosen at random from some function ensemble)? Many examples of such "natural" practices exist for which no evidence of security exists outside the Random Oracle Model.

   In particular, an example that we are interested in is the non-interactive CS-proofs, constructed by Micali [Mi94] in the Random Oracle Model. Does there exist a language L for which there is no function ensemble $\mathcal{H}$ (replacing the Random Oracle), for which CS-proofs for L remain sound (or remain a proof-of-knowledge).

4. In the ID schemes which we constructed to demonstrate the failure of the FS paradigm, soundness is based on the prover being computationally bounded (i.e it is an argument rather than an interactive proof). Can one show that the Fiat-Shamir paradigm fails for an ID scheme for which soundness holds unconditionally? Note that [FS86] is of this latter type, whereas [Sch91] is an argument.

5. Our proof technique can be viewed as a way to reduce interaction in argument systems while preserving some security properties. Can this be extended to show that there exist 3 round zero knowledge arguments? Currently it is known that using the black box zero-knowledge definition they do not exist.

6 Acknowledgements

We are gratefull to Oded Goldreich and Ran Canetti for useful comments on this work.

References

[AABN02] M. Abdalla, J. An, M. Bellare and C. Namprempre. From identification to signatures via the Fiat-Shamir transform: minimizing assumptions for security and forward-security. *Advances in Cryptography-EUROCRYPT 02, Lecture Notes in Computer Science, Springer-Verlag*, 2002.

[Bar01] B. Barak. How to go beyond the black-box simuation barrier. In *Proc. of the 42nd FOCS*, 2001.

[BF01] D. Boneh and M. Franklin. Identity-based encryption from Weil pairing. Preliminary version in *Crypto*, 2001.

[BG02] B. Barak and O. Goldreich. Universal arguments and their applications. *Proceedings of the 17th IEEE Annual Conference on Computational Complexity*, 2002.

[BGI+01] B. Barak, O. Goldreich, R. Rudich, A. Sahai, S. Vadhan, and K.Yang. On the (im)possibility of obfuscating programs. In *Crypto* 2001.

[BM84] M. Blum, S. Micali. How to Generate Cryptographically Strong Sequences of Pseudo-Random Bits. *SIAM J. Comput. 13(4): 850-864* 1984.

[BR93] Mihir Bellare and Phillip Rogaway. Random oracles are practical: a paradigm for designing efficient protocols. In *Proceedings of the First Annual Conference on Computer and Communications Security*. ACM, November 1993.

[CGH98] Ran Canetti, Oded Goldreich, and Shai Halevi. The random oracle methodology, revisited. In *Proceedings of the 32nd Annual ACM Symposium on Theory of Computing*, pages 209-218, Dalllas, 23-26 May, 1998.

[CS99] R. Cramer and V. Shoup. Signature schemes based on the strong RSA assumption. In *5th ACM Conference on Computer and Communications Security*, pages 46-51. Singapore, Nov. 1999. ACM Press.

[DH76] W. Diffie and M.E. Hellman. New directions in cryptography. *IEEE Transactions on Information Theory*, IT-22 (Nov.), pages 644-654, 1976.

[DNRS99] C. Dwork, M. Naor, O. Reingold and L.Stockmeyer. Magic funcitons. In *IEEE, editor, 40th Annual Symposium of Foundations of Computer Science*: October 17-19, 1999, New York City, New York, pages 523-534. *IEEE Computer Society Press*, 1999.

[FFS88] U. Feige, A. Fiat, and A. Shamir. Zero knowledge proofs of identity. *Journal of Cryptology*, 1(2), pp. 77-94, 1988.

[FS86] Amos Fiat and Adi Shamir. How to prove to yourself: practical solutions to identification and signature problems. In *Advances in Cryptology–Crypto 86,* pages 186-194, Springer, Berlin, 1987.

[Gol01] Oded Goldreich. Foundations of Cryptography, volume 1 – Basic Tools. *Cambridge University Press*, 2001.

[GHR99] R. Gennaro, S. Halevi and T. Rabin. Secure hash-and-sign signatures without the random oracle. *Advances in Cryptology - EUROCRYPT 99*, Lecture Notes in Computer Science Vol. 1592, J. Stern ed., Springer-Verlag, 1999.

[GGM86] Oded Goldreich, Shafi Goldwasser and Silvio Micali. How to construct random functions. *Journal of the Association of Computing Machinery*, 33(4): 792-807, 1986.

[GMR88] S. Goldwasser, S. Micali, and R. Rivest. A digital signature scheme secure against adaptive chosen-message attacks. *SIAM Journal of Computing*, 17(2):281-308, April 1988.

[GQ88] L. Guillou and J. J. Quisquater. A "paradoxical" identity-based signature scheme resulting from zero-knowledge. *Advances in Cryptology-CRYPTO 88*, Lecture Notes in Computer Science Vol. 403, S. Goldwasser ed., Springer- Verlag, 1988.

[Ki92] J. Kilian. A note on efficient zero-knowledge proofs and arguments. In *24'th STOC*, pages 723-732, 1992.

[Mer90] R.C. Merkle. A certified digital signature. *Proceedings on Advances in Cryptology*, pages 218-238, July 1989, Santa- Barbara, California.

[Ok92] T. Okamoto. Provably secure and practical identification schemes and corresponding signature schemes. *Advances in Cryptology - CRYPTO 92*, Lecture Notes in Computer Science Vol. 740, E. Brickell ed., Springer-Verlag, 1992.

[MD5] R. Rivest. The MD5 message-digest algorithm. *RFC* 1321, April 1992.

[Mi94] Silvio Micali. Computationally sound proofs. *SICOMP*, vol. 30(4), pages 1253-1298, 2000. Preliminary version in *35th FOCS*, 1994.

[MR02] S. Micali and L. Reyzin. Improving the exact security of digital signature schemes. *Journal of Cryptology*, 15(1):1-18, 2002.

[Na91] M. Naor. Bit commitment using pseudorandom generators. *Journal of Cryptology*, Vol.4, pages 151-158, 1991.

[NY89] M. Naor and M. Yung. Universal one-way hash functions and their cryptogracphic applications. *STOC 89*.

[PS96] D. Pointcheval and J. Stern. Security proofs for signature schemes. In *Advances in Cryptology-EUROCRYPT 96*, vol.1070 of Lecutre Notes in Compute Science, pages 387-398. Springer-Verlag, 1996.

[Rom90] John Rompel: One-Way Functions are Necessary and Sufficient for Secure Signatures. STOC 1990: 387-394.

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

Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science (FOCS'03)
0272-5428/03 $17.00 © 2003 **IEEE**