Challenges in Eliminating Interaction from Proof Systems

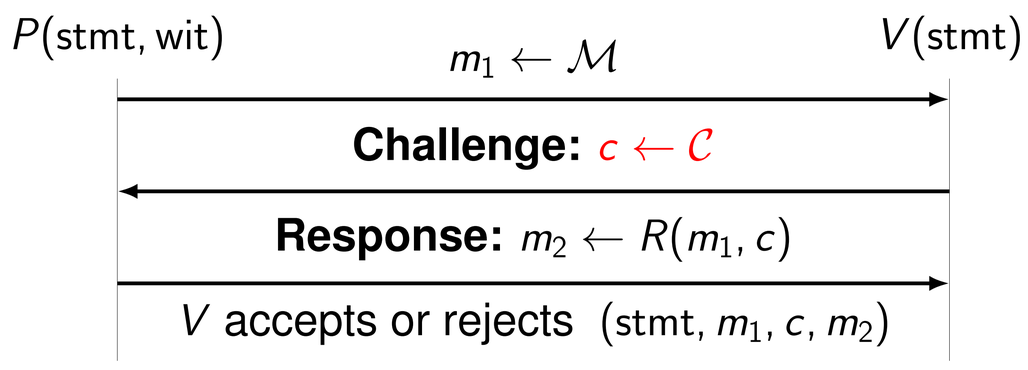
Aleksi Kalsta

Introduction

An interactive proof is a protocol between a prover P and a verifier V in which they exchange messages, and the prover aims to convince the verifier of the validity of a statement. A statement of interest could be, for instance, "I know the discrete logarithm of this group element" or "the value of the circuit $\mathcal{C}$ on input x is y".

The minimum requirements for proof systems are completeness and soundness:

- **Completeness:** An honest prover is able to convince the verifier if the statement is true.
- **Soundness:** A cheating prover cannot convince the verifier of a false statement.

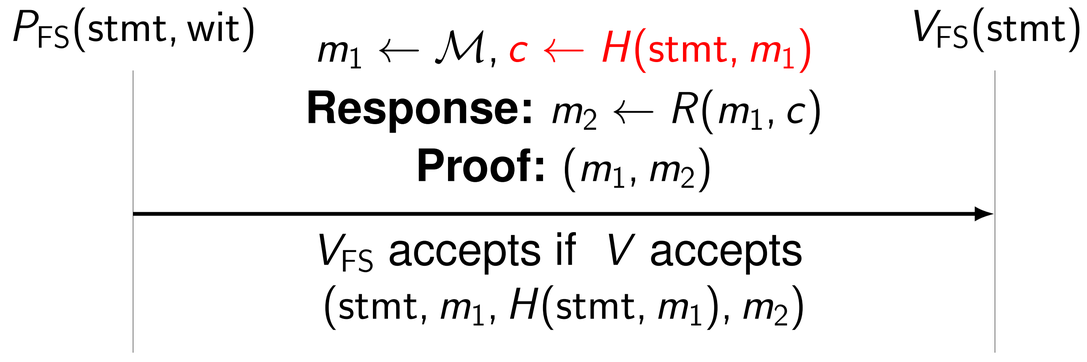


Arbitrary three-message interactive proof.

We would like to avoid interaction because:

- Communication increases latency and bandwidth usage.
- The verifier must be online during the proof generation.
- Each verifier must be convinced separately.
- Interactive proofs are harder to implement correctly and securely.

The Fiat–Shamir transform converts a public-coin interactive protocol into a non-interactive one by replacing random challenges of the verifier with the output of a cryptographic hash function $H \leftarrow \mathcal{H}$ applied to the intermediate transcript.



Fiat-Shamir transformed three-message non-interactive proof.

Question: Does this transformation preserve the properties of the original proof system?

Random Oracle Model (ROM)

The ROM is an idealized model for security analysis, where hash functions are modeled as truly random functions. Intuitively, a function H returns a uniformly random output for each new input, while responding consistently to repeated queries.

- **Security:** Many protocols remain secure under the FS transform in the ROM, when the random oracle is sampled independently of the underlying proof system.
- **Uninstantiability:** Prior works show that the random oracle cannot be instantiated by any efficient algorithm in practice.

Known limitations of Fiat-Shamir transform

A long line of work shows that there exist (contrived) proof systems for which any efficient hash function produces an insecure non-interactive proof after the FS transform.
A recent breakthrough result by Khovratovich et al. (KRS, [1]) shows that even a deployed succinct variant of the GKR protocol becomes insecure under the FS transform.

- **The KRS attack** embeds the FS hash function into the statement being proven, allowing the circuit to predict the challenge in the FS transformed non-interactive proof system.

An attempt to patch the Fiat-Shamir transform

In response to the KRS attack, Arnon and Yogev (AY, [2]) proposed an extended Fiat-Shamir transform (XFS) to address the limitations of the standard variant. They introduce a proof-of-work (PoW) mechanism and a prefix-avoiding padder with the following goals:

- **PoW:** Produces a solution s that is computationally infeasible for the circuit (i.e., the proven statement) to compute.
- **Padder:** Generates a padding value pad which is computationally infeasible for the original verifier circuit to obtain.

The challenge in the XFS transform is computed as $H(\text{stmt}, m_1, s, \mathsf{pad})$, with the aim that neither the proven circuit nor the original interactive verifier can predict the challenge value. However, security was established only in the relativized ROM, leaving open whether it extends to the standard model.

Attack on the extended Fiat-Shamir transform [3]

We show that the XFS transform proposed by AY is not secure in the standard model in general.

> **Theorem 1** (informal)**.** *If the Learning With Errors (**LWE**) assumption holds, then any secure three-message proof system Π can be transformed into another secure three-message proof system Π' such that the XFS transform of Π' is insecure for any efficiently computable hash function H.*

The high-level idea is to use a succinct proof system to overcome the verifier's inability to compute the challenge by delegating this computation to the prover, and then leverage the challenge to define a proof system that becomes insecure after the XFS transform.

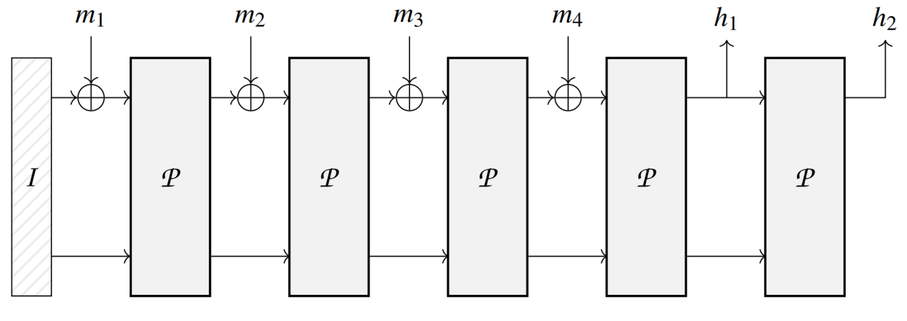
Our transformation from $\Pi = (P, V)$ to $\Pi' = (\mathcal{P}, \mathcal{V})$ relies on the following observations:

- We can add a "cheating branch" for $\mathcal{V}$, which is accepted if the XFS challenge matches the verifier's random challenge.
- In the interactive setting, the verifier's randomness is unlikely to match the XFS challenge. However, after applying the XFS transformation, the verifier's challenge matches the XFS challenge by construction, enabling a malicious prover to prove arbitrary statements in Π'.
- Using LWE assumption, one can construct a succinct non-interactive proof system Π_{SNARG} for Turing machine computations.
- $\mathcal{P}$ and $\mathcal{V}$ can employ Π_{SNARG} to delegate the computation of the XFS challenge to $\mathcal{P}$, allowing the verifier to use the challenge value without evaluating the hash function on pad directly, which would be computationally infeasible.

More practical view on FS security

- In many practical applications we have to prove the correct evaluation of a cryptographic hash function.
- In such cases, algebraic hash functions are preferable, as they require fewer constraints when represented in proof-system-friendly formats, such as arithmetic circuits.

Many hash functions are built from permutations using the Sponge construction, where the permutation is applied repeatedly to message blocks until the entire message is absorbed.



The sponge construction using the permutation $\mathcal{P}$.

The security of many algebraic hash functions, such as Poseidon, is evaluated by studying the used permutation under the so-called *Constrained-Input Constrained-Output* (CICO) problem.

Definition 1 (**CICO-k problem**)**.** *Given a permutation $\mathcal{P} : \mathbb{F}^n \to \mathbb{F}^n$, the goal is to find $\mathbf{x} \in \{0\}^k \times \mathbb{F}^{n-k}$ such that $\mathcal{P}(\mathbf{x}) \in \{0\}^k \times \mathbb{F}^{n-k}$.*

Relation Between FS Security and the CICO Problem

Question: How does the hardness of CICO problem relate to the security of Fiat-Shamir transformed proof systems if the Fiat-Shamir hash is instantiated using Poseidon?
We studied this question in my thesis [4], and we were able to give a partial answer to the other direction.

> **Theorem 2** (informal)**.** *If one can solve the restricted variant of the CICO-1 problem, then the Fiat-Shamir transformed SumCheck protocol is insecure.*

- The restricted version of the CICO-1 problem requires that the input is drawn from an affine subspace and that the output has a zero in its first coordinate.
- This restricted variant is roughly as hard as the standard CICO-1 problem.
- The attack extends naturally to a wide class of interactive protocols that use the SumCheck protocol as a subroutine.

Conclusions

- Provably secure non-interactive proof systems in the standard model are difficult to construct.
- Still, designing attacks against the FS heuristic for natural protocols is also challenging.

Open questions

- Are there natural counterexamples to the XFS transform?
- Is it possible to design an attack against a proof system that uses Poseidon to instantiate the FS transform, such that the attack's complexity is strictly lower than that of the corresponding CICO problem?

References

[1] Dmitry Khovratovich, Ron D. Rothblum, and Lev Soukhanov. How to prove false statements: Practical attacks on fiat-shamir. In Yael Tauman Kalai and Seny F. Kamara, editors, *Advances in Cryptology – CRYPTO 2025, Part VI*, volume 16005 of *Lecture Notes in Computer Science*, pages 3–26. Springer, Cham, August 2025.

[2] Gal Arnon and Eylon Yogev. Towards a white-box secure fiat–shamir transformation. In Yael Tauman Kalai and Seny F. Kamara, editors, *Advances in Cryptology – CRYPTO 2025, Part VI*, volume 16005 of *Lecture Notes in Computer Science*, pages 27–56. Springer, Cham, August 2025.

[3] Pavel Hubáček, Chris Brzuska, and Aleksi Kalsta. Simple attacks against (extended) fiat-shamir. In *Public Key Cryptography – PKC 2026*.

[4] Aleksi Kalsta. From interaction to insecurity: New attacks on the (extended) fiat–shamir transform. Msc thesis, Aalto University, Espoo, Finland, 2026.

Contact: Aleksi Kalsta, Research Trainee
Tel. +358 50 443 2975, aleksi.kalsta@vtt.fi