AICR · OPEN PROBLEM

A Quadratic-Advice Two-Block Sponge Collision Attack with Preprocessing

Lifting the multi-instance attack to the auxiliary-input random permutation model

Status. Statement: AI-written from Akshima, Xiaoqi Duan, Siyao Guo and Qipeng Liu, *On Time-Space Lower Bounds for Finding Short Collisions in Sponge Hash Functions*, TCC 2023 (IACR ePrint 2023/1444), not yet checked by a human. Settled in the multi-instance model, where the source paper's Theorem 3 gives an adversary of advantage $(\tilde{\Omega}(S^2T^4/C^2))^S$; open in the auxiliary-input model, which is what the source paper asks for. No partial result in the auxiliary-input model is reported.

Category. *Symmetric-Key Cryptography*, *Idealized Models.*

Conjecture (informal). *The best proved security bound for finding a two-block collision in sponge hashing against an attacker with S bits of preprocessed advice and T online queries has a term that is quadratic in the advice length: about S^2T^4 over the squared capacity. Nobody knows an attack that reaches it; the best known attack is only about ST over the capacity. But the same paper that proves the bound also exhibits an attack reaching it in a related setting — the multi-instance game, in which an adversary carries no advice and must solve S freshly sampled salts one after another, and where advantage of the form δ^S corresponds to advantage δ with S bits of advice. The conjecture is that the multi-instance attack can be turned into a genuine preprocessing attack of advantage about S^2T^4 over the squared capacity. If it can, the sponge analogue of the STB conjecture is false whenever ST^3 exceeds the capacity, and sponge hashing is strictly weaker than Merkle-Damgård against preprocessing at two blocks as well as at one.*

1 The setting

Sponge hashing is the domain extension standardized in SHA-3: a permutation F on $r + c$ bits is iterated, each message block is

exclusive-ored into the r-bit outer part of the state, and the c-bit inner part — the *capacity* — is initialized to a salt and never touched by the message directly. Against preprocessing the salt is what makes collision-finding non-trivial. The model is the auxiliary-input random-permutation model of Coretti, Dodis and Guo [CDG18]: an unbounded first stage sees all of the random permutation F and outputs S bits of advice; a second stage receives the advice and a uniformly random salt and makes T queries to F or to F^{-1}.

Two models are in play, and the whole question is the gap between them. In the *multi-instance* (MI) model an adversary gets no advice but is handed S independently sampled challenge salts one at a time, and must find a collision for each before seeing the next; it may reuse the queries it made in earlier rounds. The reduction that makes MI useful runs one way: if no MI adversary wins with advantage better than δ^S, then no (S,T)-AI adversary wins with advantage better than 2δ (the source paper's Theorem 6, after [AGL22]). Every known preprocessing bound for sponge is proved this way. The converse — turning an MI adversary of advantage δ^S into an AI adversary of advantage about δ — is not a theorem, and this statement is about one instance of it.

What is known at $B = 2$. Writing $R = 2^r$, $C = 2^c$, and suppressing polylogarithmic factors:

- *Best known attack.* $\tilde{\Omega}(ST/C + T^2/\min(C,R))$, due to Freitag, Ghoshal and Komargodski [FGK22].
- *Best proved security bound.* $\tilde{O}(ST/C + S^2T^4/C^2 + T^2/\min(C,R))$ — the source paper's Theorem 2, recovering [FGK22] by a simpler argument. The middle term dominates exactly when $ST^3 > C$, and in that regime the bound and the attack are a factor ST^3/C apart.
- *The middle term is not an artefact of the proof.* The source paper's Theorem 3 exhibits an MI adversary for two-block collisions with advantage $(\tilde{\Omega}(S^2T^4/C^2))^S$, so no argument routed through the MI reduction can remove it. The multi-instance side of the question is therefore settled; only the lift to auxiliary input is missing.

What the source paper asks. "*Better attacks for $B = 2$?* The current security upper bound for $B = 2$ suggests that there may exist an attack with advantage" $\Omega(S^2T^4/C^2)$ "in the auxiliary-input random permutation model. And we show an attack in the multi-instance model with advantage" $(\Omega(S^2T^4/C^2))^S$. "Can we utilize similar ideas to show a corresponding attack in the auxiliary-input random permutation model?" (p. 9).

Why it is hard. The MI attack the source paper gives is stateful across rounds by design: for $B = 2$ it spends half of each round's queries on inverse queries of the form $F^{-1}(0,*)$ and the other half trying to hit two of those from the current challenge salt, and it relies on the pairs accumulated in *earlier* rounds to make the later ones likely. That is precisely the structure an advice string of S bits cannot obviously supply: the MI adversary's advantage comes from $\tilde{\Omega}(iT^2/C)$-many useful pairs available at round i, built by iT queries already spent, whereas an AI adversary must compress the same resource into S bits chosen before the salt is known. The source paper notes in the adjacent open problem that its own MI attacks "require knowing queries from previous rounds" and therefore do not apply to stateless multi-instance games; the same feature is what stands between them and an AI attack.

Why settling it matters. Affirmatively, it would refute the sponge STB conjecture — that $\tilde{\Omega}(STB/C + T^2/\min(R,C))$ is optimal for all $B \geq 2$ — in the whole regime $ST^3 > C$, since $S^2T^4/C^2 > ST/C$ exactly there. It would also extend to two blocks the finding of [FGK22] at one block, that sponge hashing is provably less secure against preprocessing than Merkle-Damgård; for Merkle-Damgård the two-block advantage is known to be $\tilde{\Theta}(ST/N + T^2/N)$ [ACDW20, AGL22], with no S^2 term. Negatively, a proof that no such AI attack exists would have to beat the multi-instance technique, which the source paper's Theorem 3 shows cannot prove a better bound — so it would supply exactly the "other novel techniques" the paper says are needed.

2 Notation and parameters

- r is the bitrate and c the capacity, $R := 2^r$ and $C := 2^c$; the permutation is $F : [R] \times [C] \to [R] \times [C]$, message blocks lie in $[R]$ and salts in $[C]$.
- $S \in \mathbb{N}$ is the advice length in bits, $T \in \mathbb{N}$ the number of online queries, each to F or F^{-1}.
- Asymptotics are in C and R, uniformly over S and T; $\tilde{O}$ and $\tilde{\Omega}$ suppress factors polylogarithmic in R and C.
- All algorithms are information-theoretic: only queries are counted.

3 The conjecture

Definition 1 (Sponge hashing). Let $F : [R] \times [C] \to [R] \times [C]$. For $m = m_1 \| \cdots \| m_B$ with each $m_i \in [R]$ and a salt $a \in [C]$, define $\mathsf{SP}_F(m, a)$ by setting $(x_0, y_0) := (0, a)$, computing

$$(x_i, y_i) := F(x_{i-1} \oplus m_i,\ y_{i-1}) \qquad \text{for } i = 1, \ldots, B,$$

and returning x_B.

Definition 2 (Auxiliary-input two-block collision resistance). A pair $\mathcal{A} = (\mathcal{A}_1, \mathcal{A}_2)$ is an (S, T)-*AI adversary* if $\mathcal{A}_1$ has unbounded access to F and F^{-1} and outputs S bits of advice σ, and $\mathcal{A}_2$ takes σ and a challenge salt $a \in [C]$, makes T queries to F or F^{-1}, and outputs m, m'. The game $2\text{-}\mathsf{AICR}_{F,a}(\mathcal{A})$ returns 0 if m or m' has more than two blocks, returns 1 if $m \neq m'$ and $\mathsf{SP}_F(m, a) = \mathsf{SP}_F(m', a)$, and returns 0 otherwise. Write

$$\mathsf{Adv}^{\mathsf{AICR}}_{2\text{-}\mathsf{SP}}(S, T) = \max_{\mathcal{A}} \Pr\left[2\text{-}\mathsf{AICR}_{F,a}(\mathcal{A}) = 1\right],$$

the maximum over all (S, T)-AI adversaries, with F a uniformly random permutation, $a \leftarrow [C]$ uniform and independent, and the coins of $\mathcal{A}$.

Definition 3 (Multi-instance two-block collision resistance). A stateful algorithm $\mathcal{A}$ is an (S,T)-*MI adversary* if for each $i \in [S]$ it takes a salt $a_i \in [C]$, makes T queries to F or F^{-1}, and outputs m_i, m_i'. In the game $2\text{-MICR}_F^S(\mathcal{A})$ the salts $a_1, \ldots, a_S$ are sampled uniformly from $[C]$ without replacement and presented one at a time; the game returns 1 only if for every i the messages $m_i \neq m_i'$ have at most two blocks each and $\mathsf{SP}_F(m_i, a_i) = \mathsf{SP}_F(m_i', a_i)$. $\mathsf{Adv}^{\mathsf{MICR}}_{2\text{-SP}}(S,T)$ is the maximum of its winning probability over all (S,T)-MI adversaries. $\mathcal{A}$ carries no advice, but keeps its state — including the answers to all queries made in earlier rounds — from one round to the next.

Theorem 1 (the multi-instance attack; source paper, Theorem 3). *Suppose $S,T,R \geq 16$. Then*

$$\mathsf{Adv}^{\mathsf{MICR}}_{2\text{-SP}}(S,T) \;\geq\; \left(\tilde{\Omega}\left(\frac{S^2T^4}{C^2}\right)\right)^S.$$

Conjecture 1 (a matching auxiliary-input attack). *There is an absolute constant $\kappa > 0$ such that for all R,C,S,T with $S,T,R \geq 16$ and $S^2T^4 \leq C^2$,*

$$\mathsf{Adv}^{\mathsf{AICR}}_{2\text{-SP}}(S,T) \;\geq\; \kappa \cdot \frac{S^2T^4}{C^2 \cdot \mathrm{polylog}(R,C)},$$

that is, $\mathsf{Adv}^{\mathsf{AICR}}_{2\text{-SP}}(S,T) = \tilde{\Omega}(S^2T^4/C^2)$: some (S,T)-AI adversary finds two-block sponge collisions with advantage matching, up to polylogarithmic factors, the source paper's Theorem 2 security bound.

Remark 1 (the side condition). The restriction $S^2T^4 \leq C^2$ is a formalization choice, not a condition taken from the source paper, which states the target advantage as $\Omega(S^2T^4/C^2)$ with no range. It is imposed because an advantage is a probability: without it the claim would read as an assertion that a probability exceeds one, which cannot be what is meant. A reviewer who prefers a different normalization — for instance $\min\{1, S^2T^4/C^2\}$ — should say so.

Remark 2 (what the source paper does and does not claim). Theorem 1 is the source paper's, and so is the observation that its

own Theorem 2 bound "suggests that there may exist an attack". The conjecture that such an AI attack exists is a reading of the question the paper poses; the paper phrases it as a question and does not predict the answer, and its "suggests" and "may" are hedges that are reproduced here rather than removed. The attack whose optimality would be refuted is that of [FGK22], not the source paper's.

Remark 3 (relation to the sponge STB conjecture). Since $S^2T^4/C^2 > ST/C$ if and only if $ST^3 > C$, an affirmative resolution of Conjecture 1 refutes the sponge STB conjecture — that $\tilde{\Omega}(STB/C + T^2/\min(R,C))$ is optimal for every $B \geq 2$ — throughout the regime $ST^3 > C$. A negative resolution leaves that conjecture open. The two statements are therefore linked but distinct, and the source paper poses them as consecutive, separate open problems.

Remark 4 (which lift is being asked for). The MI-to-AI direction that is a theorem is the contrapositive one: a bound of δ^S on MI advantage yields a bound of 2δ on AI advantage (source paper's Theorem 6). Theorem 1 is an *attack*, and no general principle turns an MI attack of advantage δ^S into an AI attack of advantage $\Omega(\delta)$. Conjecture 1 asserts the conclusion of such a lift in this one case; it does not assert the general principle, which is false in other settings where MI and AI security are known to separate.

4 Bibliography

[ACDW20] Akshima, David Cash, Andrew Drucker, and Hoeteck Wee. *Time-space tradeoffs and short collisions in Merkle-Damgård hash functions.* In Daniele Micciancio and Thomas Ristenpart, editors, Advances in Cryptology — CRYPTO 2020, Part I, volume 12170 of Lecture Notes in Computer Science, pp. 157–186. Springer, 2020.

[ADGL23] Akshima, Xiaoqi Duan, Siyao Guo, and Qipeng Liu. *On time-space lower bounds for finding short collisions in sponge hash functions.* TCC 2023; IACR ePrint 2023/1444. (The source paper of this statement.)

[AGL22] Akshima, Siyao Guo, and Qipeng Liu. *Time-space lower bounds for finding collisions in Merkle-Damgård hash functions.* In Yevgeniy Dodis and Thomas Shrimpton, editors, Advances in Cryptology — CRYPTO 2022, Part III, volume 13509 of Lecture Notes in Computer Science, pp. 192–221. Springer, 2022.

[CDG18] Sandro Coretti, Yevgeniy Dodis, and Siyao Guo. *Non-uniform bounds in the random-permutation, ideal-cipher, and generic-group models.* In Hovav Shacham and Alexandra Boldyreva, editors, Advances in Cryptology — CRYPTO 2018, Part I, volume 10991 of Lecture Notes in Computer Science, pp. 693–721. Springer, 2018.

[FGK22] Cody Freitag, Ashrujit Ghoshal, and Ilan Komargodski. *Time-space tradeoffs for sponge hashing: attacks and limitations for short collisions.* In Yevgeniy Dodis and Thomas Shrimpton, editors, Advances in Cryptology — CRYPTO 2022, Part III, volume 13509 of Lecture Notes in Computer Science, pp. 131–160. Springer, 2022.