CONJURA · OPEN PROBLEM

The Sponge STB Conjecture

Is the short-collision sponge attack optimal for every block bound $B \geq 2$?

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 only at $B \approx T$ (by [CDG18]) and, for $B = 2$, in the regime $ST^3 \leq C$ (by the source paper's Theorem 2, recovering [FGK22]); open for every other $B \geq 2$, and the source paper poses "is the STB-conjecture true for $B = 2$ in sponge?" as the first case to attack. Note that the source paper asks for a proof *or* a refutation and takes no position on which.

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

Conjecture (informal). *Sponge hashing, the construction behind SHA-3, absorbs message blocks into the first part of a state by exclusive-or and stirs the whole state with a fixed permutation; the second part of the state, the capacity, is never touched by the message and starts at the salt. Against an attacker who may compute for free on the permutation beforehand, keep S bits of notes, and then make T online queries — forwards or backwards, since the permutation is invertible — the best known attack finding a collision on at most B blocks succeeds with probability about STB over the capacity plus about T^2 over the smaller of the bitrate and the capacity. The conjecture is that this is optimal for every B from two upwards, exactly as is conjectured for Merkle-Damgård. Two things make it hard. Sponge is genuinely different from Merkle-Damgård at $B = 1$, where it is provably* less *secure, so the analogy is not automatic. And the source paper proves that the one technique that has produced all the known bounds in this setting cannot prove the conjecture for any B at all: in the multi-instance game it exhibits attacks strong enough to block that route.*

1 The setting

The sponge paradigm is the domain extension standardized in SHA-3. A permutation F on $r + c$ bits is fixed; the state is split into an

r-bit outer part and a c-bit inner part (the *capacity*); each message block is exclusive-ored into the outer part and the whole state is pushed through F. In the auxiliary-input setting the inner part is initialized to a random *salt*, which is what stops an attacker from simply hardwiring a collision — the same role the salt plays for Merkle-Damgård.

The model is the auxiliary-input random-permutation model of Coretti, Dodis and Guo [CDG18]: F is a uniformly random permutation, an unbounded first stage sees all of it and outputs S bits of advice, and a second stage receives the advice and a uniformly random salt and makes T queries to F *or* to F^{-1}. Inverse queries are the one structural difference from the Merkle-Damgård setting, where the compression function is not invertible, and they are what the attacks below exploit.

For Merkle-Damgård, Akshima, Cash, Drucker and Wee [ACDW20] conjectured that the best (S,T)-attack finding a B-block collision succeeds with probability $\Theta((STB+T^2)/2^n)$ — the STB conjecture. The source paper transports that conjecture to sponge. The bound to be matched is the one achieved by the attack of Freitag, Ghoshal and Komargodski [FGK22]: advantage $\tilde{\Omega}(STB/C + T^2/\min(R,C))$ for $B \geq 2$.

What is proved. Throughout, $R = 2^r$ and $C = 2^c$, and $\tilde{O}$, $\tilde{\Omega}$, $\tilde{\Theta}$ suppress polylogarithmic factors.

- $B \approx T$ (no effective length bound): Coretti, Dodis and Guo [CDG18] prove $\tilde{O}(ST^2/C + T^2/R)$, which matches the known attack there. This is the only B for which the sponge STB conjecture is known.
- $B = 1$: the source paper's Theorem 1 proves $\tilde{O}(S^2T^2/C^2 + S/C + T/C + T^2/R)$, optimal for $ST^2 \leq C$, improving on [FGK22]. (Theorem 1 is stated informally on p. 3 and proved as the paper's Theorem 8 on p. 14.) The $B = 1$ case falls *outside* the conjecture: Freitag et al. [FGK22] showed sponge hashing is provably less secure at $B = 1$ than Merkle-Damgård, where $\tilde{\Theta}(S/C + T^2/R)$ holds.
- $B = 2$: the source paper's Theorem 2 proves $\tilde{O}(ST/C + S^2T^4/C^2 + T^2/\min(C,R))$, recovering the bound of [FGK22]

by a simpler and more modular argument. This equals the conjectured $\tilde{O}(ST/C + T^2/\min(C,R))$ precisely when the middle term does not dominate, that is when $ST^3 \leq C$.

- $B \geq 3$: nothing better than the general $\tilde{O}(ST^2/C + T^2/R)$ of [CDG18] is known. The source paper reports no bound at all in this range.

What is open, and where. The conjecture is open for every $B \geq 2$ except $B \approx T$. Its first open case is $B = 2$ in the regime $ST^3 > C$, where the published bound and the published attack differ by the factor ST^3/C; the source paper singles this case out. For $B \geq 3$ the gap is wider still: a factor of about T/B separates the conjectured STB/C from the only known bound ST^2/C.

Why it is hard: a proved limitation of the only working technique. Every bound listed above is obtained by reduction to a *multi-instance* (MI) game, in which an adversary with no advice must solve S independently sampled challenge salts in sequence: if no MI adversary wins with probability better than δ^S, then no (S,T)-AI adversary wins with probability better than 2δ (the source paper's Theorem 6, after [AGL22]). The source paper's Theorem 3 shows this route is exhausted. It exhibits MI adversaries achieving $(\tilde{\Omega}(S^2T^2/C^2))^S$ for $B = 1$, $(\tilde{\Omega}(S^2T^4/C^2))^S$ for $B = 2$, and $(\tilde{\Omega}(ST^2/C))^S$ for $B = 3$ when $T^2 < R$. Consequently the S^2T^2/C^2 term in its Theorem 1 cannot be removed, no bound better than S^2T^4/C^2 is provable this way for $B = 2$, and for $B \geq 3$ nothing better than ST^2/C is provable this way at all. In the paper's own words: "As the bounds in Theorem 1, Theorem 2 and the general" $O(ST^2/C + T^2/R)$ "bound are the best one can prove using the multi-instance technique, other novel techniques are required to obtain optimal bounds for collision resistance of sponge in the AI setting" (p. 5). A proof of Conjecture 1 therefore cannot go through the multi-instance reduction, and the paper suggests *stateless* multi-instance games as one place to look for a substitute.

Why settling it matters. Sponge is the standardized construction; Merkle-Damgård is the legacy one. Yet the preprocessing security of short collisions is understood for Merkle-Damgård across most

of the parameter range and, for sponge, at essentially one point. A proof would give the first tight short-collision bound for SHA-3's mode of operation. A refutation is equally live and would be the more interesting outcome: because S^2T^4/C^2 exceeds ST/C exactly when $ST^3 > C$, an AI attack matching the source paper's own $B = 2$ upper bound would refute the conjecture in that regime and extend to two blocks the message that [FGK22] established at one block, namely that sponge hashing is strictly weaker than Merkle-Damgård against preprocessing.

2 Notation and parameters

- r is the bitrate and c the capacity, with $R := 2^r$ and $C := 2^c$. $[R]$ is the set of message blocks and $[C]$ the set of salts; the permutation is $F : [R] \times [C] \to [R] \times [C]$.
- $S \in \mathbb{N}$ is the advice length in bits and $T \in \mathbb{N}$ the number of online queries, each to F or to F^{-1}.
- B is the bound on the number of blocks in each of the two colliding messages. As in the source paper, B is a function of R and C rather than a constant, which is what makes the range $B \geq 3$ meaningful.
- Asymptotics are in C and R, uniformly over S, T and B, with absolute hidden constants; $\tilde{O}$, $\tilde{\Omega}$ and $\tilde{\Theta}$ additionally suppress factors polylogarithmic in R and C.
- All algorithms are information-theoretic: only queries are counted.

3 The conjecture

Definition 1 (Sponge hashing). Let $F : [R] \times [C] \to [R] \times [C]$. A message $m = m_1 \| \cdots \| m_B$ with each $m_i \in [R]$ is a *B-block message*. For a salt $a \in [C]$ define $\mathsf{SP}_F(m, a)$ by initializing $(x_0, y_0) := (0, a)$, 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. Two distinct messages $m \neq m'$ form a *collision*

on a if $\mathsf{SP}_F(m,a) = \mathsf{SP}_F(m',a)$.

Definition 2 (Auxiliary-input collision resistance of sponge). 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 B-$\mathsf{AICR}_{F,a}(\mathcal{A})$ runs $\sigma \leftarrow \mathcal{A}_1^F$ and $m, m' \leftarrow \mathcal{A}_2^{F/F^{-1}}(\sigma, a)$; it returns 0 if m or m' consists of more than B blocks; it returns 1 if $m \neq m'$ and $\mathsf{SP}_F(m,a) = \mathsf{SP}_F(m',a)$; otherwise 0. Write

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

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

Conjecture 1 (sponge STB). *For every* $B \geq 2$,

$$\mathsf{Adv}^{\mathsf{AICR}}_{B\text{-}\mathsf{SP}}(S,T) \;=\; \tilde{\Theta}\left(\frac{STB}{C} + \frac{T^2}{\min(R,C)}\right).$$

The lower half is the attack of [FGK22], which the source paper takes as given; the open content is the matching security bound. By Section 1 it is established only for $B \approx T$ and, at $B = 2$, only when $ST^3 \leq C$. The first case the source paper asks for is therefore

$$\mathsf{Adv}^{\mathsf{AICR}}_{2\text{-}\mathsf{SP}}(S,T) \;=\; \tilde{O}\left(\frac{ST}{C} + \frac{T^2}{\min(R,C)}\right) \qquad \text{in the regime } ST^3 > C,$$

and a refutation of that special case refutes Conjecture 1.

Remark 1 (the conjecture is the source paper's; the bound it names is not). The attack whose optimality is conjectured is that of [FGK22], and the Merkle-Damgård conjecture being transported is that of [ACDW20]. What the source paper contributes is the conjecture itself, stated on its p. 8 as "It is natural to consider a similar STB-conjecture for sponge hash functions, conjecturing

the" $\Theta(STB/C + T^2/\min(R,C))$ "attack by Freitag et al. [FGK22] is optimal for $B \geq 2$", together with the Theorem 3 limitation that makes it hard. Neither the attack nor the Merkle-Damgård conjecture is claimed here for the source paper.

Remark 2 (the source takes no side). The source paper writes "It will be extremely interesting to either prove or refute the sponge STB-conjecture. To start with, is the STB-conjecture true for $B = 2$ in sponge?" (p. 9). It offers no prediction, and the sibling question it poses in the next paragraph — whether an AI attack with advantage $\Omega(S^2T^4/C^2)$ exists for $B = 2$ — would, if answered affirmatively, refute Conjecture 1 whenever $ST^3 > C$. The conjecture is recorded here in the affirmative direction because that is the form in which the paper states it; a solver should treat refutation as equally likely on this evidence.

Remark 3 ($\tilde{\Theta}$ rather than Θ). The source paper prints the conjectured quantity as $\Theta(STB/C + T^2/\min(R,C))$, without a tilde, while every bound and attack it states for sponge (including the attack of [FGK22] whose optimality is being conjectured) carries $\tilde{O}$ or $\tilde{\Omega}$. Conjecture 1 is therefore written with $\tilde{\Theta}$, that is up to polylogarithmic factors in R and C. This is a reading of the source, not a claim taken verbatim from it, and a reviewer who prefers the literal Θ should say so.

Remark 4 (what a resolution may not use). By the source paper's Theorem 3, no proof of Conjecture 1 for any $B \geq 2$ can consist of a bound on the multi-instance advantage fed through the MI-to-AI reduction of its Theorem 6: for $B = 2$ that route cannot beat S^2T^4/C^2, and for $B \geq 3$ it cannot beat ST^2/C. Note the direction of that obstruction — it blocks the reduction, not the conjecture — and note also that it does not block a refutation, since the MI attacks it exhibits are not themselves AI attacks.

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.