AICR · OPEN PROBLEM

The STB Conjecture for Short Collisions in Merkle-Damgård Hashing

Is the length-B preprocessing attack optimal for every block bound?

Status. Statement: AI-written from Akshima, David Cash, Andrew Drucker and Hoeteck Wee, *Time-Space Tradeoffs and Short Collisions in Merkle-Damgård Hash Functions*, CRYPTO 2020 (IACR ePrint 2020/770), not yet checked by a human. Settled at $B = 1$ [DGK17], at $B = 2$ [ACDW20, AGL22], for every constant B [GK22], at $B \approx T$ [CDGS18], and for every $2 < B < T$ in the regime $ST^2 \leq N$ [AGL22]; open for non-constant B with $2 < B < T$ in the regime $ST^2 > N$, where the best published security bound exceeds the conjectured one by a factor of up to S.

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

Conjecture (informal). *An attacker who is allowed unlimited offline computation on a hash function, but may carry only S bits of notes away from it, and who then gets a fresh random salt and T online evaluations, is trying to find two distinct messages of at most B blocks each that the salted hash sends to the same digest. There is a simple attack: precompute collisions at S well-chosen internal states, then walk towards one of them from the challenge salt, restarting whenever the walk would exceed the length budget. Its success probability is about STB over the digest space, and the birthday attack contributes about T^2 over the digest space. The conjecture is that nothing does better: the sum of those two terms is the right answer for every block bound, not just for the unbounded-length case and the two-block case where it is proved. What makes this hard is that the standard route to such lower bounds — replacing the attacker's notes by a list of oracle values fixed in advance — provably cannot see the length of a collision at all, so it cannot distinguish the conjectured bound from the much weaker one that holds for unbounded length.*

1 The setting

Merkle-Damgård is the domain extension used by MD5 and the SHA-1 and SHA-2 families: a compression function of fixed input length is iterated over the message blocks, starting from a salt. Against a *preprocessing* (equivalently, non-uniform) attacker the salt is what makes the question non-trivial, since without one an attacker can simply hardwire a collision. The model is Unruh's auxiliary-input random-oracle model [Unr07]: the compression function is a random oracle h, a computationally unbounded first stage sees all of h and outputs S bits of advice, and a second stage receives the advice and a uniformly random salt and makes T oracle queries.

Coretti, Dodis, Guo and Steinberger [CDGS18] showed that in this model salted Merkle-Damgård collisions are found with advantage $\Theta(ST^2/N)$, beating the birthday bound by a factor of S. The source paper's starting observation is that the colliding messages this attack produces are on the order of T blocks long. At the parameters that make the attack notable — $S = T \approx 2^{60}$ against a 180-bit digest — the colliding messages are several petabytes each, which is not a break of any deployed application. So the paper asks for the best attack that respects a *length* budget B, exhibits one achieving advantage $\tilde{\Omega}(STB/N)$ (its Theorem 3, reproduced as Theorem 1 below), and conjectures that it is optimal.

What is proved. Writing $N = 2^n$ for the digest space and suppressing polylogarithmic factors:

- $B = 1$ (a salted random function rather than an iterated one): $\Theta(S/N + T^2/N)$, by Dodis, Guo and Katz [DGK17]. Note that this is *not* the $B = 1$ instance of the conjectured formula, which would read $\Theta(ST/N + T^2/N)$; see Remark 1.
- $B = 2$: $\Theta(ST/N + T^2/N)$, by the source paper's own Theorem 7, with a considerably simpler proof later given by Akshima, Guo and Liu [AGL22]. This is the conjectured value, and it is the source paper's main technical contribution.
- $B \approx T$ (that is, no effective length bound): $\Theta(ST^2/N)$, by [CDGS18], which is again the conjectured value.

- Every constant B: confirmed by Ghoshal and Komargodski [GK22], whose bound $\tilde{O}(STB^2(\log^2 S)^{B-2}/N + T^2/N)$ matches the conjecture when $B = O(1)$ and becomes vacuous once $B > \log N$.
- Every $2 < B < T$ with $ST^2 \le N$: confirmed by [AGL22], whose bound $\tilde{O}((STB/N)\cdot\max\{1, ST^2/N\}+T^2/N)$ collapses to the conjectured one exactly in that regime.

What is open, and where. Since $T^2 \le N$ may be assumed (otherwise the birthday attack already succeeds), the factor $\max\{1, ST^2/N\}$ in the bound of [AGL22] is at most S. So for non-constant B in the regime $ST^2 > N$ the published security bound and the published attack are separated by a factor of up to S, and neither side is known to be the truth: it is open both whether the security bound can be improved to the conjectured $\tilde{O}(STB/N + T^2/N)$ and whether a better attack exists in that regime. Akshima, Guo and Liu [AGL22] pose exactly this as the first of its open problems, observing that the same regime $ST^2 > N$ is where Hellman's function-inversion attack is likewise only known to be optimal below the threshold.

Why it is hard: a named obstruction. The route by which the unbounded-length bound was obtained does not work here, and the source paper proves that it cannot. Unruh's pre-sampling technique [Unr07], tightened by [CDGS18], transfers a bound in the *bit-fixing* random-oracle model (the first stage fixes h on P points of its choice; the rest is uniform) to the auxiliary-input model. The source paper's Theorem 5 gives a bit-fixing adversary that finds *length-2* collisions with advantage $\Omega(PT/N)$ — as good as the best bit-fixing attack for unbounded length — so the bit-fixing model is insensitive to the length of the collision, and no bound proved there can yield the conjectured length-B bound in the auxiliary-input model. The source paper's own $B = 2$ proof therefore goes through a concentration argument of Impagliazzo and Kabanets [IK10] combined with compression, and its Section 7 shows that their argument must itself be modified (a fixed set of salts admits a better attack than the random set the argument needs).

The paper is explicit about how far its own technique reaches: the cases $B = 3$ and $B = 4$ look possible "in principle" but too long to write down, and "[g]oing to arbitrary length bounds seems to be out of reach, but there is no inherent obstruction in applying our technique to the general case with new ideas" (p. 5). Separately, the paper rules out one natural family of improved attacks: its Theorem 8 shows that every "zero-walk" adversary — the shape of all the best attacks known, which precompute collisions at s salts and then walk towards them by appending zero blocks — achieves at most $O(STB\ln(NB)/N)$, matching the conjecture up to logarithmic factors. That theorem carries two side conditions, $B \le T$ and $SB \ge T$, so it covers the class only where the length budget is loose enough for S precomputed targets to be reachable.

Why settling it matters. The conjecture is the quantitative form of the standard practical objection to studying non-uniform attacks at all, namely that no non-trivial non-uniform attack on a real hash function is known: "Our STB conjecture, if true, would explain the non-existence of these attacks" (p. 6), because in the regime $SB \le T$ it says that non-uniform attacks buy nothing over the birthday attack. A refutation would be more consequential still, since it would exhibit a preprocessing attack that beats the birthday bound while producing collisions short enough to matter in an application.

2 Notation and parameters

- $N = 2^n$ and $M = 2^m$ are the sizes of the compression function's chaining-value space and block space; $[N]$ denotes $\{1,\dots,N\}$, and $[M]^+$ the tuples of one or more elements of $[M]$. The compression function is $h : [N] \times [M] \to [N]$, and $\mathsf{Func}([N] \times [M], [N])$ is the set of all such functions.
- $S \in \mathbb{N}$ is the advice length in bits, $T \in \mathbb{N}$ the number of online oracle queries, and $B \in \mathbb{N}$ the bound on the number of blocks in each of the two messages output.
- Θ, O and Ω are asymptotic in N, uniformly over M, S, T and B; the constants hidden in them are absolute. $\tilde{O}$ and $\tilde{\Omega}$ additionally suppress factors polylogarithmic in N. Which of

these the conjecture is stated with is itself part of the question — see Remark 2.

- All algorithms are information-theoretic: running time is ignored and only oracle queries are counted.

3 The conjecture

Definition 1 (Merkle-Damgård). For $h : [N] \times [M] \to [N]$ define $\mathrm{MD}_h : [N] \times [M]^+ \to [N]$ by $\mathrm{MD}_h(a,\alpha) = h(a,\alpha)$ for $\alpha \in [M]$, and

$$\mathrm{MD}_h(a,(\alpha_1,\ldots,\alpha_B)) = h\big(\mathrm{MD}_h(a,(\alpha_1,\ldots,\alpha_{B-1})),\alpha_B\big)$$

for $\alpha_1,\ldots,\alpha_B \in [M]$. The elements of $[M]$ are called *blocks*, and $a \in [N]$ is the *salt*.

Definition 2 (Auxiliary-input collision resistance). A pair of algorithms $\mathcal{A} = (\mathcal{A}_1,\mathcal{A}_2)$ plays the following game against h and a salt a: the first stage computes $\sigma \leftarrow \mathcal{A}_1(h)$ from the entire function table of h; the second stage computes $\alpha,\alpha' \leftarrow \mathcal{A}_2^h(\sigma,a)$ with oracle access to h; the game returns 1 if $\alpha \neq \alpha'$ and $\mathrm{MD}_h(a,\alpha) = \mathrm{MD}_h(a,\alpha')$, and 0 otherwise. Its advantage is

$$\mathbf{Adv}^{\text{ai-cr}}_{\mathrm{MD}}(\mathcal{A}) = \Pr[\text{ the game returns 1 }],$$

the probability taken over independent uniform $h \leftarrow \mathsf{Func}([N] \times [M],[N])$ and $a \leftarrow [N]$, and over the coins of $\mathcal{A}$.
$\mathcal{A}$ is an (S,T,B)-*AI adversary* if $\mathcal{A}_1$ outputs S bits, $\mathcal{A}_2$ issues T oracle queries, and each of the two messages $\mathcal{A}_2$ outputs consists of B or fewer blocks — all three for every input and every oracle, not merely on average. Write

$$\mathbf{Adv}^{\text{ai-cr}}_{\mathrm{MD}}(S,T,B) = \max_{\mathcal{A}} \mathbf{Adv}^{\text{ai-cr}}_{\mathrm{MD}}(\mathcal{A}),$$

the maximum over all (S,T,B)-AI adversaries.

Theorem 1 (the attack; source paper, Theorem 3). *For all positive*

integers S, T, B with $B \le T < N/4$, $STB \le N/2$ and $M \ge N$,

$$\mathbf{Adv}^{\text{ai-cr}}_{\text{MD}}(S,T,B) \;\ge\; \frac{STB - 96S}{48N \log N}.$$

Conjecture 1 (STB). *For all integers B with $2 \le B \le T$,*

$$\mathbf{Adv}^{\text{ai-cr}}_{\text{MD}}(S,T,B) \;=\; \Theta\!\left(\frac{STB + T^2}{N}\right).$$

Equivalently, since Theorem 1 and the birthday attack already give the lower bound up to a logarithmic factor, the open content of Conjecture 1 is the *security* direction: that there is an absolute constant c with

$$\mathbf{Adv}^{\text{ai-cr}}_{\text{MD}}(S,T,B) \;\le\; c\,\frac{STB + T^2}{N} \qquad \text{for all } N, M, S, T \text{ and all } 2 \le B \le T.$$

By the summary in Section 1 this is settled for $B = 2$, for every constant B, for $B \approx T$, and for every $2 < B < T$ with $ST^2 \le N$; the open case is non-constant B with $2 < B < T$ and $ST^2 > N$.

Remark 1 (the range of B). The sentence the source paper prints (p. 2) attaches no range to B: "The best attack with time T and space S for finding collisions of length B in salted MD hash functions built from hash functions with n-bit outputs achieves success probability $\Theta((STB + T^2)/2^n)$". The range $2 \le B \le T$ is supplied here. The upper limit is the paper's own: Theorem 1 requires $B \le T$, and beyond it the length bound does not bind. The lower limit is forced, because $B = 1$ is known to be $\Theta(S/N + T^2/N)$ [DGK17], not the $\Theta(ST/N + T^2/N)$ the formula would give; the paper knows this and cites it. A later paper by three of the same authors restates the conjecture as being "for finding collisions of length $B \ge 2$" [ADGL23], which is the reading taken here.

Remark 2 (whether the Θ is meant up to logarithmic factors). Conjecture 1 is transcribed with the bare Θ the source paper prints. Read literally, its lower half is not established even by the paper's own attack: Theorem 1 delivers $\Omega\big((STB - 96S)/(N \log N)\big)$, which is smaller than $\Omega(STB/N)$ by a factor of $\log N$. Every subsequent treatment of the conjecture known to us, including [AGL22]

and [GK22], states it with $\tilde{\Omega}$ and $\tilde{O}$, that is up to polylogarithmic factors. A solver should treat the polylogarithmic reading as the intended one; a proof of the literal Θ would be strictly stronger and a refutation of the literal Θ that only exploits logarithmic factors would not be a refutation of the conjecture as understood. This ambiguity is in the source, not introduced here.

Remark 3 (what a resolution must not assume). The source paper's Theorem 8 already settles the conjecture's upper bound against the restricted class of zero-walk adversaries of its Definition 8, up to a factor of $\ln(NB)$, for parameters satisfying $B \leq T$ and $SB \geq T$. A resolution of Conjecture 1 in the affirmative must therefore bound adversaries outside that class, and a refutation must exhibit an attack that is not of zero-walk form — or one that lives at parameters Theorem 8 does not cover, that is with $SB < T$.

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. (The source paper of this statement; IACR ePrint 2020/770.)

[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. [Cited here only for its restatement of the conjecture's range of B; postdates the source paper and is therefore not in its reference list.]

[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. [Postdates the source paper and is therefore not in its reference list; details taken from the reference list of IACR ePrint 2023/1444.]

[CDGS18] Sandro Coretti, Yevgeniy Dodis, Siyao Guo, and John P. Steinberger. *Random oracles and non-uniformity.* In Jesper Buus Nielsen and Vincent Rijmen, editors, Advances in Cryptology — EUROCRYPT 2018, Part I, volume 10820 of Lecture Notes in Computer Science, pp. 227–258. Springer, 2018.

[DGK17] Yevgeniy Dodis, Siyao Guo, and Jonathan Katz. *Fixing cracks in the concrete: random oracles with auxiliary input, revisited.* In Jean-Sébastien Coron and Jesper Buus Nielsen, editors, Advances in Cryptology — EUROCRYPT 2017, Part II, volume 10211 of Lecture Notes in Computer Science, pp. 473–495. Springer, 2017.

[GK22] Ashrujit Ghoshal and Ilan Komargodski. *On time-space tradeoffs for bounded-length collisions in Merkle-Damgård hashing.* In Advances in Cryptology — CRYPTO 2022. Springer, 2022. [Postdates the source paper and is therefore not in its reference list; details taken from the reference list of IACR ePrint 2022/885.]

[IK10] Russell Impagliazzo and Valentine Kabanets. *Constructive proofs of concentration bounds.* In APPROX 2010 and RANDOM 2010, pp. 617–631, 2010.

[Unr07] Dominique Unruh. *Random oracles and auxiliary input.* In Alfred Menezes, editor, Advances in Cryptology — CRYPTO 2007, volume 4622 of Lecture Notes in Computer Science, pp. 205–223. Springer, 2007.