Time-Space Tradeoffs for Collisions in MD Hash Functions

Akshima
University of Chicago

Joint work with
Siyao Guo & Qipeng Liu
NYU Shanghai Simons Institute
Merkle-Damgård Hash Functions

Input \( x = x_1 \parallel \ldots \parallel x_B, x_i \in [M] \)

Salt \( a \in [N] \)

\[ h : [N] \times [M] \rightarrow [N] \] is compression function
Merkle-Damgård Hash Functions

**Input** \( x = x_1 \ || \ ... \ || x_B, \ x_i \in [M] \)

Salt \( a \) \( \in [N] \)

\[ h : [N] \times [M] \rightarrow [N] \] is compression function

Assuming \( h \) is a random function

- \( T \) queries: \( O(T^2/N) \) advantage
Merkle-Damgård Hash Functions

Input $x = x_1 || ... || x_B$, $x_i \in [M]$

Salt $a \in [N]$

$h : [N] \times [M] \rightarrow [N]$ is compression function

Assuming $h$ is a random function

• $T$ queries: $O(T^2/N)$ advantage

What if adversary can pre-learn about $h$?
Our Problem

Study *bounded length collision* finding:

1. For Salted *Merkle-Damgård* based hash functions

2. With *Pre-computation* where

   • Pre-computed advice is $S$-bits long
   
   • $T$ queries are made to $h$
   
   • $\leq B$ block collisions
## Results from Prior Work

<table>
<thead>
<tr>
<th>Work</th>
<th># Blocks in Collision</th>
<th>Security (ignores poly log factors)</th>
<th>Attack (ignores poly log factors)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[CDGS18]</td>
<td>Unbounded</td>
<td>$O\left(\frac{ST^2}{N}\right)$</td>
<td>$\Omega\left(\frac{ST^2}{N}\right)$</td>
</tr>
<tr>
<td>[ACDW20]</td>
<td>2</td>
<td>$O\left(\frac{ST}{N} + \frac{T^2}{N}\right)$</td>
<td>$\Omega\left(\frac{ST}{N} + \frac{T^2}{N}\right)$</td>
</tr>
<tr>
<td>[ACDW20]</td>
<td>B</td>
<td>$O\left(\frac{STD + T^2}{N}\right)^*$</td>
<td>$\Omega\left(\frac{STD + T^2}{N}\right)$</td>
</tr>
<tr>
<td>[GK22]</td>
<td>B</td>
<td>$O\left(\frac{STB^2(\log^2 S)^{B-2}}{N} + \frac{T^2}{N}\right)$</td>
<td>-</td>
</tr>
<tr>
<td>[GK22]</td>
<td>B</td>
<td>$O\left(\frac{S^4TB^2}{N} + \frac{T^2}{N}\right)$</td>
<td>-</td>
</tr>
</tbody>
</table>
## Results from Prior Work

<table>
<thead>
<tr>
<th>Work</th>
<th># Blocks in Collision</th>
<th>Security (ignores poly log factors)</th>
<th>Attack (ignores poly log factors)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>[CDGS18]</strong></td>
<td>Unbounded</td>
<td>(O\left(\frac{ST^2}{N}\right))</td>
<td>(\Omega\left(\frac{ST^2}{N}\right))</td>
</tr>
<tr>
<td><strong>[ACDW20]</strong></td>
<td>2</td>
<td>(O\left(\frac{ST}{N} + \frac{T^2}{N}\right))</td>
<td>(\Omega\left(\frac{ST}{N} + \frac{T^2}{N}\right))</td>
</tr>
<tr>
<td><strong>[ACDW20]</strong></td>
<td>B</td>
<td>(O\left(\frac{STB}{N} + \frac{T^2}{N}\right)^*)</td>
<td>(\Omega\left(\frac{STB}{N} + \frac{T^2}{N}\right))</td>
</tr>
<tr>
<td><strong>[GK22]</strong></td>
<td>B</td>
<td>(O\left(\frac{STB^2(\log^2 S)^{B-2}}{N} + \frac{T^2}{N}\right))</td>
<td>-</td>
</tr>
<tr>
<td><strong>[GK22]</strong></td>
<td>B</td>
<td>(O\left(\frac{S^4TB^2}{N} + \frac{T^2}{N}\right))</td>
<td>-</td>
</tr>
</tbody>
</table>
• Consider SHA2: $N = 2^{256}$, $M = 2^{512}$
  
• $S = 2^{70}$, then $ST^2/N$ bound $\implies T = 2^{93}$
• **Collisions** [CDGS18] attack finds are $2^{93}$ blocks long
Why Bounded Collisions?

- Consider SHA2: \( N = 2^{256} \), \( M = 2^{512} \)
  - \( S = 2^{70} \), then \( ST^2/N \) bound \( \implies T = 2^{93} \)
  - **Collisions** [CDGS18] attack finds are \( 2^{93} \) blocks long

Colliding messages have to be several yottabytes long for the attacker to succeed!!!
Why Bounded Collisions?

• Consider SHA2: \( N = 2^{256}, M = 2^{512} \)
• \( S = 2^{70} \), then \( ST^2/N \) bound \( \implies T = 2^{93} \)
• Collisions [CDGS18] attack finds are \( 2^{93} \) blocks long

Colliding messages have to be several yottabytes long for the attacker to succeed!!!

For \( B = 2^{20} \), then the best known attack needs \( T = 2^{166} \)

Best attack: \( \tilde{\Omega}(STB/N) \) instead of \( \tilde{\Omega}(ST^2/N) \)
# Blocks in Collision

<table>
<thead>
<tr>
<th>Work</th>
<th># Blocks in Collision</th>
<th>Security (ignores poly log factors)</th>
<th>Attack (ignores poly log factors)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[CDGS18]</td>
<td>Unbounded</td>
<td>$O\left(\frac{ST^2}{N}\right)$</td>
<td>$\Omega\left(\frac{ST^2}{N}\right)$</td>
</tr>
<tr>
<td>[ACDW20]</td>
<td>2</td>
<td>$O\left(\frac{ST}{N} + \frac{T^2}{N}\right)$</td>
<td>$\Omega\left(\frac{ST}{N} + \frac{T^2}{N}\right)$</td>
</tr>
<tr>
<td>[ACDW20]</td>
<td>B</td>
<td>$O\left(\frac{STB}{N} + \frac{T^2}{N}\right)^*$</td>
<td>$\Omega\left(\frac{STB}{N} + \frac{T^2}{N}\right)^*$</td>
</tr>
<tr>
<td>[GK22]</td>
<td>B</td>
<td>$O\left(\frac{STB^2(\log^2 S)^{B-2}}{N} + \frac{T^2}{N}\right)$</td>
<td>-</td>
</tr>
<tr>
<td>[GK22]</td>
<td>B</td>
<td>$O\left(\frac{S^4TB^2}{N} + \frac{T^2}{N}\right)$</td>
<td>-</td>
</tr>
</tbody>
</table>

Results from Prior Work
# Blocks in Collision

<table>
<thead>
<tr>
<th>Work</th>
<th># Blocks in Collision</th>
<th>Security (ignores poly log factors)</th>
<th>Attack (ignores poly log factors)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[CDGS18]</td>
<td>Unbounded</td>
<td>$O\left(\frac{ST^2}{N}\right)$</td>
<td>$\Omega\left(\frac{ST^2}{N}\right)$</td>
</tr>
<tr>
<td>[ACDW20]</td>
<td>2</td>
<td>$O\left(\frac{ST}{N} + \frac{T^2}{N}\right)$</td>
<td>$\Omega\left(\frac{ST}{N} + \frac{T^2}{N}\right)$</td>
</tr>
<tr>
<td>[ACDW20]</td>
<td>B</td>
<td>$O\left(\frac{STB}{N} + \frac{T^2}{N}\right)^*$</td>
<td>$\Omega\left(\frac{STB}{N} + \frac{T^2}{N}\right)$</td>
</tr>
<tr>
<td>[GK22]</td>
<td>B</td>
<td>$O\left(\frac{STB^2(\log^2 S)^{B-2}}{N} + \frac{T^2}{N}\right)$</td>
<td>-</td>
</tr>
<tr>
<td>[GK22]</td>
<td>B</td>
<td>$O\left(\frac{S^4TB^2}{N} + \frac{T^2}{N}\right)$</td>
<td>-</td>
</tr>
</tbody>
</table>
# Results from Prior Work

<table>
<thead>
<tr>
<th>Work</th>
<th># Blocks in Collision</th>
<th>Security (ignores poly log factors)</th>
<th>Attack (ignores poly log factors)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[CDGS18]</td>
<td>Unbounded</td>
<td>$O\left(\frac{ST^2}{N}\right)$</td>
<td>$\Omega\left(\frac{ST^2}{N}\right)$</td>
</tr>
<tr>
<td>[ACDW20]</td>
<td>2</td>
<td>$O\left(\frac{ST}{N} + \frac{T^2}{N}\right)$</td>
<td>$\Omega\left(\frac{ST}{N} + \frac{T^2}{N}\right)$</td>
</tr>
<tr>
<td>[ACDW20]</td>
<td>B</td>
<td>$O\left(\frac{STB}{N} + \frac{T^2}{N}\right)^*$</td>
<td>$\Omega\left(\frac{STB}{N} + \frac{T^2}{N}\right)$</td>
</tr>
<tr>
<td>[GK22]</td>
<td>B</td>
<td>$O\left(\frac{STB^2\left(\log^2 S\right)^{B-2}}{N} + \frac{T^2}{N}\right)$</td>
<td>-</td>
</tr>
<tr>
<td>[GK22]</td>
<td>B</td>
<td>$O\left(\frac{S^4TB^2}{N} + \frac{T^2}{N}\right)$</td>
<td>-</td>
</tr>
</tbody>
</table>
## Results from Prior Work

<table>
<thead>
<tr>
<th>Work</th>
<th># Blocks in Collision</th>
<th>Security (ignores poly log factors)</th>
<th>Attack (ignores poly log factors)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[CDGS18]</td>
<td>Unbounded</td>
<td>$O\left(\frac{ST^2}{N}\right)$</td>
<td>$\Omega\left(\frac{ST^2}{N}\right)$</td>
</tr>
<tr>
<td>[ACDW20]</td>
<td>2</td>
<td>$O\left(\frac{ST}{N} + \frac{T^2}{N}\right)$</td>
<td>$\Omega\left(\frac{ST}{N} + \frac{T^2}{N}\right)$</td>
</tr>
<tr>
<td>[ACDW20]</td>
<td>B</td>
<td>$O\left(\frac{STB}{N} + \frac{T^2}{N}\right)^*$</td>
<td>$\Omega\left(\frac{STB}{N} + \frac{T^2}{N}\right)$</td>
</tr>
<tr>
<td>[GK22]</td>
<td>B</td>
<td>$O\left(\frac{STB^2}{N}(\log^2 S)^{B-2} + \frac{T^2}{N}\right)$</td>
<td>-</td>
</tr>
<tr>
<td>[GK22]</td>
<td>B</td>
<td>$O\left(\frac{S^4TB^2}{N} + \frac{T^2}{N}\right)$</td>
<td>-</td>
</tr>
</tbody>
</table>
### Our Result

<table>
<thead>
<tr>
<th>Work</th>
<th># Blocks in Collision</th>
<th>Security</th>
<th>Attack</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S: advice size</td>
<td>(ignores poly log factors)</td>
<td>(ignores poly log factors)</td>
</tr>
<tr>
<td>[CDGS18]</td>
<td>Unbounded</td>
<td>$O\left(\frac{ST^2}{N}\right)$</td>
<td>$\Omega\left(\frac{ST^2}{N}\right)$</td>
</tr>
<tr>
<td>[ACDW20]</td>
<td>2</td>
<td>$O\left(\frac{ST}{N} + \frac{T^2}{N}\right)$</td>
<td>$\Omega\left(\frac{ST}{N} + \frac{T^2}{N}\right)$</td>
</tr>
<tr>
<td>[ACDW20]</td>
<td>B</td>
<td>$O\left(\frac{STB}{N} + \frac{T^2}{N}\right)^*$</td>
<td>$\Omega\left(\frac{STB}{N} + \frac{T^2}{N}\right)$</td>
</tr>
<tr>
<td>[GK22]</td>
<td>B</td>
<td>$O\left(\frac{STB^2(\log^2 S)^{B-2}}{N} + \frac{T^2}{N}\right)$</td>
<td>-</td>
</tr>
<tr>
<td>[GK22]</td>
<td>B</td>
<td>$O\left(\frac{S^4TB^2}{N} + \frac{T^2}{N}\right)$</td>
<td>-</td>
</tr>
<tr>
<td>This Work</td>
<td>B</td>
<td>$O\left(\max\left{1, \frac{ST^2}{N}\right} \cdot \frac{STB}{N} + \frac{T^2}{N}\right)$</td>
<td>-</td>
</tr>
</tbody>
</table>
Our Bound

What does our bound $O\left(\max\left\{1, \frac{ST^2}{N}\right\} \cdot \frac{STB}{N} + \frac{T^2}{N}\right)$ mean?

$$\frac{ST^2}{N} \leq 1 \implies \text{Our bound is } O\left(\frac{STB}{N}\right)$$

We prove $STB$ conjecture of [ACDW20] when $ST^2 \leq N$. 

Our Bound

What does our bound $O\left(\max\left\{1, \frac{ST^2}{N}\right\} \cdot \frac{STB}{N} + \frac{T^2}{N}\right)$ mean?

\[
\frac{ST^2}{N} \leq 1 \implies \text{Our bound is } O\left(\frac{STB}{N}\right)
\]

\[
\frac{ST^2}{N} > 1 \implies \text{Our bound is } O\left(\frac{STB}{N} \cdot \frac{ST^2}{N}\right)
\]

Confirms bounded length collisions are harder.
Salted Collision Finding in MD with Pre-computation

$h : [N] \times [M] \rightarrow [N]$

$\sigma \in \{0,1\}^S$

$a \leftarrow [N]$

$q_1 \leftrightarrow q_2 \leftrightarrow \cdots \leftrightarrow q_T$

$(x, x')$

To succeed:
1. $x \neq x'$
2. $\text{MD}^h(a, x) = \text{MD}^h(a, x')$
3. $x, x' \in [M]^{\leq B}$
## Comparison of Techniques

<table>
<thead>
<tr>
<th>Work</th>
<th>Technique</th>
</tr>
</thead>
<tbody>
<tr>
<td>[CDGS18]</td>
<td>Reduction to Pre-sampling</td>
</tr>
<tr>
<td>[ACDW20]</td>
<td>Reduction to Multi-instance Problem + Compression</td>
</tr>
<tr>
<td>[GK22]</td>
<td>Reduction to Multi-instance Problem + Compression</td>
</tr>
<tr>
<td>This Work</td>
<td>Reduction to modified Multi-instance Problem</td>
</tr>
</tbody>
</table>
Pre-Sampling Model

- Adversary hard-codes some points before oracle chosen
- Online phase gets oracle, no advice

Phase 1

Phase 2

\[ a \leftarrow \$ [N] \]

\[ (x, x') \]

\[ \begin{array}{ccc}
  a_1 & a_1' \\
  \vdots & \vdots \\
  a_p & a_p' \\
\end{array} \]

\[ \begin{array}{c}
  1 \\
  \vdots \\
  N \\
\end{array} \rightarrow 
\begin{array}{c}
  h(1) \\
  \vdots \\
  h(N) \\
\end{array} \]
Pre-Sampling Model

Advantage in Pre-sampling model
with $ST$ pre-fixed points and $T$ queries is $\delta$

$\Longrightarrow$ Advantage in Pre-computation model
with $S$-bit advice and making $T$ queries is $O(\delta)$
Techniques from Prior Works

[CDGS18]

• Give reduction to Pre-sampling model
• show $O\left(\frac{ST^2}{N}\right)$ advantage for collision-finding in the Pre-sampling model

$\implies O\left(\frac{ST^2}{N}\right)$ advantage in the Pre-computation model
Techniques from Prior Works

[CDGS18]

- Give reduction to Pre-sampling model
- Show $O\left(\frac{ST^2}{N}\right)$ advantage for collision-finding in the Pre-sampling model

[ACDW20]: Impossible to get better bounds for bounded-length collision finding in pre-sampling
## Comparison of Techniques

<table>
<thead>
<tr>
<th>Work</th>
<th>Technique</th>
</tr>
</thead>
<tbody>
<tr>
<td>[CDGS18]</td>
<td>Reduction to Pre-sampling</td>
</tr>
<tr>
<td>[ACDW20]</td>
<td>Reduction to Multi-instance Problem + Compression</td>
</tr>
<tr>
<td>[GK22]</td>
<td>Reduction to Multi-instance Problem + Compression</td>
</tr>
<tr>
<td>This Work</td>
<td>Reduction to modified Multi-instance Problem</td>
</tr>
</tbody>
</table>
Multi-instance Game

Given \((a_1, \ldots, a_S) \in [N]^S\)

For \(i \in [S]\), do:

\(A\) should find collisions on each salt in \(\{a_1, \ldots, a_S\}\)
Multi-instance Game

[ACDW20]

Given \((a_1, \ldots, a_S) \in [N]^S\)

For \(i \in [S]\), do:

\[ a_i \]

\(A\)

\((x_i, x'_i)\)

\[q_1, q_2, \ldots, q_T\]

\(h\)

\(A\) should find collisions on each salt in \(\{a_1, \ldots, a_S\}\)

Advantage in Multi-instance game is \(\leq \delta^S\)

\[\implies\]

Advantage in Pre-computation model is \(\leq 2 \cdot \delta\)
Techniques from Prior Works

[ACDW20]

• Give reduction to multi-instance game

• Show $O\left(\left(\frac{ST + T^2}{N}\right)^s\right)$ bound on 2-block collision finding multi-instance game via compression

$\implies O\left(\frac{ST + T^2}{N}\right)$ bound on 2-block collision finding in the Pre-computation model
Techniques from Prior Works

[ACDW20]
- Give reduction to multi-instance game
- Show $O \left( \left( \frac{ST + T^2}{N} \right)^S \right)$ bound on 2-block collision finding
  multi-instance game via compression

[GK22] uses a similar approach

For more details
- Full talk on YouTube
- eprint: 2022/309
## Comparison of Techniques

<table>
<thead>
<tr>
<th>Work</th>
<th>Technique</th>
</tr>
</thead>
<tbody>
<tr>
<td>[CDGS18]</td>
<td>Reduction to Pre-sampling</td>
</tr>
<tr>
<td>[ACDW20]</td>
<td>Reduction to Multi-instance Problem + Compression</td>
</tr>
<tr>
<td>[GK22]</td>
<td>Reduction to Multi-instance Problem + Compression</td>
</tr>
<tr>
<td>This Work</td>
<td>Reduction to modified Multi-instance Problem</td>
</tr>
</tbody>
</table>
Simplifying the Model

$X_i$: indicator of succeeding on salt $a_i$

\[ \Pr \left[ \bigwedge_{i=1}^{u} X_i = 1 \right] \]

[ACDW20,GK22] bound
Simplifying the Model

$X_i$: indicator of succeeding on salt $a_i$

[ACDW20,GK22] bound

$$\Pr \left[ \bigwedge_{i=1}^{u} X_i = 1 \right]$$

This work bounds

$$\Pr \left[ X_i = 1 \mid X_{<i} = 1 \right]$$

for any $i \in [S]$
Simplifying the Model

\[
(a_1, \ldots, a_i) \leq ST
\]

‘Offline’ queries

\[
T
\]

‘Online’ queries

Suffices to bound \( \text{Pr} [X_i | X_{<i}] \) to

\[
O \left( \max \left\{ 1, \frac{ST^2}{N} \right\} \cdot \frac{STB}{N} + \frac{T^2}{N} \right)
\]
Consider collision type

\[ m_1 \]
\[ m_2 \]
\[ m_3 \]

\[ a_i \rightarrow a'_i \]

such that

\[ q_1 := (a_i, m_1) \]
\[ q_2 := (a'_i, m_2) \]
\[ q_3 := (a'_i, m_3) \]

The output of \( q_1 \) is limited to certain values

‘Offline’ queries

‘Online’ queries
Consider collision type

\[ q_1 := (a_i, m_1) \]
\[ q_2 := (a'_i, m_2) \]
\[ q_3 := (a'_i, m_3) \]

such that

It should be input salt of one of \( ST \) queries in ‘Offline’ phase

\[ \implies \Pr[ \exists \text{ } q_1\text{-like Online query}] \leq ST^2/N \]
Consider collision type

\[
\begin{align*}
  q_1 &:= (a_i, m_1) \\
  q_2 &:= (a'_i, m_2) \\
  q_3 &:= (a'_i, m_3)
\end{align*}
\]

such that

\[
\Pr[\exists q_1\text{-like Online query}] \leq \frac{ST^2}{N}
\]

It should be input salt of one of \(ST\) queries in ‘Offline’ phase

\[
\implies \Pr[\exists q_1\text{-like Online query}] \leq \frac{ST^2}{N}
\]

But we can bound better!
Consider collision type

\[ a_i \xrightarrow{m_1} a'_i \xrightarrow{m_2} a_i \xrightarrow{m_3} a_{i-1} \]

such that

\[ q_1 := (a_i, m_1) \]
\[ q_2 := (a'_i, m_2) \]
\[ q_3 := (a'_i, m_3) \]

How many \((q_2, q_3)\) such pairs can there be in ‘Offline’ queries?

We refer to this as ‘**useful knowledge gain**’ from offline queries.
Consider collision type

\[ a_i \overset{m_1}{\rightarrow} a'_i \overset{m_2}{\rightarrow} a'_i \overset{m_3}{\rightarrow} \]

such that

\[ q_1 := (a_i, m_1) \]
\[ q_2 := (a'_i, m_2) \]
\[ q_3 := (a'_i, m_3) \]

**Example**

How many \((q_2, q_3)\) such pairs can there be in ‘Offline’ queries?

There can be at most \(ST/2\) pairs starting from distinct salts.
Consider collision type

\[
\begin{align*}
q_1 &:= (a_i, m_1) \\
q_2 &:= (a'_i, m_2) \\
q_3 &:= (a'_i, m_3)
\end{align*}
\]

such that

\[
\begin{align*}
a_1 &\quad \ldots \quad a_i \\
q_2 &\quad q_3
\end{align*}
\]

How many \((q_2, q_3)\) such pairs can there be in ‘Offline’ queries?

There can be at most \(ST/2\) pairs starting from distinct salts. \(\Rightarrow O(ST^2/N)\)

This is \textbf{worst-case analysis}, similar to \textbf{Pre-sampling}
Consider collision type

\[ a_i \rightarrow m_1 \rightarrow a'_i \rightarrow m_2 \rightarrow a_i \]

such that

\[ q_1 := (a_i, m_1) \]
\[ q_2 := (a'_i, m_2) \]
\[ q_3 := (a'_i, m_3) \]

How many \((q_2, q_3)\) such pairs can there be in ‘Offline’ queries?

There can be at most \(ST/2\) pairs starting from distinct salts. \(\Rightarrow O(ST^2/N)\)

We bound better via **average-case** analysis
Consider collision type

\[ q_1 := (a_i, m_1) \]
\[ q_2 := (a'_i, m_2) \]
\[ q_3 := (a'_i, m_3) \]

such that

\[ a_i \rightarrow a'_i \]

How many \((q_2, q_3)\) such pairs can there be in ‘Offline’ queries?

We show: The probability of finding \(\geq S\) pairs \((q_2, q_3)\) in \(ST\) queries is ‘small’

\[ \Rightarrow \Pr[\exists \text{ } q_1\text{-like Online query}] \leq \frac{ST}{N} \]
Proof Overview

1. We identify all types of “useful knowledge gains” from Offline queries
1. We Identify all types of “useful knowledge gains” from Offline queries

2. For each type, we show the probability of ‘high’ knowledge gain is ‘small’ even conditioned on winning in all previous rounds
1. We Identify all types of “useful knowledge gains” from Offline queries

2. For each type, we show the probability of ‘high’ knowledge gain is ‘small’ even conditioned on winning in all previous rounds

3. When none of the knowledge gain is high, we can easily bound $\Pr [X_i | X_{<i}]$ as required
Future Work

1. For $ST^2 \geq N$ is there a better attack or security bound?

2. Time-space trade-offs for collision finding in the quantum setting
Thank you