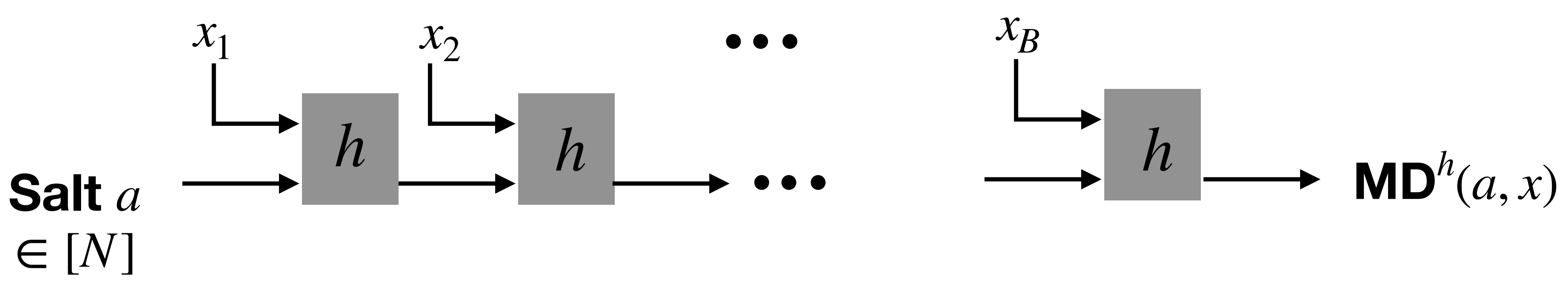
Time-Space Tradeoffs for Collisions in MD Hash Functions

Akshima
University of Chicago

Joint work with
Siyao Guo & Qipeng Liu
NYU Shanghai *Simons Institute*

Merkle-Damgård Hash Functions

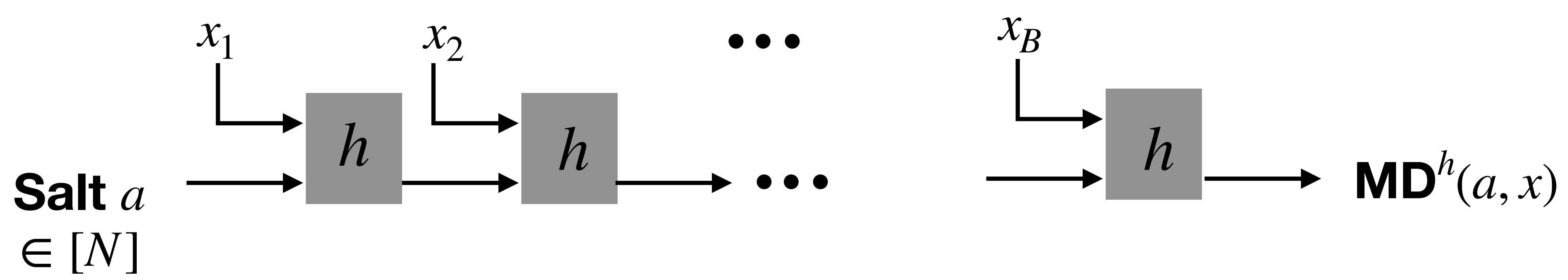
Input $x = x_1 || \ldots || x_B$, $x_i \in [M]$



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

Merkle-Damgård Hash Functions

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



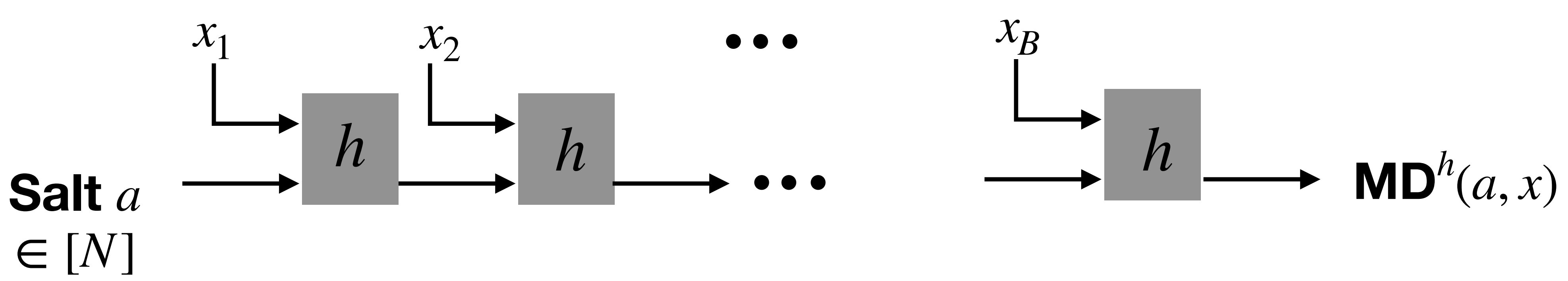
$h : [N] \times [M] \to [N]$ is compression function

Assuming h is a random function

- T **queries:** $O(T^2/N)$ advantage

Merkle-Damgård Hash Functions

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



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

Work	**# Blocks in Collision** **S: advice size** **T: Queries**	**Security** (ignores poly log factors)	**Attack** (ignores poly log factors)
[CDGS18]	Unbounded	$O\left(\frac{ST^2}{N}\right)$	$\Omega\left(\frac{ST^2}{N}\right)$
[ACDW20]	2	$O\left(\frac{ST}{N}+\frac{T^2}{N}\right)$	$\Omega\left(\frac{ST}{N}+\frac{T^2}{N}\right)$
[ACDW20]	B	$O\left(\frac{STB}{N}+\frac{T^2}{N}\right)^*$	$\Omega\left(\frac{STB}{N}+\frac{T^2}{N}\right)$
[GK22]	B	$O\left(\frac{STB^2(\log^2 S)^{B-2}}{N}+\frac{T^2}{N}\right)$	-
[GK22]	B	$O\left(\frac{S^4TB^2}{N}+\frac{T^2}{N}\right)$	-

Results from Prior Work

Work	# Blocks in Collision S: advice size T: Queries	**Security** (ignores poly log factors)	**Attack** (ignores poly log factors)
[CDGS18]	**Unbounded**	$O\left(\frac{ST^2}{N}\right)$	$\Omega\left(\frac{ST^2}{N}\right)$
[ACDW20]	2	$O\left(\frac{ST}{N}+\frac{T^2}{N}\right)$	$\Omega\left(\frac{ST}{N}+\frac{T^2}{N}\right)$
[ACDW20]	B	$O\left(\frac{STB}{N}+\frac{T^2}{N}\right)^*$	$\Omega\left(\frac{STB}{N}+\frac{T^2}{N}\right)$
[GK22]	B	$O\left(\frac{STB^2(\log^2 S)^{B-2}}{N}+\frac{T^2}{N}\right)$	-
[GK22]	B	$O\left(\frac{S^4TB^2}{N}+\frac{T^2}{N}\right)$	-

Why Bounded Collisions?

[ACDW20]

- 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?

[ACDW20]

- 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?

[ACDW20]

- 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** $\tilde{\Omega}(ST^2/N)$

Results from Prior Work

Work	# Blocks in Collision S: advice size T: Queries	**Security** (ignores poly log factors)	**Attack** (ignores poly log factors)
[CDGS18]	Unbounded	$O\left(\frac{ST^2}{N}\right)$	$\Omega\left(\frac{ST^2}{N}\right)$
[ACDW20]	2	$O\left(\frac{ST}{N}+\frac{T^2}{N}\right)$	$\Omega\left(\frac{ST}{N}+\frac{T^2}{N}\right)$
[ACDW20]	**B**	$O\left(\frac{STB}{N}+\frac{T^2}{N}\right)^*$	$\Omega\left(\frac{STB}{N}+\frac{T^2}{N}\right)$
[GK22]	B	$O\left(\frac{STB^2(\log^2 S)^{B-2}}{N}+\frac{T^2}{N}\right)$	-
[GK22]	B	$O\left(\frac{S^4TB^2}{N}+\frac{T^2}{N}\right)$	-

Results from Prior Work

Work	# Blocks in Collision S: advice size T: Queries	Security (ignores poly log factors)	Attack (ignores poly log factors)
[CDGS18]	Unbounded	$O\left(\frac{ST^2}{N}\right)$	$\Omega\left(\frac{ST^2}{N}\right)$
[ACDW20]	**2**	$O\left(\frac{ST}{N}+\frac{T^2}{N}\right)$	$\Omega\left(\frac{ST}{N}+\frac{T^2}{N}\right)$
[ACDW20]	B	$O\left(\frac{STB}{N}+\frac{T^2}{N}\right)^*$	$\Omega\left(\frac{STB}{N}+\frac{T^2}{N}\right)$
[GK22]	B	$O\left(\frac{STB^2(\log^2 S)^{B-2}}{N}+\frac{T^2}{N}\right)$	-
[GK22]	B	$O\left(\frac{S^4TB^2}{N}+\frac{T^2}{N}\right)$	-

Results from Prior Work

Work	# Blocks in Collision S: advice size T: Queries	Security (ignores poly log factors)	Attack (ignores poly log factors)
[CDGS18]	Unbounded	$O\left(\frac{ST^2}{N}\right)$	$\Omega\left(\frac{ST^2}{N}\right)$
[ACDW20]	2	$O\left(\frac{ST}{N}+\frac{T^2}{N}\right)$	$\Omega\left(\frac{ST}{N}+\frac{T^2}{N}\right)$
[ACDW20]	B	$O\left(\frac{STB}{N}+\frac{T^2}{N}\right)^*$	$\Omega\left(\frac{STB}{N}+\frac{T^2}{N}\right)$
[GK22]	**B**	$O\left(\frac{STB^2(\log^2 S)^{B-2}}{N}+\frac{T^2}{N}\right)$	-
[GK22]	B	$O\left(\frac{S^4TB^2}{N}+\frac{T^2}{N}\right)$	-

Results from Prior Work

Work	# Blocks in Collision S: advice size T: Queries	**Security** (ignores poly log factors)	**Attack** (ignores poly log factors)
[CDGS18]	Unbounded	$O\left(\frac{ST^2}{N}\right)$	$\Omega\left(\frac{ST^2}{N}\right)$
[ACDW20]	2	$O\left(\frac{ST}{N}+\frac{T^2}{N}\right)$	$\Omega\left(\frac{ST}{N}+\frac{T^2}{N}\right)$
[ACDW20]	B	$O\left(\frac{STB}{N}+\frac{T^2}{N}\right)^*$	$\Omega\left(\frac{STB}{N}+\frac{T^2}{N}\right)$
[GK22]	B	$O\left(\frac{STB^2(\log^2 S)^{B-2}}{N}+\frac{T^2}{N}\right)$	-
[GK22]	**B**	$O\left(\frac{S^4TB^2}{N}+\frac{T^2}{N}\right)$	-

Our Result

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

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?

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

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?

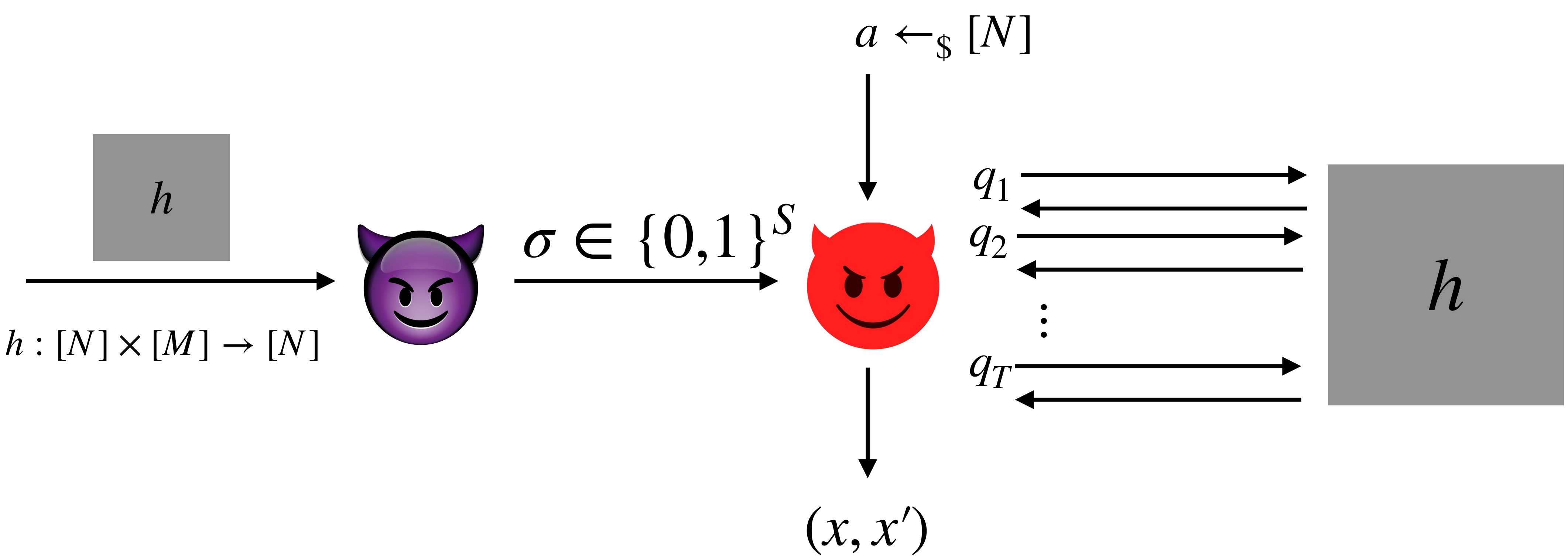
$$\boxed{\frac{ST^2}{N} \leq 1 \implies \text{Our bound is } O\left(\frac{STB}{N}\right)}$$

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

Confirms bounded length collisions are harder

Salted Collision Finding in MD with Pre-computation

[ACDW20]



To succeed:

1. $x \neq x'$

2. $\mathsf{MD}^h(a, x) = \mathsf{MD}^h(a, x')$

3. $x, x' \in [M]^{\leq B}$

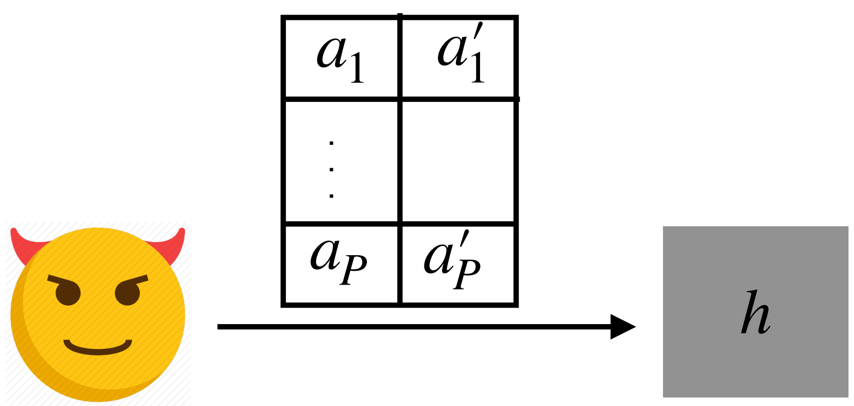
Comparison of Techniques

Work	Technique
[CDGS18]	**Reduction to Pre-sampling**
[ACDW20]	Reduction to Multi-instance Problem + Compression
[GK22]	Reduction to Multi-instance Problem + Compression
This Work	Reduction to modified Multi-instance Problem

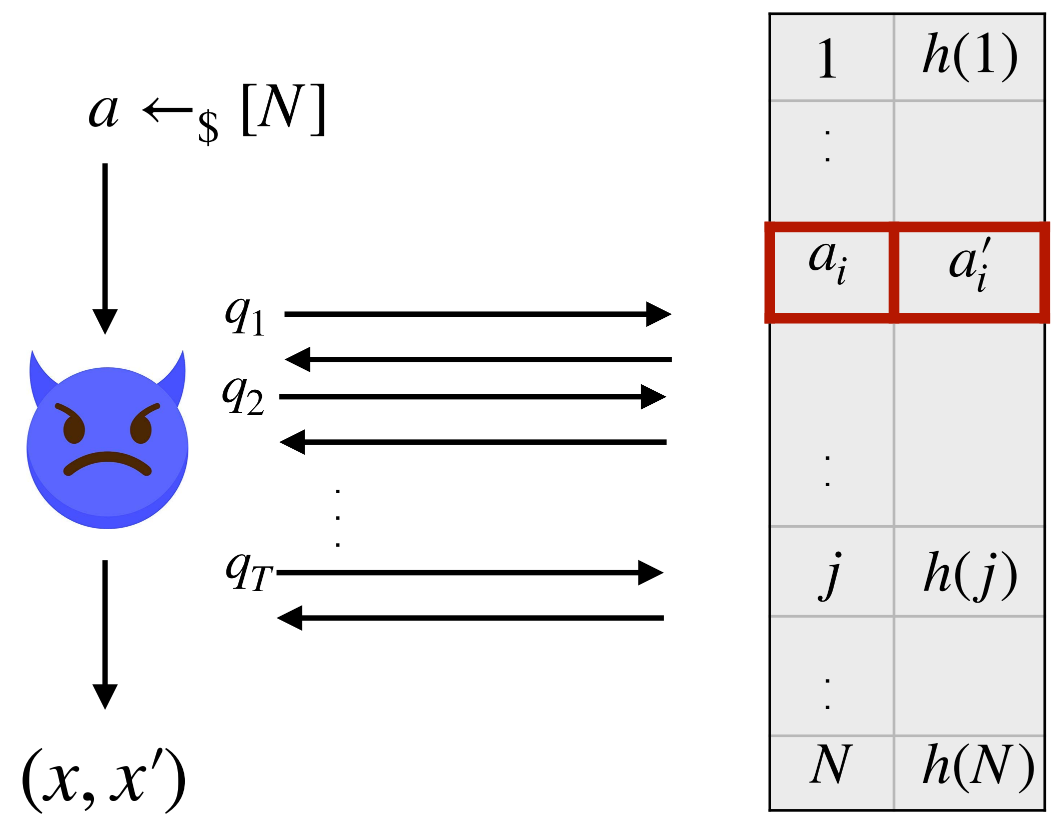
Pre-Sampling Model [CDGS18]

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

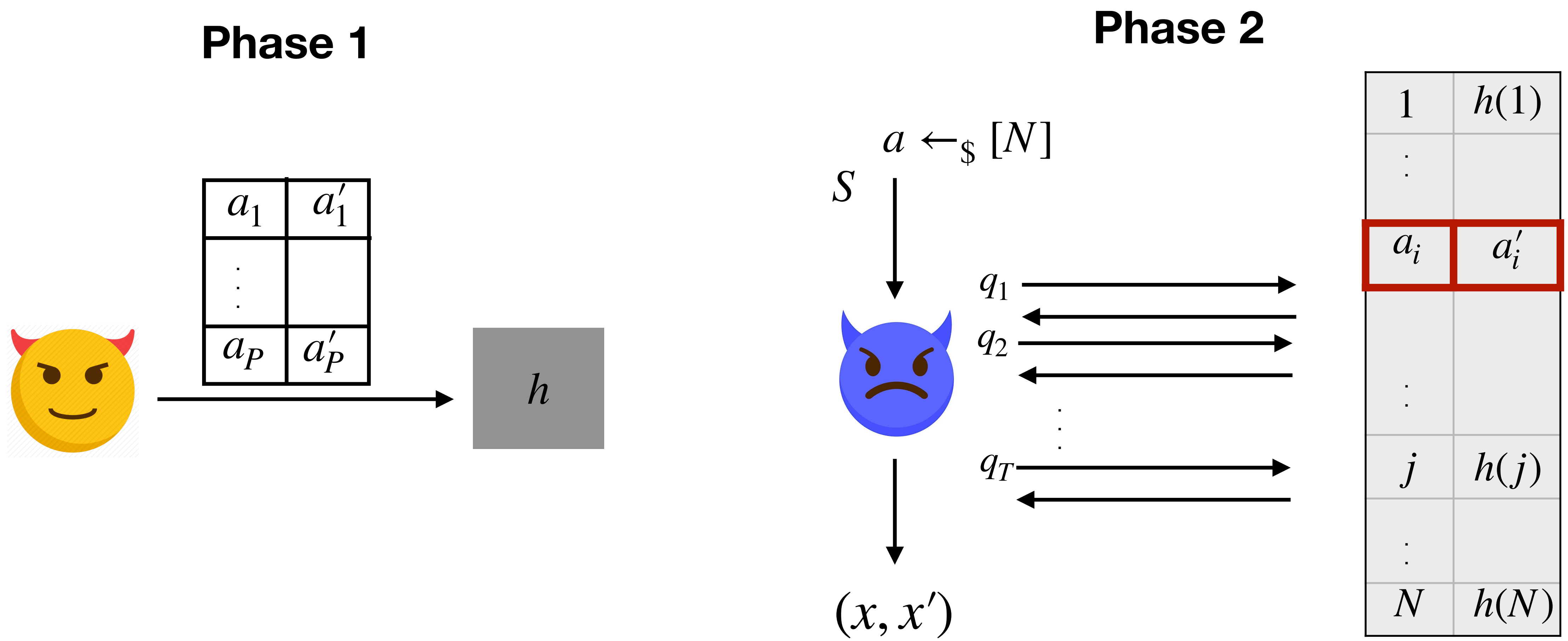
Phase 1



Phase 2



Pre-Sampling Model [CDGS18]



Advantage in **Pre-sampling model**
with ST **pre-fixed points** and T **queries** is δ

$\Longrightarrow$ 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**

$\Longrightarrow$ $O\left(\frac{ST^2}{N}\right)$ 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

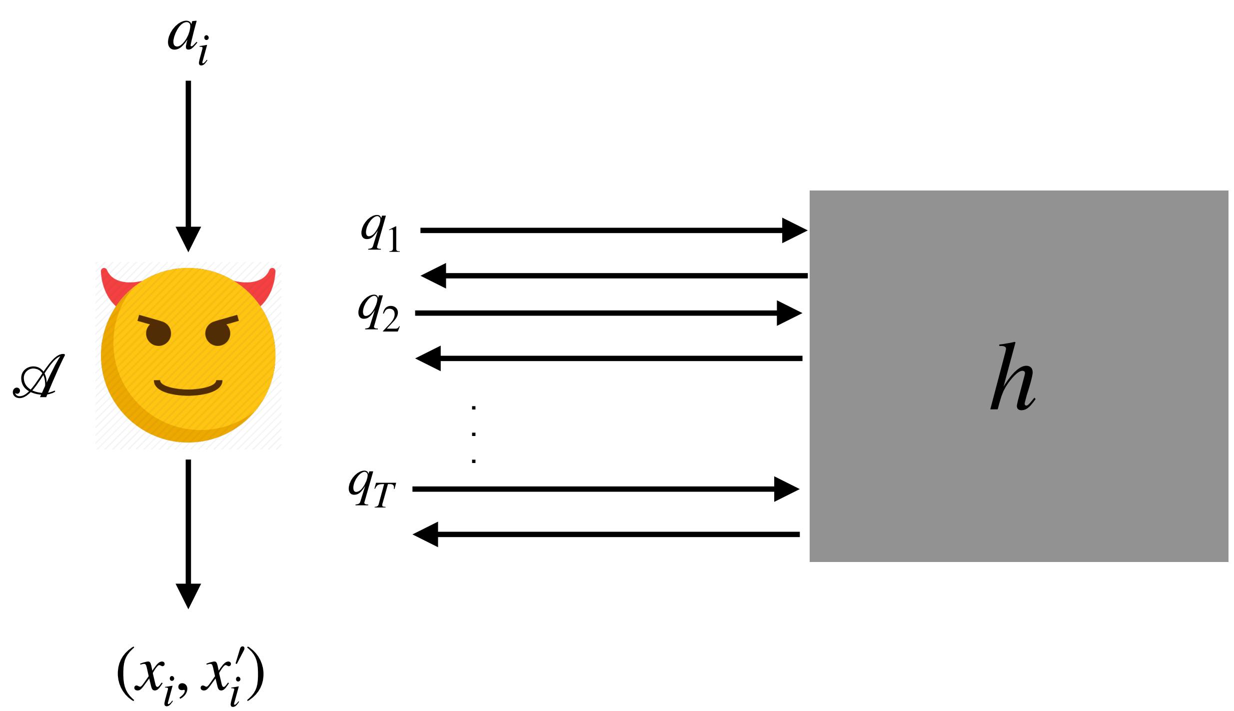
Work	Technique
[CDGS18]	Reduction to Pre-sampling
[ACDW20]	**Reduction to Multi-instance Problem + Compression**
[GK22]	Reduction to Multi-instance Problem + Compression
This Work	Reduction to modified Multi-instance Problem

Multi-instance Game

[ACDW20]

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

For $i \in [S]$**, do:**



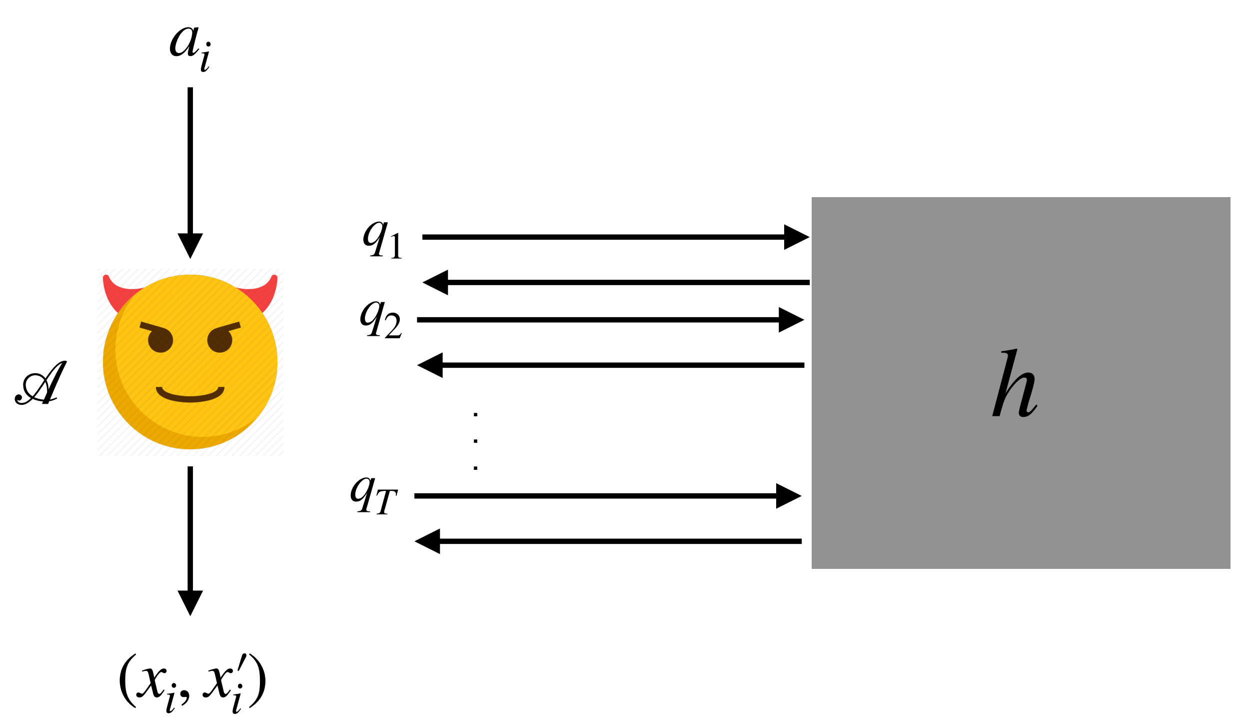
$\mathscr{A}$ **should find collisions on each salt in** $\{a_1, \ldots, a_S\}$

Multi-instance Game

[ACDW20]

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

For $i \in [S]$, **do:**



Advantage in **Multi-instance game** is $\leq \delta^S$

$\implies$ Advantage in **Pre-computation model** is $\leq 2 \cdot \delta$

$\mathscr{A}$ **should find collisions on each salt in** $\{a_1, \ldots, a_S\}$

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

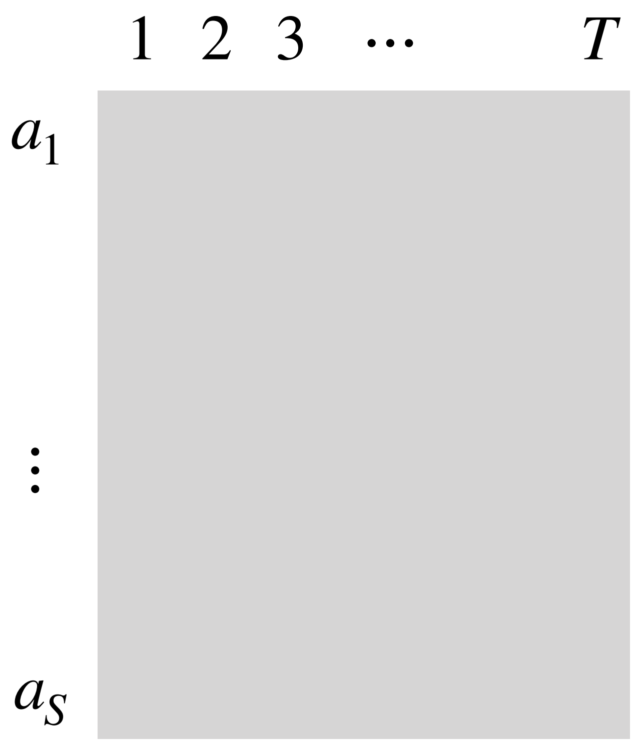
- Full talk on YouTube
- eprint: 2022/309

Comparison of Techniques

Work	Technique
[CDGS18]	Reduction to Pre-sampling
[ACDW20]	Reduction to Multi-instance Problem + Compression
[GK22]	Reduction to Multi-instance Problem + Compression
This Work	**Reduction to modified Multi-instance Problem**

Simplifying the Model

X_i : indicator of succeeding on salt a_i

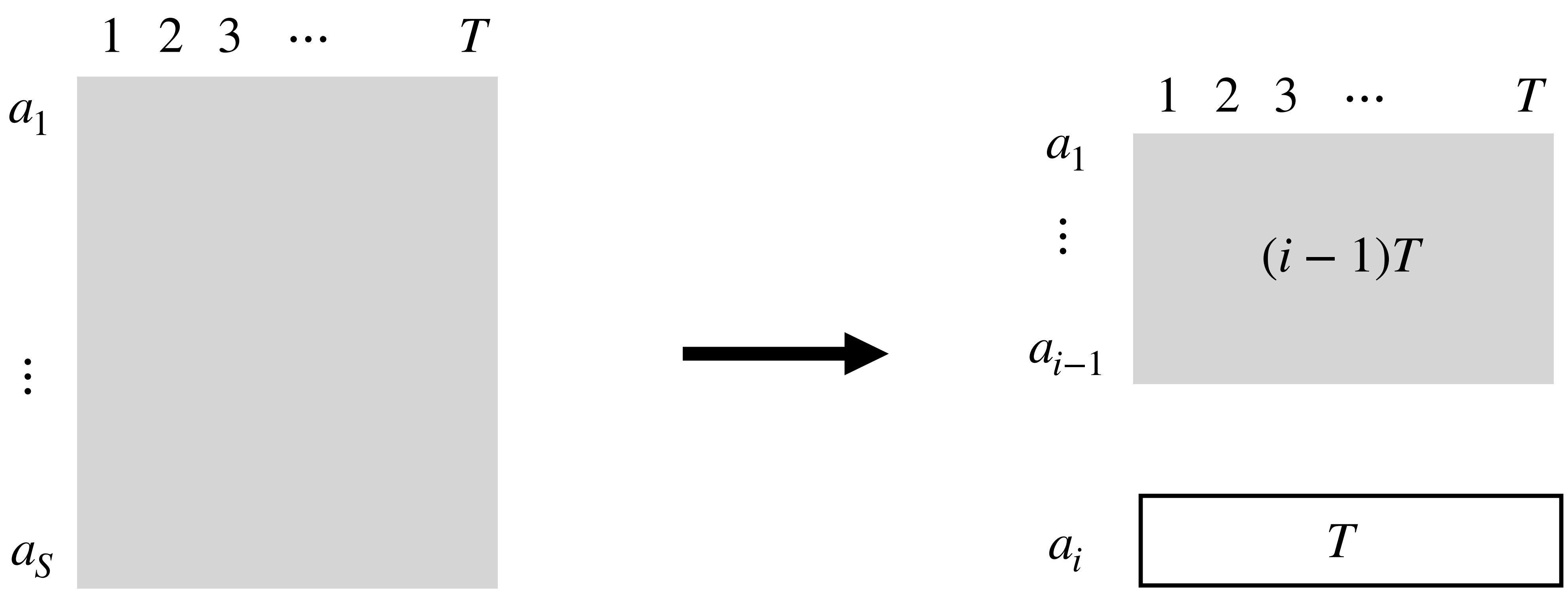


[ACDW20,GK22] bound

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

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

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

for any $i \in [S]$

Simplifying the Model

a_1

$\vdots$

a_{i-1}

$(i-1)T$

$\leq ST$

'Offline' queries

a_i

T

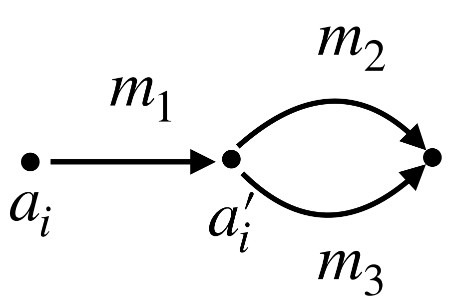
'Online' queries

Suffices to **bound** Pr $\left[X_i \mid X_{<i}\right]$ to

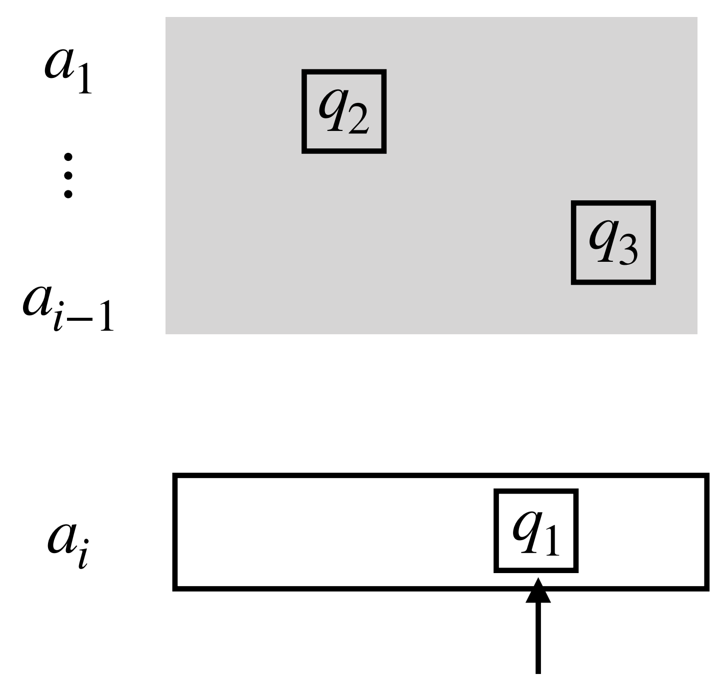
$$O\left(\max\left\{1, \frac{ST^2}{N}\right\} \cdot \frac{STB}{N} + \frac{T^2}{N}\right)$$

Example

Consider collision type



such that



'Offline' queries

'Online' queries

$$q_1 := (a_i, m_1)$$

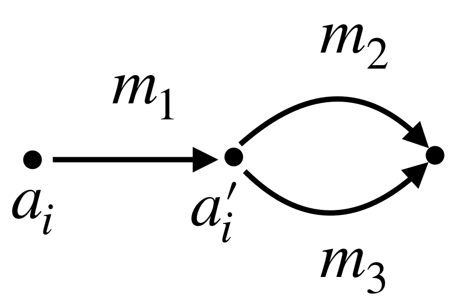
$$q_2 := (a_i', m_2)$$

$$q_3 := (a_i', m_3)$$

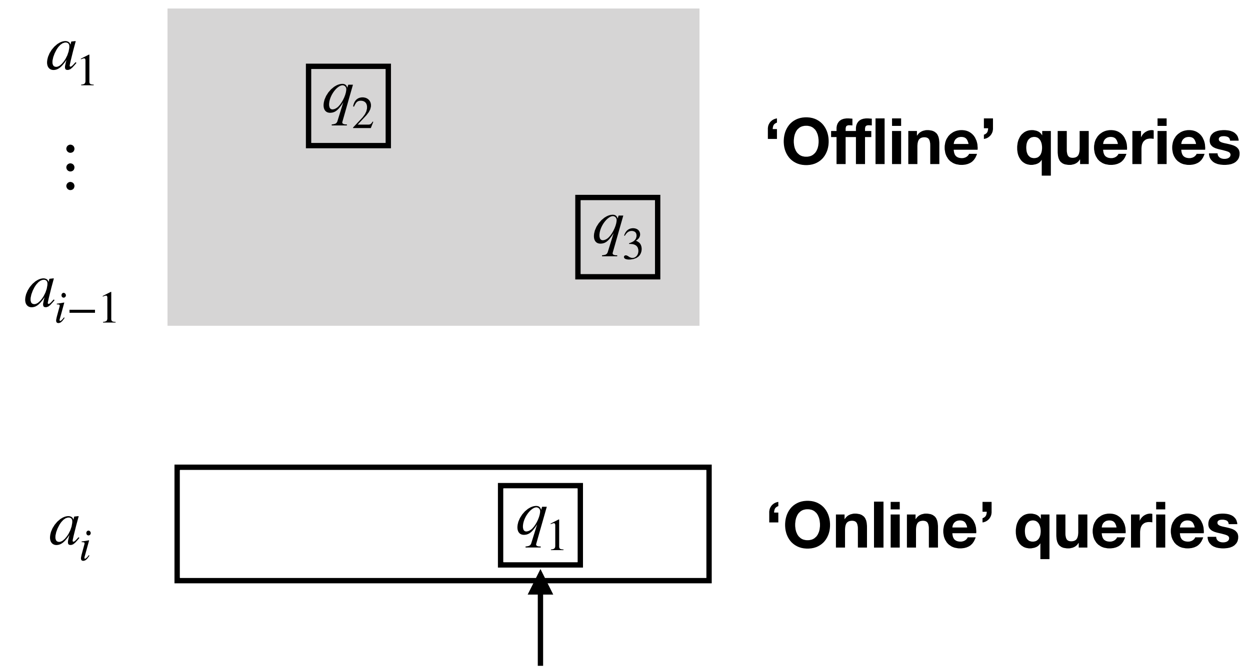
The output of q_1 is limited to certain values

Example

Consider collision type



such that



$$q_1 := (a_i, m_1)$$

$$q_2 := (a_i', m_2)$$

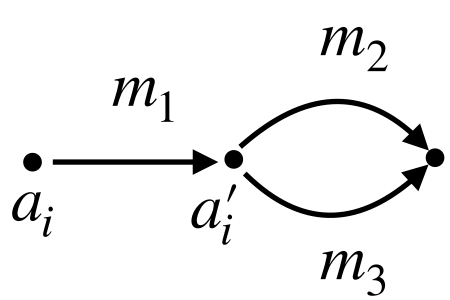
$$q_3 := (a_i', m_3)$$

It should be input salt of one of ST queries in 'Offline' phase

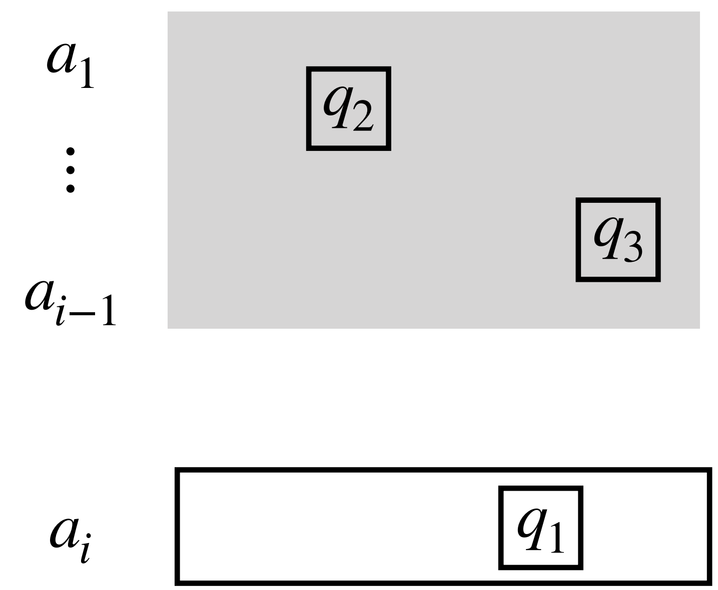
$$\Longrightarrow \ \Pr[\exists \ q_1\text{-like Online query}] \leq ST^2/N$$

Example

Consider collision type



such that



'Offline' queries

'Online' queries

$$q_1 := (a_i, m_1)$$

$$q_2 := (a_i', m_2)$$

$$q_3 := (a_i', m_3)$$

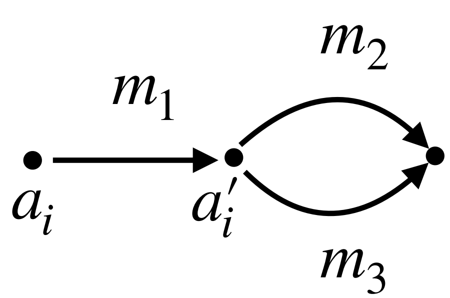
It should be input salt of one of ST queries in 'Offline' phase

$$\Longrightarrow \Pr[\exists\ q_1\text{-like Online query}] \le ST^2/N$$

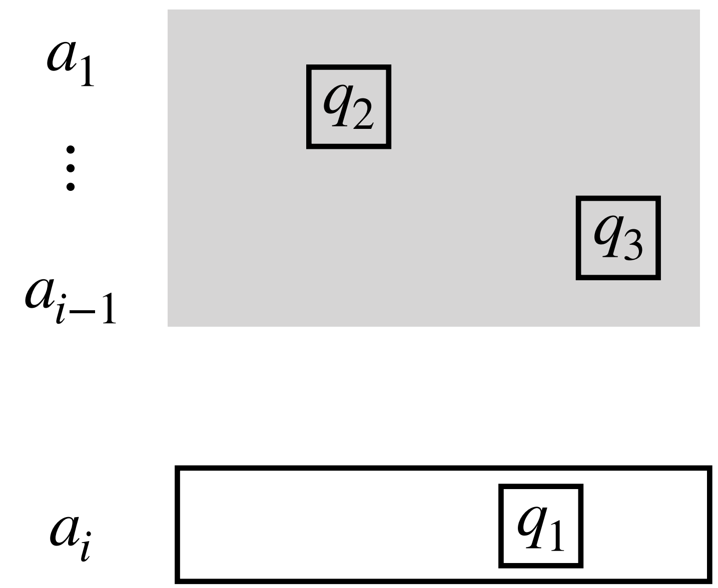
But we can bound better!

Example

Consider collision type



such that



'Offline' queries

'Online' queries

$$q_1 := (a_i, m_1)$$

$$q_2 := (a_i', m_2)$$

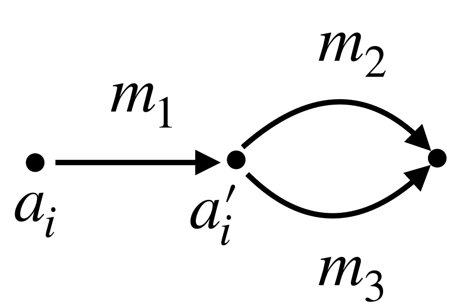
$$q_3 := (a_i', m_3)$$

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

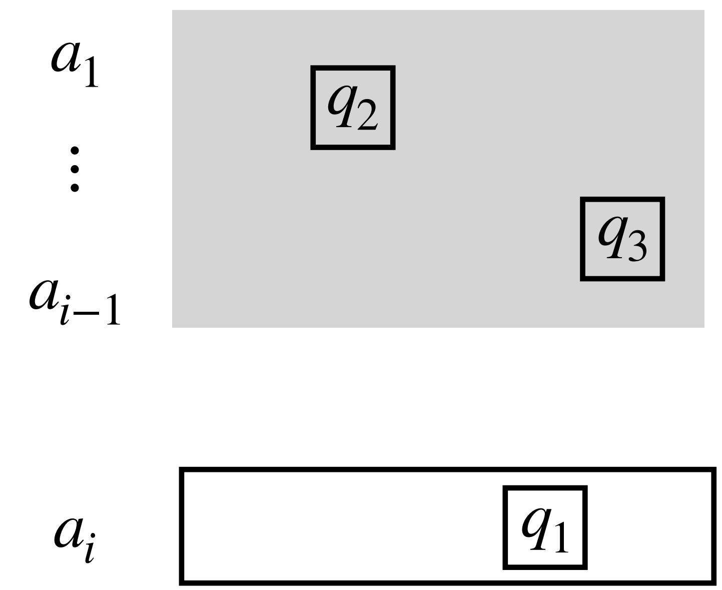
We refer to this as **<u>'useful knowledge gain'</u>** from offline queries

Example

Consider collision type



such that



‘Offline’ queries

‘Online’ queries

$q_1 := (a_i, m_1)$

$q_2 := (a_i', m_2)$

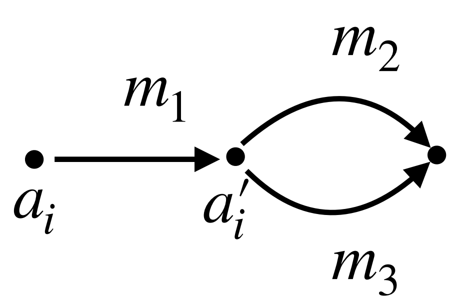
$q_3 := (a_i', m_3)$

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

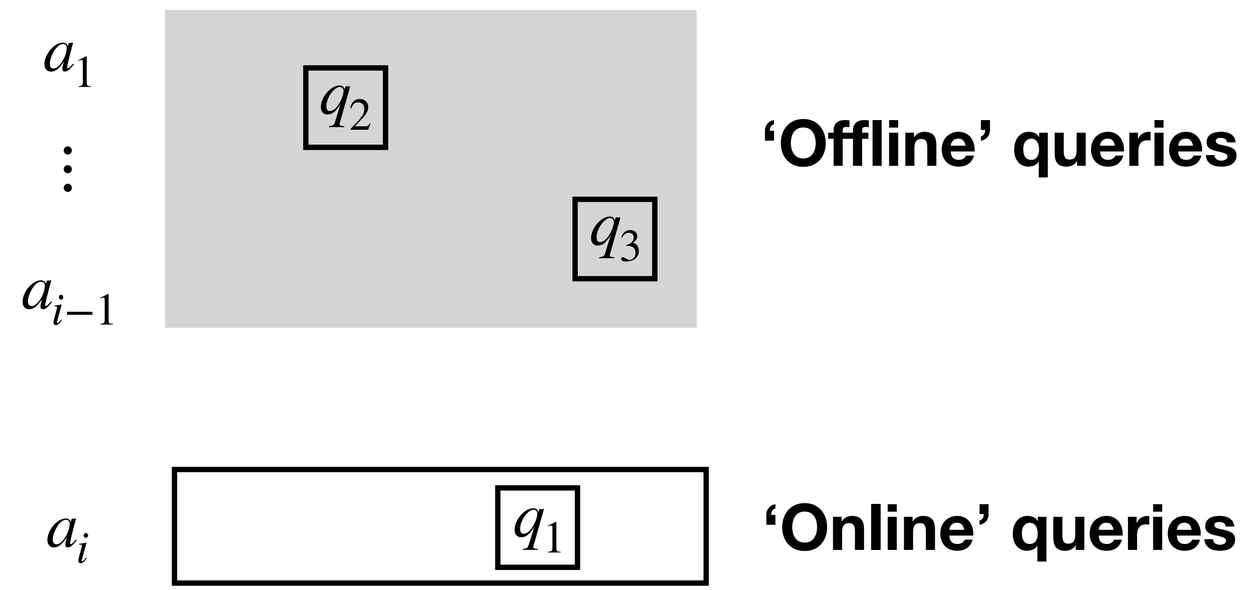
There can be at most $ST/2$ pairs starting from distinct salts.

Example

Consider collision type



such that



$q_1 := (a_i, m_1)$

$q_2 := (a_i', m_2)$

$q_3 := (a_i', m_3)$

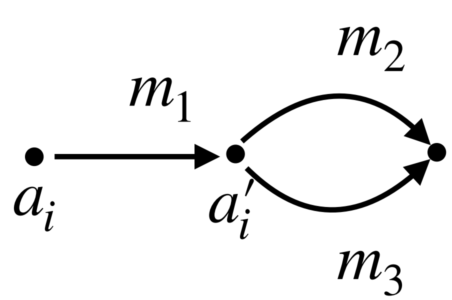
How many (q_2, q_3) such pairs can there be in 'Offline' queries?

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

This is **worst-case analysis**, similar to **Pre-sampling**

Example

Consider collision type

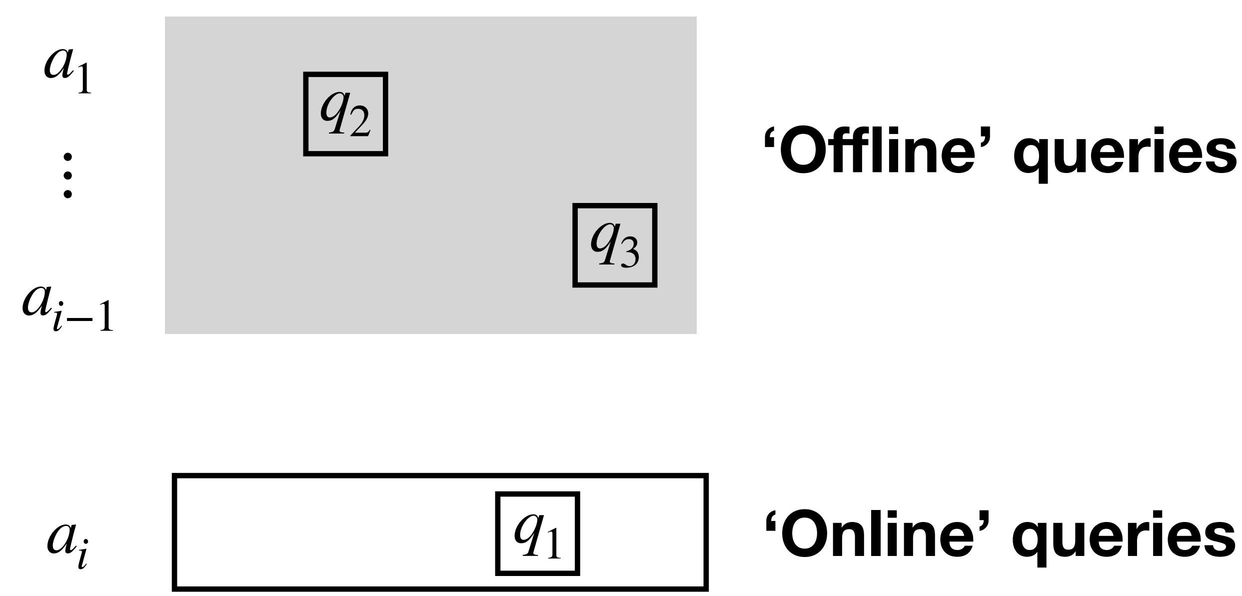


$q_1 := (a_i, m_1)$

$q_2 := (a_i', m_2)$

$q_3 := (a_i', m_3)$

such that



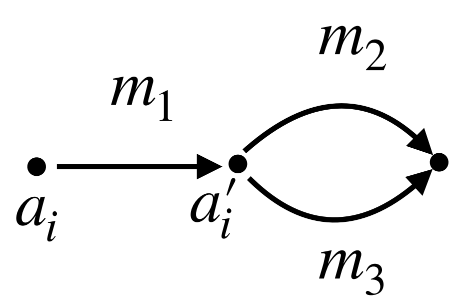
How many (q_2, q_3) such pairs can there be in 'Offline' queries?

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

We bound better via **average-case** analysis

Example

Consider collision type

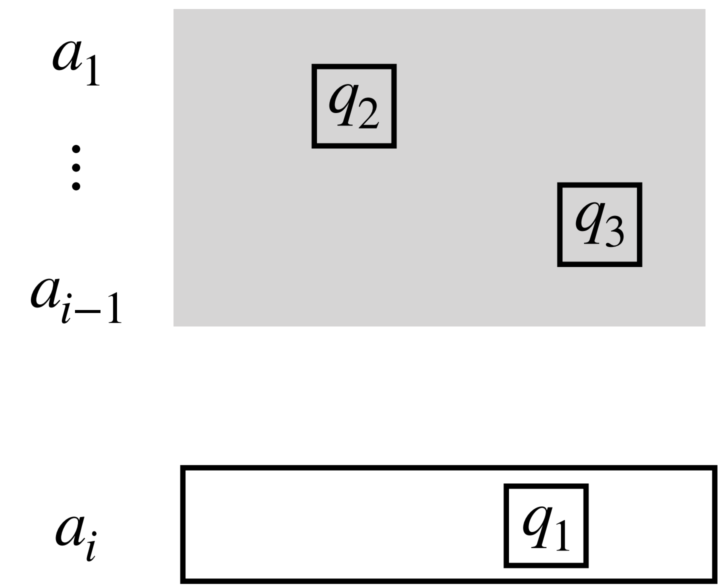


$q_1 := (a_i, m_1)$

$q_2 := (a_i', m_2)$

$q_3 := (a_i', m_3)$

such that



'Offline' queries

'Online' queries

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'**

$\Longrightarrow$ $\Pr[\exists\ q_1$-like Online query$] \leq ST/N$

Proof Overview

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

Proof Overview

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

Proof Overview

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

Paper https://eprint.iacr.org/**2022/885**.pdf