Time-Memory Tradeoffs for Short Hash Collisions

Akshima
University of Chicago

Joint work with David Cash, Andrew Drucker, Hoeteck Wee

This Talk

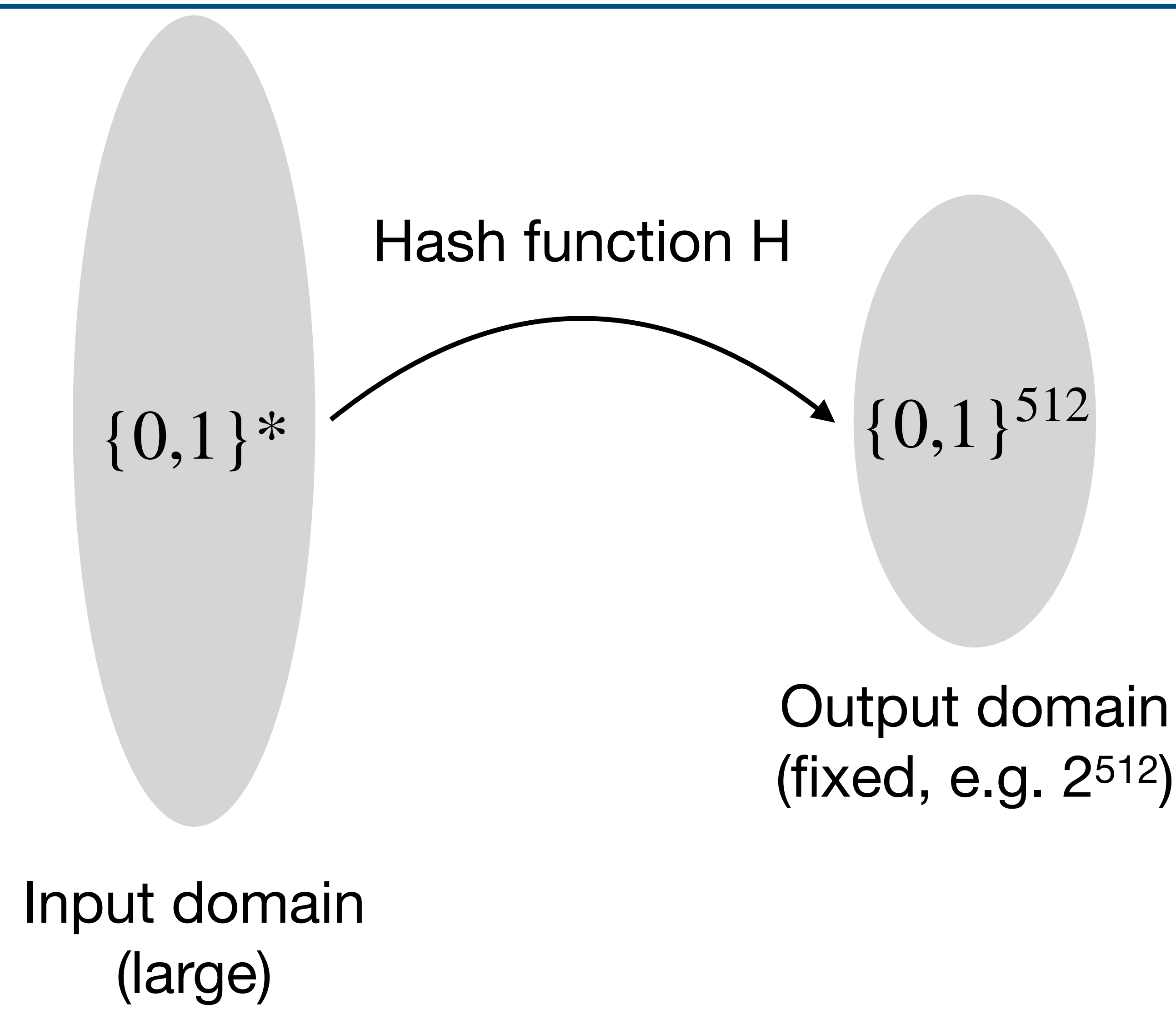
- Inspects time-space tradeoffs for finding ***short*** collisions in Merkle-Damgård hash functions.

- Shows gaps in complexity of finding 1, 2 and B-block collisions.

Talk Outline

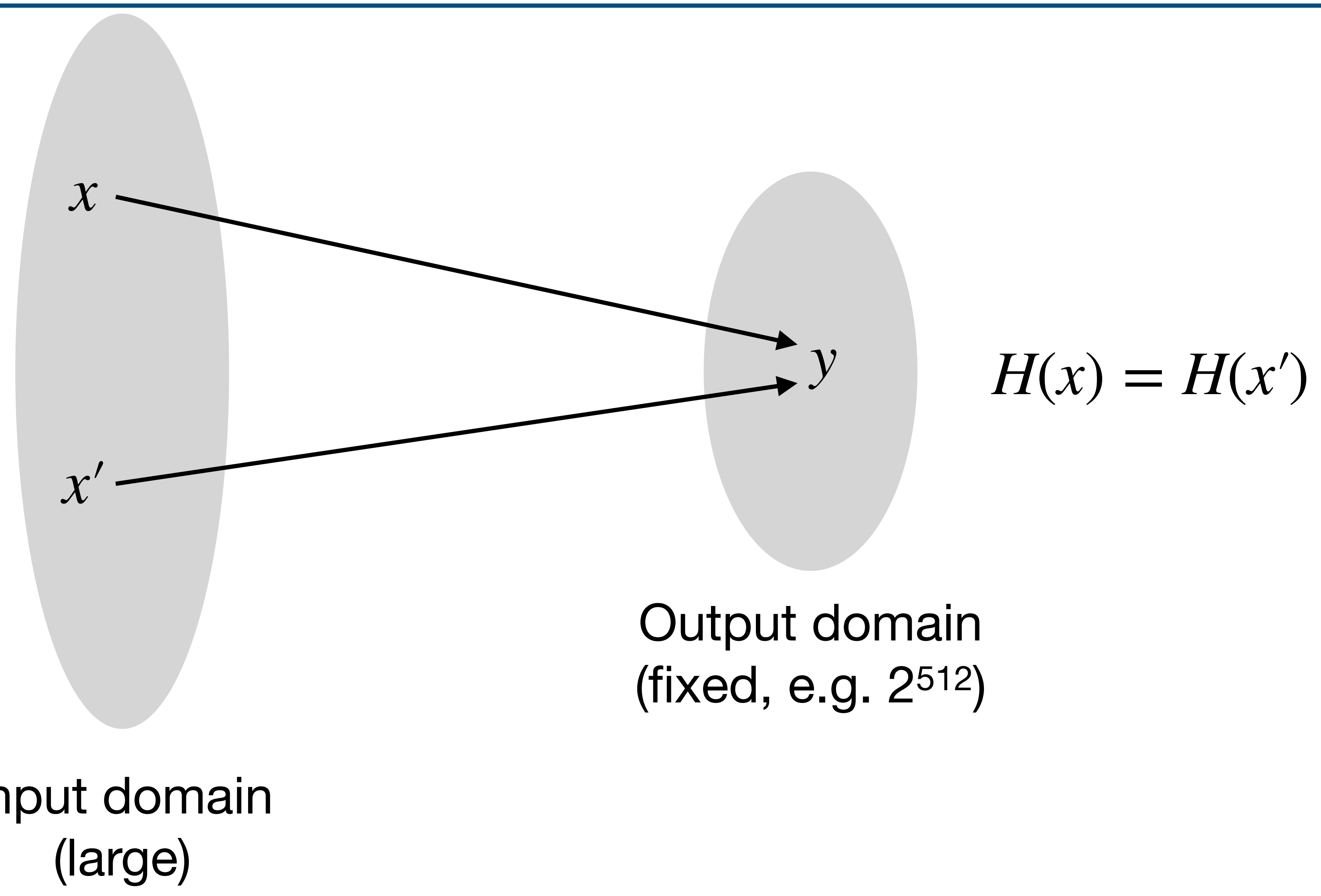
- **Basic definitions**
- Our work and comparison with prior work
- Why prior techniques cannot extend to *short* collisions
- Our technique for
  - Bound on 2-block collisions
  - Bound on zero-walk adversaries
- Conclusion

Cryptographic Hash Functions



- Widely deployed practical hashes (SHA512, SHA3)
- Many security properties required

Collisions in Hash Functions

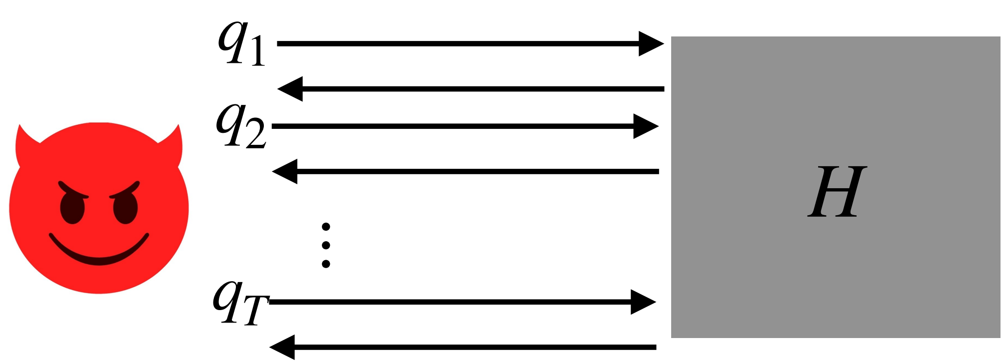


- Collisions damaging in practice (e.g. in authentication)
- Finding collisions should be very hard (e.g. 2^{256} time)

Modeling Hashes: The ROM

[Bellare-Rogaway,96]

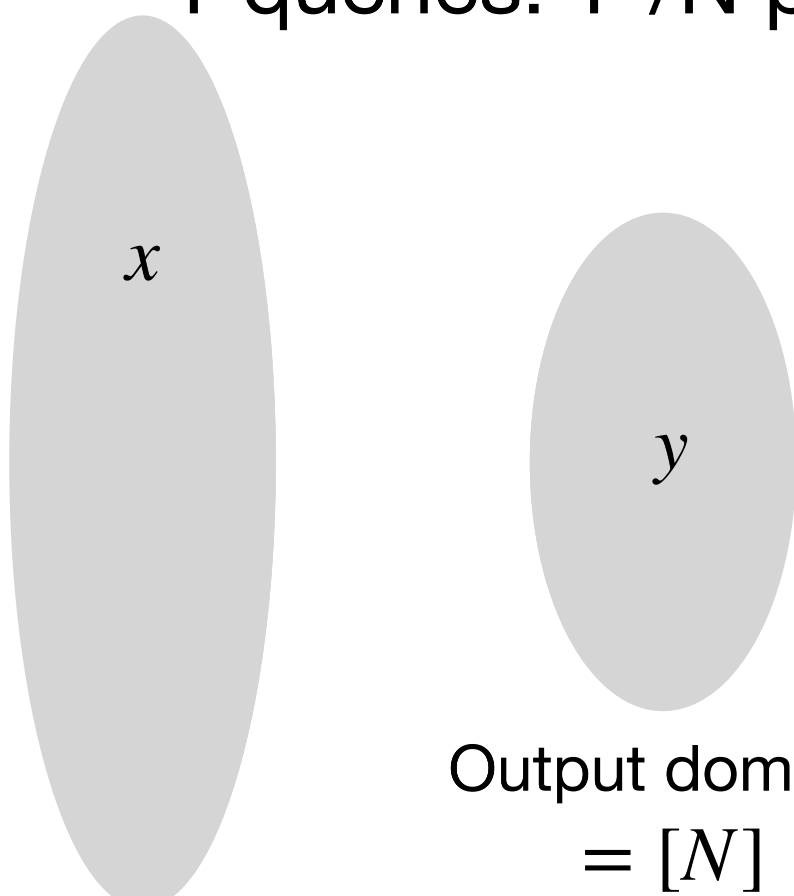
- Can't actually prove collisions are hard to find (P vs NP)
- Instead, pretend H is a random function and give proofs
  - Called the "random oracle model" (ROM)
- Adversary is computationally unbounded and deterministic.



T: # queries

Finding Collisions in the ROM

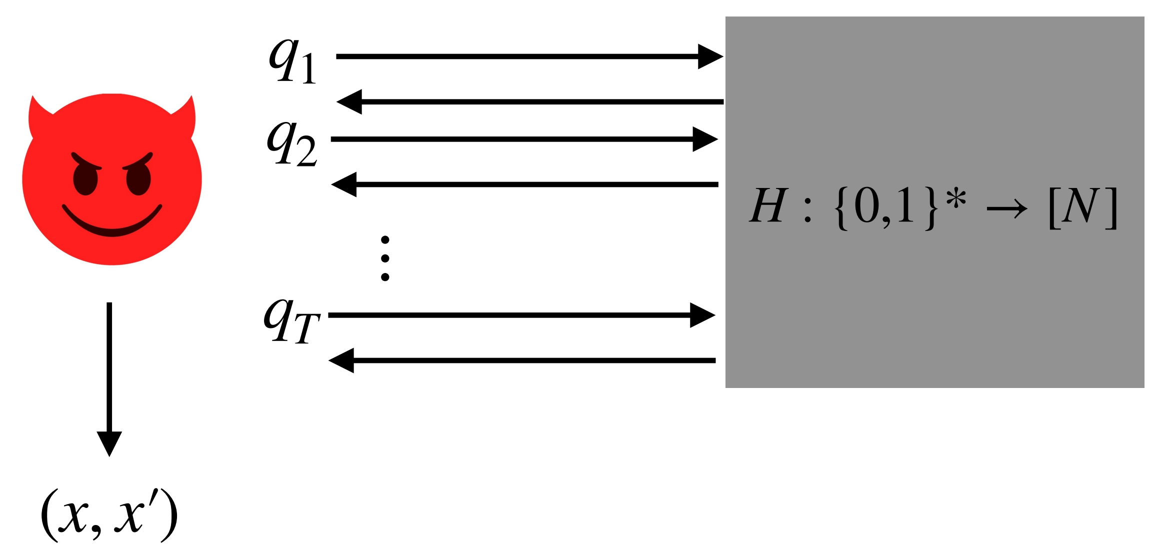
- Can prove unconditionally that a random function is collision resistant
- T queries: T^2/N probability of success



Output domain
$= [N]$

where $[N] = \{1,2,\ldots,N\}$

Input domain
$= \{0,1\}^*$

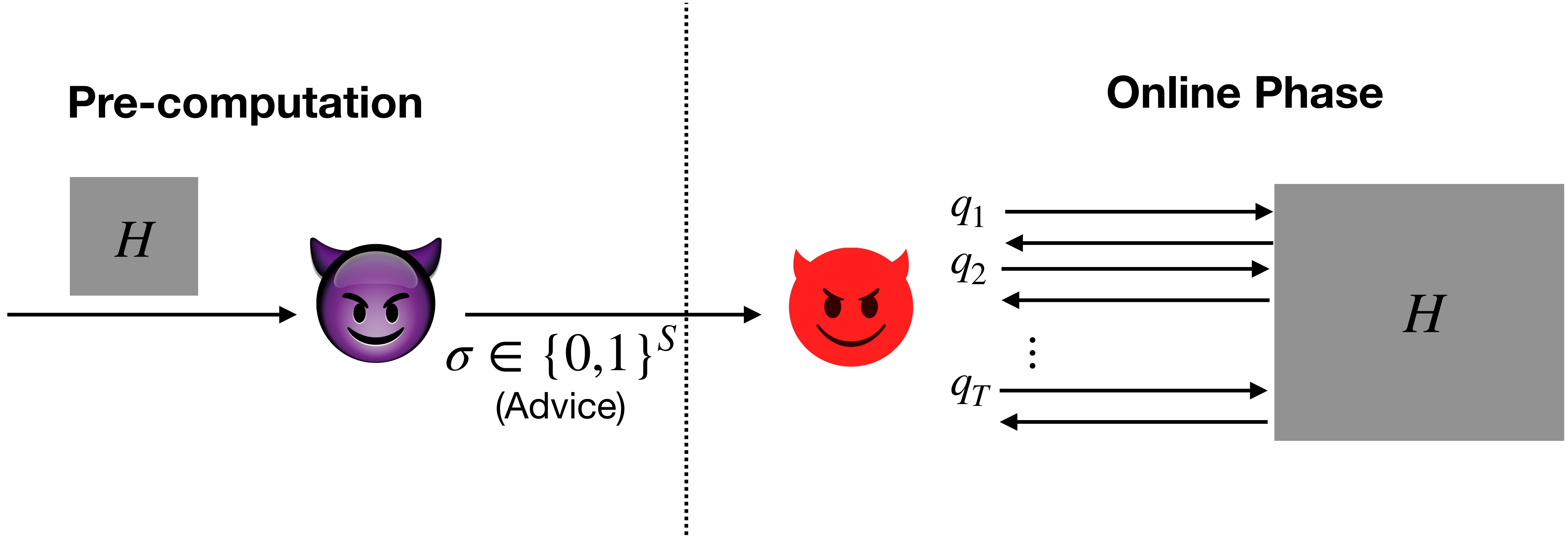


$\Pr[x \neq x' \text{ and } H(x) = H(x')] \leq T^2/N$

Pre-Computation in the ROM

[Unruh,07]

- **Unbounded** pre-computation produces S bits of advice
- **Bounded** T number of queries in online phase

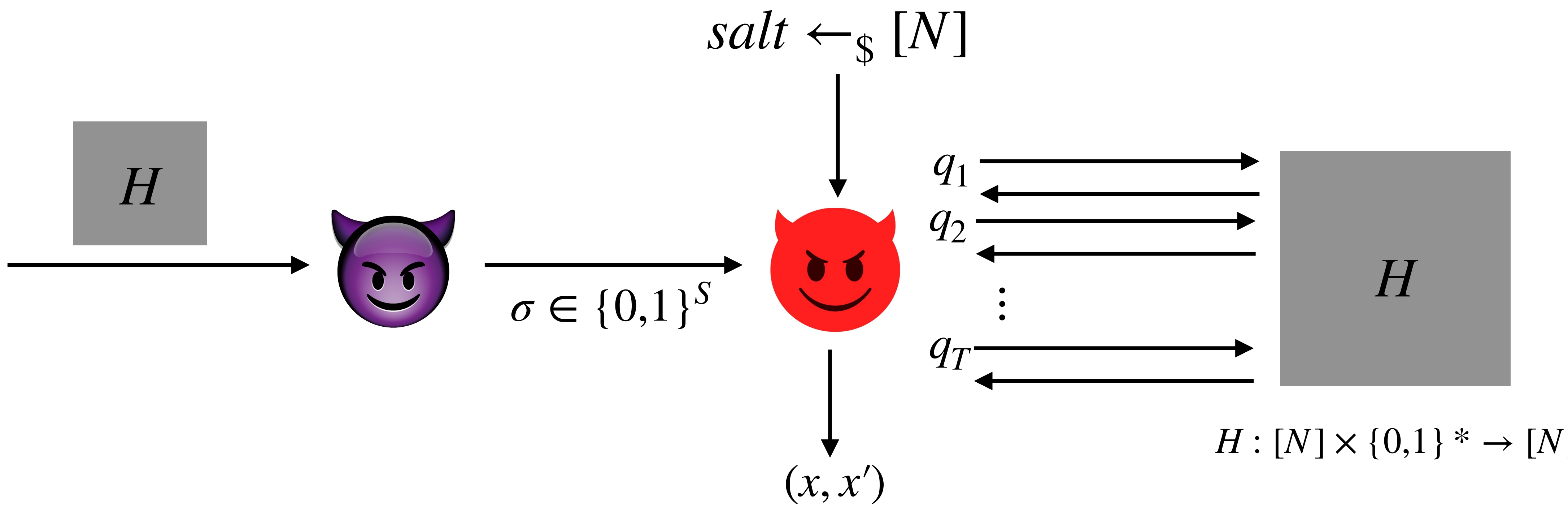


- Trivial attack: Just precompute a collision.

Salting to Confound Pre-Computation

[Dodis-Guo-Katz,17]

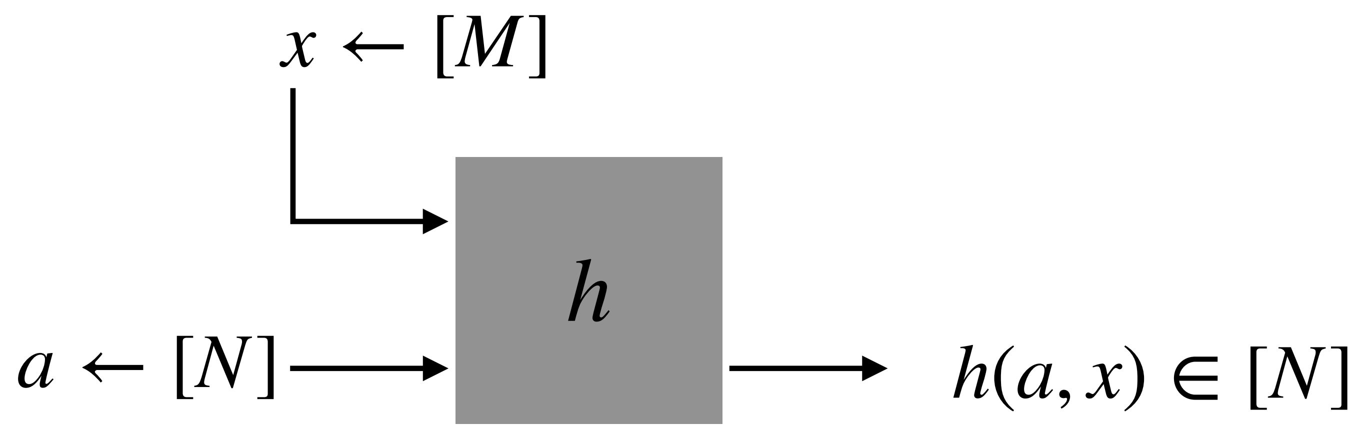
- Require adversary to find collision with a random prefix, called a *salt*
  - Adversary learns salt only in online phase
  - Defeats trivial attack



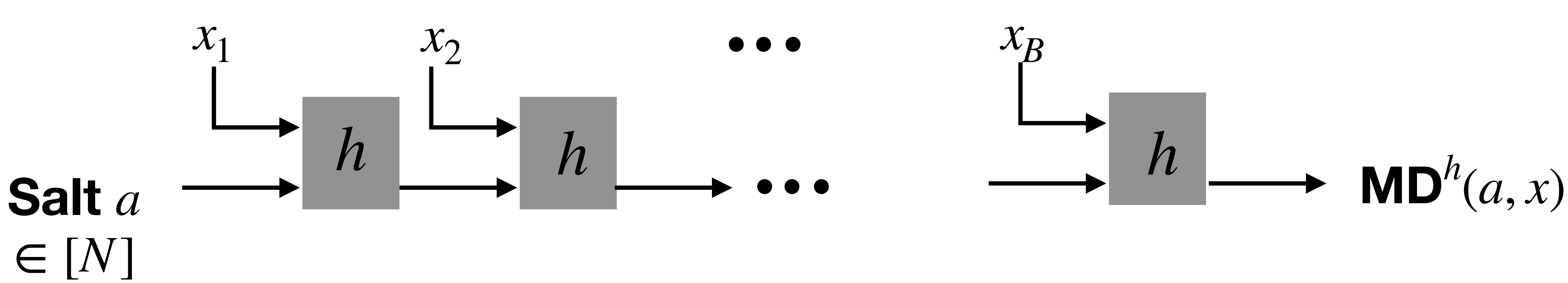
$$\Pr[x \neq x' \textbf{ and } H(salt, x) = H(salt, x')] = \tilde{\theta}\left((S + T^2)/N\right)$$

- Showed optimal attack is to write down S collisions and hope there is a collision for input $salt$ or perform birthday.

Merkle-Damgård Hash Functions



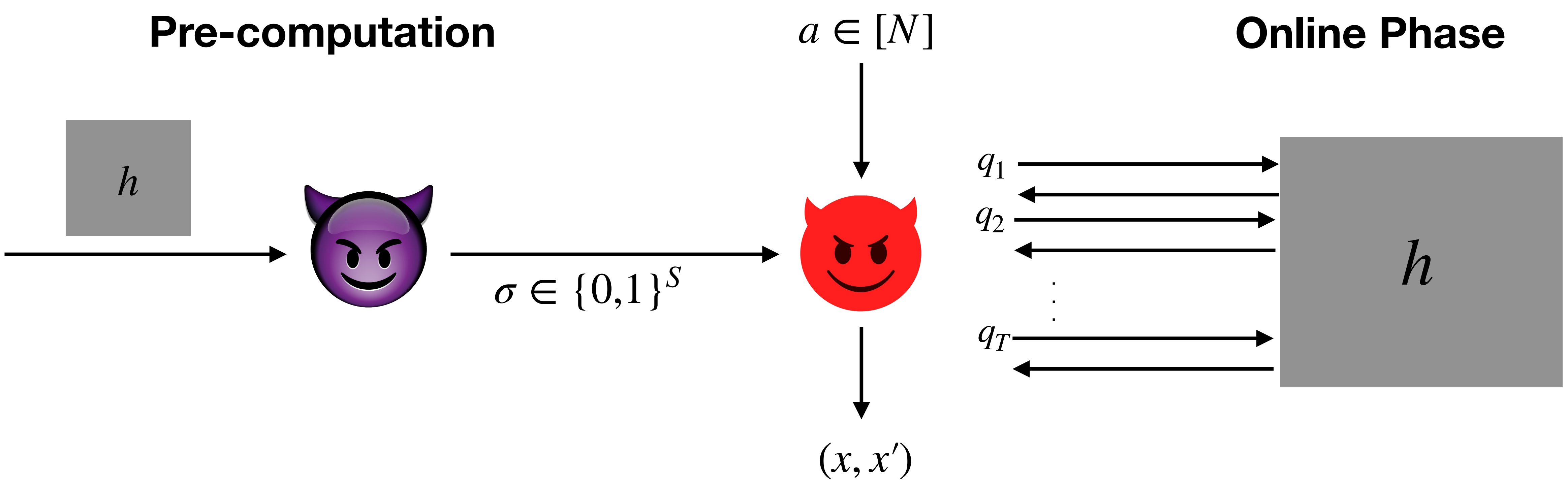
Input $x = x_1 || \ldots || x_B,\ x_i \in [M]$



Salting Merkle-Damgård

[Coretti-Dodis-Guo-Steinberger,18]

- h is modeled as RO
- Adversary must find salted collision in H = MD^h



$$\Pr[x \neq x' \text{ and } \mathbf{MD}^h(a,x) = \mathbf{MD}^h(a,x')] = \tilde{\theta}(ST^2/N)$$

- Non-trivial *time-space tradeoffs* improve over birthday using advice ($T = S = N^{1/3}$)

Talk Outline

- Basic definitions
- **Our work and comparison with prior work**
- Why prior techniques cannot extend to *short* collisions
- Our technique for
  - Bound on 2-block collisions
  - Bound on zero-walk adversaries
- Conclusion

Our Work

Initiate study of *short* collision-finding in Merkle-Damgård hashes with pre-computation

- Same model as before, but adversary is required to find colliding messages with B or fewer blocks.

Our Work

Initiate study of *short* collision-finding in Merkle-Damgård hashes with pre-computation

- Same model as before, but adversary is required to find colliding messages with B or fewer blocks.

Result 1: Qualitative time-space hardness jumps from $B = 1$, $B = 2$, and unbounded B lengths.

- Via new concentration+compression-based techniques

Our Work

Initiate study of *short* collision-finding in Merkle-Damgård hashes with pre-computation

- Same model as before, but adversary is required to find colliding messages with B or fewer blocks.

Result 1: Qualitative time-space hardness jumps from $B = 1$, $B = 2$, and unbounded B lengths.

- Via new concentration+compression-based techniques
- **Open**: Fine-grained bounds for $B = 3,4,\ldots$

Our Work

Initiate study of *short* collision-finding in Merkle-Damgård hashes with pre-computation

- Same model as before, but adversary is required to find colliding messages with B or fewer blocks.

<u>Result 1</u>: Qualitative time-space hardness jumps from $B = 1$, $B = 2$, and unbounded B lengths.

- Via new concentration+compression-based techniques
- **Open**: Fine-grained bounds for $B = 3, 4, \ldots$

<u>Result 2</u>: Impossibility for restricted class of attacks on general B (includes all known attacks).

Our Concrete Results

Work	# Blocks in Collision	Advantage Bound S: advice size T: Queries
[DGK17]	1	$\tilde{\theta}\left(\frac{S+T^2}{N}\right)$
[CDGS18]	Unbounded	$\tilde{\theta}\left(\frac{ST^2}{N}\right)$
Our Work	B	$\tilde{\Omega}\left(\frac{STB}{N}\right)$
Our Work	B (only for restricted adversary)	$\tilde{O}\left(\frac{STB}{N}\right)$
Our Work	2	$\tilde{\theta}\left(\frac{ST}{N}\right)$

Why Short Collisions?

- Consider SHA2: $N=2^{256}$, $M=2^{512}$
  - When $S=2^{70}$, $B=T= 2^{93}$
  - Collisions have to be over 2^{93} blocks long

Why Short Collisions?

- Consider SHA2: $N=2^{256}$, $M=2^{512}$
  - When $S=2^{70}$, $B=T= 2^{93}$
  - Collisions have to be over 2^{93} blocks long

- Say we want $B= 2^{20}$, then the best known attack needs $T= 2^{166}$

Talk Outline

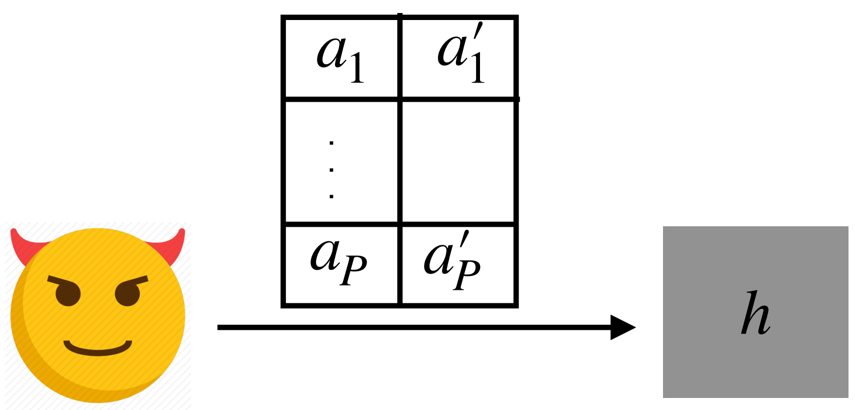
- Basic definitions
- Our work and comparison with prior work
- **Why prior techniques cannot extend to *short* collisions**
- Our technique for
  - Bound on 2-block collisions
  - Bound on zero-walk adversaries
- Conclusion

Pre-Sampling Model

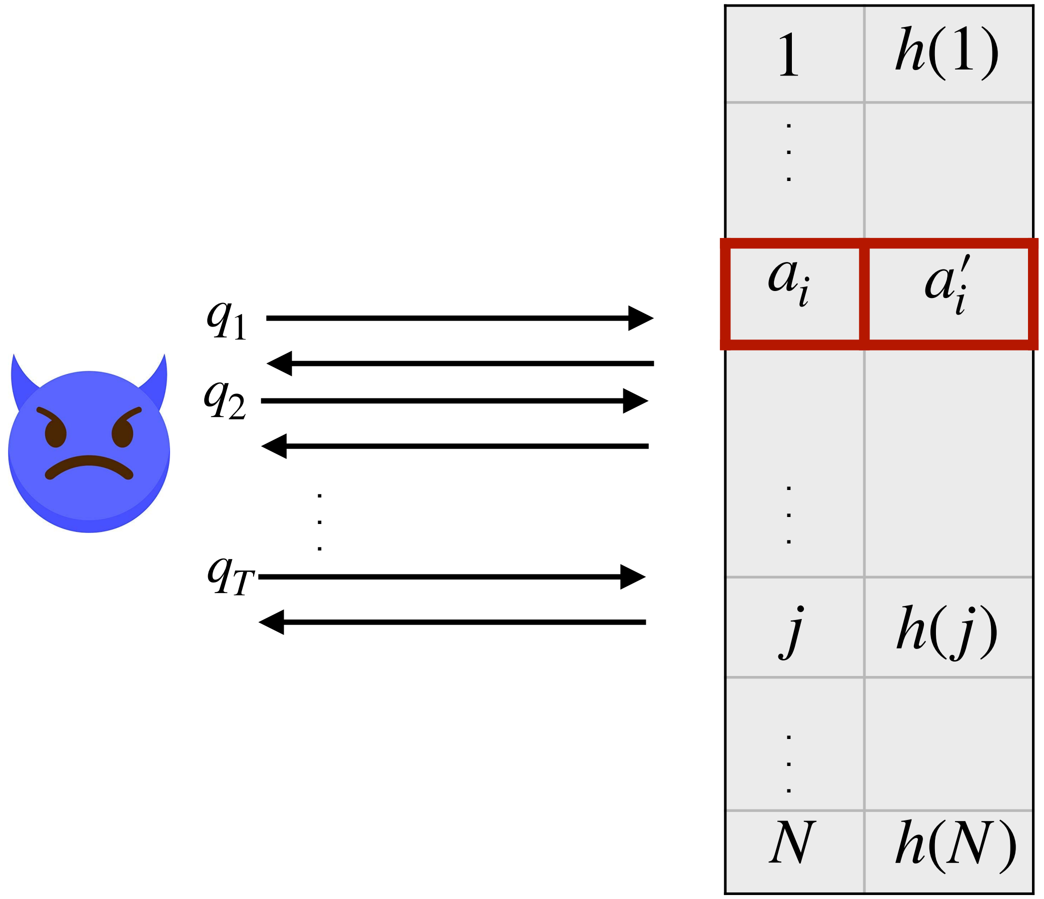
[Unruh,07]

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

Phase 1

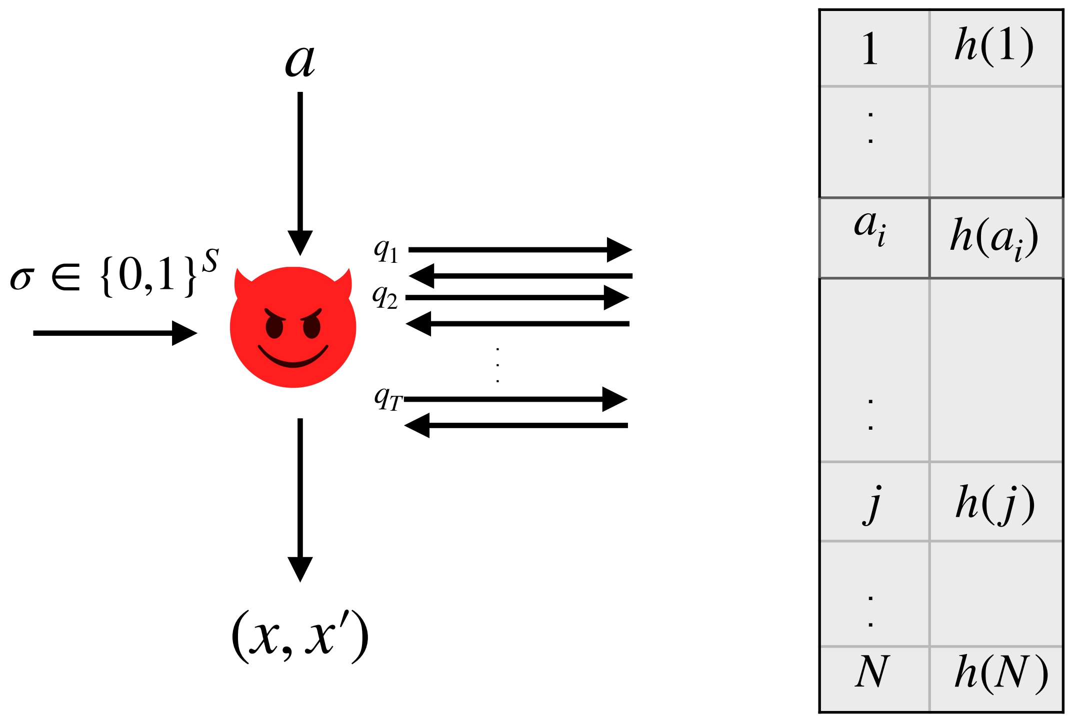


Phase 2

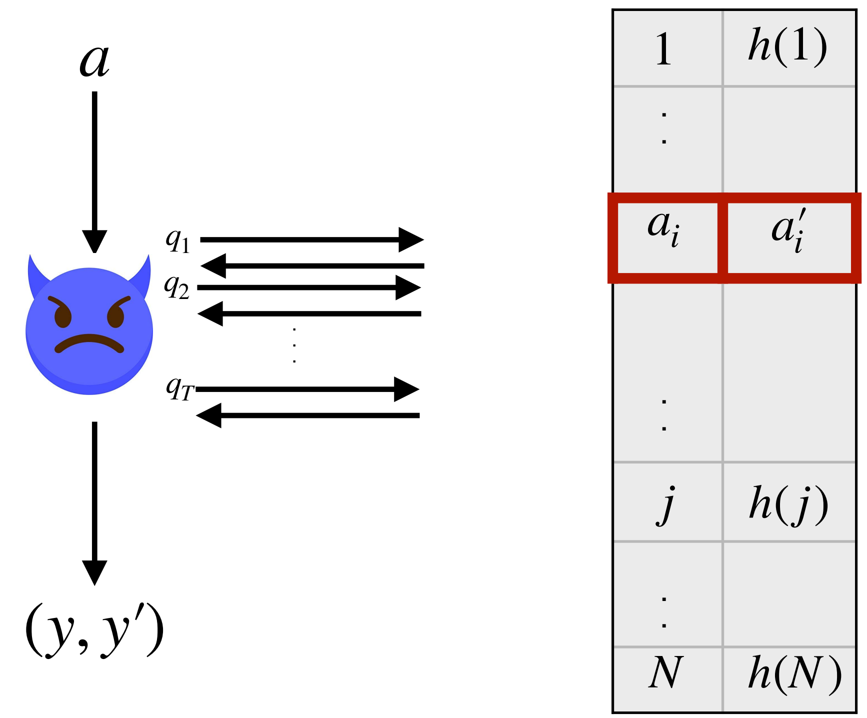


Pre-Computation to Pre-Sampling

[Unruh,07]



Adversary with Pre-computation



Adversary with Pre-sampling

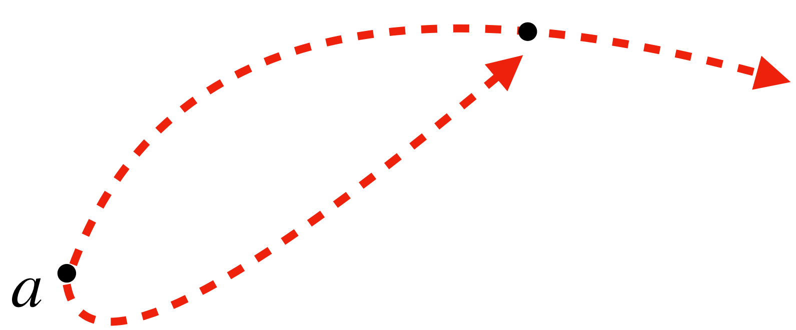
Indicates pre-fixed point

Pre-computing adversary with S-bit advice, making T queries -> Pre-sampling adversary pre-fixing ST points making T queries

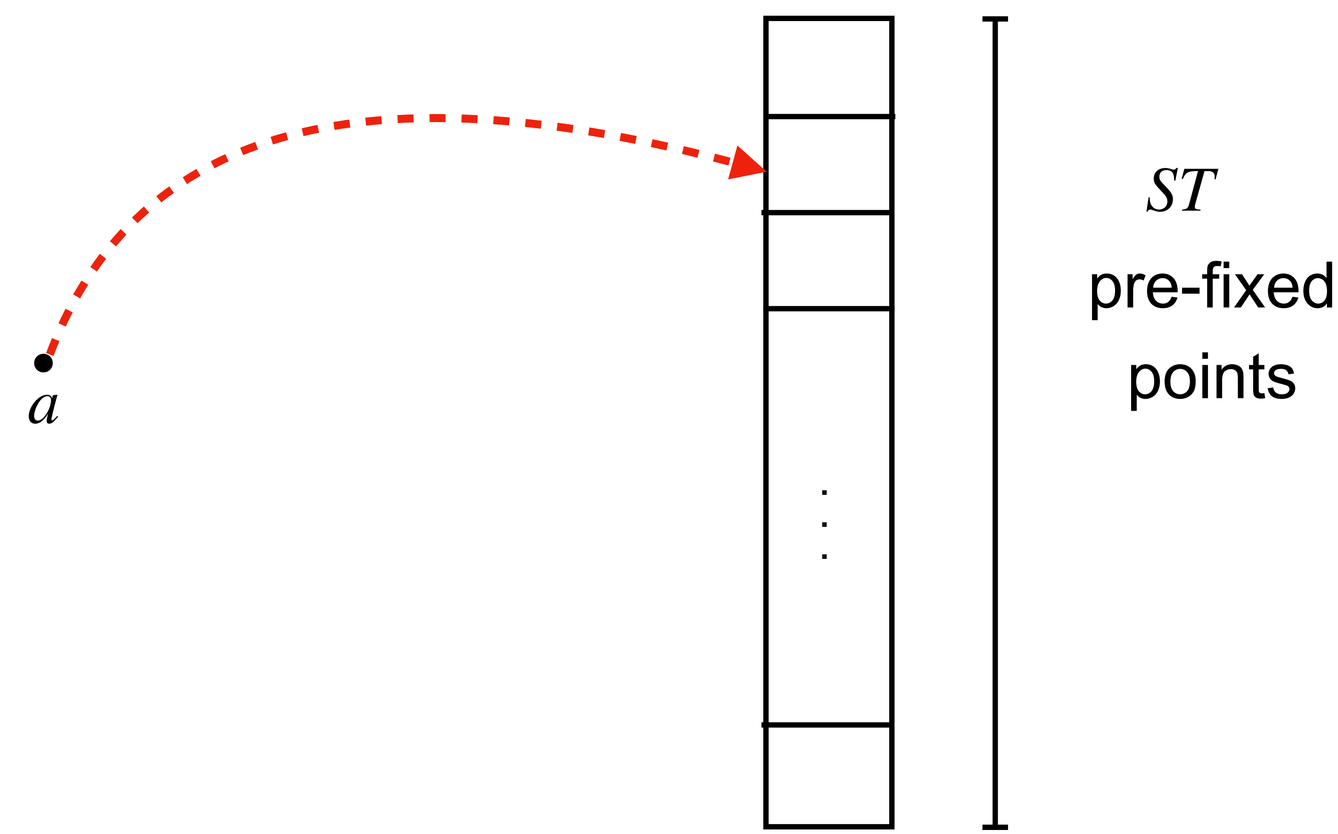
Proving impossibility of pre-sampling adversary is sufficient.

Pre-Sampling Bound, then Pre-Computation Bound [Unruh,07]

- Analyzing MD-based hash in the pre-sampling model with ST fixed points and T queries to find unbounded collisions.



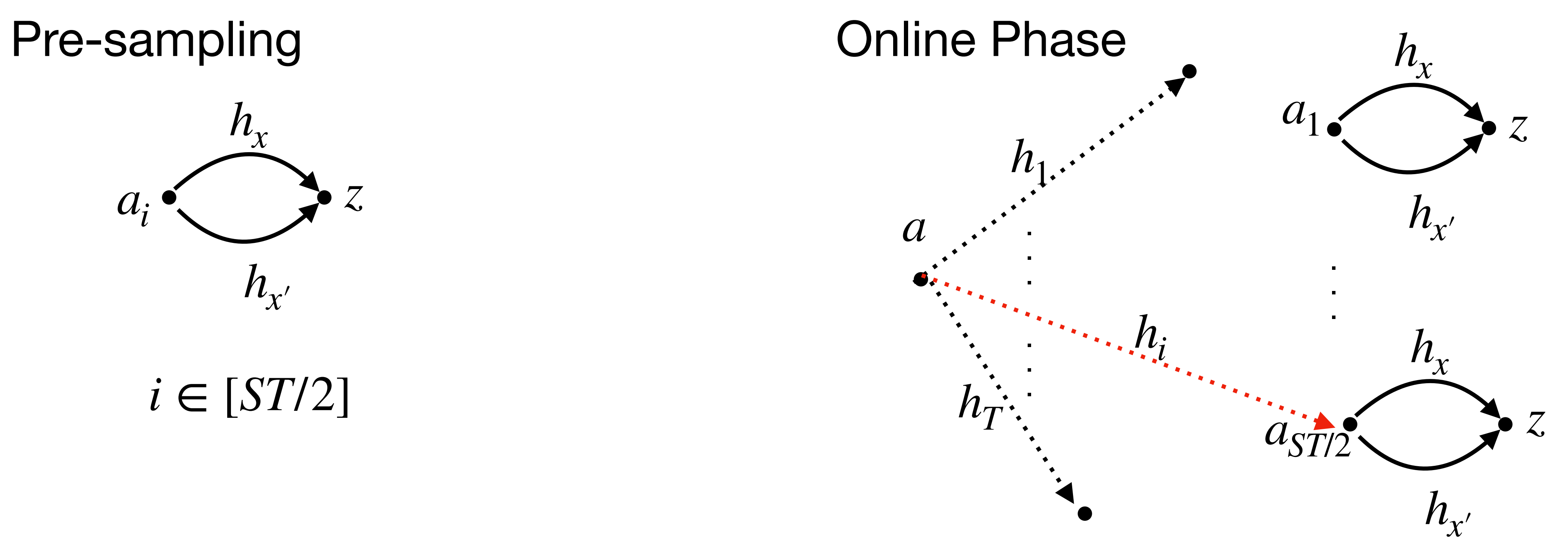
Or



This proves a bound of $O\left(ST^2/N\right)$ on finding unbounded collisions in MD hashes with Pre-computation.

Pre-Sampling is Length Insensitive

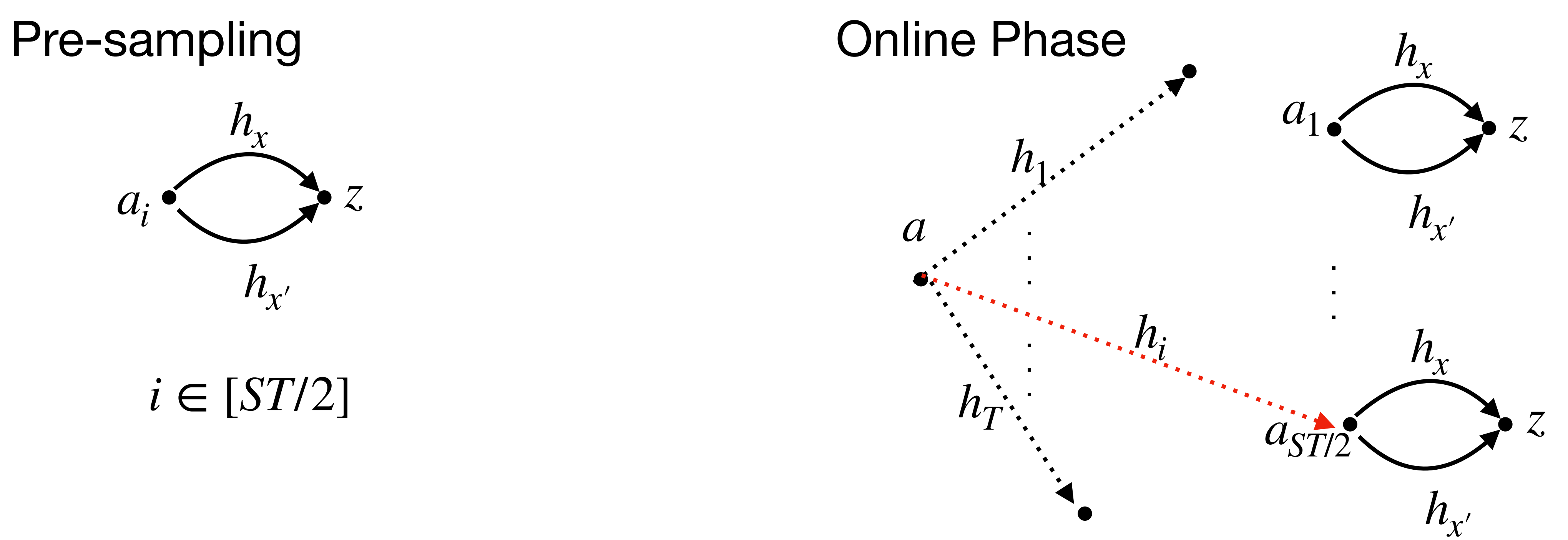
We give a 2-block collision finding attack with pre-sampling that has advantage $\Omega(ST^2/N)$.



Thus, short collisions are as easy as long collisions for pre-sampling

Pre-Sampling is Length Insensitive

We give a 2-block collision finding attack with pre-sampling that has advantage $\Omega(ST^2/N)$.

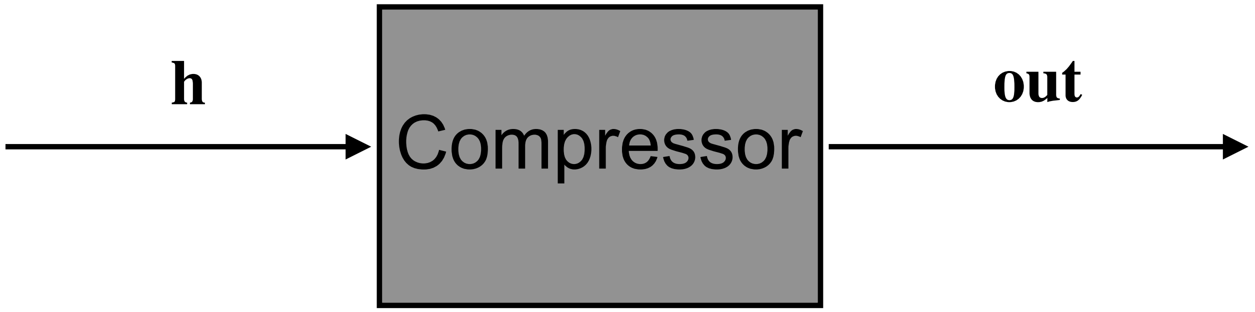


Thus, short collisions are as easy as long collisions for pre-sampling

We prove short collisions are harder than long collisions for pre-computation.

Compression Technique

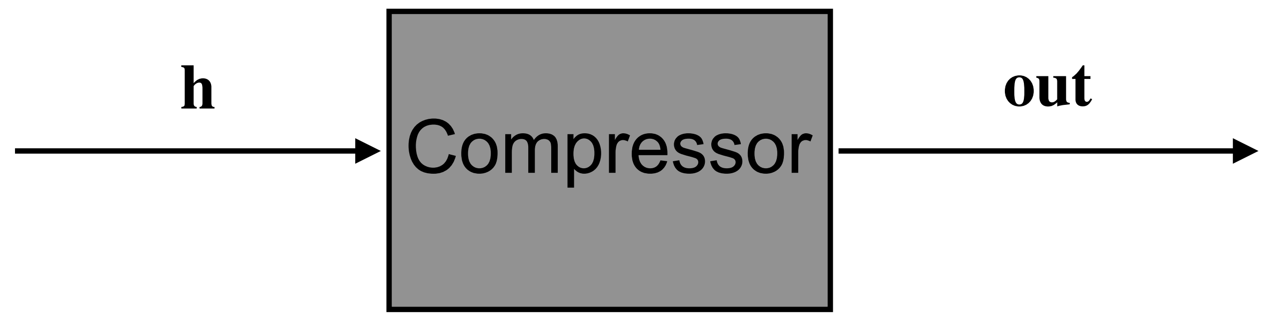
[Dodis-Guo-Katz,17]



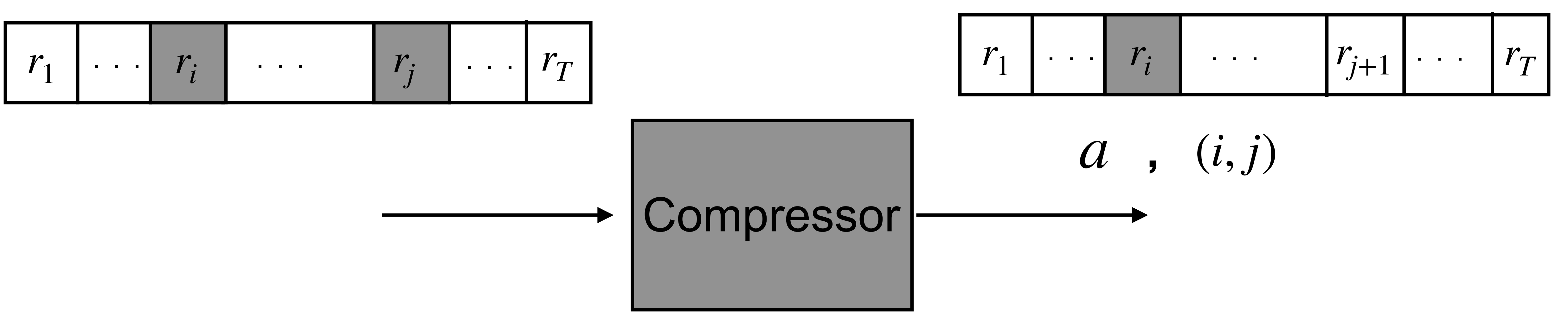
- Shannon bound: $\mathbb{E}[\,|\mathbf{out}|\,] \geq \text{entropy}(\mathbf{h})$

Compression Technique

[Dodis-Guo-Katz,17]



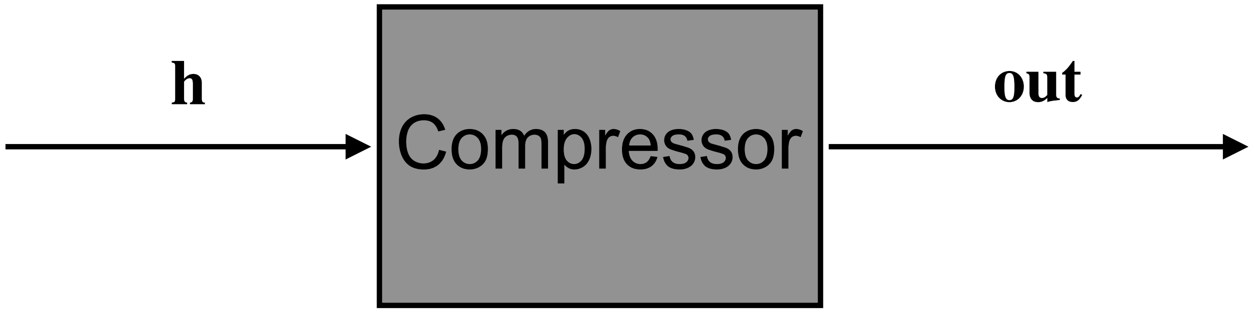
- Shannon bound: $\mathbb{E}[\,|\mathbf{out}|\,] \geq$ entropy$(\mathbf{h})$
- Say adversary $\mathscr{A}$ wins on some salt a, making queries $(q_1, \ldots, q_T)$ and getting responses $(r_1, \ldots, r_T)$. Then $\exists i, j$ such that $r_i = r_j$.



Say $\mathscr{A}$ wins on ε fraction of salts. Then compressor repeats this on every winning salt.

Compression Technique

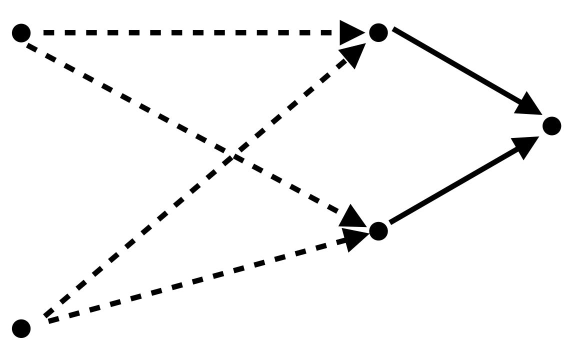
[Dodis-Guo-Katz,17]

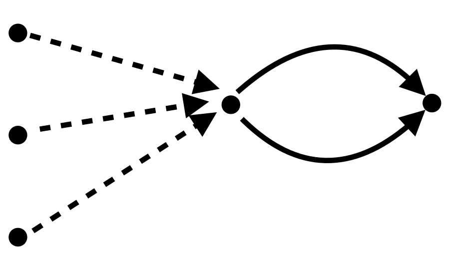


- Shannon bound: $\mathbb{E}[\,|\mathbf{out}|\,] \geq$ entropy$(\mathbf{h})$
- Say $\mathscr{A}$ wins on ε fraction of salts. Then compressor compresses $\mathbf{h}$ by at least $(\varepsilon N \cdot \log(\varepsilon N/T^2) - S)$ bits on average.
- This contradicts the Shannon bound and gives $\varepsilon \leq (S + T^2)/N$.

Extending Compression Technique Is Not Trivial

- Say some 2-block collision finding adversary $\mathscr{A}$ wins on ε fraction of salts on $\mathbf{h}$.
- Want to delete εN entries in $\mathbf{h}$ with same output as a prior entry.
- For 2-block collisions there may not be εN such unique entries.



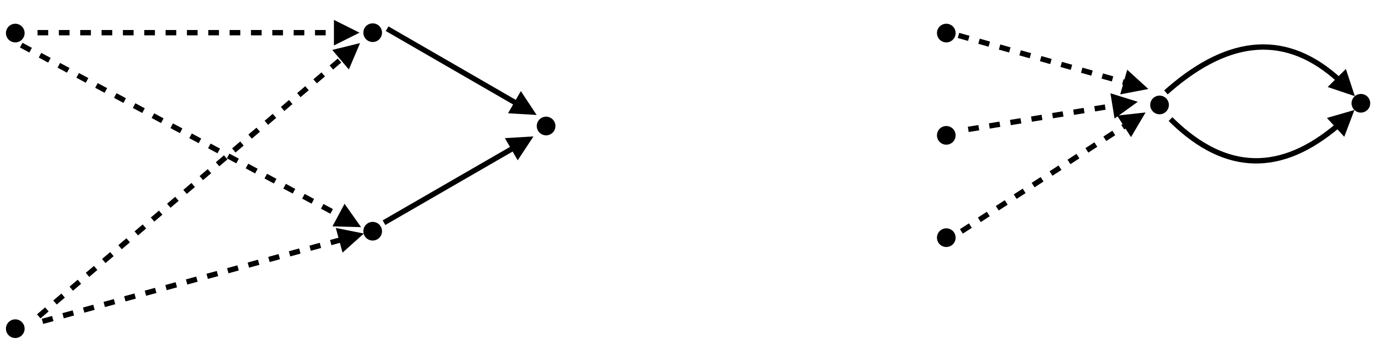


Extending Compression Technique Is Not Trivial

- Say some 2-block collision finding adversary $\mathscr{A}$ wins on ε fraction of salts on $\mathbf{h}$.

- Want to delete εN entries in $\mathbf{h}$ with same output as a prior entry.

- For 2-block collisions there may not be εN such unique entries.



Finding collision for a salt is not independent of finding collision for other salts.

Talk Outline

- Basic definitions
- Our work and comparison with prior work
- Why prior techniques cannot extend to *short* collisions
- **Our technique for**
    - **Bound on 2-block collisions**
    - Bound on zero-walk adversaries
- Conclusion

Chernoff for Dependent Indicators

Traditional (one-sided) Chernoff Bound:

Let $\mathbf{X}_1, \ldots, \mathbf{X}_N$ be i.i.d. 0/1 random variables and let $\mathbf{X} = \sum_{i \in [N]} \mathbf{X}_i$.

Assume $\Pr[\mathbf{X}_i = 1] = \delta$. Then

$$\Pr[\mathbf{X} \geq 6\delta N] \leq 2^{-\delta N}.$$

Chernoff for Dependent Indicators

Traditional (one-sided) Chernoff Bound:

Let $\mathbf{X}_1, \ldots, \mathbf{X}_N$ be i.i.d. 0/1 random variables and let $\mathbf{X} = \sum_{i \in [N]} \mathbf{X}_i$.

Assume $\Pr[\mathbf{X}_i = 1] = \delta$. Then

$$\Pr[\mathbf{X} \geq 6\delta N] \leq 2^{-\delta N}.$$

Limited-dependence, "bounded large moments" Chernoff:

Let $\mathbf{X}_1, \ldots, \mathbf{X}_N$ be any 0/1 random variables and let $\mathbf{X} = \sum_{i \in [N]} \mathbf{X}_i$.

Fix u, δ and assume for all u-sized subsets $U \subseteq [N]$ that $p_U = \Pr[\Pi_{i \in U} \mathbf{X}_i = 1] \leq \delta^u$. Then

$$\Pr[\mathbf{X} \geq 6\delta N] \leq 2^{-u}.$$

[Impagliazzo-Kabanets'10]

- Allows $\mathbf{X}_i$ to be correlated. Only requires bound on large moments of sum.

Chernoff with Even More Dependent Indicators

Limited-dependence, "bounded large moments" Chernoff:

Let $\mathbf{X}_1, \ldots, \mathbf{X}_N$ be any 0/1 random variables and let $\mathbf{X} = \sum_{i \in [N]} \mathbf{X}_i$.

Fix u, δ and assume for all u-sized subsets $U \subseteq [N]$ that $p_U = \Pr[\Pi_{i \in U} \mathbf{X}_i = 1] \leq \delta^u$.
Then

$$\Pr[\mathbf{X} \geq 6\delta N] \leq 2^{-u}.$$

[Impagliazzo-Kabanets'10]

Chernoff with Even More Dependent Indicators

Limited-dependence, “bounded large moments” Chernoff:

Let $\mathbf{X}_1, \ldots, \mathbf{X}_N$ be any 0/1 random variables and let $\mathbf{X} = \sum_{i \in [N]} \mathbf{X}_i$.

Fix u, δ and assume for all u-sized subsets $U \subseteq [N]$ that $p_U = \Pr[\Pi_{i \in U} \mathbf{X}_i = 1] \leq \delta^u$.
Then

$$\Pr[\mathbf{X} \geq 6\delta N] \leq 2^{-u}.$$

[Impagliazzo-Kabanets’10]

- In our application, some p_U may be large, so does not apply. Instead we use an easy-to-prove modification:

Chernoff with Even More Dependent Indicators

Limited-dependence, "bounded large moments" Chernoff:

Let $\mathbf{X}_1, \ldots, \mathbf{X}_N$ be any 0/1 random variables and let $\mathbf{X} = \sum_{i \in [N]} \mathbf{X}_i$.

Fix u, δ and assume for all u-sized subsets $U \subseteq [N]$ that $p_U = \Pr[\Pi_{i \in U} \mathbf{X}_i = 1] \leq \delta^u$. Then

$$\Pr[\mathbf{X} \geq 6\delta N] \leq 2^{-u}.$$

[Impagliazzo-Kabanets'10]

Our limited-dependence, "bounded *average* large moments" Chernoff:

Let $\mathbf{X}_1, \ldots, \mathbf{X}_N$ be any 0/1 random variables and let $\mathbf{X} = \sum_{i \in [N]} \mathbf{X}_i$.

Fix u, δ. Assume that $p_U = \Pr[\Pi_{i \in U} X_i = 1]$ is at most δ^u when averaged over $U \subseteq [N]$. Then

$$\Pr[\mathbf{X} \geq 6\delta N] \leq 2^{-u}.$$

Impagliazzo's Method

[Impagliazzo,11]

Step 1: Analyze adversary w/o advice on any fixed set U of salts:

$$\Pr_{\mathbf{h}}[\text{Adversary succeeds on all salts in } U] \leq \delta^u$$

Impagliazzo's Method

[Impagliazzo,11]

Step 1: Analyze adversary w/o advice on any fixed set U of salts:

$$\Pr_{\mathbf{h}}[\text{Adversary succeeds on all salts in } U] \leq \delta^u$$

Step 2: Apply dependent Chernoff ($\mathbf{X}_i$ indicates success on i-th salt):

$$\Pr_{\mathbf{h}}[\text{Adversary succeeds on any } 6\delta N \text{ salts}] \leq 2^{-u}$$

Impagliazzo's Method

[Impagliazzo,11]

Step 1: Analyze adversary w/o advice on any fixed set U of salts:

$$\Pr_{\mathbf{h}}[\text{Adversary succeeds on all salts in } U] \leq \delta^u$$

Step 2: Apply dependent Chernoff ($\mathbf{X}_i$ indicates success on i-th salt):

$$\Pr_{\mathbf{h}}[\text{Adversary succeeds on any } 6\delta N \text{ salts}] \leq 2^{-u}$$

Step 3: Apply union bound over all 2^S possible advice strings:

$$\Pr_{\mathbf{h}}[\exists\text{advice: Adversary succeeds on any } 6\delta N \text{ salts}] \leq 2^S \cdot 2^{-u}$$

Impagliazzo's Method

[Impagliazzo,11]

Step 1: Analyze adversary w/o advice on any fixed set U of salts:

$$\Pr_{\mathbf{h}}[\text{Adversary succeeds on all salts in } U] \leq \delta^u$$

Step 2: Apply dependent Chernoff ($\mathbf{X}_i$ indicates success on i-th salt):

$$\Pr_{\mathbf{h}}[\text{Adversary succeeds on any } 6\delta N \text{ salts}] \leq 2^{-u}$$

Step 3: Apply union bound over all 2^S possible advice strings:

$$\Pr_{\mathbf{h}}[\exists\text{advice: Adversary succeeds on any } 6\delta N \text{ salts}] \leq 2^S \cdot 2^{-u}$$

Conclude bound $6\delta + 2^S \cdot 2^{-u}$ on adversaries with advice.

Concretely: $u = \Omega(S + \log N)$, $\delta =$ desired bound (e.g. $O(ST/N)$).

Impagliazzo's Method, Modified

Step 1: Analyze adversary w/o advice on a random set U of salts:

$$\Pr_{\mathbf{h},\mathbf{U}}[\text{Adversary succeeds on all salts in } \mathbf{U}] \leq \delta^u$$

Step 2: Apply dependent Chernoff ($\mathbf{X}_i$ indicates success on i-th salt):

$$\Pr_{\mathbf{h}}[\text{Adversary succeeds on any } 6\delta N \text{ salts}] \leq 2^{-u}$$

Step 3: Apply union bound over all 2^S possible advice strings:

$$\Pr_{\mathbf{h}}[\exists\text{advice: Adversary succeeds on any } 6\delta N \text{ salts}] \leq 2^S \cdot 2^{-u}$$

Conclude bound $6\delta + 2^S \cdot 2^{-u}$ on adversaries with advice.

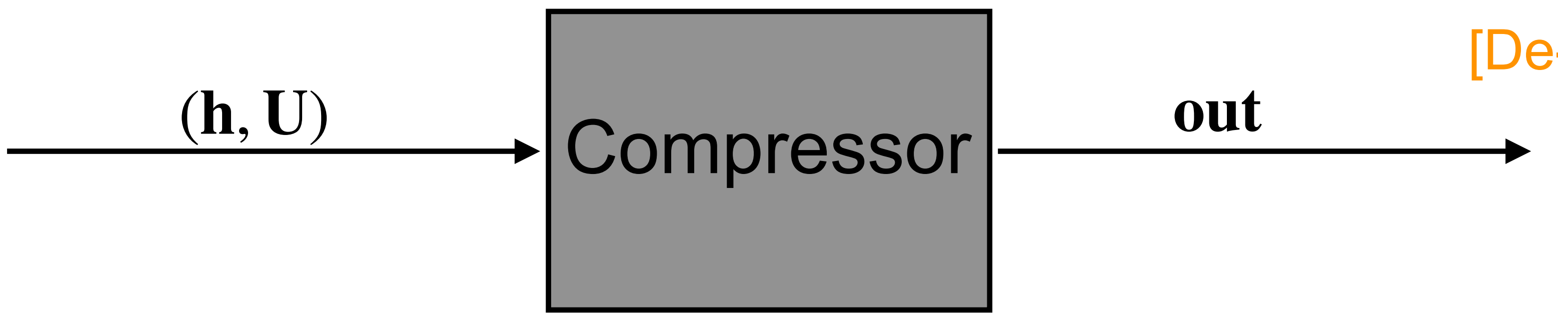
Concretely: $u = \Omega(S + \log N)$, $\delta =$ desired bound (e.g. $O(ST/N)$).

Step 1 via Compression

- Step 1: Analyze adversary w/o advice on a random set U of salts:

$$\Pr_{\mathbf{h},\mathbf{U}}[\text{Adversary succeeds on all salts in } \mathbf{U}] \leq \delta^u$$

[De-Trevisan-Tulsiani,10]



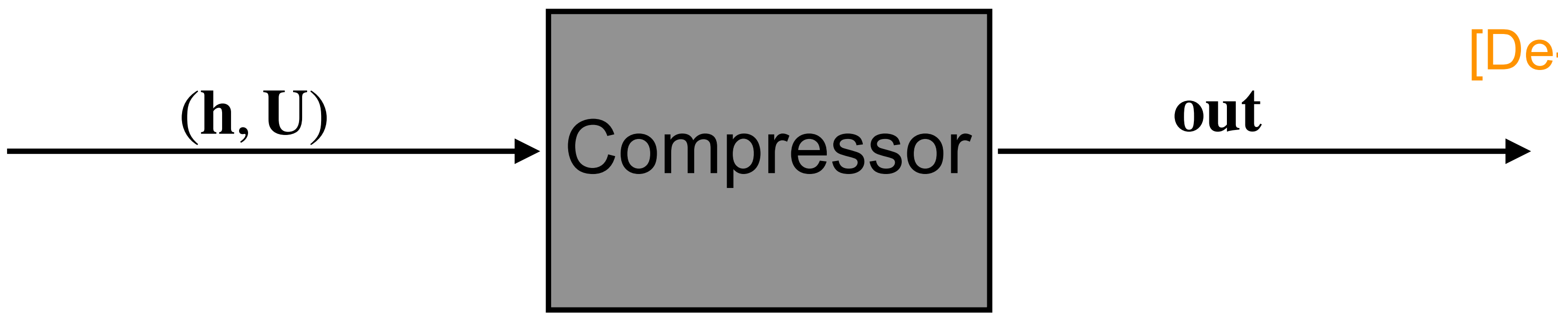
- Shannon bound: $\mathbb{E}[\,|\mathbf{out}|\,] \geq \text{entropy}(\mathbf{h},\mathbf{U})$

Step 1 via Compression

- Step 1: Analyze adversary w/o advice on a random set U of salts:

$$\Pr_{\mathbf{h},\mathbf{U}}[\text{Adversary succeeds on all salts in } \mathbf{U}] \leq \delta^u$$

[De-Trevisan-Tulsiani,10]

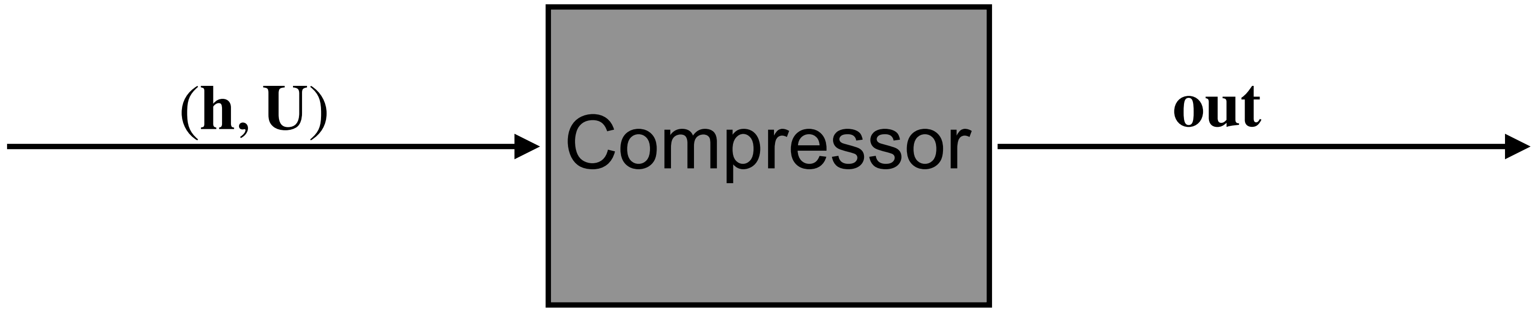


- Shannon bound: $\mathbb{E}[\,|\mathbf{out}|\,] \geq$ entropy$(\mathbf{h}, \mathbf{U})$
- Plan:
  1. Say some adversary $\mathscr{A}$ succeeds on $(\mathbf{h}, \mathbf{U})$ with large probability, say ε.
  2. Fix some (h, U) on which $\mathscr{A}$ wins.
  3. We give a compressor that uses $\mathscr{A}$ to save $\log(1/\delta)$ bits for each salt in U.
  4. This contradicts the Shannon bound and gives $\varepsilon \leq \delta^u$.

Bound on 2-block Collisions

Analyze adversary w/o advice on a random set U of salts and prove:

$$\Pr_{\mathbf{h},\mathbf{U}}[\text{Adversary finds 2-block collisions on all salts in } \mathbf{U}] \leq (ST/N)^u$$

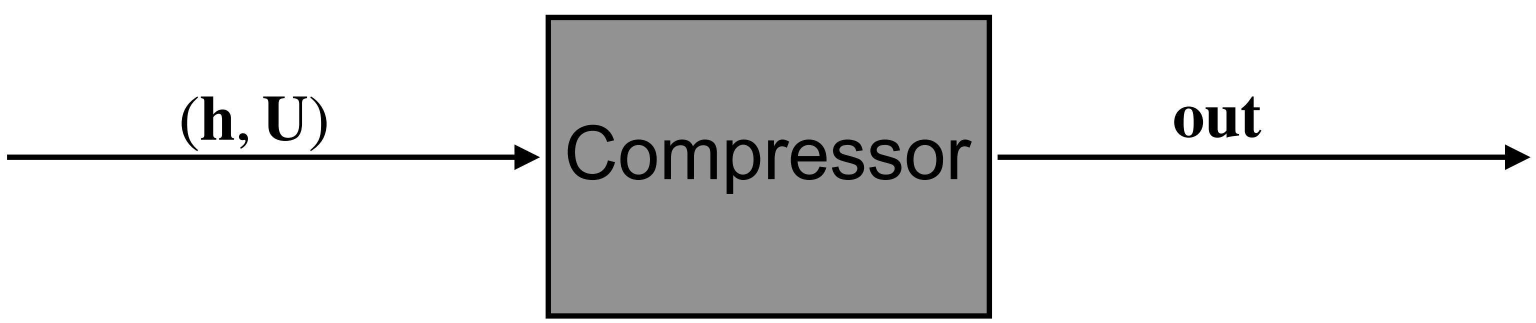


1. Fix (h, U) and consider an adversary that finds 2-block collisions on all salts in U.
2. Compress both h and U at a total of u spots. In each spot, compressor stores at most $O(\log S + \log T)$ bits to save $\log N$ bits.

Bound on 2-block Collisions

Analyze adversary w/o advice on a random set U of salts and prove:

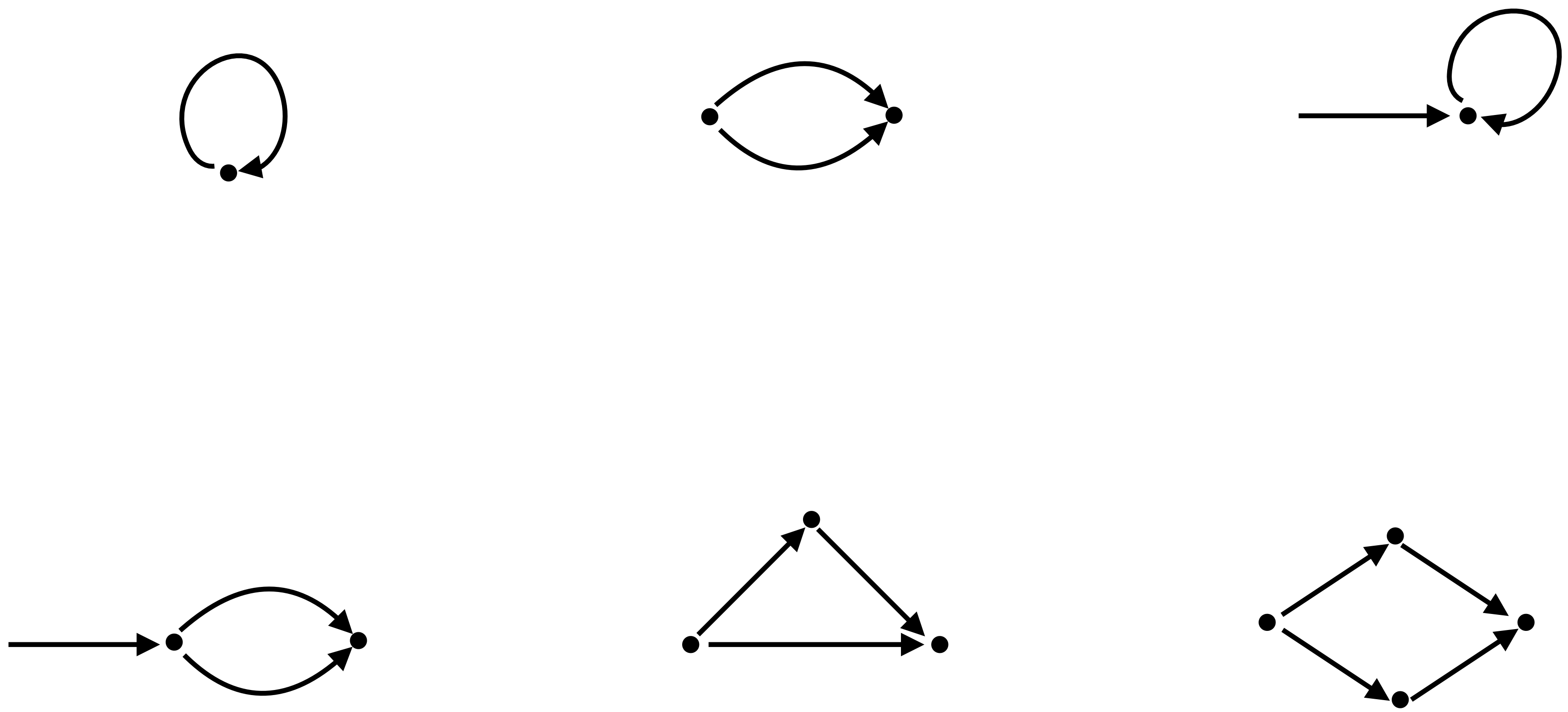
$$\Pr_{\mathbf{h},\mathbf{U}}[\text{Adversary finds 2-block collisions on all salts in } \mathbf{U}] \leq (ST/N)^u$$



1. Fix (h, U) and consider an adversary that finds 2-block collisions on all salts in U.
2. Compress both h and U at a total of u spots. In each spot, compressor stores at most $O(\log S + \log T)$ bits to save $\log N$ bits.

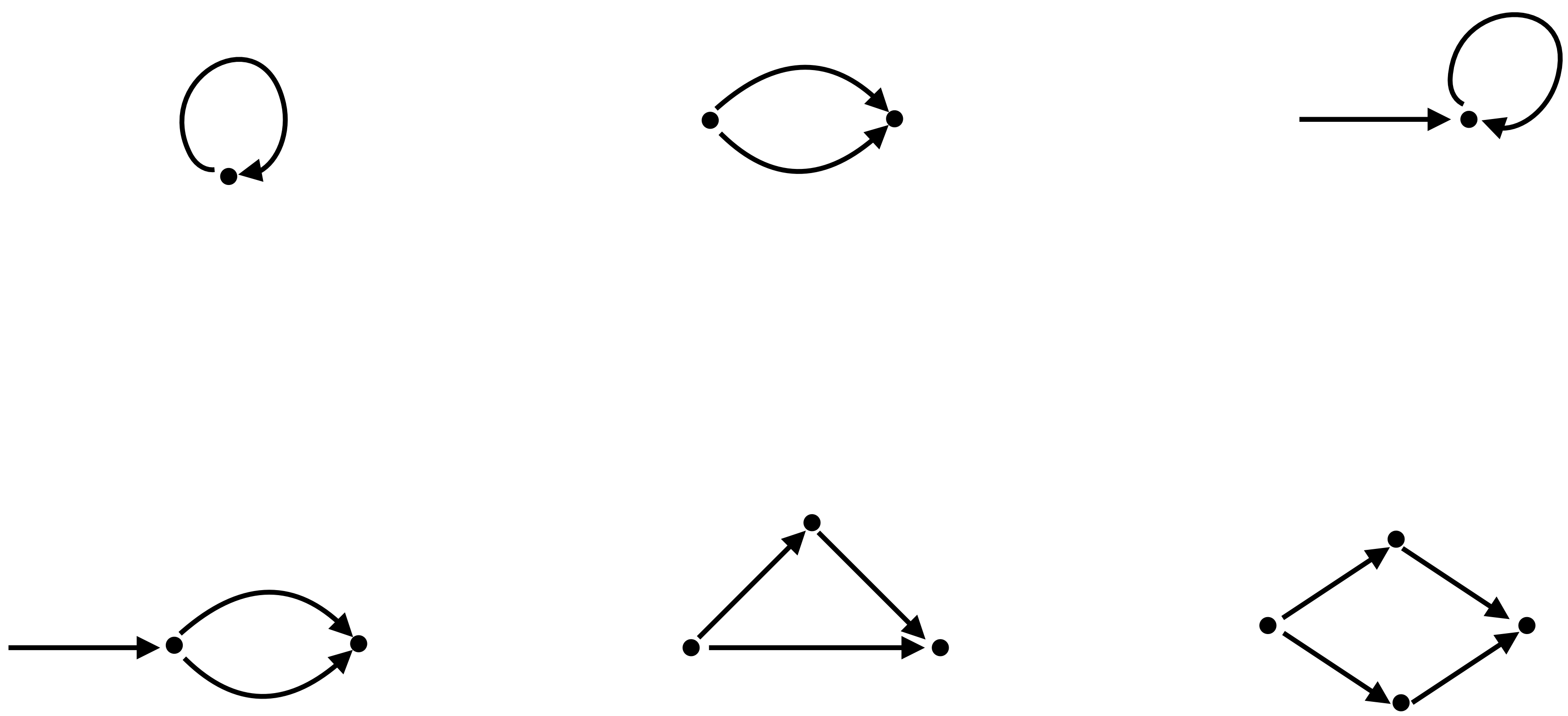
This compressor is complicated (see paper).

Types of 2-block Collisions



Compressor needs to handle each of these types differently.

Types of 2-block Collisions



Compressor needs to handle each of these types differently.

Types of B-block collisions increase exponentially with B. Thus arbitrary B is hard.

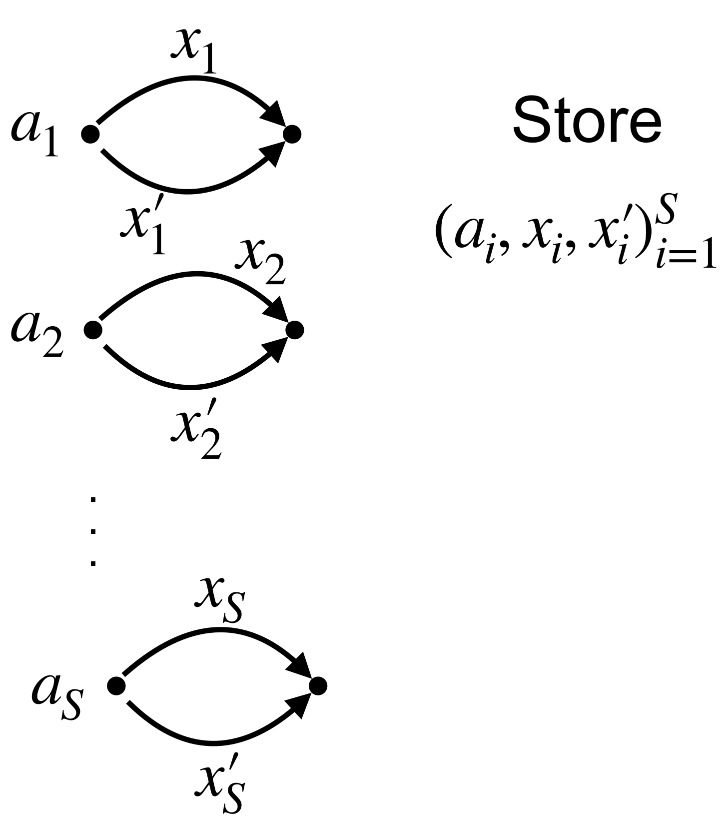
Talk Outline

- Basic definitions
- Our work and comparison with prior work
- Why prior techniques cannot extend to *short* collisions
- **Our technique for**
  - Bound on 2-block collisions
  - **Bound on zero-walk adversaries**
- Conclusion

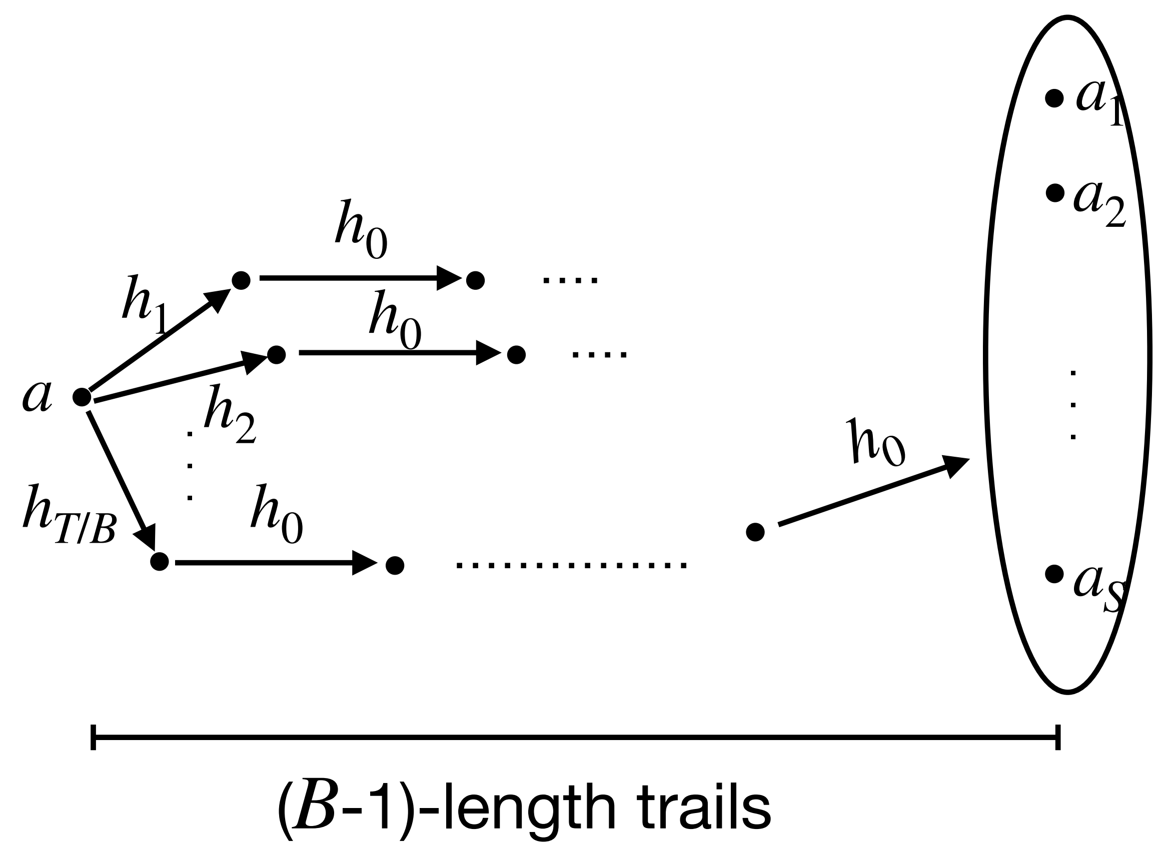
Definition of Zero-Walk Adversary

- We define a restricted class of pre-computing adversary, referred as Zero-Walk adversary.

Pre-computation

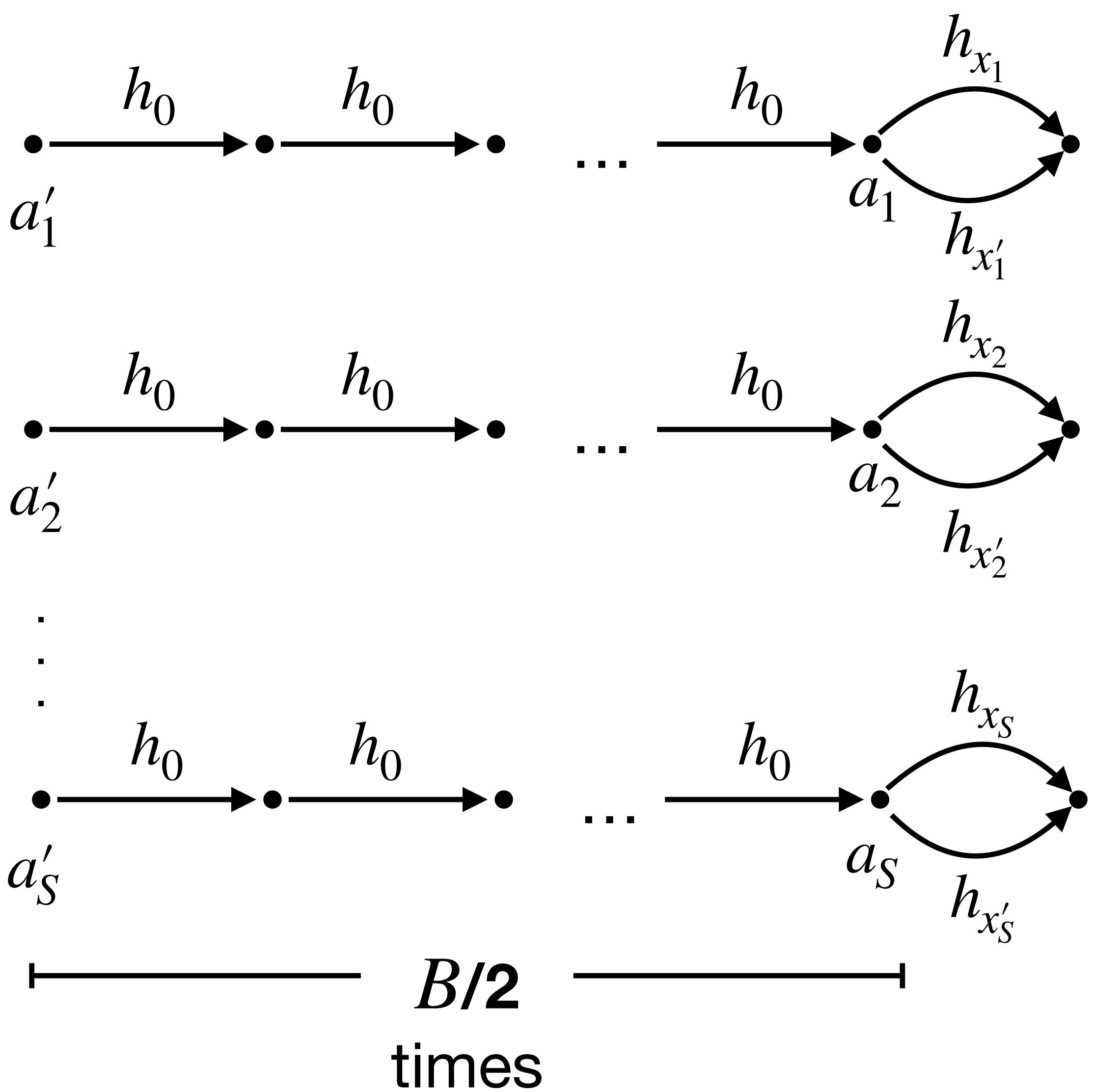


Online Phase



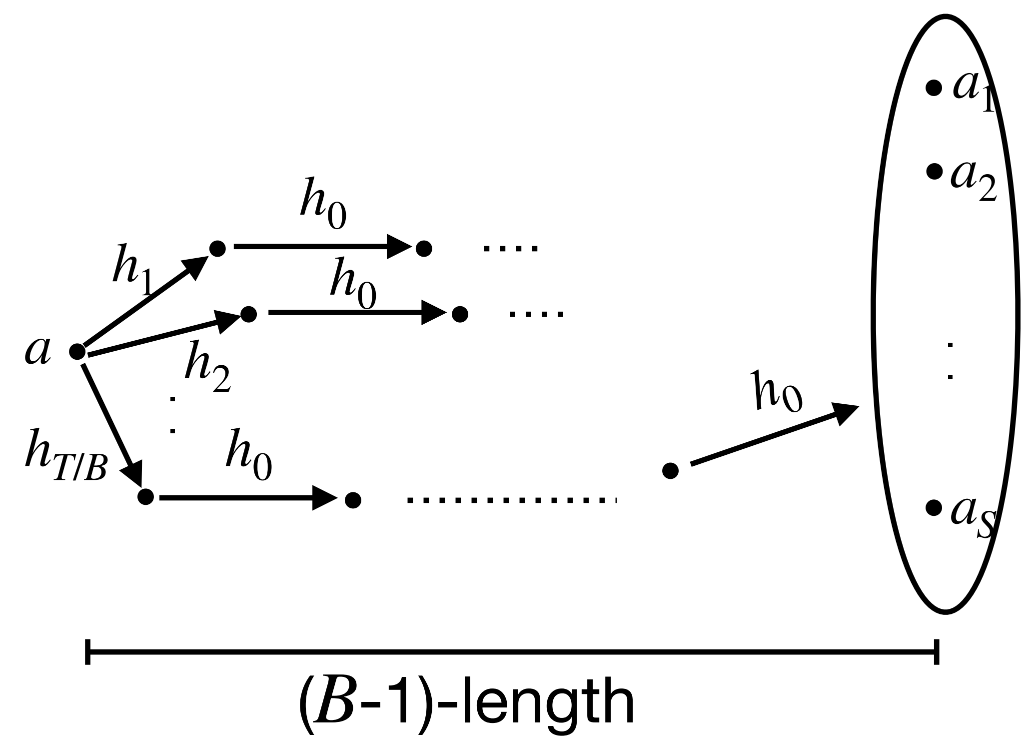
Best Known B-block Collision Finding Adversary

Pre-computation



Output all $(a_i, x_i, x'_i)_{i=1}^{S/3 \log N}$

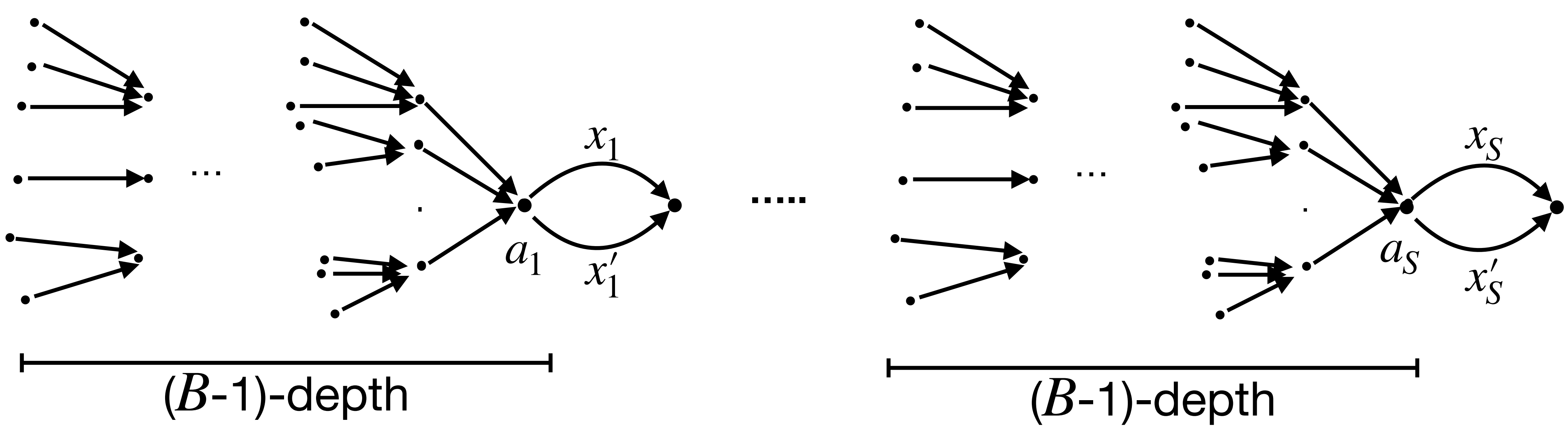
Online Phase



Achieves $\Omega(STB/N)$ advantage

Are There Better Zero-Walk Adversaries?

- Adversary could store collisions for salts with large B-depth trees leading to them
- Advantage would be $O(ST * (\text{tree-size})/BN)$



- We prove that the largest B-depth tree has size $\tilde{O}(B^2)$ with high probability, so previous strategy is optimal.

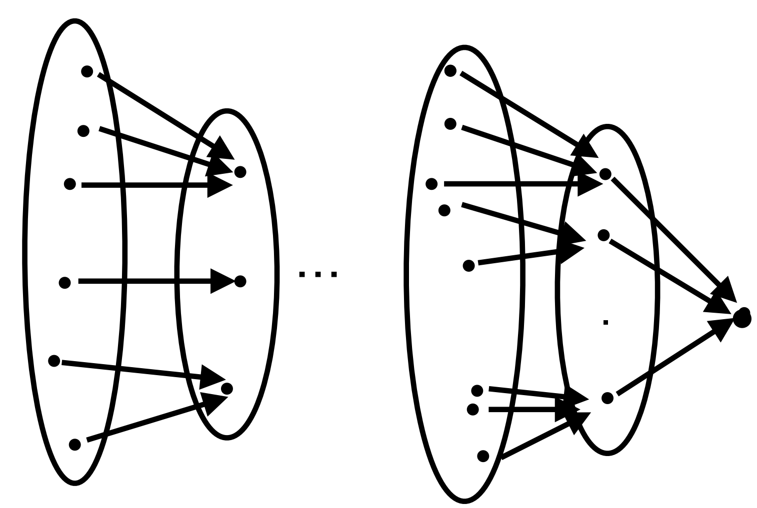
Size B-depth Trees in Random Functional Graphs

Bounded B-depth trees of Random Functional Graphs:

For a random function $\mathbf{f} : [N] \rightarrow [N]$ functions, the probability there exists a B-depth tree in the graph for $\mathbf{f}$ with $\tilde{\Omega}(B^2)$ nodes is at most $1/N$.

A naive approach would be using Chernoff and then applying union bound over B depths but that gives a loose bound of $\tilde{O}(B^3)$.

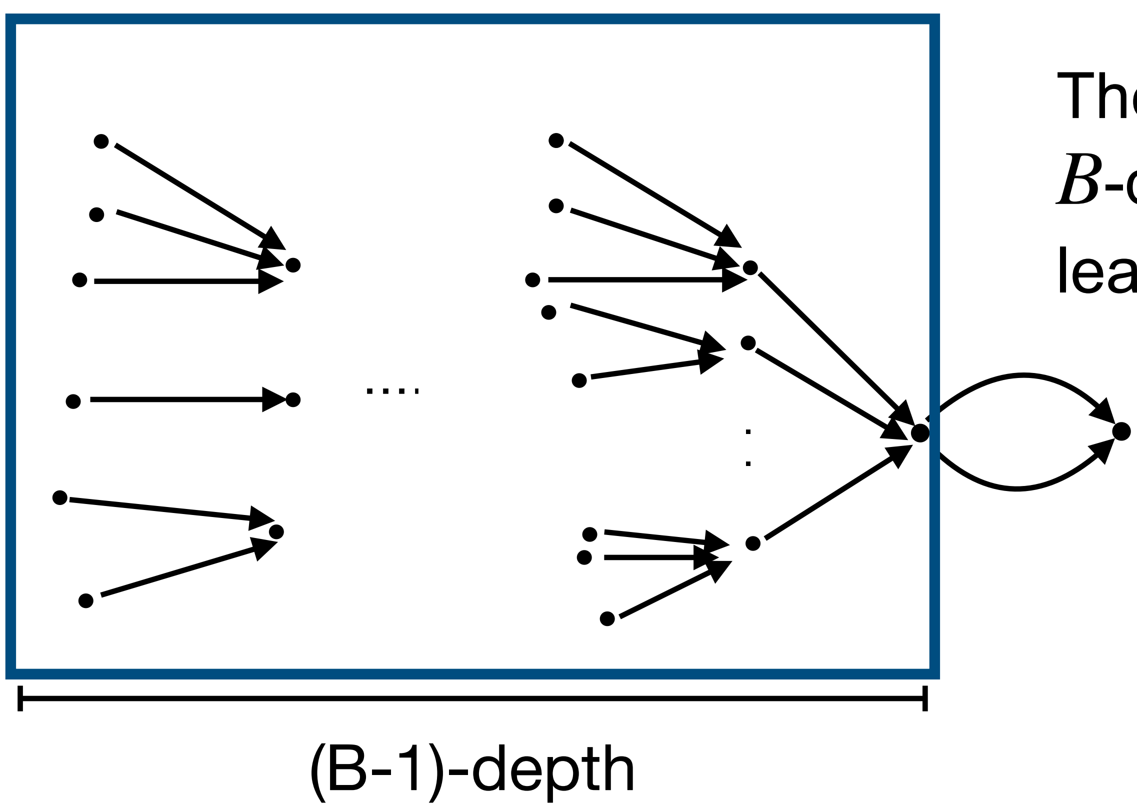
We obtain a tighter bound in the paper.



Bound on Zero Walk Adversary

Bounded B-depth trees of Random Functional Graphs:

For a random function $\mathbf{f}: [N] \to [N]$ functions, the probability there exists a B-depth tree in the graph for $\mathbf{f}$ with $\tilde{\Omega}(B^2)$ nodes is at most $1/N$.



The theorem implies the size of the largest B-depth tree is $\tilde{O}(B^2)$ with probability at least $(1 - 1/N)$.

Conclusions

- We present new techniques that gives us the following results:

Result 1: For any 2-block collision finding adversary, its advantage is $\tilde{\theta}(ST/N)$.

Result 2: For arbitrary B-block collision finding "zero walk" adversary, its advantage is $\tilde{\theta}(STB/N)$.

- **Open problem**: prove the conjectured $\tilde{O}(STB/N)$ bound on arbitrary B-block collision finding adversary's advantage, not just zero-walking adversary.

Thank you.

https://eprint.iacr.org/2020/770.pdf