Time-Space Tradeoffs for Sponge Hashing: Attacks and Limitations for Short Collisions

Cody Freitag
Cornell Tech

Ashrujit Ghoshal
University of Washington

Ilan Komargodski
Hebrew University and NTT Research

CRYPTO 2022

Iterative hashing

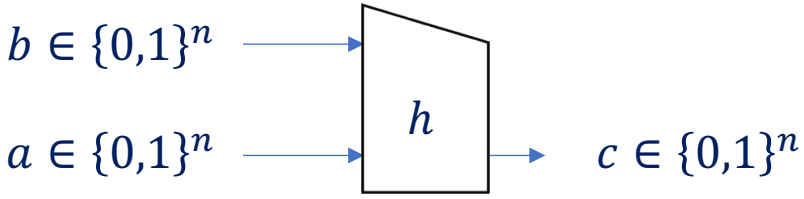
Hash functions need to handle variable input lengths

- password hashing
- hash and sign
- commitments

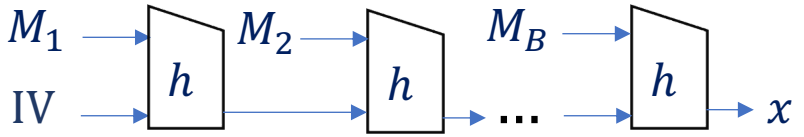
Cannot design a different hash for every length

Construct a VIL hash function from an underlying FIL primitive

e.g., Merkle Damgård hashing [Mer89, Dam89]



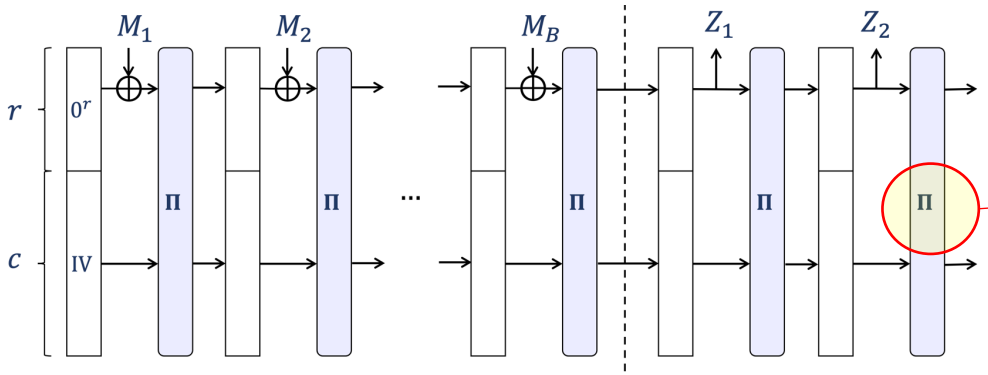
$M = (M_1, M_2, \ldots, M_B)$
$\mathrm{MD}_h(\mathrm{IV}, M) = x$



Used in MD5, SHA-1, SHA-2

SHA-3

- 2006 NIST competition after attacks on MD-5, SHA-0
- Winner: Keccak [BDPV07] became SHA-3
- New iterative hashing technique: sponge construction



permutation, instead of function

The sponge construction

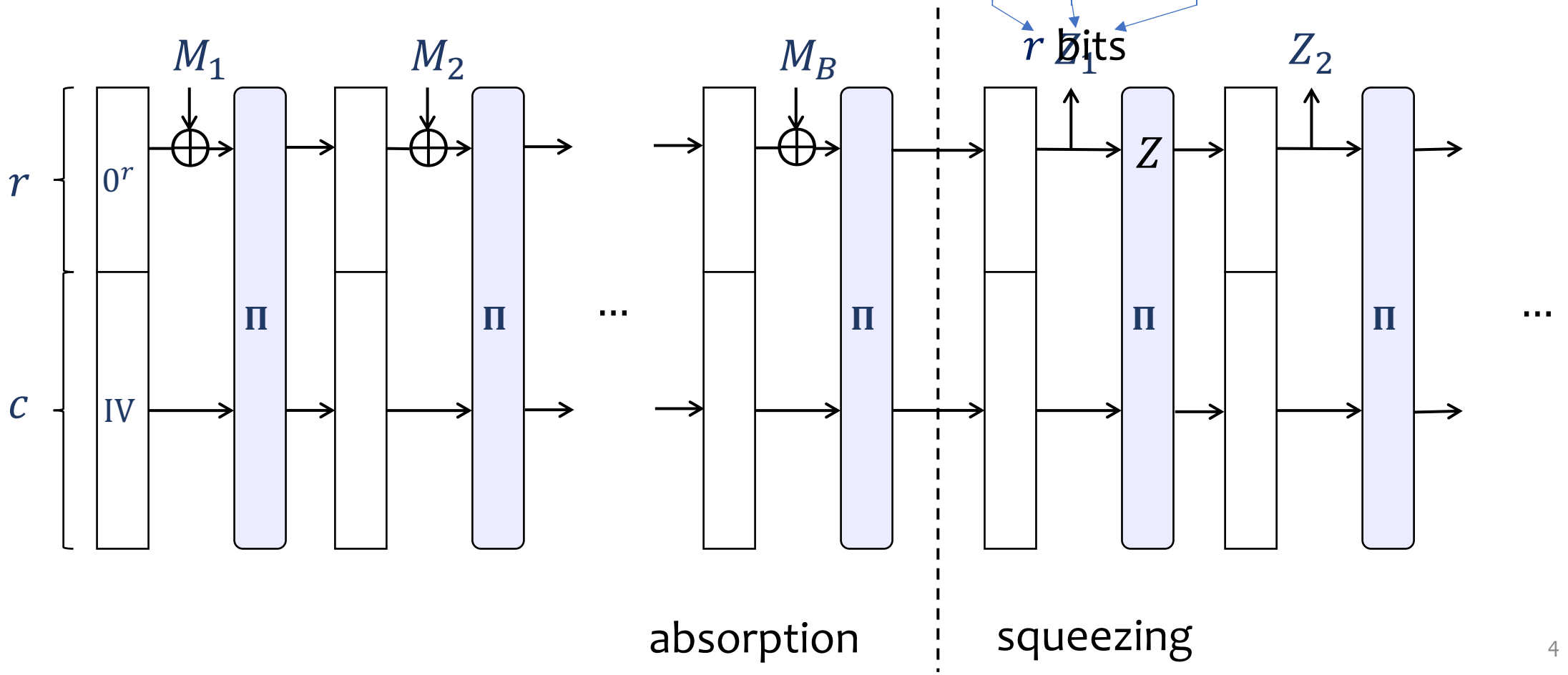
Sponge construction

Permutation $\Pi: \{0,1\}^{r+c} \rightarrow \{0,1\}^{r+c}$

r = bit-rate, c = capacity

$\boldsymbol{Sp_{\Pi}(\mathrm{IV}, M)} = (Z_1, Z_2, \ldots)$ This talk

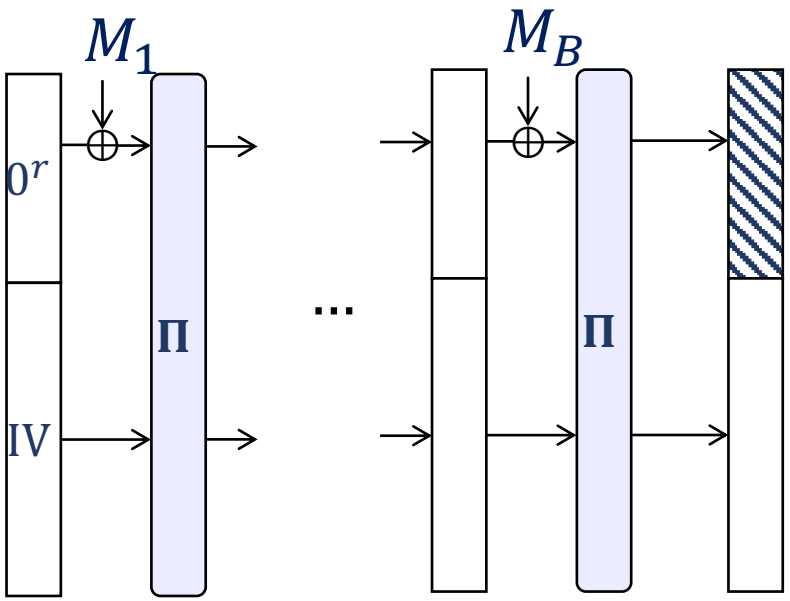
$M = (M_1, M_2 \ldots, M_B)$



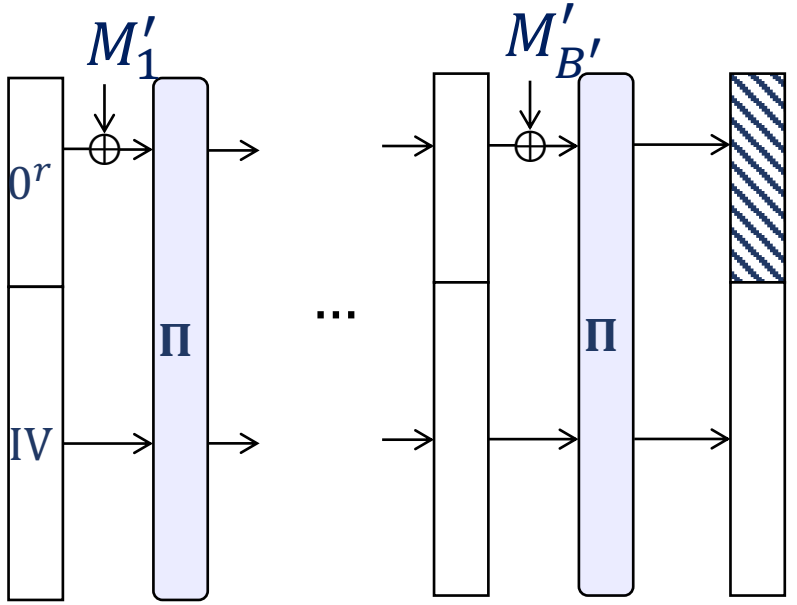
Collision resistance

Given random IV, hard to find $M \neq M'$ such that $Sp_{\Pi}(\mathrm{IV}, M) = Sp_{\Pi}(\mathrm{IV}, M')$

$M = (M_1, \ldots, M_B)$

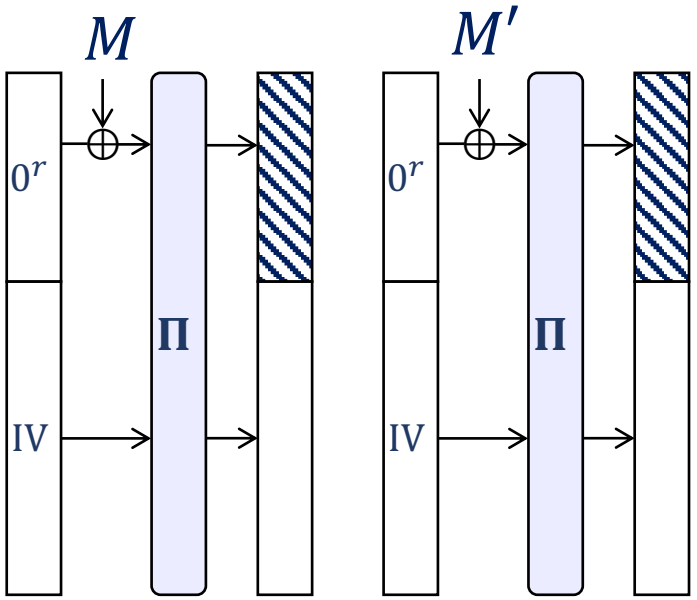


$M' = (M'_1, \ldots, M'_{B'})$



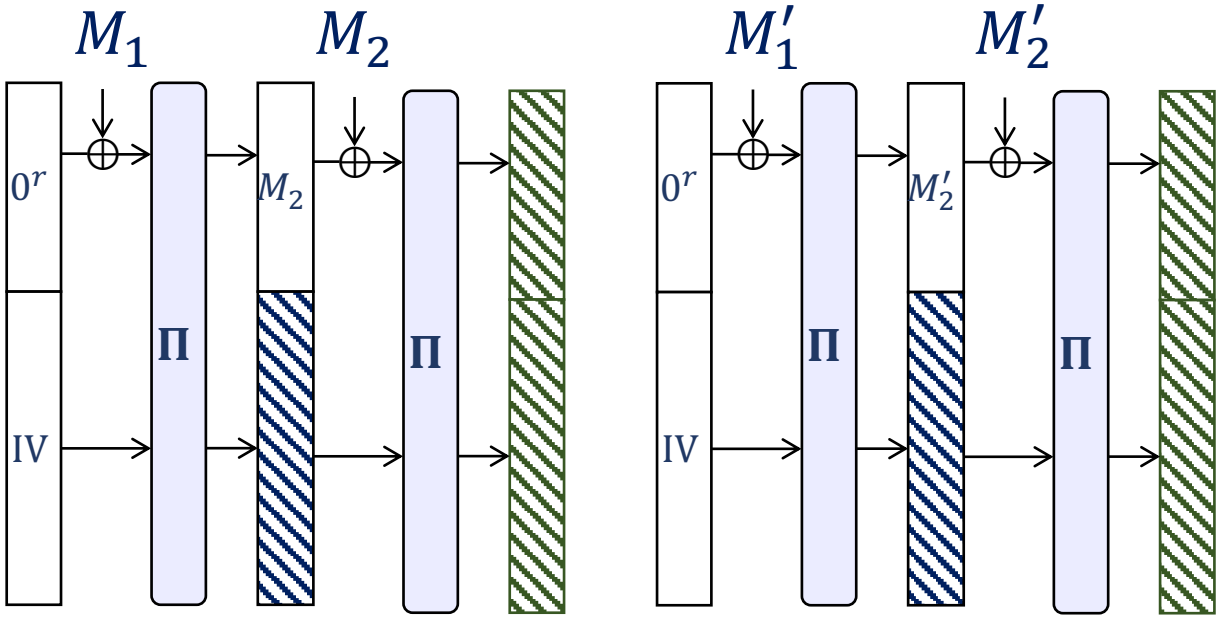
Complexity of finding collisions

- Model Π as a random permutation
- Using $T \approx \min(2^{\frac{r}{2}}, 2^{\frac{c}{2}})$ queries, can find collisions



$2^{\frac{r}{2}}$ queries

Collision: (M, M')



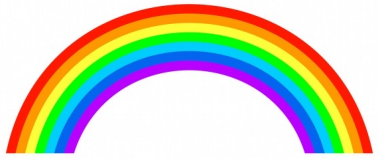
$2^{\frac{c}{2}}$ queries

Collision: $((M_1, M_2), (M_1', M_2'))$

Complexity of finding collisions

- Model Π as a random permutation
- Using $T \approx \min(2^{\frac{r}{2}}, 2^{\frac{c}{2}})$ queries, can find collisions
  - Provably optimal!
    - indifferentiability from a RO with r bit output for $\leq 2^{\frac{c}{2}}$ queries [BDPA08]

- What about adversaries that use large pre-processing?
  - Indifferentiabilty framework does not apply
  - Scenario studied by [Hellman80, Fiat-Naor99, Unruh07, …]
  - Captures non-uniform attacks



Auxiliary-input random permutation model (AI-RPM) [CDG18]

$= (A_1, A_2)$

$\Pi: \{0,1\}^{r+c} \to \{0,1\}^{r+c}$

A

$\mathrm{IV} \xleftarrow{\$} \{0,1\}^c$

T queries

A_1

Π

S bits

A_2

Π

Π^{-1}

(M, M')

"pre-processing" phase

"online" phase

A wins if $M' \neq M, Sp_\Pi(\mathrm{IV}, M) = Sp_\Pi(\mathrm{IV}, M')$

$$\mathrm{Adv}_{c,r}(S,T) = \max_{(S,T)\ \mathrm{adv}\ A} \Pr[A \text{ wins}]$$

Prior work

Theorem. [CDG18] $\mathrm{Adv}_{c,r}(S,T) = \Theta\left(\frac{ST^2}{2^c} + \frac{T^2}{2^r}\right)$

An observation: the attack finds collisions of length $\Omega(T)$!

Say, $T \approx 2^{60} \Rightarrow$ petabytes sized collision!

Shorter collisions seem harder to find

Can we characterize hardness of finding B-block collisions for sponge?

Question recently studied recently for MD.

Takeaway: easier as B grows. See next talk for details

This work:

Attacks and limitations for B-block sponge collisions

Our results, in a nutshell

- **New attacks**
  - for $B = 1$
  - for $B \geq 2$
- **New limitations** on attacks
  - for $B = 1$
  - for $B = 2$

Bounds for attacks and limitations do not match. Many open problems!

Π^{-1} queries lead to new $B = 1$ attack, make harder to prove limitations!

Our results: new attacks

1. New attack for $B = 1$

1-block collision

$$\mathrm{Adv}_{c,r,1}(S,T) \geq \Omega\left(\min\left\{\left(\frac{S^2T}{2^{2c}}\right)^{\frac{2}{3}}, \left(\frac{ST}{2^c}\right)^2\right\}\right)$$

The "trivial" attack. For MD, provably optimal for $B = 1$

Prev best known attack for $B = 1$ has advantage $\Omega\left(\frac{S}{2^c} + \frac{T^2}{2^r}\right)$

New attack better for some regimes e.g., $S = 2^{\frac{4c}{5}}, T = 2^{\frac{c}{5}}$ for $c = r$

$$\Omega\left(\frac{S}{2^c} + \frac{T^2}{2^r}\right) = \Omega\left(2^{-\frac{c}{5}}\right), \qquad \Omega\left(\frac{S^2T}{2^{2c}}\right)^{\frac{2}{3}} = \Omega(1)$$

Our results: new attacks (2)

2. New attack for $B \geq 2$

$$\mathrm{Adv}_{c,r,B}(S,T) \geq \Omega\left(\frac{STB}{2^c} + \frac{T^2}{2^c} + \frac{T^2}{2^r}\right)$$

Analogue of MD attack for $B \geq 2$

Our results: limitations

1. Limitation for $B = 1$

$$\mathrm{Adv}_{c,r,1}(S,T) \leq O\left(\frac{ST}{2^c} + \frac{T^2}{2^r}\right)$$

Proof using bit-fixing [Unruh07, CDGS18, CDG18]

not believed to be tight

intermediate model where

- adversary does not have pre-processing
- instead, can fix $\approx ST$ points of Π

Our results: limitations

2. Limitation for $B = 2$

$$\mathrm{Adv}_{c,r,2}(S,T) \le O\left(\frac{ST}{2^c} + \frac{T^2}{2^c} + \frac{T^2}{2^r} + \frac{S^2T^4}{2^{2c}}\right)$$

Proof via multi-instance framework [IK10, CGLQ20, ACDW20] + compression argument

not believed to be tight

Reduction to a game where
- adversary does not have pre-processing
- has to find collisions wrt S random IVs

Our $B = 2$ attack has advantage $\Omega\left(\frac{ST}{2^c} + \frac{T^2}{2^r} + \frac{T^2}{2^c}\right)$, optimal when $ST^3 < 2^c$

Our results: the sponge state of affairs

	Best attack*		Advantage upper bound
	Next		
$B = 1$	$\Omega\left(\min\left\{\left(\frac{S^2T}{2^{2c}}\right)^{\frac{2}{3}}, \left(\frac{ST}{2^c}\right)^2\right\} + \frac{T^2}{2^r}\right)$	gap	$O\left(\frac{ST}{2^c} + \frac{T^2}{2^r}\right)$
$B = 2$	$\Omega\left(\frac{ST}{2^c} + \frac{T^2}{2^r} + \frac{T^2}{2^c}\right)$	gap	$O\left(\frac{ST}{2^c} + \frac{T^2}{2^r} + \frac{S^2T^4}{2^{2c}}\right)$
$B > 2$	$\Omega\left(\frac{STB}{2^c} + \frac{T^2}{2^c} + \frac{T^2}{2^r}\right)$	gap	$O\left(\frac{ST^2}{2^c} + \frac{T^2}{2^r}\right)$ [CDG18]

See paper

*Hiding factors poly in c, r

Attack for $B = 1$

Theorem. [this work]

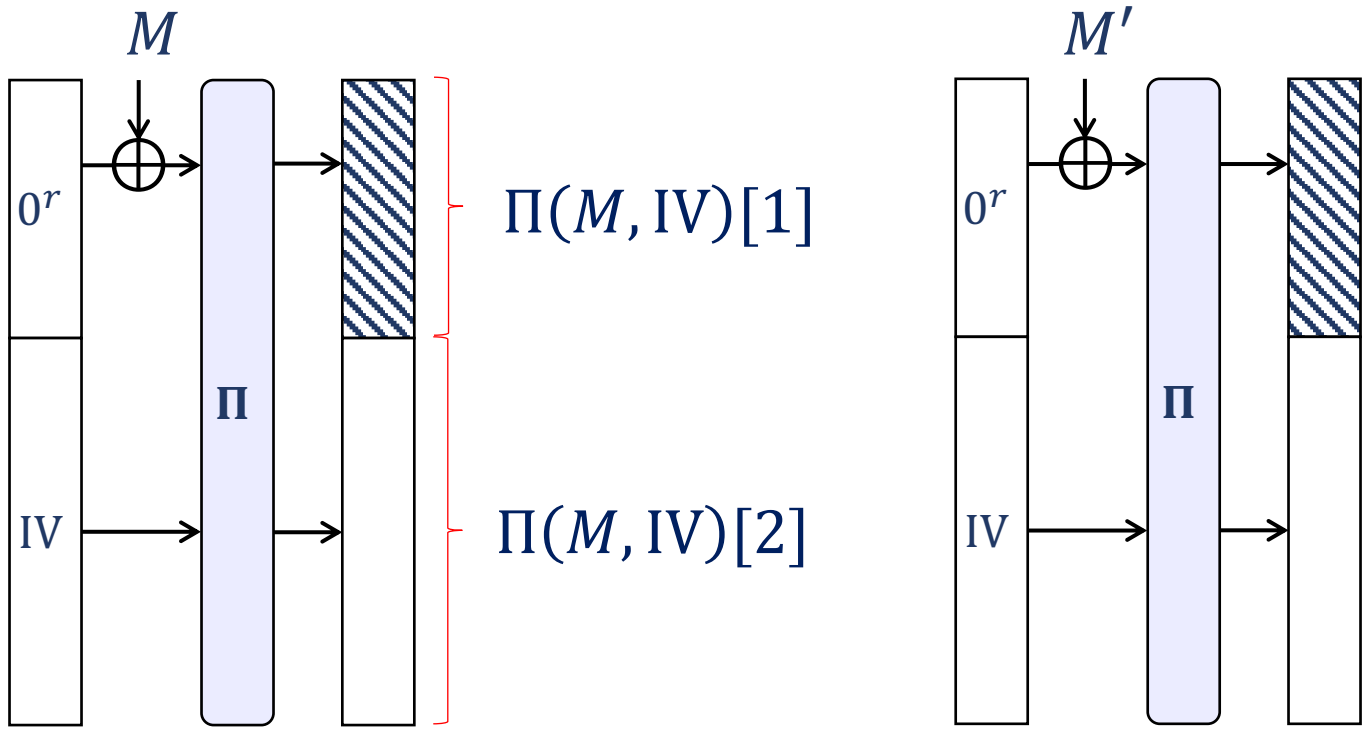
$$\mathrm{Adv}_{c,r,1}(S,T) \geq \Omega\left(\min\left\{\left(\frac{S^2T}{2^{2c}}\right)^{\frac{2}{3}}, \left(\frac{ST}{2^c}\right)^2\right\}\right) = \Omega(\varepsilon_H^2)$$

$$\varepsilon_H := \min\left\{\left(\frac{S^2T}{2^{2c}}\right)^{\frac{1}{3}}, \frac{ST}{2^c}\right\}$$

advantage for Hellman's attack for random function inversion

Attack for $B = 1$

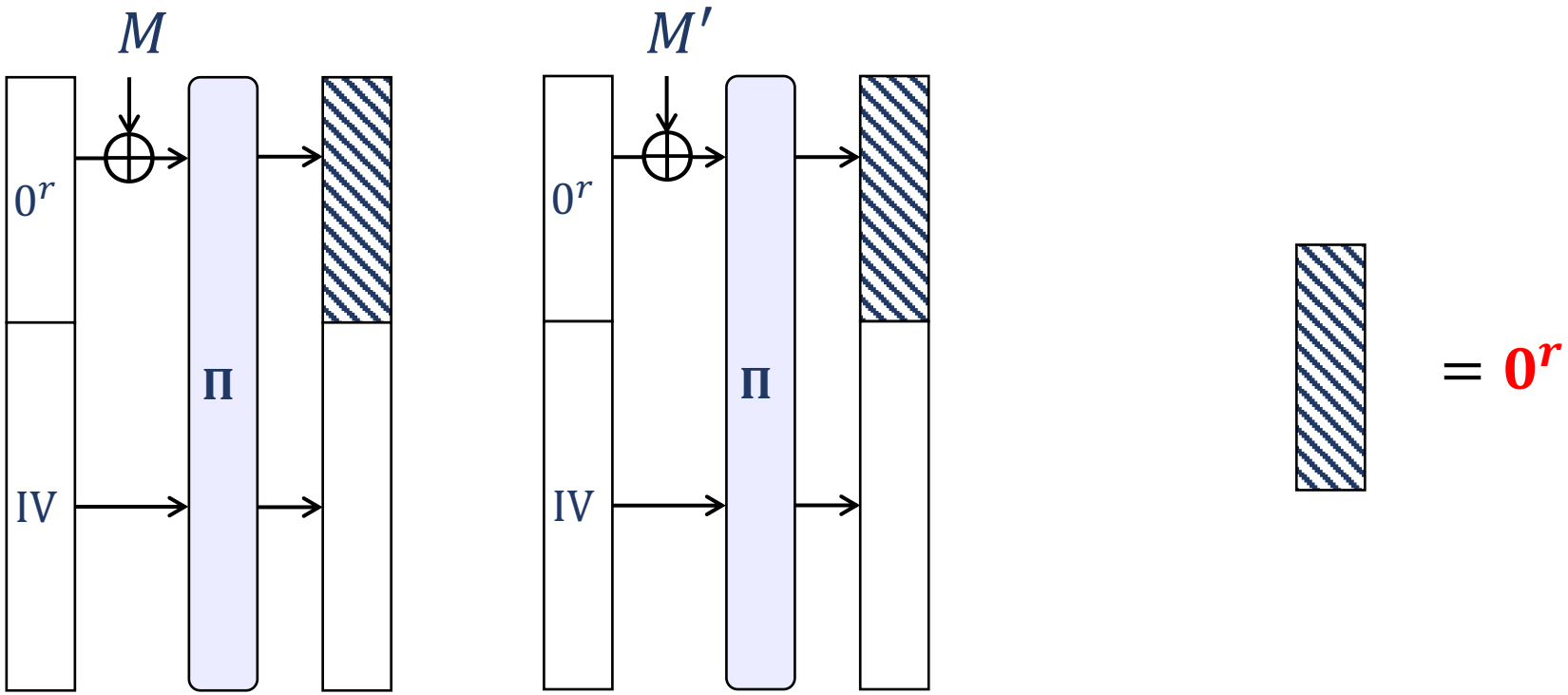
Goal: Find M, M' s.t. $\Pi(M, \mathrm{IV})[1] = \Pi(M', \mathrm{IV})[1]$



Attack for $B = 1$

Solve a harder problem!

Goal: Find M, M' s.t. $\Pi(M, \mathrm{IV})[1] = \Pi(M', \mathrm{IV})[1] = \mathbf{0^r}$



Attack for $B = 1$

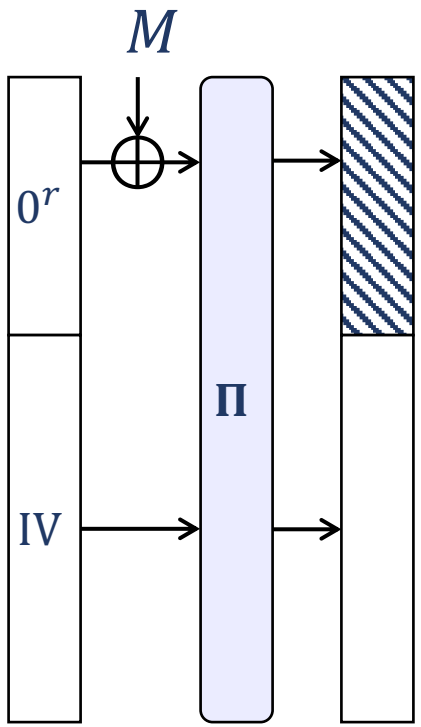
$= 0^r$

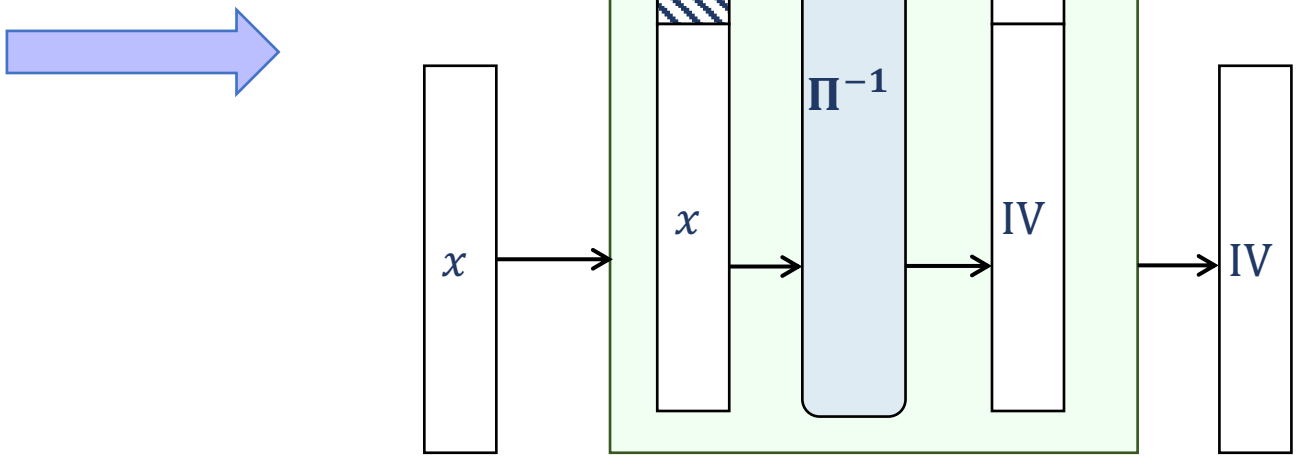
An alternate view

$f: \{0,1\}^c \to \{0,1\}^c$

$f(x) = \Pi^{-1}(0^r, x)[2]$

f does not depend on IV





Attack strategy

Strategy:
Given IV find $x \neq x' \in f^{-1}(\mathrm{IV})$

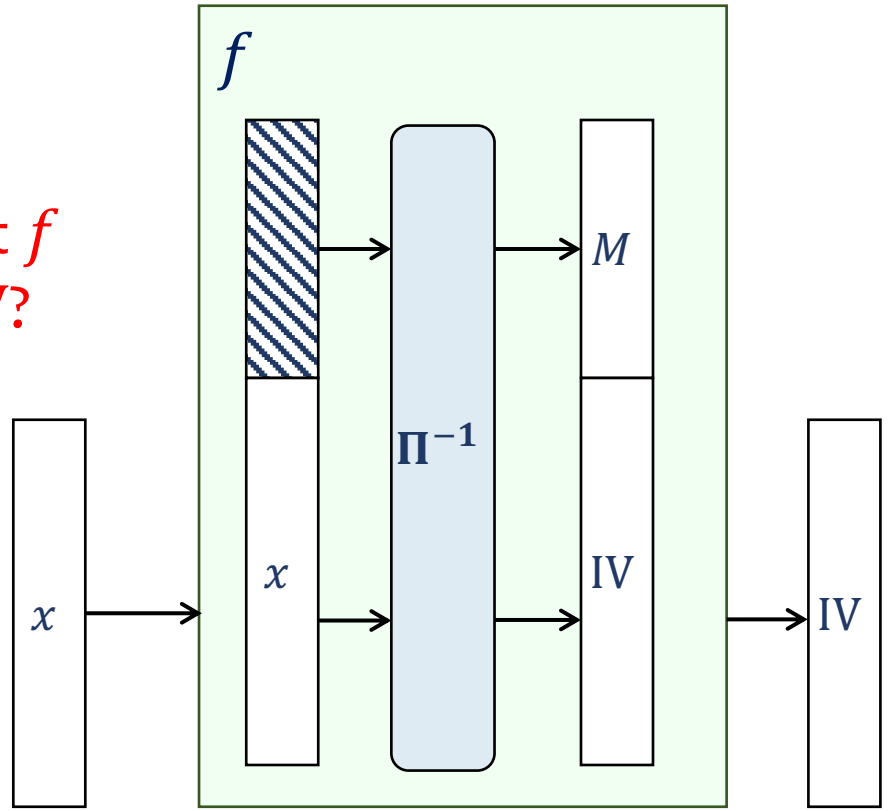
$M = \Pi^{-1}(0^r, x)[1]$
$M' = \Pi^{-1}(0^r, x')[1]$

How to invert f on random IV?

Output M, M'

Observe:
$f(x) = \mathrm{IV} \Rightarrow \Pi^{-1}(0^r, x)[2] = \mathrm{IV}$
$M = \Pi^{-1}(0^r, x)[1]$
$\Rightarrow \Pi(M, \mathrm{IV})[1] = 0^r$
Similarly, $\Pi(M', \mathrm{IV})[1] = 0^r$

$f(x) = \Pi^{-1}(0^r, x)[2]$



Hellman's function inversion [Hellman80, FN99]

Random function $f: \{0,1\}^c \to \{0,1\}^c$

$= (A_1, A_2)$

A

$y \xleftarrow{\$} \text{Image}(f)$

T queries

A_1

f

S bits

A_2

f

x

A wins if $f(x) = y$

$$\varepsilon_H = \Omega\left(\min\left\{\left(\frac{S^2T}{2^{2c}}\right)^{\frac{1}{3}}, \frac{ST}{2^c}\right\}\right)$$

Technical challenges

$f(x) = \Pi^{-1}(0^r, x)[2]$

1. f is **not a random function**!
2. the challenge (random IV) **may not be in the image** of f!
3. need to find **2 distinct pre-images** for the challenge under f

Challenge 1: f is **not a random function**!

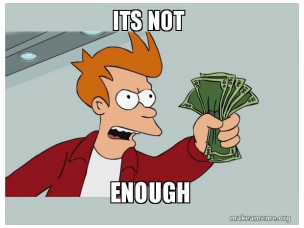
Running Fiat-Naor's extension for general functions too expensive!

Our solution

Tl;dr: $f \approx$ random function, can adapt Hellman's analysis!

Challenge 2: the challenge (random IV) **may not be in the image** of f!

Can show $\Omega(1)$ fraction of co-domain has ≥ 2 pre-images. Does it suffice?



No, Hellman's attack might fail for this $\Omega(1)$ fraction!

Our solution

We show for a fixed $y \in \{0,1\}^c$, the attack succeeds w.p.

$$\Omega\left(\min\left\{\left(\frac{S^2T|f^{-1}(y)|}{2^{2c}}\right)^{\frac{1}{3}}, \frac{ST|f^{-1}(y)|}{2^c}\right\}\right)$$

Challenge 3: Need to find **2 distinct pre-images** for the challenge under f

It does!

Does running the algorithm twice work? ~~Not immediately clear!~~

Our solution

We prove Hellman's algorithm finds a uniform pre-image in $f^{-1}(y)$!

Running Hellman twice makes the succ. prob ε_H^2

Conclusions

- **Inverse queries** are useful for attacks!
- **2-block collisions harder to find** than arbitrary length collisions (like in MD)

Open problems

- **Tight bounds** for $B = 1, 2$
- Attacks that **exploit the inverse queries** for $B \geq 2$
- Limitations for $B \geq 3$

https://eprint.iacr.org/2022/1009