

Bitslice Masking and Improved Shuffling: How and When to Mix Them in Software?

Melissa Azouaoui, <u>Olivier Bronchain</u>, Vincent Grosso Kostas Papagiannopoulos, François-Xavier Standaert

CHES 2022, Leuven, Belgium







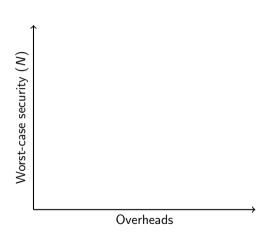
Contents

Introduction

Linear layers

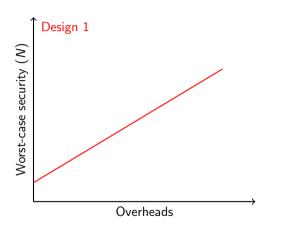
Non-linear layers

Perf. vs security



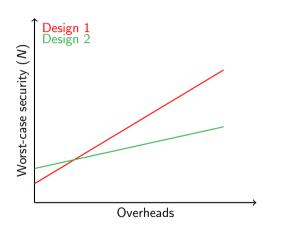
Countermeasures compared on:

- Run time overheads.
- ► Worst-case security (*N*).



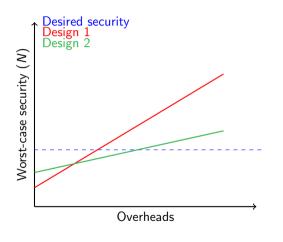
Countermeasures compared on:

- Run time overheads.
- ► Worst-case security (*N*).



Countermeasures compared on:

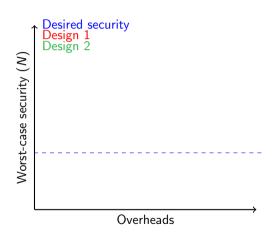
- Run time overheads.
- ► Worst-case security (*N*).



Countermeasures compared on:

- Run time overheads.
- ► Worst-case security (*N*).

 \rightarrow Best design depends on desired security.



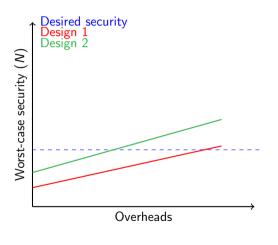
Countermeasures compared on:

- Run time overheads.
- ► Worst-case security (*N*).

 \rightarrow Best design depends on desired security.

Best design is device dependent:

- ► Noise level.
- Platform architecture.



Countermeasures compared on:

- Run time overheads.
- ► Worst-case security (*N*).

 \rightarrow Best design depends on desired security.

Best design is device dependent:

- Noise level.
- Platform architecture

Masking:

- ► Randomized the data processed.
- ► Sharing of $x := (x^0, x^1, ..., x^{d-1})$
- ► Noise amplification:

$$N \approx \frac{c}{\prod_i \operatorname{MI}(X^i; L)} \approx \frac{c}{\operatorname{MI}(X^i; L)^d}$$

Data Layout:

$$x^2$$

Olivier Bronchain

(MI(X;L)<1)

Masking:

- Randomized the data processed.
- ► Sharing of $x := (x^0, x^1, ..., x^{d-1})$
- Noise amplification:

$$N pprox rac{c}{\prod_i \mathrm{MI}(X^i; L)} pprox rac{c}{\mathrm{MI}(X^i; L)^d}$$

Data Layout:

1-th share

2-th share

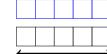
Shuffling:

- Randomized processing order.
- Execution based on a perm. of size n
- Noise addition:

$$N \approx \frac{\eta \cdot c}{\operatorname{MI}(X; L)}$$

Data Layout:

Perm:



data:

(MI(X; L) < 1)

Masking:

- ► Randomized the data processed.
- ► Sharing of $x := (x^0, x^1, ..., x^{d-1})$
- ► Noise amplification:

$$N \approx \frac{c}{\prod_i \mathrm{MI}(X^i; L)} \approx \frac{c}{\mathrm{MI}(X^i; L)^d}$$

► Data Layout:

0-th share
$$x^0$$
1-th share x^1
2-th share x^2

Shuffling:

- ► Randomized processing order.
- \triangleright Execution based on a perm. of size n
- Noise addition:

$$N \approx \frac{\eta \cdot c}{\operatorname{MI}(X; L)}$$

► Data Layout:

Perm: 2 4 1

data: x₂

 $\left(\operatorname{MI}(X;L)<1\right)$

Masking:

- ► Randomized the data processed.
- ► Sharing of $x := (x^0, x^1, ..., x^{d-1})$
- ► Noise amplification:

$$N \approx \frac{c}{\prod_{i} \operatorname{MI}(X^{i}; L)} \approx \frac{c}{\operatorname{MI}(X^{i}; L)^{d}}$$

► Data Layout:

0-th share
$$x^0$$
1-th share x^1
2-th share x^2

Shuffling:

- ► Randomized processing order.
- \triangleright Execution based on a perm. of size n
- ► Noise addition:

$$N \approx \frac{\eta \cdot c}{\operatorname{MI}(X; L)}$$

► Data Layout:

Perm: 2 4 1 3 0 data: x₂ x₄

indep. data

(MI(X; L) < 1)

Masking:

- ► Randomized the data processed.
- ► Sharing of $x := (x^0, x^1, ..., x^{d-1})$
- ► Noise amplification:

$$N \approx \frac{c}{\prod_{i} \operatorname{MI}(X^{i}; L)} \approx \frac{c}{\operatorname{MI}(X^{i}; L)^{d}}$$

► Data Layout:

0-th share
$$x^0$$
1-th share x^1
2-th share x^2

Shuffling:

- ► Randomized processing order.
- \triangleright Execution based on a perm. of size n
- Noise addition:

$$N \approx \frac{\eta \cdot c}{\operatorname{MI}(X; L)}$$

► Data Layout:

Perm: 2 4 1 3 0 data: $x_1 x_2 x_4$

indep. data

(MI(X; L) < 1)

Masking:

- ► Randomized the data processed.
- ► Sharing of $x := (x^0, x^1, ..., x^{d-1})$
- ► Noise amplification:

$$N \approx \frac{c}{\prod_{i} \operatorname{MI}(X^{i}; L)} \approx \frac{c}{\operatorname{MI}(X^{i}; L)^{d}}$$

► Data Layout:

0-th share
$$x^0$$
1-th share x^1
2-th share x^2

Shuffling:

- ► Randomized processing order.
- \triangleright Execution based on a perm. of size n
- ► Noise addition:

$$N \approx \frac{\eta \cdot c}{\operatorname{MI}(X; L)}$$

► Data Layout:

 $\left(\operatorname{MI}(X;L)<1\right)$

Masking:

- ► Randomized the data processed.
- ► Sharing of $x := (x^0, x^1, ..., x^{d-1})$
- ► Noise amplification:

$$N \approx \frac{c}{\prod_{i} \operatorname{MI}(X^{i}; L)} \approx \frac{c}{\operatorname{MI}(X^{i}; L)^{d}}$$

► Data Layout:

0-th share
$$x^0$$
1-th share x^1
2-th share x^2

Shuffling:

- ► Randomized processing order.
- \triangleright Execution based on a perm. of size n
- Noise addition:

$$N \approx \frac{\eta \cdot c}{\operatorname{MI}(X;L)}$$

► Data Layout:

indep. data

Existing side-channel countermeasures

 $\left(\operatorname{MI}(X;L)<1\right)$

Masking:

- ► Randomized the data processed.
- ► Sharing of $x := (x^0, x^1, ..., x^{d-1})$
- ► Noise amplification:

$$N \approx \frac{c}{\prod_i \operatorname{MI}(X^i; L)} \approx \frac{c}{\operatorname{MI}(X^i; L)^d}$$

▶ Data Lavout:

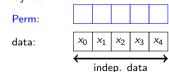
0-th share
$$x^0$$
1-th share x^1
2-th share x^2

Shuffling:

- ► Randomized processing order.
- ightharpoonup Execution based on a perm. of size η
- ► Noise addition:

$$N \approx \frac{\eta \cdot c}{\mathrm{MI}(X;L)}$$

► Data Layout:



 \rightarrow How to amplify shuffling thanks to masking ? (η^d)

1. Security:

- ► Explore design space for shuffling + masking.
- Evaluate the security:
 - ► Paper & pencil.
 - Confirmed with simulations.

Rivain et al. [RPD09]:

- ▶ Linear layers: $\binom{d \cdot \eta}{d}$
- Non-linear layers: η

1. Security:

- ► Explore design space for shuffling + masking.
- Evaluate the security:
 - ► Paper & pencil.
 - Confirmed with simulations.

2. Performances:

- Explore perf. bitslice and shuffle.
- Benchmarks on Cortex-M4.

Rivain et al. [RPD09]:

- ► Linear layers: $\binom{d \cdot \eta}{d}$
- ightharpoonup Non-linear layers: η

1. Security:

- Explore design space for shuffling + masking.
- Evaluate the security:
 - ► Paper & pencil.
 - Confirmed with simulations.

Rivain et al. [RPD09]:

- ▶ Linear layers: $\binom{d \cdot \eta}{d}$
- ightharpoonup Non-linear layers: η

2. Performances:

- Explore perf. bitslice and shuffle.
- Benchmarks on Cortex-M4.

3. Performances vs. security:

▶ Pertinence of masking and shuffling combination.

Contents

Introduction

Linear layers

Non-linear layers

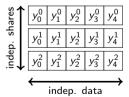
Perf. vs security

Protecting masked linear layers

x_{0}^{0}	x_{1}^{0}	x_{2}^{0}	x ₃ ⁰	x ₄ ⁰
x_0^1	x_1^1	x_{2}^{1}	x_{3}^{1}	x_{4}^{1}
x_0^2	x_1^2	x_{2}^{2}	x_3^2	x ₄ ²

Setting:

- $ightharpoonup \eta = 5$ independent data x_j and y_j
- ightharpoonup d = 3 shares x^i and y^i .



Protecting masked linear layers



$$z_i^j = x_i^j \oplus y_i^j \bigoplus$$

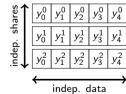


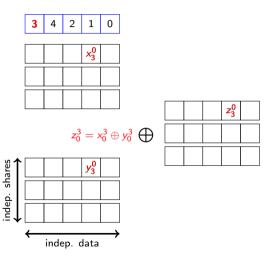
Setting:

- $ightharpoonup d = 3 \text{ shares } x^i \text{ and } y^i.$

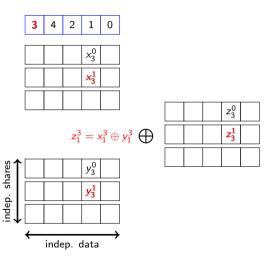
Goal:

► Compute all: $z_j^i = x_j^i \oplus y_j^i$

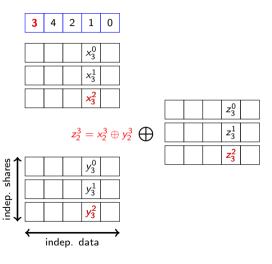




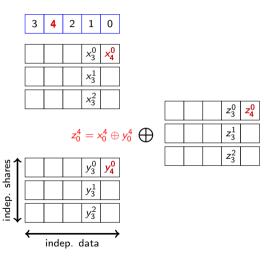
- ► Shuffle between variables
- ▶ Permutations:
 - Number: 1.
 - ightharpoonup Size: η .



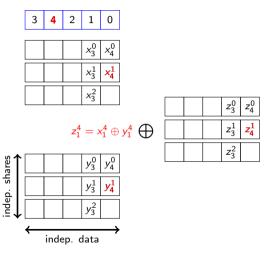
- Shuffle between variables
- ► Permutations:
 - Number: 1.
 - ightharpoonup Size: η .



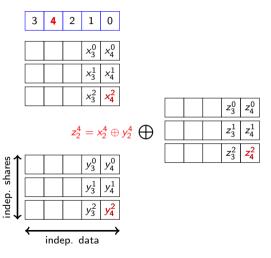
- Shuffle between variables
- ► Permutations:
 - Number: 1.
 - ightharpoonup Size: η .



- Shuffle between variables
- Permutations:
 - Number: 1.
 - ightharpoonup Size: η .



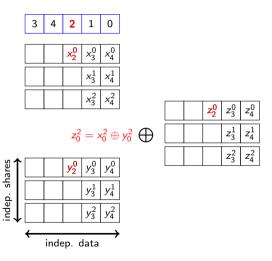
- Shuffle between variables
- ▶ Permutations:
 - Number: 1.
 - ightharpoonup Size: η .



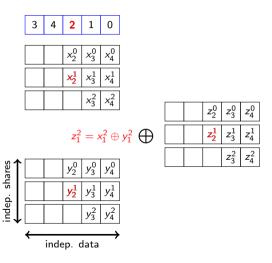
Description:

- Shuffle between variables
- Permutations:
 - Number: 1
 - ightharpoonup Size: η .

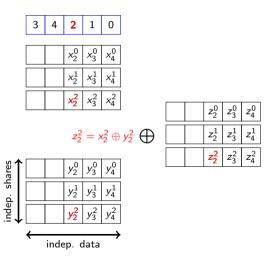
shares



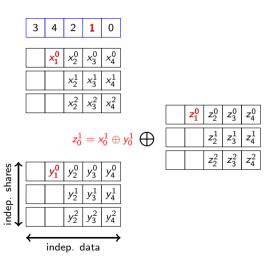
- Shuffle between variables
- ▶ Permutations:
 - Number: 1.
 - ightharpoonup Size: η .



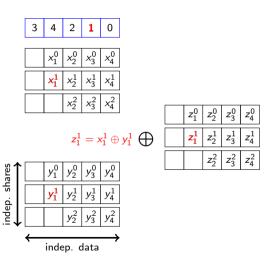
- Shuffle between variables
- Permutations:
 - Number: 1.
 - ightharpoonup Size: η .



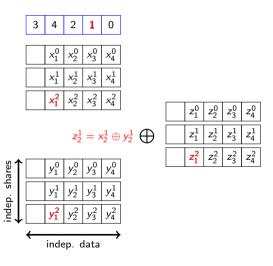
- Shuffle between variables
- Permutations:
 - Number: 1.
 - ightharpoonup Size: η .



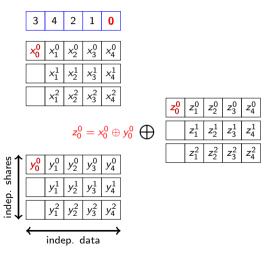
- Shuffle between variables
- Permutations:
 - Number: 1.
 - ightharpoonup Size: η .



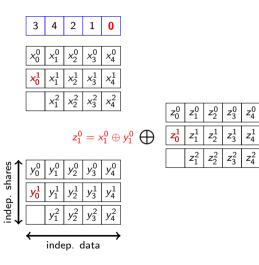
- ► Shuffle between variables
- Permutations:
 - Number: 1.
 - ightharpoonup Size: η .



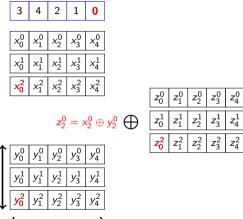
- ► Shuffle between variables
- Permutations:
 - Number: 1.
 - ightharpoonup Size: η .



- Shuffle between variables
- ▶ Permutations:
 - Number: 1.
 - ightharpoonup Size: η .



- Shuffle between variables
- ▶ Permutations:
 - Number: 1.
 - ightharpoonup Size: η .



Description:

- Shuffle between variables
- Permutations:

Number: 1.

ightharpoonup Size: η .

Decrease MI(X; L) by a factor η .

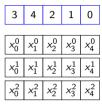
$$N pprox rac{c \cdot \eta}{\prod_i \operatorname{MI}(X^i; L)}$$

→ Masking does not amplify shuffling.

indep. data

shares

indep.



shares	•	<i>y</i> ₀ ⁰	y_1^0	y_{2}^{0}	<i>y</i> ₃ ⁰	<i>y</i> ₄ ⁰	
		y_0^1	y_1^1	y_{2}^{1}	y_3^1	<i>y</i> ₄ ¹	
indep		y_0^2	y_1^2	y_{2}^{2}	y_3^2	<i>y</i> ₄ ²	
indep. data							

z_{0}^{0}	z_1^0	z_{2}^{0}	z_{3}^{0}	z_{4}^{0}	
z_{0}^{1}	z_1^1	z_{2}^{1}	z_{3}^{1}	z_4^1	
z_0^2	z_1^2	z_{2}^{2}	z_3^2	z_4^2	

Description:

- ► Shuffle between variables
- Permutations:

Number: 1.

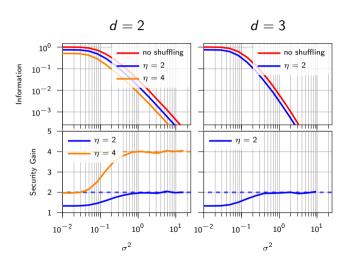
ightharpoonup Size: η .

Decrease MI(X; L) by a factor η .

$$N \approx \frac{c \cdot \eta}{\prod_i \operatorname{MI}(X^i; L)}$$

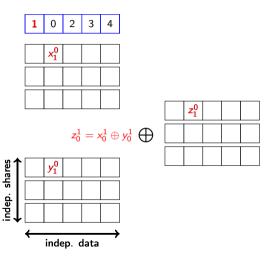
→ Masking does not amplify shuffling.

Shuffling-shares on linear layers: simulations



Expected security:

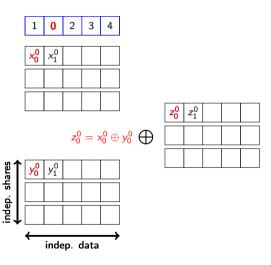
$$N pprox rac{c \cdot \eta}{\prod_i \operatorname{MI}(X^i; L)}$$



Description:

- ▶ Shuffle the *i*-th share of each x_i .
- ► Permutations:

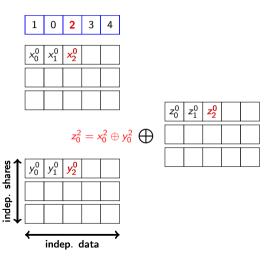
► Number: *d*.



Description:

- ▶ Shuffle the *i*-th share of each x_i .
- ► Permutations:

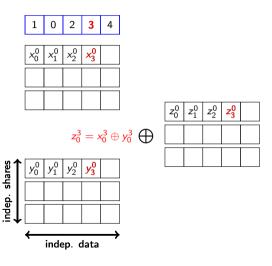
Number: d.



Description:

- ▶ Shuffle the *i*-th share of each x_i .
- ► Permutations:

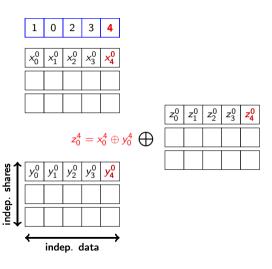
► Number: d.



Description:

- ▶ Shuffle the *i*-th share of each x_i .
- Permutations:

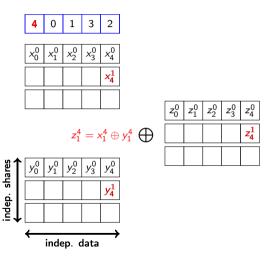
► Number: d.



Description:

- ▶ Shuffle the *i*-th share of each x_i .
- Permutations:

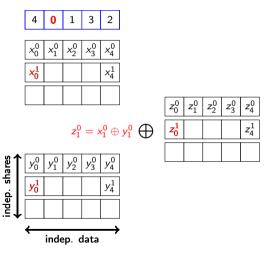
Number: d.



Description:

- \triangleright Shuffle the *i*-th share of each x_i .
- Permutations:
 - Number: d.
 - ightharpoonup Size: η .

shares

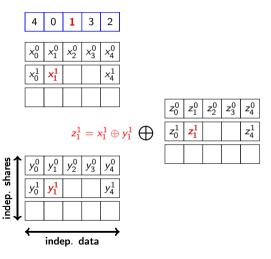


Description:

- ▶ Shuffle the *i*-th share of each x_i .
- ► Permutations:

► Number: d.

ightharpoonup Size: η .



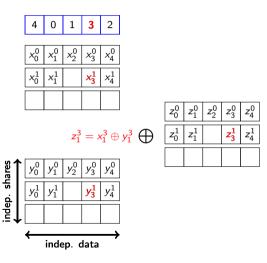
Description:

- \triangleright Shuffle the *i*-th share of each x_i .
- Permutations:

Number: d.

ightharpoonup Size: η .

shares

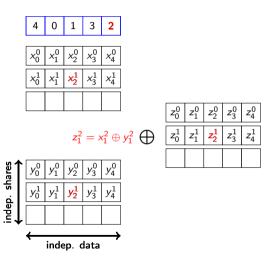


Description:

- ▶ Shuffle the *i*-th share of each x_i .
- ► Permutations:

► Number: d.

ightharpoonup Size: η .

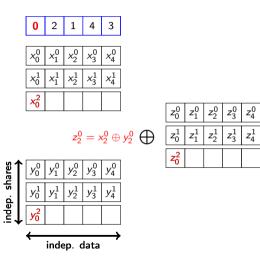


Description:

- ▶ Shuffle the *i*-th share of each x_i .
- Permutations:

Number: d.

ightharpoonup Size: η .



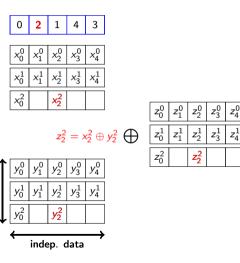
Description:

- Shuffle the *i*-th share of each x_i .
- Permutations:

Number: d.

Size: η .

shares



Description:

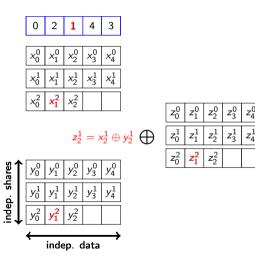
- ▶ Shuffle the *i*-th share of each x_i .
- ► Permutations:

Number: d.

ightharpoonup Size: η .

shares

indep.



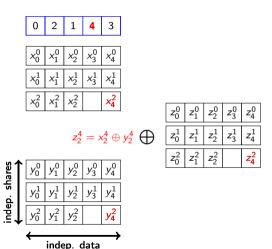
Description:

- Shuffle the *i*-th share of each x_i .
- Permutations:

Number: d.

Size: η .

shares



- ▶ Shuffle the *i*-th share of each x_i .
- ► Permutations:
 - ► Number: d.
 - ightharpoonup Size: η .



z_0^0	z_1^0	z_{2}^{0}	z_{3}^{0}	z ₄ ⁰
z_0^1	z_1^1	z_2^1	z_3^1	z_4^1
z_0^2	z_1^2	z_{2}^{2}	z_3^2	z_4^2

Description:

- ▶ Shuffle the *i*-th share of each x_i .
- Permutations:

Number: d.

ightharpoonup Size: η .

Decrease $MI(X^i; L)$ by a factor η .

$$N pprox rac{c}{\prod_{i} \operatorname{MI}(X^{i}; L)/\eta} pprox rac{c \cdot \eta^{d}}{\operatorname{MI}(X^{i}; L)^{d}}$$

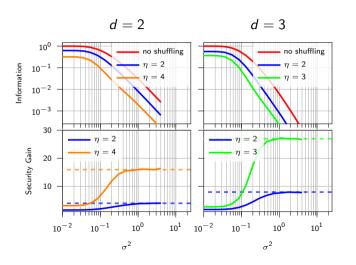
ightarrow Masking amplifies shuffling.

indep. data

shares

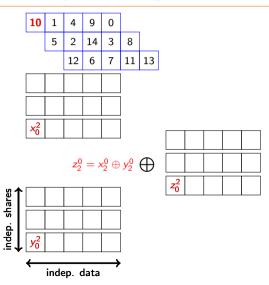
indep.

Shuffling-shares on linear layers: simulations



Expected security:

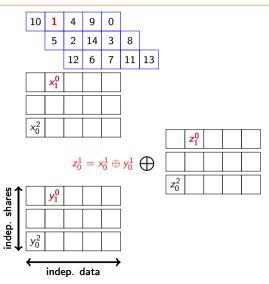
$$N pprox rac{c \cdot oldsymbol{\eta^d}}{\operatorname{MI}(X^i; L)^d}$$



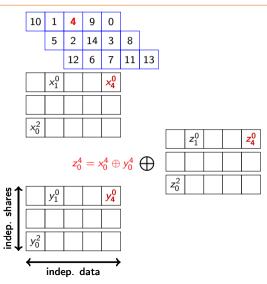
Description:

- ► Shuffle all the possible operations.
- Permutations:
 - Number: 1.
 - ightharpoonup Size: $d \cdot n$.

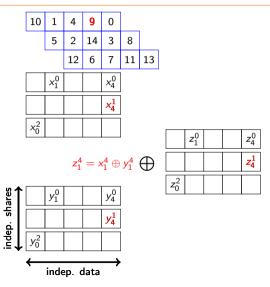
shares



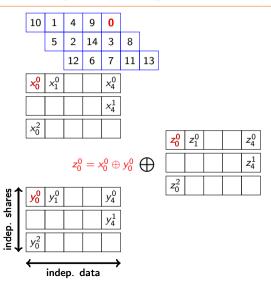
- ► Shuffle all the possible operations.
- ▶ Permutations:
 - Number: 1.
 - ► Size: $d \cdot \eta$.



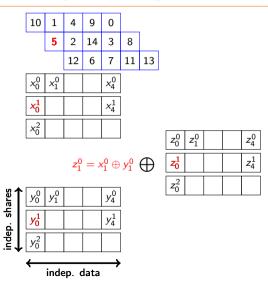
- ► Shuffle all the possible operations.
- ► Permutations:
 - Number: 1.
 - ► Size: $d \cdot \eta$.



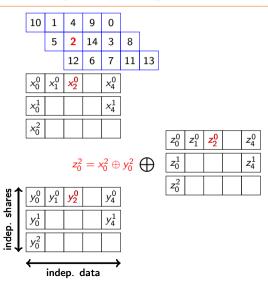
- ► Shuffle all the possible operations.
- ► Permutations:
 - Number: 1.
 - ► Size: $d \cdot \eta$.



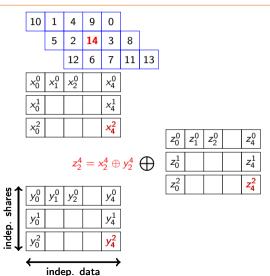
- ► Shuffle all the possible operations.
- ► Permutations:
 - Number: 1.
 - ► Size: $d \cdot \eta$.



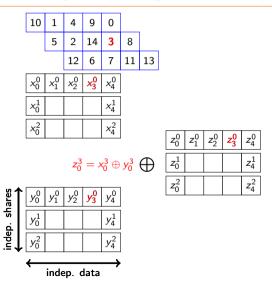
- ► Shuffle all the possible operations.
- ► Permutations:
 - Number: 1.
 - ► Size: $d \cdot \eta$.



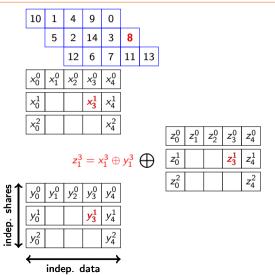
- ► Shuffle all the possible operations.
- Permutations:
 - Number: 1.
 - ► Size: $d \cdot \eta$.



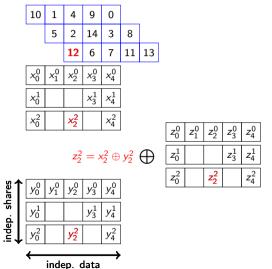
- ► Shuffle all the possible operations.
- ► Permutations:
 - Number: 1.
 - ► Size: $d \cdot \eta$.



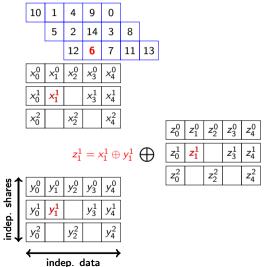
- ► Shuffle all the possible operations.
- Permutations:
 - Number: 1.
 - ► Size: $d \cdot \eta$.



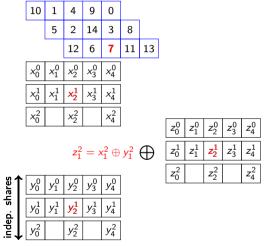
- ► Shuffle all the possible operations.
- ► Permutations:
 - Number: 1.
 - ► Size: $d \cdot \eta$.



- ► Shuffle all the possible operations.
- ► Permutations:
 - Number: 1.
 - Size: $d \cdot \eta$.



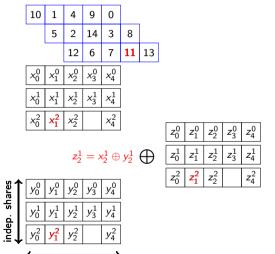
- ► Shuffle all the possible operations.
- ► Permutations:
 - Number: 1.
 - Size: $d \cdot \eta$.



Description:

- ► Shuffle all the possible operations.
- ► Permutations:
 - Number: 1.
 - ► Size: $d \cdot \eta$.

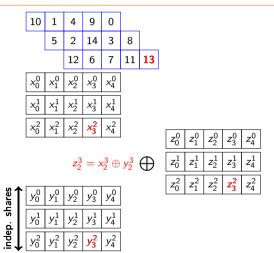
indep. data



Description:

- ► Shuffle all the possible operations.
- ► Permutations:
 - Number: 1.
 - ► Size: $d \cdot \eta$.

indep. data



Description:

- Shuffle all the possible operations.
- Permutations:
 - Number: 1.
 - ightharpoonup Size: $d \cdot n$.

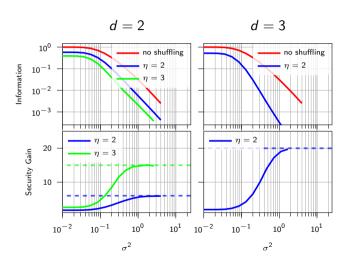
$$N \approx \frac{c \cdot \binom{d \cdot \eta}{d}}{\prod_i \operatorname{MI}(X^i; L)}$$

 \rightarrow Masking amplifies shuffling.

indep. data

shares

Shuffling-everything on linear layers: simulations



Expected security:

$$N pprox rac{c \cdot inom{d \cdot \eta}{d}}{\prod_i \operatorname{MI}(X^i; L)}$$

Contents

Introduction

Linear layers

Non-linear layers

Perf. vs security

Non-linear layers: summary of the results

For shuffled multiplications:

- ► Shuffling-shares and shuffling-tuples still apply with similar gain.
- ► Shuffling-everything could not be analyzed with paper & pencil:
 - Permutation on the output shares is not uniform.

	Linear layer			Non-linear layer		
	Gain	perm.	# perm.	Gain	perm.	# perm.
shuffling-tuples	η	η	1	η	η	1
shuffling-shares	η^d	η	d	η^d	η	d^2
shuffling-everything	$\begin{pmatrix} d \cdot \eta \\ d \end{pmatrix}$	$d\cdot \eta$	1	?	?	?

Table: Summary of the shuffling + masking combinations.

Non-linear layers: summary of the results

For shuffled multiplications:

- ► Shuffling-shares and shuffling-tuples still apply with similar gain.
- ► Shuffling-everything could not be analyzed with paper & pencil:
 - Permutation on the output shares is not uniform.

	Linear layer			Non-linear layer		
	Gain	perm.	# perm.	Gain	perm.	# perm.
shuffling-tuples	η	η	1	η	η	1
shuffling-shares	η^d	η	d	η^d	η	d^2
shuffling-everything	$\binom{d \cdot \eta}{d}$	$d\cdot \eta$	1	?	?	?

Table: Summary of the shuffling + masking combinations.

\rightarrow Next focus on shuffling-shares.

Contents

Introduction

Linear layers

Non-linear layers

Perf. vs security

Time versus security for shuffled ISW: open questions

Bitslice masking:

- ► Favors large #ANDs.
- ▶ Profits from parallelism.
- Randomness usage:

$$\# AND \cdot \frac{d \cdot (d-1)}{2}$$

Time versus security for shuffled ISW: open questions

Bitslice masking:

- ► Favors large #ANDs.
- Profits from parallelism.
- Randomness usage:

$$\# \text{AND} \cdot \frac{d \cdot (d-1)}{2}$$

Shuffling:

- ► Favors large #ANDs.
- ▶ Profits from serialization.
- Randomness usage:

$$d^2 \cdot \eta \cdot \log_2 \eta$$

Time versus security for shuffled ISW: open questions

Bitslice masking:

- ► Favors large #ANDs.
- ▶ Profits from parallelism.
- ► Randomness usage:

$$\# \text{AND} \cdot \frac{d \cdot (d-1)}{2}$$

Shuffling:

- ► Favors large #ANDs.
- ▶ Profits from serialization.
- Randomness usage:

$$d^2 \cdot \eta \cdot \log_2 \eta$$

Challenges when protecting ISW:

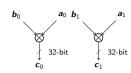
- Should we favor parallelism or serialization.
- ▶ Does it depend on the platform ?
- ▶ Does it depend on the primitive to protect ?

Time versus security for shuffled ISW: design space

(#AND = 64)

Option 1:

- Only bitsliced ISW.
- ▶ 32 bits per reg (full para.).



Time versus security for shuffled ISW: design space

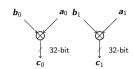
(#AND = 64)

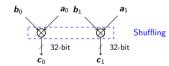
Option 1:

- ► Only bitsliced ISW.
- ▶ 32 bits per reg (full para.).

Option 2:

- ► Shuffled bitsliced ISW.
- ▶ 32 bits per reg (full para.).





Time versus security for shuffled ISW: design space

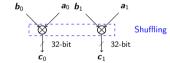
(#AND = 64)

Option 1:

- ► Only bitsliced ISW.
- ▶ 32 bits per reg (full para.).

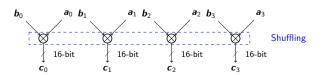
Option 2:

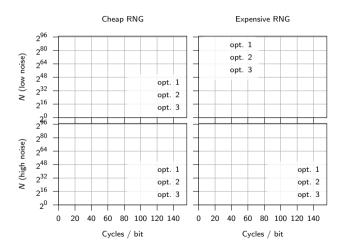
- Shuffled bitsliced ISW.
- ▶ 32 bits per reg (full para.).

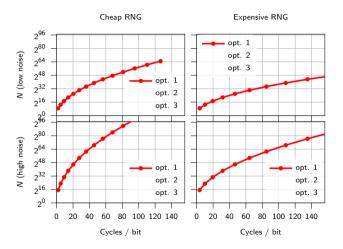


Option 3:

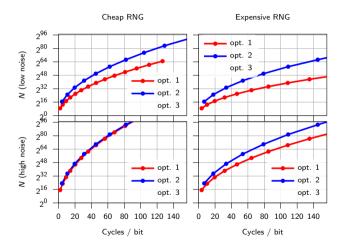
- ► Shuffled bitsliced ISW.
- ▶ 16 bits per reg (inc. ser.).





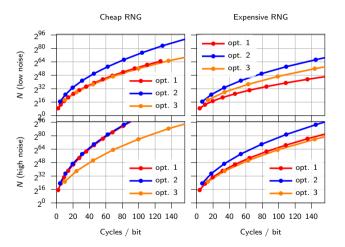


Opt 1: mask. only



Opt 1: mask. only

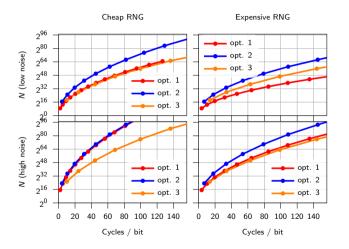
Opt 2: mask. & shuffl.



Opt 1: mask. only

Opt 2: mask. & shuffl.

Opt 3: lager perm.



Opt 1: mask. only Opt 2: mask. & shuffl. Opt 3: lager perm.

<u>Take home</u>: Use fully the registers and then shuffle

- masking is faster.
- ► : masking + shuffling is faster.

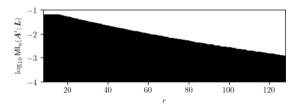


Figure: #AND=128, $N = 2^{64}$

- masking is faster.
- ► : masking + shuffling is faster.

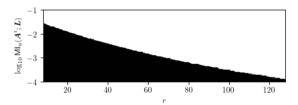


Figure: #AND=256. $N = 2^{64}$

- masking is faster.
- ► : masking + shuffling is faster.

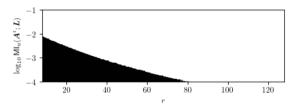


Figure: #AND=512. $N = 2^{64}$

- masking is faster.
- ► : masking + shuffling is faster.

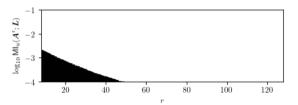


Figure: #AND=1024, $N = 2^{64}$

When to favor shuffling + masking:

- ► large of independent #AND.
- ightharpoonup expensive randomness r.
- relatively low noise.

- masking is faster.
- ▶ ☐: masking + shuffling is faster.

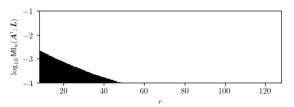


Figure: #AND=1024, $N = 2^{64}$

When to favor shuffling + masking:

- ► large of independent #AND.
- ightharpoonup expensive randomness r.
- relatively low noise.

Thanks!

Masking or Masking + shuffling:

- masking is faster.
- ► : masking + shuffling is faster.

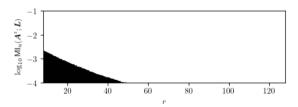


Figure: #AND=1024, $N = 2^{64}$

https://github.com/uclcrypto/bitslice_masking_and_shuffling