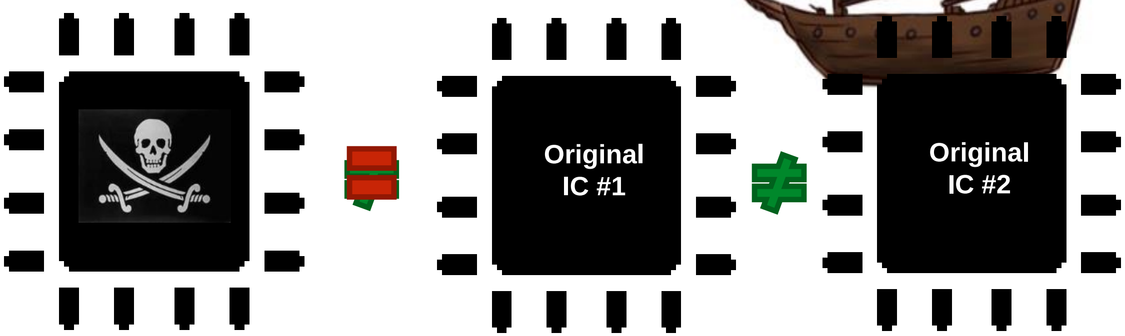
Strong Machine Learning Attack against PUFs with No Mathematical Model

Fatemeh Ganji, Shahin Tajik, Fabian Faessler, Jean-Pierre Seifert

Motivation

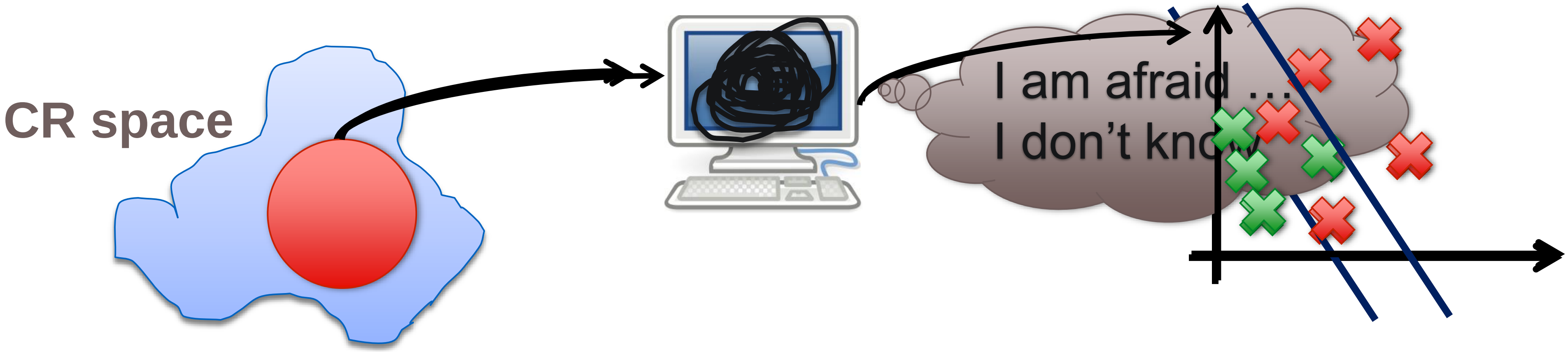
- Integrated circuits (ICs): vulnerability to piracy and overbuilding attacks [1]
- PUFs: Physically Unclonable Functions
  - Inspired by the characteristics of human finge
    inherent, unclonable



- Strong and weak PUFs

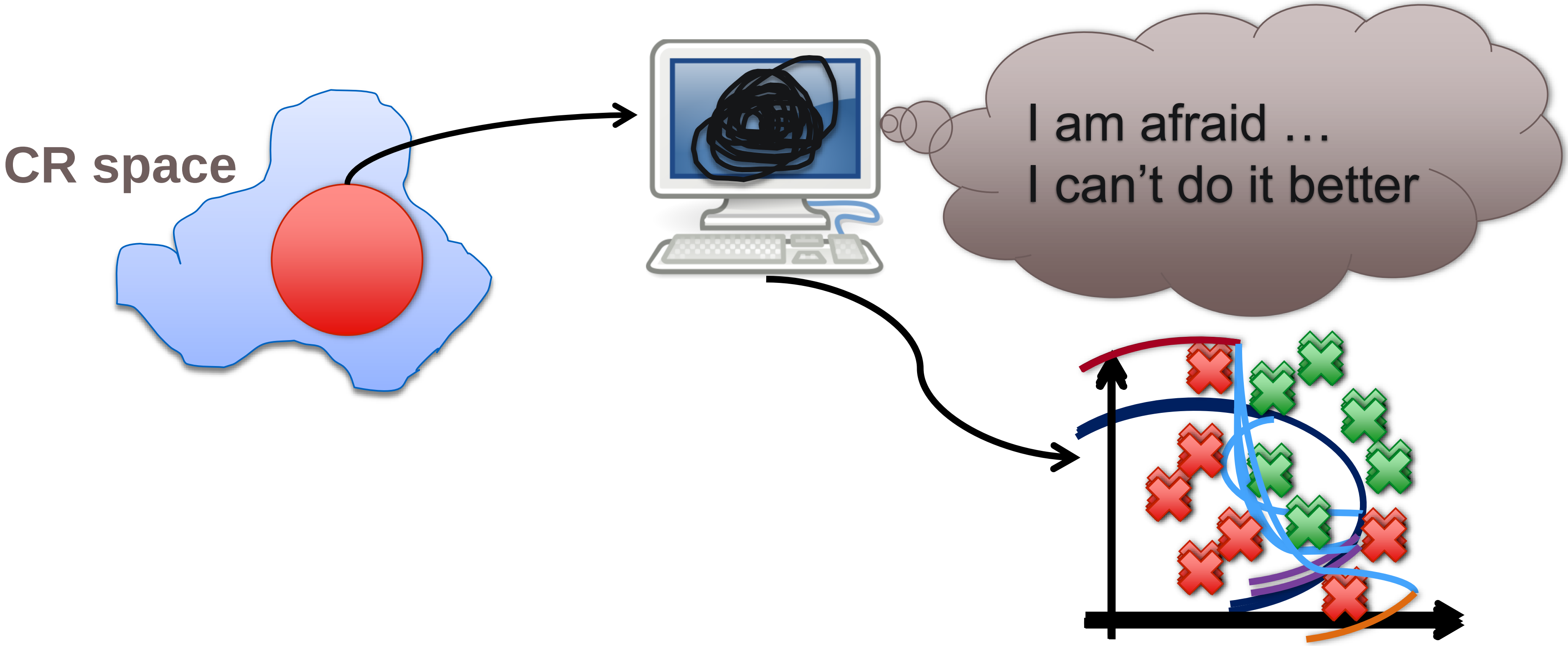
Unclonable?!

Empirical vs. PAC learning attacks



- Empirical learning approaches
  - No pre-defined levels of accuracy and confidence
- PAC learning approaches
  - For given levels of accuracy and confidence

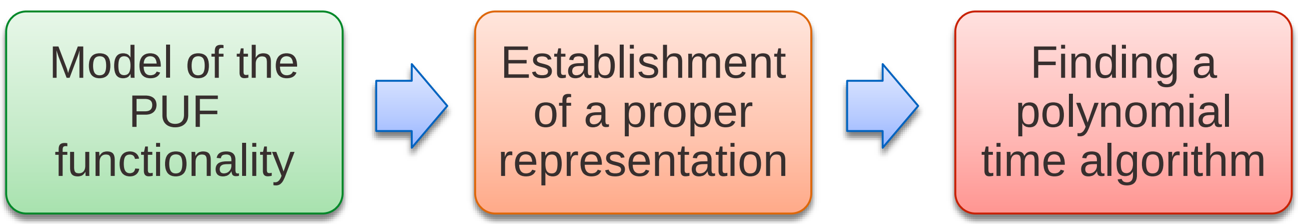
Strong vs. weak PAC learning



- A Weak learner: the accuracy of the model delivered is only slightly better that 50%
- Weak PAC learning and strong PAC learning are equivalent [3]

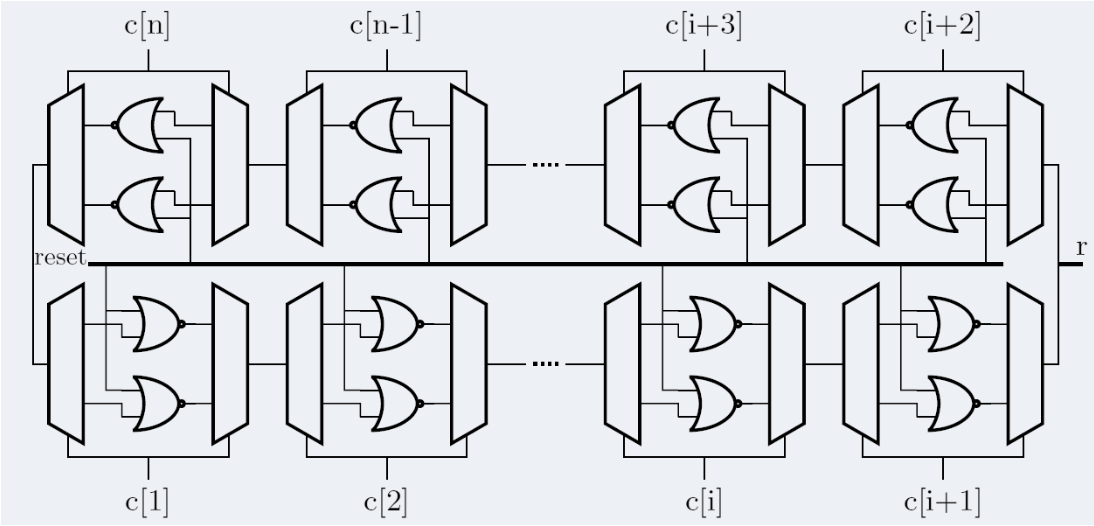
Why attackers win

- Linear behavior of Arbiter PUFs, cf. [4,5]: an example of the model representing the internal functionality of the respective PUF



- What happens if this model is unknown?
- Prime example: Bistable Ring PUFs

BR PUFs



- No precise mathematical model of the BR PUF functionality

Model of the PUF functionality → Establishment of a proper representation → Finding a polynomial time algorithm

PUF as a Boolean function

$$c=c_1\ldots c_n \longrightarrow \boxed{f_{PUF}} \xrightarrow{\;r\;}$$

- f_{PUF}: a Boolean function from $\{0,1\}^n$ to $\{0,1\}$, shown as

$$f_{PUF}: \mathbb{F}_2^{\,n} \rightarrow \mathbb{F}_2$$

- Linear Boolean functions
  - $f(c+c') = f(c) + f(c')$

c	$\boldsymbol{r = f_{PUF}(c)}$
$c=1\ldots0$	1
$c'=1\ldots1$	1
$c+c'=0\ldots1$	**???**

Linearity over $\mathbb{F}_2$

- Linear function over $\mathbb{F}_2$: ONLY parity function

  No PUF represented as a Boolean function over $\mathbb{F}_2$ is linear

  - Unequal influence of challenge bit positions on the respective responses

How many influential bits?

- Determined by the notion of **average sensitivity *I(f*** $\boldsymbol{I(f_{PUF})}$

  - c_1 is chosen uniformly at random

- Friedgut’s theorem rel on to the number of relevant bits [6]

$$I(f_{PUF}) := \sum_{i=1}^{n} \Pr[r_2 \neq r_1]$$

Learning juntas

c	***r***
*c=**1**…00*	*1*
*c'=**1**…11*	*1*
*c''=**1**…01*	***1***

Is 'K' a constant value?

- Example of a *1*-junta
- K-junta learning: finding the relevant coordinates
  - Algorithm presented by, e.g., Angluin [7]

What we know about BR PUFs

- Practical observations
  - Statistical analysis of the 2048 CRPs, given to a 64-bit BR-PUF: 5 influential bits [8]
    - Our experiments on 64-bit BR PUFs implemented on Altera Cyclone IV FPGAs
      - results for 30000 CRPs: 7 influential bits
- Mathematical, more precise observation
  - Computation of the average sensitivity

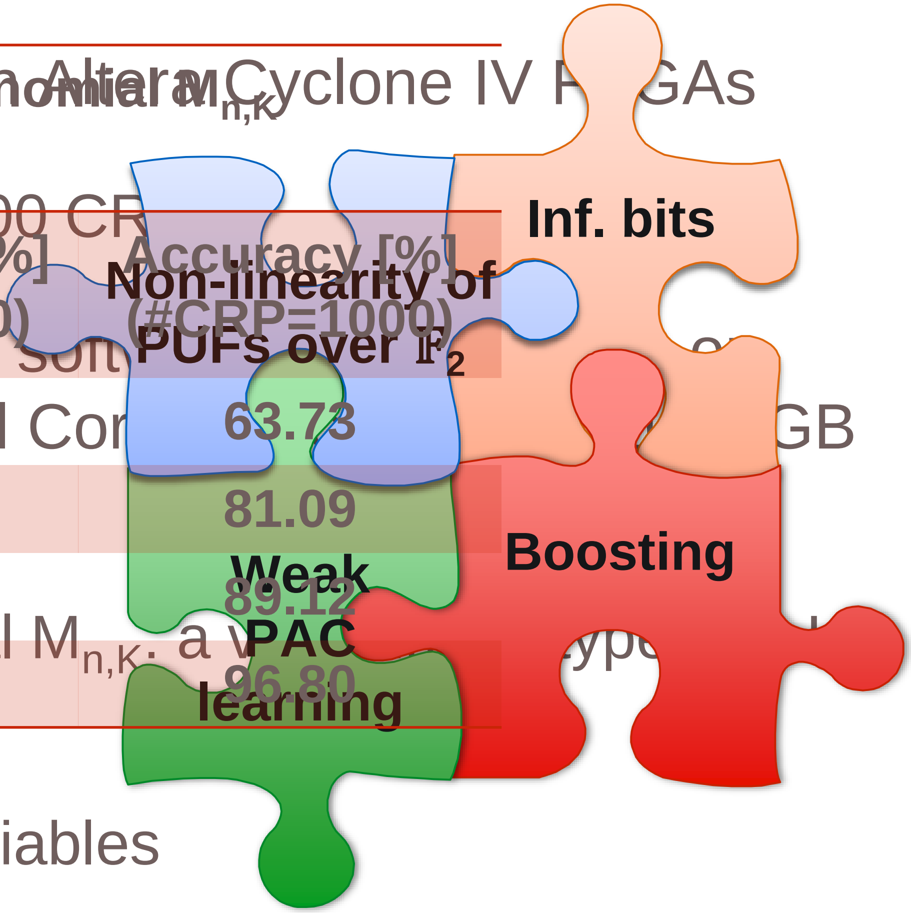
n	$I(f_{PUF})$
4	1.25
8	1.86
16	2.64
32	3.6
64	5.17

Experimental setup and results

- 64-bit BR PUFs implemented on Altera Cyclone IV F GAs
  - Size of training set: 100 and 1000 CR
- **Open source** machine learning soft ... MacBook Pro with 2.6 GHz Intel Cor ... GB of RAM
  - Learning algorithm for Monomial $M_{n,K}$: a w ... typ ... junta
    - Conjunction of the relevant variables
  - Adaboost

Learning of Monomial $M_{n,K}$

#boosting iterations	Accuracy [%] (#CRP=100)	Accuracy [%] (#CRP=1000)
0	54.48	63.73
10	67.12	81.09
20	77.53	89.12
50	82.65	96.80



More complex representation, e.g., Decision Lists (DL): 98.32% accurate final model

Conclusion

- Successful attack against PUFs with no mathematical model
  - Spectral properties of Boolean functions
  - Boosting technique
- Introduction of a new metric to assess the security of PUFs: the average sensitivity
- **In practice?**

References

[1] Koushanfar.: Hardware metering: A survey. Introduction to Hardware Security and Trust, 2012.

[2] Ruehrmair et al.: Modeling Attacks on Physical Unclonable Functions. In: Proc. of CCS 2010.

[3] Schapire, R.E.: The Strength ofWeak Learnability. Machine learning, 1990.

[4] Ganji et al.: PAC Learning of Arbiter PUFs. Journal of Cryptographic Engineering, 2016.

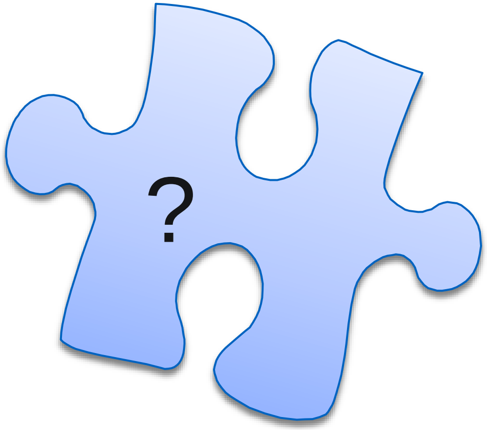
[5] Angluin: Learning regular sets from queries and counterexamples. Information and computation, 1987.

[6] Friedgut, E.: Boolean Functions with Low Average Sensitivity Depend on Few Coordinates. Combinatorica , 1988.

[7] Angluin: Queries and Concept Learning. Machine Learning, 1988.

[8] Yamamoto et al.: Security Evaluation of Bistable Ring PUFs on FPGAs using Differential and Linear Analysis. In Proc. of FedCSIS, 2014.

Thank you for you attention!



Outline

- Introduction and motivation
  - Let’s talk about PAC learning!
- Why having a mathematical Model matters
- PAC learning with no mathematical model
  - Example of BR PUFs
- Conclusion

Digital intrinsic PUFs

- Key idea: Manufacturing process variations on different chips used to generate PUFs

- Physically unclonable functions
  - Input to output mappings

Challenge c → PUF f_{PUF} → **response** r

- Strong and weak PUFs

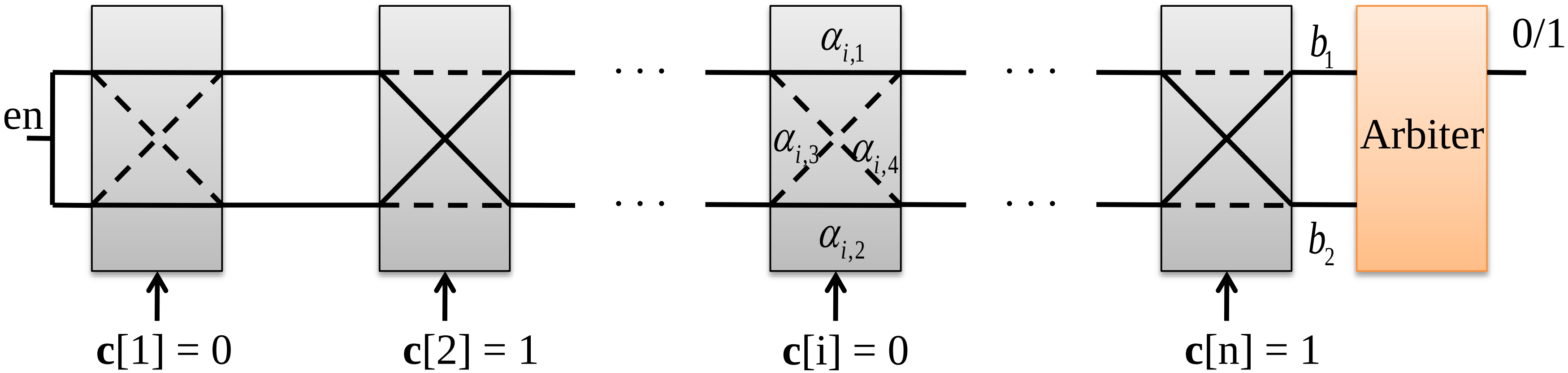
Modeling attacks [3]

- In practice: two phases, namely, enrolment and verification

Motivation (1)

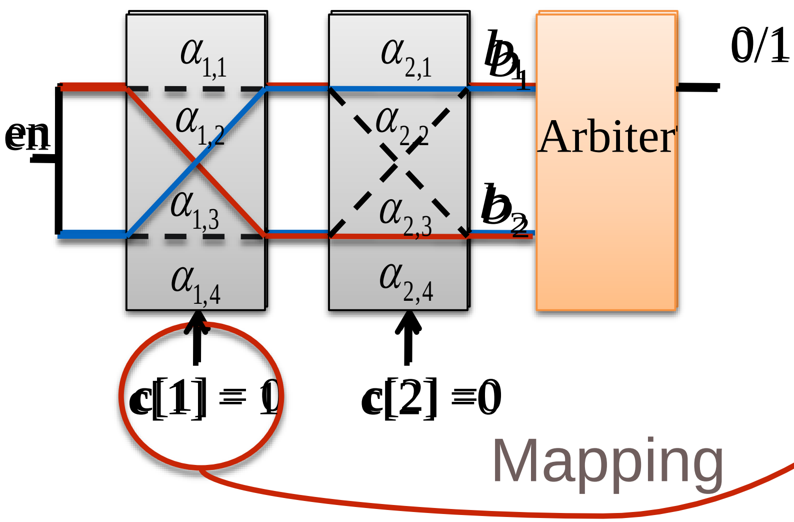
- Wide-spread use of Integrated Circuits (ICs) in different applications
  - Authentication, Identification, Transaction, Communication
    - Key generation, key storing, and device fingerprinting

What we have learned: an example of PAC learning attacks



- The security is relying on an assumption:
  - The attacker cannot measure the delays in each stage

Arbiter PUFs and its linear behavior

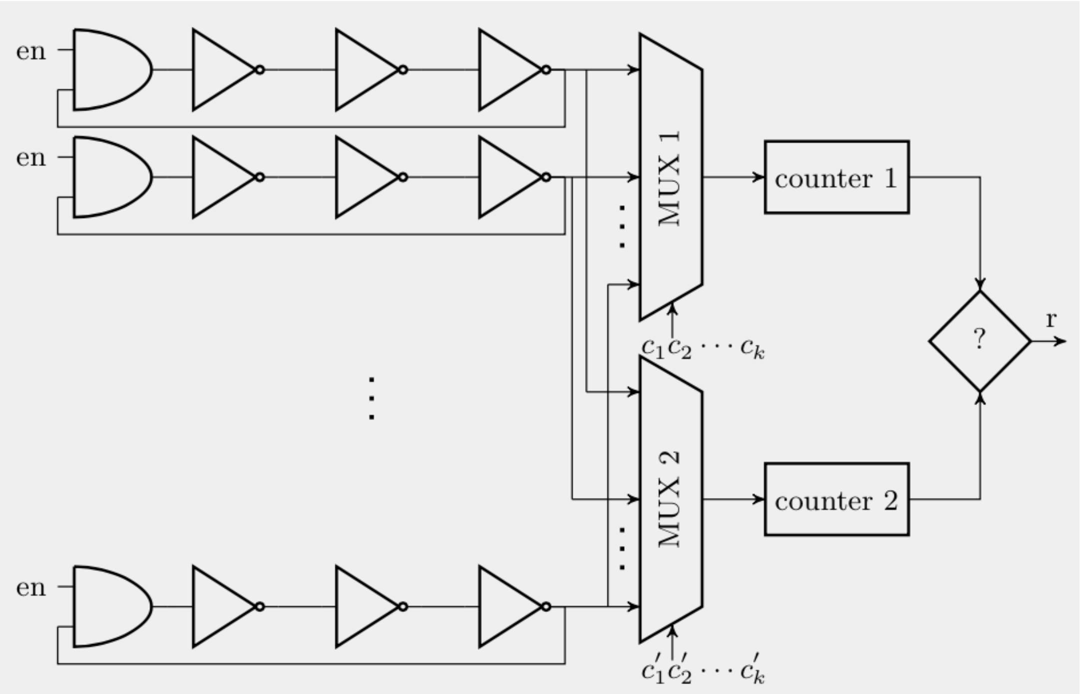


$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} 0,0,1,0,1,0,0,0 \\ 0,1,0,0,0,0,0,1 \end{pmatrix} \cdot \left(\alpha_{1,1}, \alpha_{1,2}, \alpha_{1,3}, \alpha_{1,4}, \alpha_{2,1}, \alpha_{2,2}, \alpha_{2,3}, \alpha_{2,4}\right)^t$$

$$\mathbf{C}\boldsymbol{\alpha} = \mathbf{B}$$

- PAC learning for given levels of accuracy and confidence [4]
  - Representation: polynomial-size Deterministic Finite Automata (DFA)
  - Algorithm presented by Angluin [5]

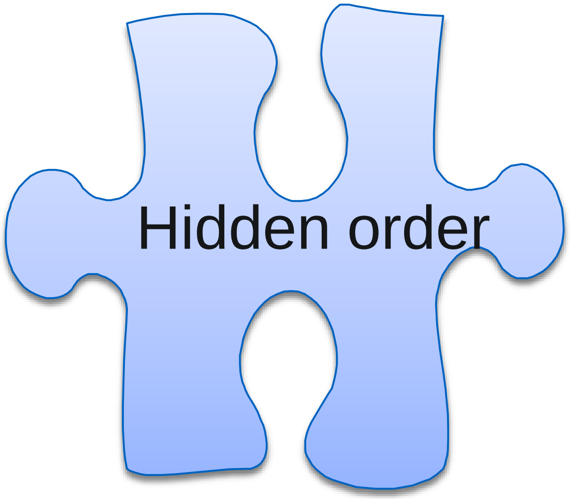
RO PUFs



- The security is relying on an assumption:
  - The attacker cannot measure the frequencies of the rings

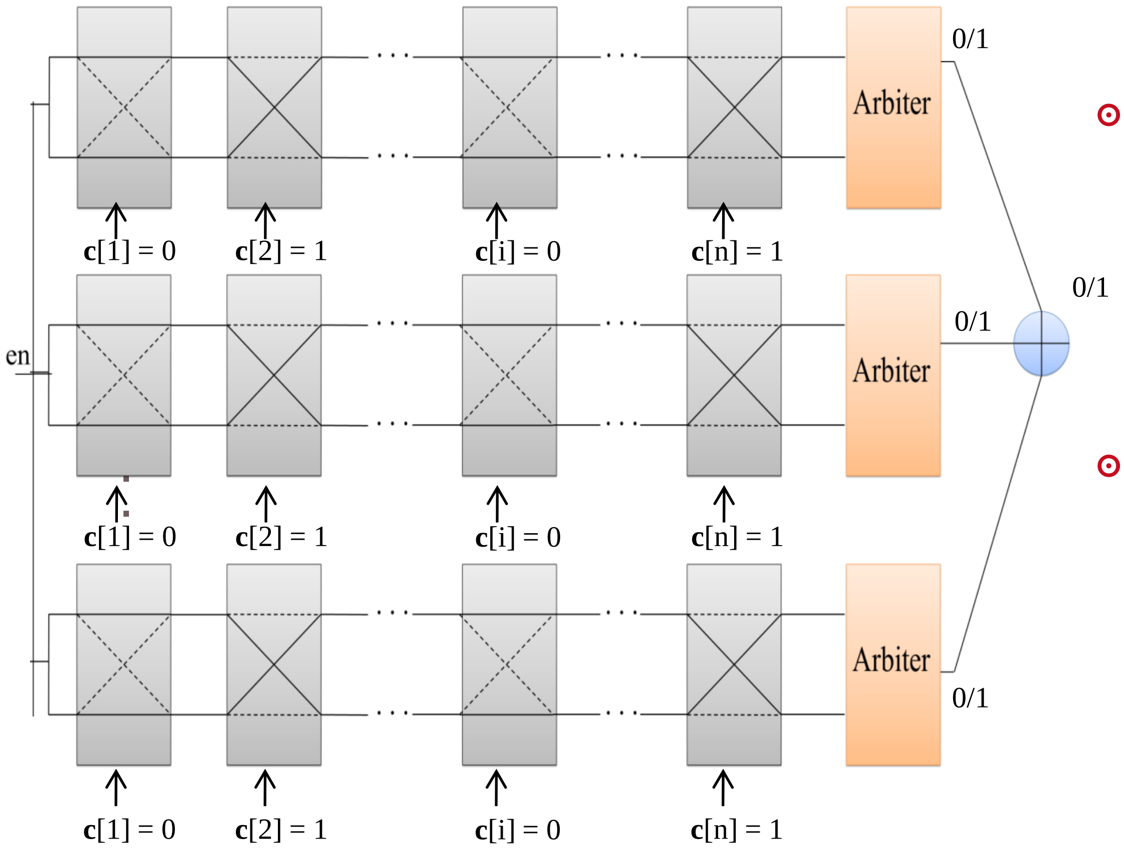
Fragile security of RO PUFs

- N ring-oscillators → N(N-1)/2 pairs are possible
  - Non-exponential CRP space!
- PAC learning for given levels of accuracy and confidence [6]
  - Representation: polynomial-size Decision List (DL)
  - Algorithm presented by Rivest [7]
- The reason for success:
  - A hidden order of frequencies



Refined architectures

Let’s XOR k arbiter chains [8]

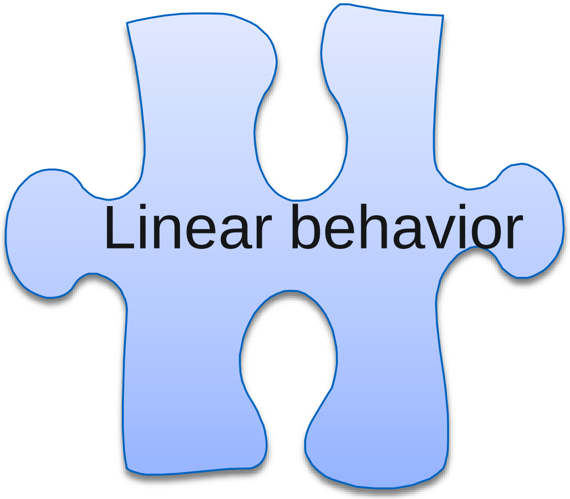


- Modeling attacks
  - Applicable only up to a certain number of chains [9,10]
- Side channel analysis
  - Successful but requires access to the challenges

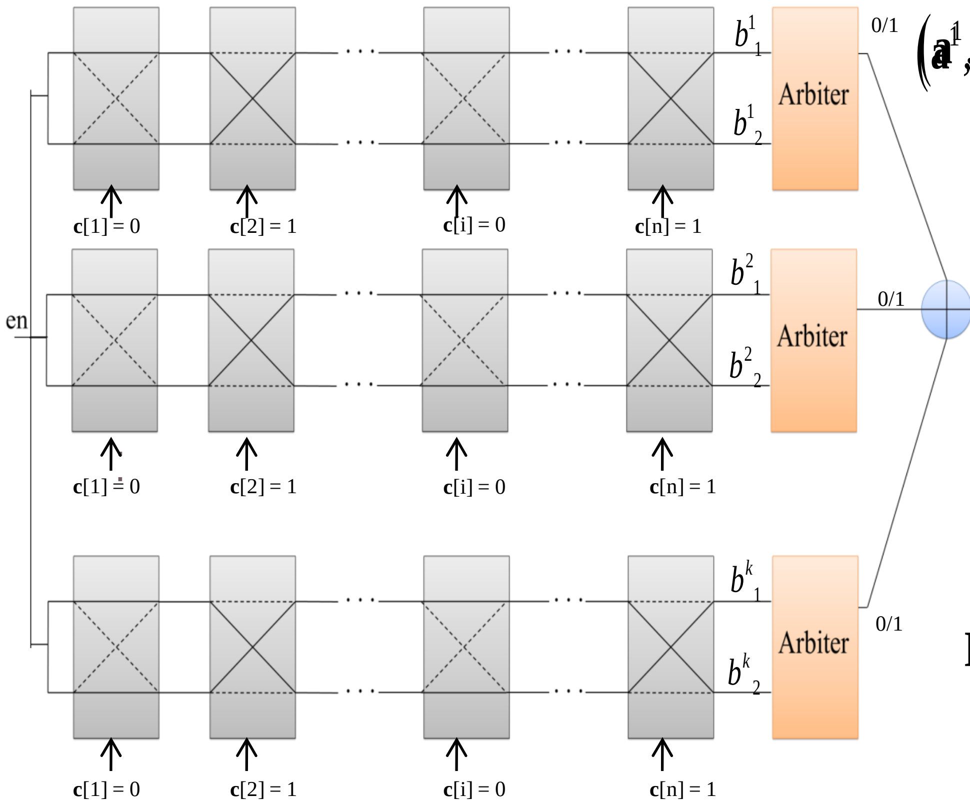
Controlled PUFs [4]

Our successful hybrid attack

- Combination of a lattice basis reduction attack and a photonic side-channel analysis [14]
  - Disclosing the hidden challenges, and delays
  - Applicable to unlimited number of arbiter chains



Controlled XOR PUFs



$$\left(\mathbf{a}^1, \mathbf{a}^2\right) = \begin{pmatrix} \alpha^1_{1,1},\ldots,\alpha^1_{1,4} & \cdots & \alpha^1_{n,1},\ldots,\alpha^1_{n,4} \\ \vdots & & \vdots \\ \alpha^k_{1,1},\ldots,\alpha^k_{1,4} & \cdots & \alpha^k_{n,1},\ldots,\alpha^k_{n,4} \end{pmatrix}^t$$

$$\begin{pmatrix} b^1_1, b^1_2 \\ \vdots \\ b^k_1, b^k_2 \end{pmatrix}$$

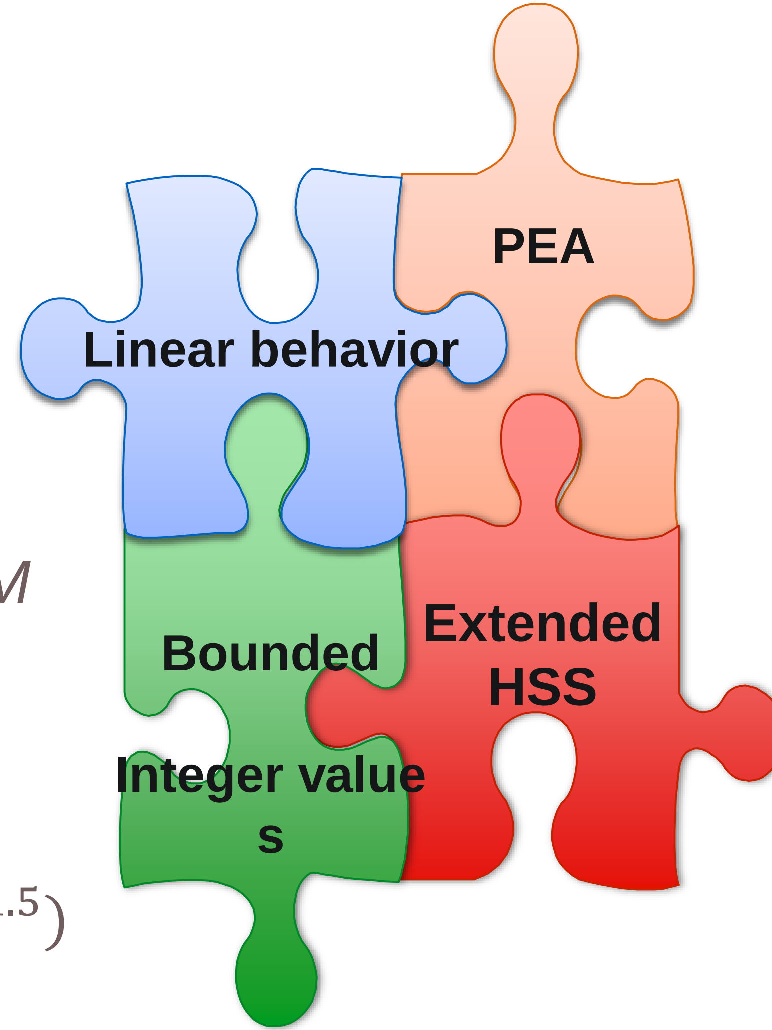
- *m measurements*

$$\mathbf{B} := \left(\mathbf{b}^1, \mathbf{b}^2, \ldots, \mathbf{b}^k\right) = \begin{pmatrix} b^{1,1}_1 & \ldots & b^{1,m}_1 & b^{1,1}_2 & \ldots & b^{1,m}_2 \\ \vdots & & & \vdots & & \vdots \\ b^{k,1}_1 & \ldots & b^{k,m}_1 & b^{k,1}_2 & \ldots & b^{k,m}_2 \end{pmatrix}^t$$

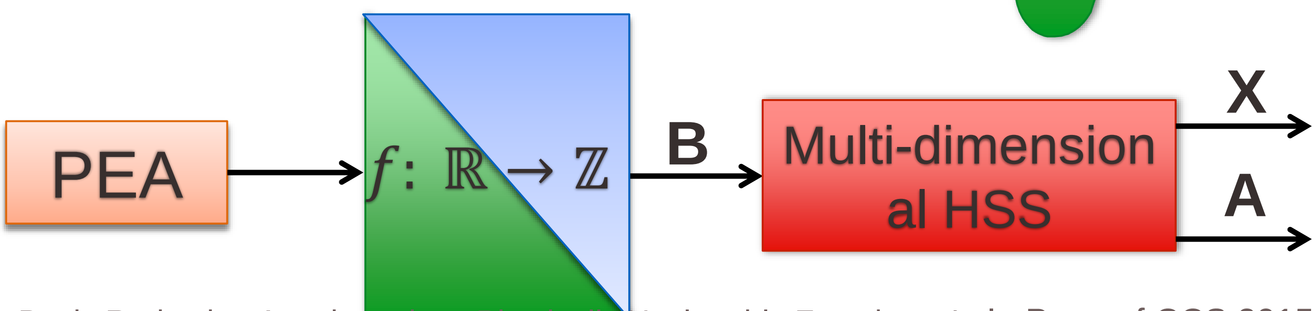
$$\mathbf{CA} = \mathbf{B}$$

Hidden Subset Sum! [12]

Hybrid attack [1]

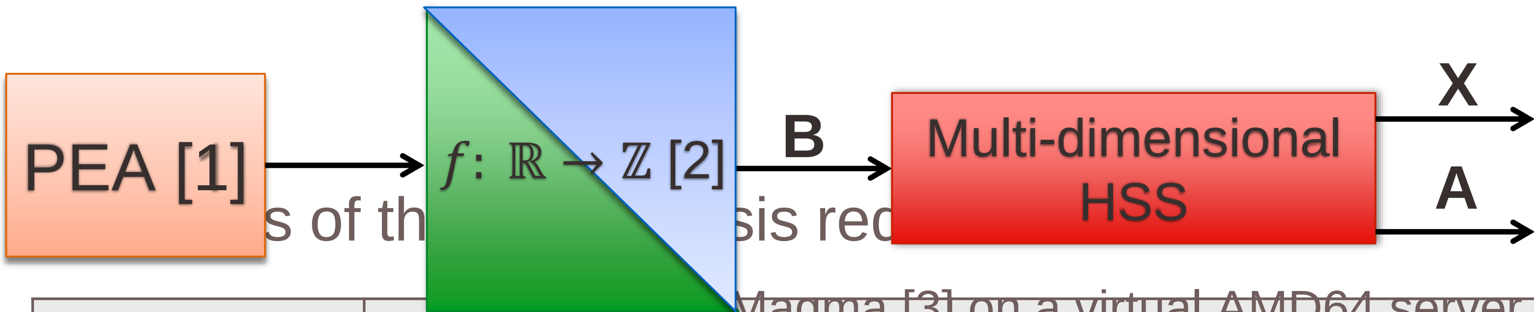


- Extension to Multi-dimensional HSS
  - In comparison to the HSS: smaller *M*
    - HSS: $M \gg \left(\frac{\sqrt{mn(m-n-1)}}{4}\right)^n$
    - Multi-dimensional HSS: $M \gg O(m^{1.5})$



[1] Ganji et al.: Lattice Basis Reduction Attack against Physically Unclonable Functions, In In Proc. of CCS 2015.

Experimental setup and results



s of th ... sis rec

Magma [3] on a virtual AMD64 server (1 core and 32 GB of RAM)

Setting	Approach	M	Total number of disclosed coefficients
n=11, k=11, m=78 (the number of hidden coefficients=44)	HSS	2^{160}	44
	Multi-dimensional HSS	2^{6}	44
n=32, k=32, m=370	HSS	---	---
	Multi-dimensional HSS	2^{15}	123

Arbiter PU

http://magm

Noise: a real enemy?

Noisy PUFs

- Applying the same challenge → Different responses
  - Due to the environmental variations
- Failure of the conventional learning methods

Is it possible to apply a PAC learning framework?

Yes! New PAC learning framework containing principles of learning theory and Boolean analysis