

Predicting PUFs faster: using high entropy input codes to improve machine learning attacks on a category of Physical Unclonable Functions

R. C. Surita, M. L. Côrtes, G. Araujo and D. F. Aranha

Abstract

PUFs are a class of security primitives that rely on statistical variations of integrated circuits production processes to provide authentication without the need explicitly storing cryptographic keys. Nevertheless several PUF architectures have shown to be predictable using machine learning techniques. This work extends a code published on related work that maximizes the output entropy for a single PUF architecture to a class of delay based PUF architectures and provides a method to make machine learning attacks more efficient against these devices.

Key words:

Hardware security, information theory, machine learning.

Introduction

Physical Unclonable Functions (PUFs) are devices capable of exploring process variations on manufacturing technologies to produce random functions that are different between instances produced without the need of explicitly setting keys [1, 2]. These devices provide low power, cheap authentication as required for IoT devices. Nevertheless, many PUF architectures shown to be vulnerable to several attacks, more critically, modelling attacks using black box machine learning [1]. Besides, there is not much information on why these attacks are effective or how to produce modelling resistant PUF designs.

Results and Discussion

Assuming gaussian distribution of random parameters, related work states that the output of a loop PUF can be evaluated as the sign of the inner product $y = \text{sgn}(\vec{w} \cdot \vec{x})$, where \vec{w} is a vector of random delays and \vec{x} is an input challenge and suggests that using Hadamard matrixes as a loop PUF input could maximize the output entropy [2]. This work extends these properties for a class of delay PUFs such as the Arbiter PUF and the XOR PUF given that they can have their outputs given by the sign of an inner product by applying a parity transform over the input vector [1].

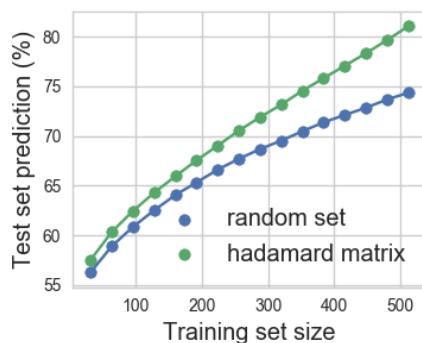


Figure 1. Logistic regression accuracy for an average of 10,000 digitally simulated 256 Arbiter PUFs trained with a Hadamard matrix vs a same size random set.

By plotting the precision for several linear PUF device sizes, it also indicates that the precision for a Hadamard matrix $N \times N$ is constant for a PUF of input size N , changing by the PUF design related with \vec{w} dimension.

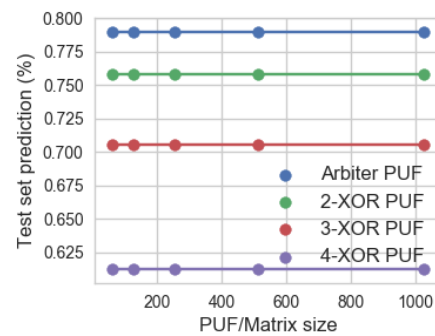


Figure 2. Logistic regression accuracy for PUFs trained with a Hadamard matrix of size $N \times N$ by input dimension N for some PUF architectures averaged from 10,000 digitally simulated devices.

Conclusions

The results for the learning model of linear PUFs using maximum output entropy sets vs random sets suggests that high entropy sets can considerably increase the learning rate of machine learning algorithms. The results for the $N \times N$ performance vs PUF size also indicate that increasing the dimension of such PUFs has a constant learning effort for the Hadamard set, thus discouraging designing architectures that can be represented as an inner product and where the hardware footprint grows linearly with the input vector dimension.

Acknowledgement

We thank CNPq and PIBIC for founding this project.

¹ U. Rührmair, F. Sehnke, J. Sölter, G. Dror, S. Devadas, and J. Schmidhuber. "Modeling attacks on physical unclonable functions". In Proceedings of the 17th ACM conference on Computer and communications security (CCS '10). ACM, New York, NY, 2010, pp. 237-249.
² O Rioul, P. Solé, S. Guilley and J. L. Danger, "On the entropy of Physically Unclonable Functions," 2016 IEEE International Symposium on Information Theory (ISIT), Barcelona, 2016, pp. 2928-2932.