Abstract

Physically Unclonable Functions (PUFs) promise to be a critical hardware primitive to provide unique identities to billions of connected devices in Internet of Things (IoTs). An ideal PUF should be unclonable (by manufacturing or by mathematical modeling). However, due to the gap between the ideal characteristics of a hypothetical PUF and the real characteristics of practically implementable PUF, several attacks have been launched. Specifically, modeling attacks are considered as greatest threat to the PUF. In modeling attacks, typically a small fraction of the CRP dataset of a particular PUF instance is made available to the adversary, based on which she builds an accurate computational model of the PUF instance, capable of predicting the response for an arbitrary challenge with high probability of success. The primary goal of the thesis is to design, analyse and implement the modeling attacks against PUF.

In this thesis, we develop and implement novel modeling attacks for PUFs using deep feedforward neural networks (DFNN). Through our work, we demonstrate the power of DFNN based modeling attacks on PUFs by launching attacks on well-known robust PUFs like *Multiplexer PUF* and *Interpose PUF*. Our next investigation focuses on the better modeling methods for launching the attack on XOR APUF and its variants. For this, we developed a novel machine learning model using tensor regression for improving the modeling attack. In this study, we demonstrate that our new model is computationally more efficient than previously known attacks. As an interesting thought, we then explore on the constructive side of modeling attacks i.e. Boolean function representation of PUFs. We propose a novel CAD framework to generate a combinational circuit representation of an APUF instance and highlight its significance. As a final objective, the thesis aims to investigate the crucial problem associated with PUF, that is, to understand and analyze its robustness against modeling attacks. our research applies a data-driven empirical metric termed the *Correlation Integral based Intrinsic Dimension* (ID), to estimate the *inherent complexity* of the relationship between the challenges and responses of a given PUF variant. This metric is computed from the linear projection layer of a Deep Neural Network (DNN) aimed at modeling the PUF.

Keywords: Physically Unclonable Functions, Modeling Attacks, Neural Networks, Robustness, Internet-of-Things.