Abstract

The widespread adoption of the Internet of Things (IoT) has ensued a significant increase in the connectivity of billions of smart devices to the Internet. Many of these devices are resource-constrained and cannot perform complex cryptographic operations. Securing these devices relies on a hardware "root of trust" (RoT) that is a foundation for various applications. A Physically Unclonable Function (PUF) is a hardware-based primitive that exploits the intrinsic physical characteristics of a device to generate a pseudorandom, unique, and unclonable device identifier, making it a promising candidate for a hardware RoT. Despite extensive research on PUF utilization in security applications like device identification, code attestation, key generation, and authenticated key exchange protocols, the formal analysis of this construct with respect to different attack strategies remains inadequate. Most silicon-based PUF designs are guided by attacks on former constructions and lack a rigorous mathematical analysis. Furthermore, the inherent uniqueness of each instance poses challenges for analysis using traditional verification strategies employed for VLSI systems. This necessitates the exploration of new test methodologies and the development of specialized formal analysis techniques.

This thesis focuses on the formal analysis of Silicon PUFs, aiming to address these challenges. The first objective is to develop a black-box testability analysis framework for PUFs to evaluate the presence of any aberrations in the circuit implementation. PUFs realize pseudorandom Boolean mappings in hardware. To assess a given PUF implementation, we leverage the correlation spectra properties of Boolean functions. This framework does not assume any specific design and applies to all PUFs, regardless of their architecture. The next objective is to assess the non-linearity of the mappings realized by different PUF instances. Non-linearity is a crucial property mandated for cryptographic Boolean functions. We propose a systematic framework for estimating the non-linearity of Strong PUFs, characterized by large challenge lengths, using a small fraction of its challenge-response space.

In the next part of the thesis, we focus on the unpredictability property of PUFs and analyze various PUF constructions and their compositions using provable and empirical machine learning-based modeling techniques. First, we propose a CAD framework for formal learnability assessment of PUF constructions in the Probably Approximately Correct (PAC) model. Herein, we propose a formal language named `PUF-G` for the uniform representation of Silicon PUF constructions. The CAD framework takes a PUF design represented in `PUF-G` language and returns the sample complexity of the PAC learning algorithm. Next, we extend the

capability of the tool that allows provable learnability assessment of generic PUF compositions in the basic and uniform PAC model. It also analyses various PUF design strategies based on the asymptotic complexity of the learnability bounds. Besides this class of provable attacks requiring only challenge-response pairs, we also assess a contemporary Silicon PUF construction against a reliability-based modeling attack that leverages response bias, reliability information, and CRPs. This is followed by a formal analysis of the reliability-based modeling attacks using Gradient-based optimization techniques, depicting the impact of various objective functions on the modeling attack performance. By providing theoretical guarantees of learnability using CRPs and reliability information, the thesis provides formal insight into the impact of various design strategies on the modeling robustness of PUF compositions, thereby aiding designers to make informed design choices to construct robust Silicon PUF constructions.

Keywords: Physically Unclonable Functions, Boolean Functions, Probably Approximately Correct Learning, Machine Learning, Modeling attack, Reliability-based modeling attack.