Challenging the Unclonability and Unlearnability Claims of Physical Unclonable Functions

Sergio Vinagrero Gutierrez[1,2], Cédric Marchand
[1]Centrale Lyon, INSA Lyon, CNRS, Université Claude Bernard Lyon 1, CPE Lyon, INL, UMR5270, Ecully 69134, France
sergio.vinagrero@ec-lyon.fr

Elena-Ioana Vatajelu, Giorgio Di Natale
[2]Univ. Grenoble Alpes, CNRS, Grenoble INP*, TIMA, 38000 Grenoble, France

Speaker: Sergio Vinagrero Gutierrez

The security of electronic systems has traditionally relied on cryptographic primitives that store secret information in Non-Volatile Memories (NVMs). These made NVMs the prime targets for adversaries and their security can no longer be guaranteed. Physical Unclonable Functions (PUFs) emerged as a promising alternative, exploiting inherent manufacturing variability to generate unique device-specific responses dynamically, thereby eliminating the need for embedded secrets stored in NVM. To guarantee the system security, PUFs are often evaluated using a limited set of canonical metrics, mainly Uniformity, Bit-aliasing, Uniqueness, and Reliability. While these metrics provide a basic assessment of statistical behavior, they frequently fall short in capturing deeper structural dependencies between challenges and responses. As a result, potential vulnerabilities can remain undetected.

The selling point of PUFs originated from the claims of unclonability and resistance to modeling, supported by their exploitation of uncontrollable process variations and absence of stored data. However, these claims have been repeatedly challenged in the literature. Numerous studies have demonstrated the feasibility of modeling and cloning attacks, revealing that the very statistical and physical properties leveraged for PUF design can also be exploited by adversaries.

In this work, we investigate a broad range of analytical and statistical frameworks aimed at understanding and modeling the challenge-response behavior of PUFs. Specifically, we analyze Boolean function analysis, Bayesian inference models, and complex neural network architectures. In addition to evaluating their effectiveness as attack strategies, we highlight how these models can be leveraged by hardware designers to identify design limitations, enabling more secure and resilient PUF implementations.