A Survey on the Security of PUFs

Shuqin Su\textsuperscript{1}, Min Zhu\textsuperscript{2}, Hanning Wang\textsuperscript{1}, Bohan Yang\textsuperscript{1} and Leibo Liu\textsuperscript{1}

\textsuperscript{1} School of Integrated Circuits, Tsinghua University, Beijing China, 100084
\textsuperscript{2} Wuxi Micro Innovation Integrated Circuit Design Company Ltd., Wuxi 214028, China.

ssq18@tsinghua.org.cn

Abstract. Physical Unclonable Function (PUF) is an essential security block in security systems, which can be used for authentication and key generation. Its security is related to the security of the entire security system, thus it has received widespread attention and it is necessary for designers and users to know the security threats and solutions for PUFs before working on PUFs. The paper mainly introduces and summarizes a variety of side-channel attacks, machine learning attacks, and corresponding countermeasures for PUFs. Side-channel attacks are usually combined with machine learning attacks to improve the effectiveness of attacks. Attackers can use machine learning methods to achieve mathematical modeling of PUFs and most of the existing PUFs are vulnerable to machine learning attacks.

1. Introduction

For modern communication systems, mobile computing, and Internet of Things (IoT) which are widely applied in our modern life, security is significant. Various cryptographic algorithms and protocols are utilized to ensure the security of these applications. PUFs can be used to realize identity authentication and generate keys for cryptographic algorithms and protocols. Compared with the traditional secure authentication method where a physically secure Non-volatile Memory (NVM) is used to store the identity authentication code and keys, PUFs have advantages in resource consumption and security due to their lightweight structures and the information loss after power failure.

The security problems of most security algorithms are not mainly derived from the algorithms but from the security blocks in the algorithms. PUF is an essential security block in security systems whose security is related to the security of the whole system directly. To ensure the security of PUFs, it is necessary to understand the existing attack methods and corresponding countermeasures. The attacks on PUFs mainly include side-channel attacks and machine learning attacks. Side-channel attacks are conducted based on the obtained side-channel information, such as power attack, electromagnetic attack, and fault attack. Machine learning attacks model PUFs to predict unknown outputs with various machine learning algorithms. For different attack methods, corresponding countermeasures are proposed to resist the attacks and improve the security of PUFs. For side-channel attacks, the principle of countermeasures is to prevent side-channel information leakages, such as shielding and obfuscation operations. For machine learning attacks, the increasing of non-linearity in PUFs can increase the difficulty of machine learning attacks, such as XOR operations. Side-channel attacks are usually combined with machine learning attacks to conduct more effective attacks for PUFs.
with higher accuracy and less attack time, where the complexity of machine learning attacks is reduced with side-channel information. In the paper, popular attack methods and corresponding countermeasures are introduced and summarized to help designers and users have an overall understanding of the security of PUFs.

2. Background of PUF

PUFs are based on manufacturing variations without precise control [1]. Silicon PUF is the main category in PUFs, which are implemented based on the process variations in integrated circuits and can be called chip fingerprint. With the same manufacturing process and parameters, each PUF instance with the same implementation in different chips has its unique outputs for the same inputs due to the process variations. The outputs and inputs of PUFs are called responses and challenges. Four essential metrics are applied to measure the quality of a PUF and judge whether a design can be applied as a PUF, including uniformity, uniqueness, reliability, and Strict Avalanche Criterion. The uniformity represents the proportion of 1 or 0 in all the responses measured with Hamming weight. The ideal value of uniformity is 50% and a larger deviation between actual uniformity and the ideal value means larger bias with lower uncertainty. Uniqueness represents the difference among different PUF instances and is measured with the average value of Hamming distances between any two response sets obtained from two PUF instances with the same challenge set. The ideal value of uniqueness is 50%, which illustrates that the responses are determined by the random process variations in chips. Reliability represents the stability of responses to the same challenges in multiple runs, which can be measured by the bit error rate with the ideal value 0. Strict Avalanche Criterion is a critical metric for security, which is used to measure the randomness in a PUF design reflecting the influence of one challenge bit on the response. Ideally, once one challenge bit is complementary, the response will change with a probability of 50% representing the response is generated randomly.

PUFs can be divided into weak PUFs and strong PUFs according to the number of challenge-response pairs (CRPs). The number of CRPs of a weak PUF is limited and has a polynomial relationship with the circuit size, which can be exhaustively tabulated by attackers if there is no restriction on access to the PUF. The limited number of CRPs causes weak PUFs can be attacked with a software clone. However, the number of CRPs of strong PUF has an exponential relationship with circuit size and it is impractical to traverse all CRPs. Thus, strong PUFs can not be broken by tabulation, but most of the existing strong PUFs can be broken by side-channel attacks and machine learning attacks. Arbiter PUF (APUF) [2] and RO PUF [3] are the classic strong PUF and weak PUF, the structures of which are shown in figure 1 and figure 2 respectively. PUFs can also be divided into delay-based PUFs and memory-based PUFs according to the operating mechanism. The delay-based PUFs utilize the delay differences in the PUF circuits to generate responses, such as APUF. Memory-based PUFs generate responses based on bistable memory components, such as SRAM PUF [4].

![Figure 1. The structure of Arbiter PUF.](image-url)
3. Conventional Side-Channel Attacks

3.1. Attack method

3.1.1. Power attack. Power attack is to obtain the leaked information of circuit state by detecting the change of power consumption or current trajectory during circuit state transition. It can be divided into simple power analysis and differential power analysis. Simple power analysis can be conducted for APUF, XOR APUF [5], and lightweight PUF [6]. The principle of the attack is to confirm the transition of the latch in the arbiter PUF based on the power trajectory, and the power trajectory can be obtained based on the measurement of the amount of current drawn from the power supply during the transition. Based on the power analysis, the status of the latch as the arbiter in an APUF can be judged, and then the corresponding output can be revealed.

For the XOR APUFs and lightweight PUFs composed of multiple parallel APUFs, simple power analysis can be used to determine the number of 0s and 1s stored in the latch or arbiter, but the specific output of an APUF can not be determined. Thus, different analysis methods need to be applied to attack the PUFs according to different structural characteristics [7]. For lightweight PUFs, a divide-and-conquer method is utilized where each of the parallel APUFs is attacked with selected CRPs after using simple power analysis to obtain the total number of 1s or 0s in all the responses. The selected CRPs make the outputs of all parallel APUFs be 1s or 0s to determine the output of each PUF. With the attack method, attackers can achieve high-precision modeling of each APUF, and then realize the high-accuracy attack for lightweight PUF [7]. For XOR APUFs, differential power analysis can be used to conduct an attack based on the comparison between the power trajectories obtained before and after the generation of a response, and the gradient optimization algorithm to obtain the fitted model with the smallest variance [8].

3.1.2. Fault attack. Fault attacks are conducted based on the fault information or behaviors. The reliability of actual PUFs can not reach 100% due to noise, such as thermal noise and environmental changes, and it can be used as the side-channel information in fault attacks. The reliability-based CMA-ES machine learning attack is an effective attack method for XOR APUFs, which uses the divide-and-conquer method to achieve the attack. Each APUF in an XOR APUF is attacked separately and the unreliability introduced by other APUFs is treated as noise. Thus, the addition of an APUF is considered as an addition of extra noise, which makes the relationship between the complexity of machine learning attacks and the number of APUFs change from an exponential relationship to a
linear relationship greatly reducing the difficulty of the attack and breaking the security of XOR PUFs [9].
If attackers can access the PUF and operate it, more fault information can be obtained to improve the fault attack by adjusting the power supply voltage or changing the environmental conditions, which is called fault injection. The increase in the proportion of unstable CRPs from active adjustment accelerates fault attacks and improves the accuracy of the attack.

3.1.3. Electromagnetic attack. Electromagnetic attacks are mainly applied to the PUFs based on oscillators, such as RO PUFs. The information about oscillator frequency is leaked from electromagnetic information. For the electromagnetic attacks of RO PUFs, the frequency amplitude spectrum can be calculated from the obtained electromagnetic trajectories. The frequency range of a RO is identified by the high difference peak in all the frequency amplitude differences, where a small frequency amplitude difference close to 0 represents no oscillation only with noise and a high frequency amplitude difference represents the occurrence of oscillation of a RO. The area with ROs placed can also be identified by the average frequency difference. The frequency of a RO is obtained by comparing the frequency amplitude spectrum where the same frequency with high frequency amplitude difference in two frequency amplitude spectrums corresponds to the RO which is used for the two corresponding comparisons to generate responses. The obtained frequency and the corresponding ROs can be combined to generate a model of the RO PUF [10].

3.2. Countermeasures

3.2.1. Power attack countermeasures. To prevent power attacks, we can balance the power for different outputs of the critical gates. For example, a dual-rail circuit [11] is used to charge and discharge the two data rails of each gate in each clock cycle. Each rail is charged equally, and it is completely discharged for each cycle regardless of the data. The dual-rail circuit involves pre-discharge and calculation cycles, and the overall capacitor charging and discharging is always a constant.
Besides, complementary operations can be applied to resist power attacks where two complementary circuits perform complementary operations. The complementary operations corresponding to two different outputs 0 and 1 are performed at the same time, so the total power corresponding to different actual outputs is consistent. For example, two symmetrical latches with two inverted outputs can be used to balance the power for APUFs.

3.2.2. Fault attack countermeasures. Based on the principle of fault attacks, fault detection can be applied to detect faults in time to prevent information leakage, and then improve the resistance to fault attacks, such as spatial redundancy and temporal redundancy. For the PUFs with restrictions on access and control of attackers, fault attacks can be prevented by limiting the number of reused challenges to reduce the available fault information. For example, if each challenge is only used once, attackers can not get the fault or reliability to conduct fault attacks.
Besides, different PUF designs have been proposed to overcome the issue of reliability with stronger resistance against reliability-based attacks, such as the MPUF [12] based on a single MUX and multiple APUFs as shown in figure 3. The rate of reliability decline of MPUF is lower than XOR APUFs as the number of APUF increases, which has less fault information and higher resistance to reliability-based attacks than XOR APUFs.
3.2.3. Electromagnetic attack countermeasures. To reduce the electromagnetic leakage to improve the resistance to electromagnetic attacks, transistor-level countermeasures can be used, such as transistor-level balance. However, transistor-level countermeasures increase the design cost and reduce the circuit performance. In the worst case, a large number of balanced units is needed for each critical component with careful placement and routing, which greatly increases the design cost. Another countermeasure for the electromagnetic attacks based on micro-probes can be achieved by preventing micro-probes from approaching the PUFs. Detecting package opening may be a solution, which is limited by the demand for special packaging materials and the greatly increased manufacturing cost. Another possible method is to install an active shield on or around the chip with PUFs, which is limited by the power required to drive the signal through the shield. Besides, an electromagnetic sensor can be designed to detect the proximity of the probe to resist electromagnetic attacks, such as a sensor based on an LC oscillator. When the probe is close to the PUF, the electrical coupling will occur between the probe and the nearby sensor, which will interfere with the original magnetic field of the sensor causing the proximity of the probe detected by the sensor [13].

Besides, the resistance to electromagnetic attacks can be improved through a special compact layout, such as placing adjacent RO chains according to sine waves and cosine waves to make the electromagnetic leakage of adjacent ROs overlap. The overlap makes the electromagnetic detector cannot distinguish the frequency of each RO, which reduces the vulnerability to electromagnetic attacks [14]. The special working modes of RO PUFs can also improve the resistance against electromagnetic attacks, such as only using a RO once or comparing all ROs at the same time to prevent the frequency of each RO from being distinguished. However, both the special working modes will increase hardware overhead.

4. Machine learning attack

4.1. Attack method

Machine learning attack is a powerful attack method, which can be divided into three types: white-box attack, gray-box attack, and black-box attack according to the attacker's knowledge. The black-box attacks are the simplest method based on the collected inputs and outputs, without the demands of the knowledge of the attacked party. However, the black-box attacks have the lowest attack accuracy due to the lack of information about the attacked party. On the contrary, white-box attacks require a full understanding of the internal structure and working conditions of the attacked party. Attackers need to model the attacked party and conducted machine learning attacks based on the model for white-box
attacks. White-box attacks have the highest attack complexity and accuracy. The complexity and accuracy of gray-box attacks are between the white-box attacks and black-box attacks.

Machine learning attacks are effective attack methods for strong PUFs and most of the existing PUFs are vulnerable to machine learning attacks, including Neural Network (NN), Logistic Regression (LR), Evolutionary Strategies (ES), and Support Vector Machine (SVM). For example, APUFs can be broken easily by machine learning attacks, especially by LR attacks with an attack accuracy larger than 99% [15].

4.2. Countermeasure
To improve the resistance of PUFs against machine learning attacks, the main countermeasure is to increase the difficulty of the attacks by increasing the nonlinearity in the PUFs. Many variants are proposed to improve the resistance, such as XOR PUFs [5] as shown in figure 4. The outputs of multiple identical PUFs are XORed to produce the final output in XOR PUFs with higher resistance to machine learning attacks. However, XOR PUFs are still vulnerable to machine learning attacks, such as LR attack [15] and reliability-based CMA-ES attack [9]. Besides, other variants, such as lightweight PUF, feedforward APUF [16] and interpose PUF (IPUF) [17], can also be broken by machine learning attacks. For example, the accuracy of LR attack on the 128-bit lightweight PUF with 5XORs can reach 99%, the accuracy of CMA-ES attack on the 128-bit feedforward APUF with 10 feedforward loops can reach 97% [18], and the accuracy of Deep Learning based model building attack on the 128-bit (4,4) IPUF can reach 97.68%. Therefore, the mathematical unclonability of PUFs is still a critical issue in security.

![Figure 4. The structure of XOR PUF.](image)

5. Conclusion
To ensure the security of a PUF design, it is necessary to consider the potential security threats of PUFs. It is important to understand the existing attacks and corresponding countermeasures for PUFs. The paper introduces the popular side-channel attacks, machine learning attacks, and corresponding countermeasures for PUFs as shown in table 1, which helps designers and users clarify the security issues and corresponding solutions of PUFs to work on the security of PUFs more effectively.

| Attack types          | Specific attacks | Countermeasures                                           |
|-----------------------|------------------|----------------------------------------------------------|
| Power attack          |                  | Balancing power, such as dual-rail circuit and complementary operations. |
| Side-channel attacks  | Fault attack     | Detecting fault, limiting access, improving reliability.  |
|                       | Electromagnetic attack | Transistor-level balance, detecting the probe, special compact layout, special working mode. |
| Machine learning attacks | NN, LR, ES, SVM  | Increasing nonlinearity, such as XOR and feedforward operations. |
Side-channel attacks and machine learning attacks are the two main attack methods for PUFs. Power attack, fault attack, and electromagnetic attack are three popular side-channel attacks and the basic principle of all the countermeasures is to prevent information leakage. NN, LR, ES, and SVM are four popular machine learning attacks for PUFs and the principle of countermeasures is to increase nonlinearity to increase the difficulty of machine learning attacks. Most of the existing PUFs lack resistance against machine learning attacks. Side-channel attacks and machine learning attacks are often combined to improve the effectiveness of the attacks, where side-channel information can reduce the complexity of machine learning attacks.

6. References
[1] V Rozic, B Yang, J Vliegen, N Mentens and I Verbauwhede, “The monte carlo PUF,” in 27th International Conference on FieldProgrammable Logic and Applications, FPL 2017, Ghent, Belgium, September 4-8, 2017 (M. D. Santambrogio, D. Göhringer, D. Stroobandt, N. Mentens, and J. Nurmi, eds.), pp. 1–6, IEEE, 2017.
[2] S Devadas, E Suh, S Paral, R Sowell, T Ziola and V Khandelwal, “Design and implementation of PUF-based ‘unclonable’ RFID ICs for anti-counterfeiting and security applications,” in Proc. IEEE Int. Conf. RFID, Apr. 2008, pp. 58–64.
[3] G E Suh and S Devadas, “Physical unclonable functions for device authentication and secret key generation,” in Proceedings of the 44th Design Automation Conference, DAC 2007, San Diego, CA, USA, June 4–8, 2007, pp. 9–14, IEEE, 2007.
[4] J Guajardo, S S Kumar, G J Schrijen and P Tuly, “FPGA intrinsic PUFs and their use for IP protection,” in Proc. Int. Conf. CHES, 2010, pp. 63–80.
[5] G E Suh and S Devadas, “Physical unclonable functions for device authentication and secret key generation,” in Proceedings of the 44th Design Automation Conference, DAC 2007, San Diego, CA, USA, June 4–8, 2007, pp. 9–14, IEEE, 2007.
[6] M Majzoobi, F Koushanfar and M Potkonjak, “Lightweight secure pufs,” in 2008 International Conference on Computer-Aided Design, ICCAD 2008, San Jose, CA, USA, November 10–13, 2008 (S R Nassif and J S Roychowdhury, eds.), pp. 670–673, IEEE Computer Society, 2008.
[7] Mahmoud A, Räuhrmair U, Majzoobi M and Koushanfar F: Combined modeling and side channel attacks on strong PUFs. IACR Cryptol. ePrint Arch. (2013).
[8] Räuhrmair U, et al.: Efficient power and timing side channels for physical unclonable functions. In: Batina, L., Robshaw, M. (eds.) CHES 2014. LNCS, vol. 8731, pp. 476–492. Springer, Heidelberg (2014).
[9] Delvaux J and Verbauwhede I: Side channel modeling attacks on 65 nm arbiter PUFs exploiting CMOS device noise. In: IEEE International Symposium on Hardware-Oriented Security and Trust, pp. 137–142. IEEE (2013).
[10] Merli D, Schuster D, Stumpf F and Sigl G: Semi-invasive EM attack on FPGA RO PUFs and countermeasures. In: Proceedings of the Workshop on Embedded Systems Security, pp. 1–9 (2011).
[11] Sokolov D, Murphy J, Bystrov A and Yakovlev A. Improving the Security of Dual-Rail Circuits (2004): 282-297.
[12] D P Sahoo, D Mukhopadhyay, R S Chakraborty and P H Nguyen, "A Multiplexer-Based Arbiter PUF Composition with Enhanced Reliability and Security," in IEEE Transactions on Computers(2018): 403-417.
[13] Naofumi H, Yu-ichi H, Noriyuki M, Daisuke F, Daichi T, Makoto N and Takafulmi A. EM Attack Is Non-invasive? - Design Methodology and Validity Verification of EM Attack Sensor. CHES(2014)
[14] Zhaojun L, Dongfang L, Hailong L, Mingyang G and Zhenglin L. An Anti-Electromagnetic Attack PUF Based on a Configurable Ring Oscillator for Wireless Sensor Networks. Sensors(2017).
[15] 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 2010, Chicago, Illinois, USA, October 4-8, 2010 (E. Al-Shaer, A D Keromytis, and V Shmatikov, eds.), pp. 237–249, ACM, 2010.

[16] J W Lee, D Lim, B Gassend, G E Suh, M van Dijk and S Devadas, “A Technique to Build a Secret Key in Integrated Circuits for Identification and Authentication Applications”, Symposium on VLSI Circuits Digest of Technical Papers, pp. 176-179, June 2004.

[17] P H Nguyen, D P Sahoo, C Jin, K Mahmood, U Rührmair and M van Dijk, “The interpose PUF: secure PUF design against state-of-the-art machine learning attacks.” IACR Trans.Cryptogr. Hardw. Embed. Syst., vol. 2019, no. 4, pp. 243–290, 2019.

[18] J Ye, Q Guo, Y Hu, H Li and X Li, "Modeling attacks on strong physical unclonable functions strengthened by random number and weak PUF," 2018 IEEE 36th VLSI Test Symposium (VTS), San Francisco, CA, 2018, pp. 1-6

Acknowledgments
This work was supported by the National Key Research and Development Program of China under Grant 2018YFB2202101. We also thank the editors and reviewers for their valuable comments.