A Machine Learning Resistant Strong PUF using Subthreshold Voltage Divider Array in 65nm CMOS

Abilash Venkatesh and Arindam Sanyal
Department of Electrical Engineering, University at Buffalo, Buffalo, NY, USA; {abilashv, arindams}@buffalo.edu

Abstract—**Physically Unclonable Functions (PUFs) are extensively used in hardware security blocks as key-generators and light-weight authentication. With recent advances in machine learning (ML), most existing PUFs are shown to be vulnerable to modeling attacks based on ML algorithms. We present a novel silicon strong PUF architecture that cascades three strong PUFs to implement a single strong PUF that is resistant to ML based modeling attacks. Designed in 65nm CMOS technology, the proposed PUF with 2^{60} challenge response pairs consume 0.43pJ/bit energy consumption from a power supply of 0.8V. The simulated inter-HD and intra-HD of the PUF are 0.5065 and 0.0696 respectively. When subjected to ML based modeling attacks, the prediction accuracy is 60% for logistic regression, artificial neural networking and support vector machine with nonlinear RBF kernel.**

I. Introduction

The dependence on electronic devices has proliferated in almost everyday activities making them easy targets and threatening the security and privacy of an individual or a group. The traditional practice of using a secret binary key stored in non-volatile memory (NVM) for authentication is less secure against hardware or software based attacks. In contrast to NVMs, a physical unclonable function (PUF) does not store a physical key but rather derives its unique signature from random variations. A silicon PUF exploits random variations introduced during IC fabrication process to generate a unique key, akin to fingerprint for the silicon PUF which can be used to uniquely identify it. The random variation introduced into each PUF is hard to predict and unknown even to the manufacturer, thus making it a useful tool in hardware security. PUFs implement unique complex functions that are very difficult to model mathematically. When interrogated with a challenge, a PUF generates a response which is unique to that PUF. These challenge response pairs (CRP) can be used to classify a PUF into 3 groups: strong PUF, weak PUF and controlled PUF [1]. Strong PUFs have CRPs which grow exponentially with the number of challenges while weak PUFs have CRPs which increase linearly with number of challenges. Controlled PUFs are strong PUFs whose CRPs are protected through a control logic block. Strong PUFs are mostly used for authentication purposes due to large number of CRPs, making it more complex to analyze as the attacker has access to only limited CRPs over a short amount of time [2].

The promise of using PUF for authentication due to its unique properties of *unclonability* and *unpredictability* is debatable in recent times due to various modeling attacks based on machine learning (ML) algorithms. While ML algorithms are not new, the recent advances in computing power has enabled mounting of complicated ML attacks on PUF cells. For a given number of CRPs the attack is successful when the PUF's complex functions are digitally cloned, providing high accurate predictions for the response developed through ML algorithms from unknown challenges. [2] shows that use of support-vector machine (SVM) attack on the well-known arbiter PUF [3] can predict PUF response with an accuracy $>$90% and the prediction accuracy improves as the attacker has access to more CRPs. [1] shows successful modeling attacks (prediction accuracy of 99%) on various PUF architectures using logistic regression (LR) algorithm. [4] shows arbiter PUF and 2-XOR arbiter PUF are broken through SVM model with accuracy $>$95% and $>$80% respectively. [5], [6] show an accuracy of $>$95% and $>$97% for prediction when ML modeling attacks based on evolution strategies (ES) were made against current-based PUFs and arbiter-PUFs respectively. [7], [8] employs artificial neural network (ANN) based ML modeling attack on feed forward PUFs and 64bit/128bit XOR PUFs resulting in prediction accuracy of $>$84% and $>$98% respectively. These cases show that PUFs can be broken through modeling attacks using ML algorithms.

In this work, we propose a strong PUF that can resist ML attacks. As shown later in the paper, when attacked with LR, ANN and SVM with nonlinear RBF kernel, the prediction accuracy of PUF response does not increase with increase in number of CRPs that are available to the attacker. The proposed strong PUF is based on a voltage divider array comprising of MOSFETs that operate in subthreshold region to exploit large random variation in threshold voltage. Three nominally identical voltage divider arrays are then cascaded to ensure the required amount of entropy in CRP, which makes the whole circuit immune to ML attacks. The rest of the paper is organized as follows: in Section II, the architecture of the proposed strong PUF circuit is presented. Section III presents simulation results which validate the proposed architecture. Finally, the conclusion is made in Section IV.

II. Proposed Architecture

The proposed cascaded strong PUF design is shown in Fig. 1. The unit PUF cell is designed using a CMOS inverter with the gate and drain shorted together allowing it to act as a voltage divider circuit. Each PUF cell is biased in weak inversion through a tail current source. Challenges are provided as inputs to each PUF cell which either connect/disconnect the PUF cell to the comparator input depending on the challenge

978-1-7281-0397-6/19/$31.00 ©2019 IEEE

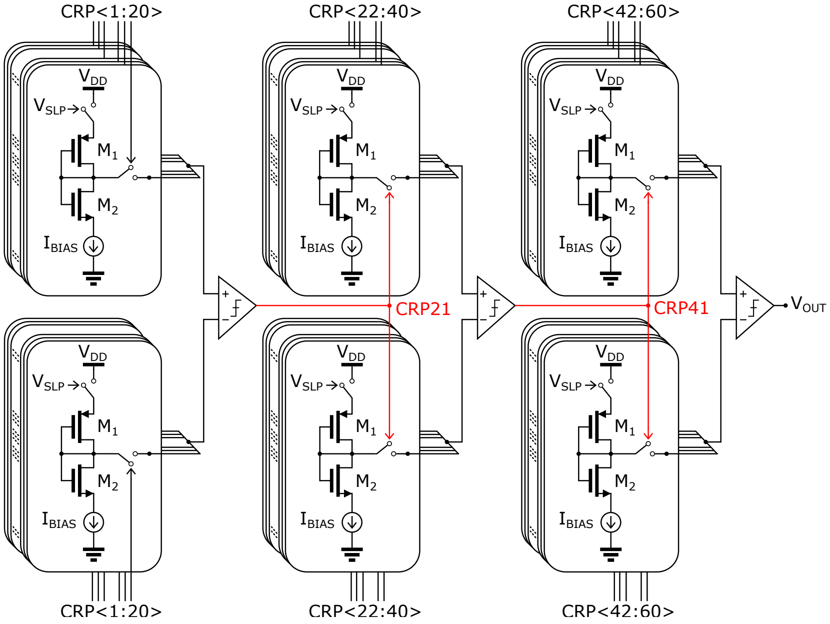

Fig. 1. Proposed cascaded strong PUF architecture

input being '1' or '0' respectively. Each PUF cell also has a switch V_{SLP} to disconnect it from power supply and transitioning the PUF into sleep mode to save power when the PUF is not in use. Each stage of the cascade is formed by connecting an array of 20 unit PUFs differentially to the two inputs of a comparator. Three such stages are cascaded to form the overall strong PUF with 2^{58} challenge inputs, with the first stage accepting 2^{20} external challenges and the other two stages accepting 2^{19} external challenges. The comparator output of the first stage provides the $20-th$ challenge input to the second stage and comparator output of the second stage provides the $20-th$ challenge input to the third stage.

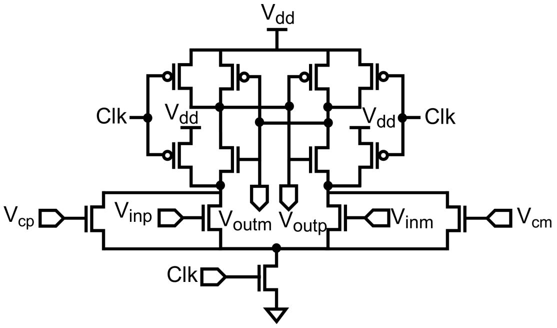

Fig. 2. Comparator schematic

A strong-arm latch is used as the comparator as shown in Fig. 2. Offset of the three comparators should be calibrated to ensure that the PUF has a normalized inter-HD close to 0.5. In this design, the comparator offset is calibrated through the pins V_{cp} and V_{cm} which control the bias voltage of two auxiliary input transistors as shown in Fig. 2. Simulation results indicate that upto 8mV of comparator offset can be canceled by varying the bias voltages V_{cp} and V_{cm}.

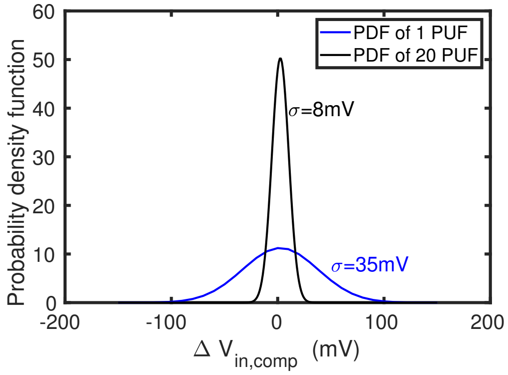

Fig. 3. PDF of differential voltage between PUF cells

Modeling attack accuracy reduces drastically with the increase in number of cascade stages [9]. In addition, each stage needs at least 7 unit PUF cells to improve resistance against ML attacks [9]. However, if there are many stages involved the design becomes more complex and hardware implementation requires more area. Also, increase in number of stages, each with its own comparator, increases the total comparator noise thus necessitating more energy consumption to improve PUF reliability. Keeping these considerations in mind, we have chosen 3 cascade stages with 20 unit PUFs in each stage.

Since each PUF cell is biased in subthreshold region, the component MOS transistors exhibit large variation in threshold voltages. The probability distribution of the differential voltages between two PUF units is shown in Fig. 3. Based on monte-carlo simulations across process and mismatch corners, voltage difference between two unit PUFs has a standard deviation of 35mV. For each cascade stage, the worst case condition in terms of intra-HD occurs when all 20 PUF cells are connected to the comparator inputs resulting in a reduced mismatch standard deviation of 8mV. Reliability of the PUF is degraded due to noise from the comparator and the unit PUF cells. Noise from PUF cell can be suppressed by reducing its bandwidth through capacitive loading. For this design, capacitive loading due to parasitics at the comparator input is sufficient to adequately suppress PUF noise. The comparator thermal noise is the dominant noise source for this design. If the PUF mismatch for a certain challenge input is smaller than comparator noise level, the response for that challenge will vary with time or be temporally unstable. Cascading of stages increases temporal instability and thus places greater constraints on allowable comparator noise.

Comparator noise depends on input common mode voltage V_{cmi} and the power consumed by it. Fig. 4 shows variation of comparator noise with V_{cmi}. Keeping the power consumption constant, the comparator noise can be lowered by reducing the V_{cmi} as shown in Fig. 4. The variation of comparator power with respect to the comparator input common mode voltage is shown in Fig. 5. Comparator power reduces with increase in V_{cmi} while noise increases with increase in V_{cmi}. Thus, there is a trade-off between PUF reliability and power and requires a judicial choice of V_{cmi}. For the present design, V_{cmi} of 520mV is chosen for a comparator power of 0.8μW and noise standard deviation of 600μV.

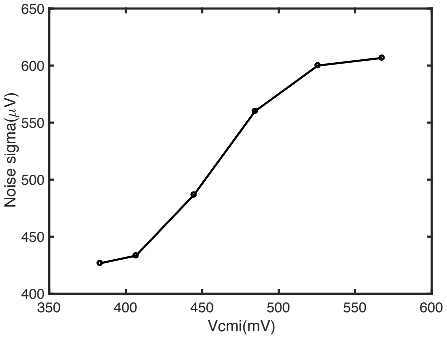

Fig. 4. Comparator noise vs V_{cmi}

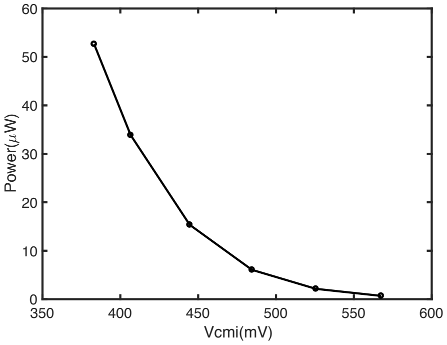

Fig. 5. PUF power vs V_{cmi}

For good reliability of PUF output, standard deviation of noise should be much smaller than standard deviation of random mismatch in PUF cells. Since the distribution of PUF mismatch and noise are gaussian, probability of a CRP being temporally stable can be written as

$$P = \left[1 - erf\left(\frac{\sigma_n}{\sigma_{mis}\sqrt{2}}\right)\right]^3 \quad (1)$$

where σ_n is the standard deviation of comparator noise, σ_{mis} is the standard deviation of random mismatch in each stage and the factor 3 comes from the fact that there are 3 stages in the design. For a target average reliability of 95% and σ_{mis} of 8mV, σ_n can be calculated to be 170μV. Since the comparator noise standard deviation is 600μV, we use averaging in the form of majority voting to reduce comparator noise. Averaging a random variable by n times reduces its standard deviation by $\sqrt{n}$. The comparator output has to be averaged 13 times to reduce its noise to 170μV. A counter clocked by comparator positive output is used to implement majority voting of comparator output. We choose a majority voting of 15 in this design, as for this case the MSB of the counter can be directly used as the averaged comparator output without additional hardware.

III. Simulation Results

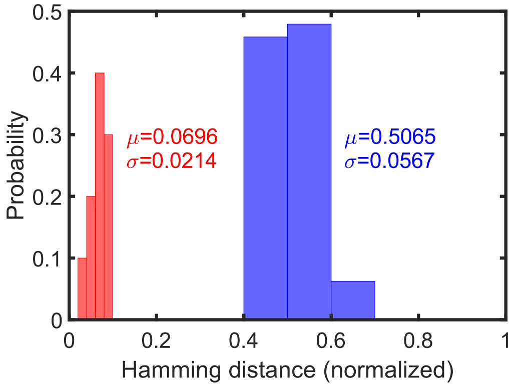

Fig. 6. Normalized hamming distance

The inter-HD and intra-HD histograms are shown in Fig. 6. For intra-HD simulations, the supply voltage is varied from 750mV to 900mV and the temperature is varied from $-20°$C to $85°$C. Intra-HD simulation is performed for 1000 challenges and averaged over 2 different PUF instances. For inter-HD simulation, 50 monte-carlo runs of 1000 challenges at $27°$ C and 0.8V is performed. The normalized intra-HD has a mean of 0.0696 and standard deviation of 0.0214, while the normalized inter-HD has a mean of 0.5065 and standard deviation of 0.0567.

We have used logistic regression, artificial neural network (ANN) and SVM with RBF kernel for implementing ML modeling attacks on the proposed PUF. These three algorithms are widely used for testing resistance of strong PUF against modeling attacks. Open source python package *scikit-learn* is used for logistic regression and ANN [14] and open source libraries LIBLINEAR and LIBSVM are used for SVM [15]. We have set inverse regularization strength and penalty parameter

TABLE I
COMPARISON WITH STATE-OF-THE-ART PUFS

	This work	[10]	[11]	[12]	[3]	[13]
Technology(nm)	**65**	130	90	40	180	28
Type of PUF	**Strong**	Strong	Strong	Strong	Strong	Strong
Possible CRPs	**1.15 x 10^{18}**	$\approx 3.7 \times 10^{19}$	523776	$\approx 5.5 \times 10^{28}$	$\approx 1.4 \times 10^{20}$	1.17×10^{11}
ML attack accuracy for 10^4 CRPs	**60%**	60%	99%	–	99%	89.4%
Energy/bit (pJ/bit)	**0.43**	11	–	17.75	–	0.097
Voltage range (V)	**0.75–0.9**	1.08 – 1.32	1.08 – 1.2	0.7 – 1.2	1.75 – 1.85	0.5 – 0.9
Temperature range (°C)	**-20 to 85**	-20 to 80	20 to 120	-25 to 125	20 to 70	0 to 80
Inter-HD	**0.5065**	0.499	0.4615	0.5007	0.4	0.481-0.495
Intra-HD	**0.0696**	0.058	0.0048	0.0101	0.0357	0.0317

to 0.01 for optimizing LR and SVM algorithms respectively. ANN attack was performed using a hidden layers consisting of 50 neurons with 1000 iterations at learning rate of 0.001. Fig. 7 shows the prediction accuracy for all ML modeling attacks to be around 60% for the proposed design and the prediction accuracy does not improve beyond 60% as the number of CRPs are increased. This indicates a strong resilience of the proposed PUF against ML attacks.

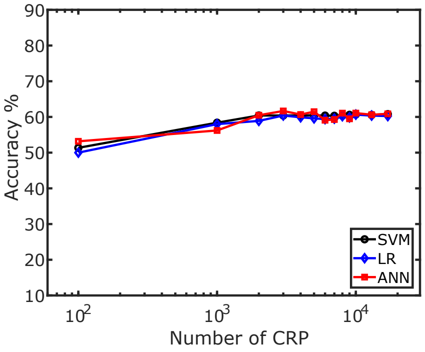

Fig. 7. Prediction accuracy for ML attacks on proposed design

In order to more completely investigate the ML resistance characteristic of the proposed PUF, we adopted 5-fold cross-validation technique for SVM based ML attack in which for the same number of CRPs, the training and test set are chosen randomly 5 times. Fig. 8 shows the results of the cross-validation simulation. The prediction accuracy remained within 58%-61% over the different cross-validation folds. Principal component analysis (PCA) is performed with 1000 CRP. The data set provided to the ML algorithms are visualized through PCA while converting the input challenges to 2-dimension as principal component 1 & 2 while the output response is mapped to a binary 0 or 1 shown in Fig. 9. No defined clusters are formed due to the randomness of the designed PUF's CRP. This randomness and the non linearity of the design results in robustness of the proposed PUF design against ML based modeling attacks. Table I compares the proposed work with state-of-the-art PUFs. The proposed PUF has similar robustness against ML attacks as [10] but with 20X better energy-efficiency. The proposed PUF has a high worst-case bit-error rate (BER) of 14.75% but this can be reduced by either removal of worst-case CRPs like in [10] or by reducing comparator noise at the cost of increased power consumption.

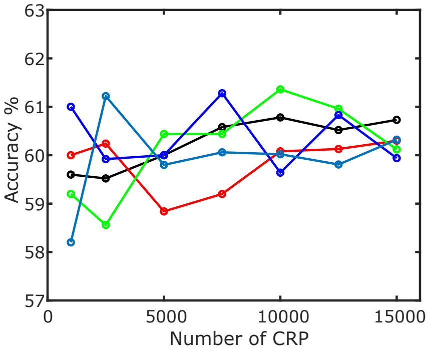

Fig. 8. Prediction accuracy of SVM based ML attacks for five-fold cross-validation

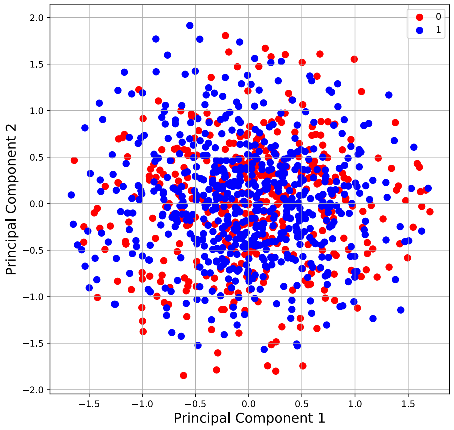

Fig. 9. PCA on 1000 CRP

IV. CONCLUSION

In this work, we have presented a novel, cascaded strong PUF using voltage divider arrays that can resist ML based modeling attacks limiting the prediction accuracy to 60% even with large CRP data. The proposed PUF consumes an energy of 0.43pJ/bit which compares favorably to state-of-the-art strong PUFs.

REFERENCES

[1] U. Rührmair, J. Sölter, F. Sehnke, X. Xu, A. Mahmoud, V. Stoyanova, G. Dror, J. Schmidhuber, W. Burleson, and S. Devadas, “PUF modeling attacks on simulated and silicon data,” in *IEEE transactions on information forensics and security*, vol. 8, 2013, pp. 1876–1891.

[2] S. S. Zalivaka, A. A. Ivaniuk, and C.-H. Chang, “Low-cost fortification of arbiter PUF against modeling attack,” in *IEEE International Symposium on Circuits and Systems (ISCAS)*, 2017, pp. 1–4.

[3] 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,” in *IEEE Symposium on VLSI Circuits*, 2004, pp. 176–179.

[4] G. Hospodar, R. Maes, and I. Verbauwhede, “Machine learning attacks on 65nm arbiter PUFs: Accurate modeling poses strict bounds on usability,” in *IEEE international workshop on Information forensics and security (WIFS)*, 2012, pp. 37–42.

[5] R. Kumar and W. Burleson, “Hybrid modeling attacks on current-based PUFs,” in *IEEE 32nd International Conference on Computer Design (ICCD)*, 2014, pp. 493 – 496.

[6] G. T. Becker, “On the Pitfalls of Using Arbiter-PUFs as Building Blocks,” in *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 34, 2015, pp. 1295 – 1307.

[7] A. O. Aseeri, Y. Zhuang, and M. S. Alkatheiri, “A Machine Learning-Based Security Vulnerability Study on XOR PUFs for Resource-Constraint Internet of Things,” in *IEEE International Congress on Internet of Things (ICIOT)*, 2018, pp. 49 – 56.

[8] M. S. Alkatheiri and Y. Zhuang, “Towards fast and accurate machine learning attacks of feed-forward arbiter PUFs,” in *IEEE Conference on Dependable and Secure Computing*, 2017, pp. 181 – 187.

[9] A. Vijayakumar, V. C. Patil, C. B. Prado, and S. Kundu, “Machine learning resistant strong PUF: Possible or a pipe dream?” in *IEEE International Symposium on Hardware Oriented Security and Trust (HOST)*, 2016, pp. 19 – 24.

[10] X. Xi, H. Zhuang, N. Sun, and M. Orshansky, “Strong subthreshold current array PUF with 2^{65} challenge-response pairs resilient to machine learning attacks in 130nm CMOS,” in *IEEE Symposium on VLSI Circuits*. IEEE, 2017, pp. C268–C269.

[11] G. E. Suh and S. Devadas, “Physical unclonable functions for device authentication and secret key generation,” in *Proceedings of the 44th annual design automation conference*. ACM, 2007, pp. 9–14.

[12] K. Yang, Q. Dong, D. Blaauw, and D. Sylvester, “A physically unclonable function with BER $<10^{-8}$ for robust chip authentication using oscillator collapse in 40nm CMOS,” in *IEEE International Solid-State Circuits Conference-(ISSCC)*, 2015, pp. 1–3.

[13] S. Jeloka, K. Yang, M. Orshansky, D. Sylvester, and D. Blaauw, “A sequence dependent challenge-response PUF using 28nm SRAM 6T bit cell,” in *IEEE Symposium on VLSI Circuits*, 2017, pp. C270–C271.

[14] “scikit-learn: Machine learning in python. [online],” available at http://scikit-learn.org/stable/.

[15] Chang, Chih-Chung, Lin, and Chih-Jen, “LIBSVM: A library for support vector machines,” *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011, software available at http://www.csie.ntu.edu.tw/∼cjlin/libsvm.