Determining the Quality Metrics for PUFs and Performance Evaluation of Two RO-PUFs

Giray Kömürcü
National Research Institute of Electronics and Cryptology,
TÜBİTAK, 41470, Kocaeli, TURKEY
Email: girayk@uekae.tubitak.gov.tr

Günhan Dündar
Bogazici University, Dept. of Electrical and Electronics Eng.
34342 Bebek, Istanbul, Turkey
Email: dundar@boun.edu.tr

***Abstract*—Physical Unclonable Functions (PUFs) are circuit primitives that generate chip specific signatures depending on the uncontrollable components present in the manufacturing process. Authentication, key generation and IP protection are three important usage areas of PUF circuits. Beside unclonability, uniqueness and robustness are the main properties that every PUF should provide. Although a number of PUF types are presented in the literature, standard and satisfactory performance evaluation metrics for these properties or testing methodologies have not been proposed yet. In this work a complete set of quality metrics have been developed for a fair and detailed performance evaluation of PUFs. Secondly, a testing methodology is proposed and confidence interval and confidence level concepts are adopted to PUF evaluation in order to maintain the reliability of the results. We have implemented two Ring Oscillator(RO) based PUF circuits on FPGA and evaluated their performance in varying environmental conditions in detail, according to the quality metrics that are proposed.**

***Keywords*-PUF, Physical Unclonable Functions, Uniqueness, Robustness, Ring Oscillator, confidence level, quality metrics, FPGA.**

I. Introduction

PUF was first introduced by Pappu et al. in 2001 [1]. These functions have the unique capability of generating chip specific signatures during operation. Their unclonability is a result of uncontrollable components present in the manufacturing process such as threshold voltage and doping concentrations. Since it is impossible to replicate these process variations for another die, the generated signature is unique and chip specific.

Although optical PUFs were the first structures presented [1], silicon PUFs drew more attention with less fabrication cost and easy integration with integrated circuits. Ring Oscillator (RO) type PUFs, which depend on the delay differences of identical structures, were first presented by Gassend et. al in 2002 [2], [3]. In regular RO PUFs, the output basically depends on the oscillation frequencies of two ring oscillators with the same number of identical delay elements. By using two ROs, one bit response is generated. RO structures generally suffer from high power consumption and speed limitations. However, they are strong in terms of robustness with less sensitivity to environmental variations.

Arbiter type PUF structure was presented by Lim et. al [4], [5] based on the differing timing behaviour of elements on chips [6]. In arbiter PUFs, a number of delay elements that construct two parallel paths are connected serially and a rising signal is applied to both paths synchronously. At the end of these lines, an arbiter decides which signal passed the lines faster and outputs a 1 bit response. Arbiter type PUFs suffer from sensitivity to environmental variations, modeling attacks and symmetrical routing requirement. Their strong side is fast bit generation capability and exponential number of challenge response pair support. In addition to RO and arbiter type PUFs, there are also SRAM PUFs, glitch PUFs, sense amplifier PUFs and reconfigurable PUFs that again depend on the process variations of integrated circuits.

Even though quite a number of different PUF structures exist and results are presented in the literature, no detailed evauation of robustness and uniqueness is constituted. Works presented on performancce evaluation by Hori et. al and Maiti et. al [7], [8] defines quality metrics but evaluates robustness and uniqueness with the straight forward approach. This prevents comparing the PUFs and choosing the best fitting structure for a specific application. Similarly, a testing methodology for PUF circuits is not defined and each PUF is tested with a different set of parameters, again preventing a meaningful comparison of circuits.

In this work, our contribution is threefold. Firstly, we focus on developing a set of metrics with a solid background for a fair evaluation of the structures that are already developed and that will be presented in the future. For each key property of PUFs, more than one quality metric is determined. Next, a testing methodology is determined and the PUF results are presented with a confidence level and confidence interval to show their reliability. Finally, two PUF implementations realized on FPGA circuits are described and their performance evaluation results are presented with the set of quality metrics that are proposed in our work.

II. Derivation of Quality Metrics for Uniqueness

Uniqueness is inter-chip variation of PUFs. In the ideal case, all PUF outputs from different chips should be uniformly distributed and statistically independent. If the set of measurements are statistically independent, their Hamming distances (HD) would be 50%. This quality measure, named as U_QM1, is used widely in the literature and can be calculated as shown

in Equation 1.

$$U\_QM1 = \frac{2}{k(k-1)} \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \frac{HD(Ri, Rj)}{n} * 100\% \quad (1)$$

$$U\_QM2 = Corr(DIS\_HD, Gaus(Mn(HD\_PUF), \sigma)) \quad (2)$$

Even if the quality metric U_QM1 stated above gives information about the performance of the system, it does not guarantee uniform distribution since non-uniform data may also produce 50% HD. Two qualitatively different performing PUFs may be evaluated as the same if the first quality metric is used as the only performance parameter. In a uniformly distributed set of outputs, their Hamming distances will be distributed according to Gaussian distribution. At this point, we propose defining another quality metric to evaluate the uniqueness of PUFs. This second quality metric U_QM2 should check how Gaussian the distribution of Hamming distances is. This is calculated via (2) correlating the HD distribution of PUF data (DIS_HD) with the ideal Gaussian distribution. Standard deviation and mean of the ideal Gaussian distribution is also the standard deviation and mean of HD's of the collected data (HD_PUF) from the real implementation. The closer result is to 1, the more the distribution is Gaussian; hence the circuit is better performing in terms of uniqueness.

In [9], Gilbert-Varshamov bound (GVB) is used to determine the security of PUF outputs against exhaustive search attacks. This is achieved via calculating the minimum Hamming distance between two random outputs within a uniformly distributed set of outputs. We claim that this bound can be used to determine the uniqueness of the structure as well. After collecting a certain number of data, the minimum distance *dHm* among them is calculated and *R'* is determined via (3) and (4). Ideal *R* is calculated via (5) using the number of measurments *M* and PUF length *N*. Proportion of *R'* to the ideal *R* can serve as a quality metric, U_QM3, for uniqueness too as shown in (6). If the outputs are uniformly distributed, meaning uniqueness is ideal, minimum HD is compatible with the GVB and U_QM3 converges to unity. Otherwise the minimum HD is worse than the bound states and U_QM3 is less then 1.

$$\underline{d}_{Hm} = \frac{d_{Hm}}{N} \quad (3)$$

$$R' \leq 1 - H2(\underline{d}_{Hm}) =$$
$$1 + \underline{d}_{Hm} \log_2(\underline{d}_{Hm}) + (1 - \underline{d}_{Hm}) log_2(1 - \underline{d}_{Hm}) \quad (4)$$

$$R = \frac{log_2 M}{N} \quad (5)$$

$$U\_QM3 = \frac{R}{R'} \quad (6)$$

In addition to these, U_QM3 and GVB can be used to determine the number of circuits that can be identified with a previously set security level and a certain length PUF output. Similarly the required PUF length can be determined for a previously set security level and number of chips to be identified using the U_QM3. Here, the security level is determined by the user and means the minimum distance between two device IDs generated by PUF.

For this purpose, U_QM3 is calculated. Then, a security level is set as *dHm* which is the minimum distance percentage between two PUF outputs in the system. (3) is calculated again with this *dHm* and result is multiplied by U_QM3 and *R* is determined. Finally via (4) the designer can either set the number of PUF bits *N* and calculate the maximum number of circuits *M* that can be identified or set the number of circuits and calculate the minimum length of PUF output.

III. Derivation of Quality Metrics for Robustness

Robustness is the intra-die variation that should be ideally zero for best performing PUF circuits. However, due to environmental variations and internal characteristics of structures, some bits of the output may differ from measurement to measurement. Robustness is measured by taking a number of measurements from a single IC and calculating the mean error rate, R_QM1, in the previous works on subject. Mean error rate is calculated as shown in (7).

$$R\_QM1 = \frac{1}{x} \sum_{y=1}^{x} \frac{HD(Ri, R'i, y)}{n} * 100\% \quad (7)$$

Since some of the systems that use PUF outputs, require error free data, Error Correction Codes are used to generate the same output at every measurement. The complexity, hence the cost of error correction codes depend on the maximum number of erroneous bits they can recover. Thus, we claim that mean error rate is not critical for such systems and maximum error rate within a certain number of measurements named as R_QM2 can be presented as another quality metric for robustness as shown in (8).

$$R\_QM2 = \max \frac{HD(Ri, R'i, y)}{n} \quad (1 \leq y \leq \#ofmeas.) \quad (8)$$

Another set of data presented in [10] is the distribution of errors on response bits. This data is used to mask the most erroneous bits and calculate the improved error rate. This approach may be helpful in practice if it is convenient to detect the most problematic bits in each circuit and eliminate them in each measurement afterwards. Thus, error reduction rate with masking a certain number of bits (3 in this case) may serve as another quality metric, R_QM3, for robustness in PUF circuits. This is calculated as shown in (9), where R'_QM1 represents the mean error rate after masking the most erroneous 3 bits.

$$R\_QM3 = \frac{R\_QM1 - R'\_QM1}{R\_QM1} \quad (9)$$

A common method to improve the robustness of PUF circuits is majority voting [10], [11]. Each bit of output is generated via majority voting. This method increases the robustness especially in normal operating conditions. Thus,

mean error rate after majority voting, R_QM4, can serve as an important quality metric as well. This is calculated as shown in (10), where the R"_QM1 represents the mean error rate after majority voting for 3 times.

$$R\_QM4 = \frac{R\_QM1 - R"\_QM1}{R\_QM1} \tag{10}$$

Stable bit count, the bits that generate the same output at each measurement, is also an important parameter. If stable bits are selected and used, the need for error correction codes is eliminated. Therefore, stable bit count, R_QM5, for a PUF structure can serve as a quality metric for robustness. Since PUF outputs are very vulnerable to changes in the environment, we will present the results according to the quality metrics both at Normal Operating Conditions (NOC) and at Varying Temperature (VT).

IV. Confidence Interval Concept for PUF Evaluation

In previous sections, quality metrics that will be calculated from the measurements taken were presented. However, the number of measurements to be taken for a reliable performance evaluation is still questionable. Thus confidence interval and confidence level concepts are adopted to PUF performance evaluation in order to present the trustworthiness of the results as well. In this method, confidence level and confidence interval is set to determine the number of measurements that have to be taken. This is calculated via Chebyshev inequality [12] as shown in (11)-(14). As the number of measurements increases, confidence level is increased and/or confidence interval is diminished. For instance 99.9% confidence within 0.1% confidence interval can be achieved by using 1000 measurements whereas 25 measurements only provide 95% confidence level within 2% confidence interval. Relation between the number of measurements, confidence interval, confidence level and the standard deviation of measurements is presented in Figures 1 and 2. As seen from the figures, confidence level of the results increases as the number of measurements increase. Confidence interval is ten times wider for the first figure, ensuring a high confidence level with fewer measurements.

$$P[M_n(X) - c \leq \mu x \leq M_n(X) + c] = \tag{11}$$

$$P[\frac{-c}{\sigma\chi/\sqrt{n}} \leq \frac{M_n(X)}{\sigma\chi/\sqrt{n}} \leq \frac{+c}{\sigma\chi/\sqrt{n}}] = \tag{12}$$

$$1 - 2Q(\frac{c\sqrt{n}}{\sigma\chi}) = \tag{13}$$

$$Q(A) = 1 - 2\phi(A) \tag{14}$$

V. Implementation of two basic RO PUF structures

In order to evaluate the performances of previously presented PUF structures according to our new set of quality metrics, two RO based structures are implemented on FPGA that are presented by Gassend et.al. in [3]. In both designs, the same ring oscillator structure that is composed of 4 inverter stages and 1 nand stage that enables optional oscillation is used as shown in Fig. 3. To maintain equal wire loads and hence minimize the systematic variation, RO is built as hard macro. 1 bit of PUF output is generated by comparing the oscillation frequencies of two RO's. Two counters are connected to outputs of the two RO structures and a limit to the counters are set to a value such as 1024. The counter that reaches to its limit first raises a flag. The bit is set to 0 if the first counter reaches the limit first and bit is set to 1 if the second counter reaches the limit first as shown in Fig. 4. In the first structure, n+1 ROs are implemented. Each 2 RO that are placed next to each other are compared to generate 1 bit output. This structure generates n bit outputs by n+1 ROs. In the second structure, 2n ROs are used to generate n bit outputs. In this structure each RO is used only once and again beside ROs are used to generate each bit.

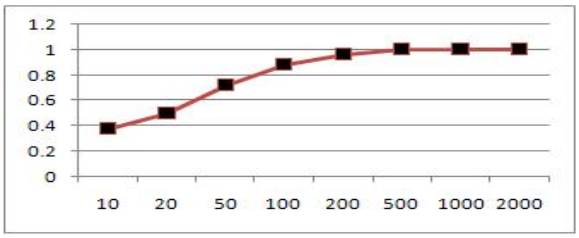

Fig. 1. Number of meas. - confidence level relation with confidence interval of 0.01 and standard deviation of 0.064

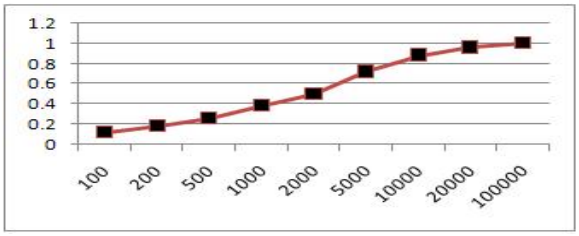

Fig. 2. Number of meas. - confidence level relation with confidence interval of 0.001 and standard deviation of 0.064

VI. Analysis of Experimental Data

In the system we have set up, Xilinx 3S5000 is used and outputs are collected via Matlab. The number of outputs that will be collected is determined by adopting confidence interval approach to the PUF structures. For uniqueness, 95% confidence is achieved within 2% confidence interval with 25 measurements. Uniqueness measurements are done by mapping the PUF structure to different parts of FPGA since we did not have enough number of chips to use for measurement. For robustness 1000 outputs are used providing 99.9% confidence within 0.1% confidence interval. Robustness is measured both at normal operating conditions (NOC) and at varying temperature (VT). For VT, 1000 measurements are taken each at 0, 20, 40, 60, 80 and 100 C^o.

Uniqueness and robustness results of two PUF structures are presented in table 1. For uniqueness, RO_PUF2 seems better

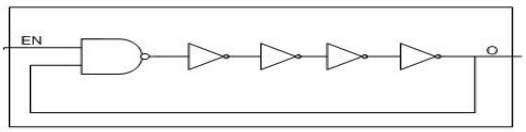

Fig. 3. Ring oscillator structure.

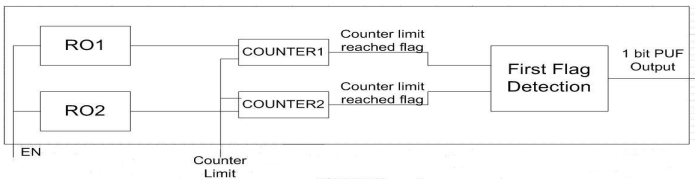

Fig. 4. PUF output bit generation.

performing than RO_PUF1 slightly in terms of all quality measures we have determined. This result is expectable since each RO is used only once in RO_PUF2, whereas each RO is used twice in RO_PUF1. This decreases the entropy of the system and hence the uniqueness.

For robustness according to R_QM1 under NOC, error rate of RO_PUF1 is 0.8% and error rate of RO_PUF2 is 1.3%. If the temperature changes, error rates are almost tripled for both of the implementations. R_QM2 states that the maximum error rate is 3.9% for both structures. According to R_QM3, bitwise masking reduces the errors significantly for both structures. Effect of majority voting is also presented as R_QM4. Robustness increases significantly if the majority voting is applied at NOC. However, its effect diminishes as the temperature varies in the system. Finally, according to R_QM5, 85-88% of the bits are stable for the two structures even the temperature changes.

The number of circuits that can be identified with a certain security level and PUF bits is presented in Table 3 by making use of U_QM3. In addition to this, the required number of PUF bits is calculated with a certain security level and number of chips to be identified is also presented. RO_PUF2 enables to identify more chips than RO_PUF1 with the same PUF length and at the same security level. For instance, at a security level of 0.2, RO_PUF1 identifies roughly 900000 circuits, whereas RO_PUF2 identifies 5800000 circuits. Similarly, fewer PUF bits are enough to identify 10000000 chips by using RO_PUF2. It is obvious that this quality metric will help system designers to choose the best fitting PUF securely.

VII. Future Work and Conclusion

We have introduced a complete set of quality metrics for the robustness and uniqueness properties of PUF and adopted the confidence interval and confidence level concepts to PUF performance evaluation for the first time. Two RO type PUFs from the literature are implemented on FPGA and evaluated according the set of quality metrics proposed. Our future work will focus on performance evaluation of other PUF structures.

TABLE I
Uniqueness and Robustness results of RO_PUF1 and RO_PUF2.

Uniqueness Analysis	**Time per bit μs**	**# of Meas.**	**Conf. Int.**	**Conf. Level**
RO_PUF1	81,92	25	2	96
RO_PUF2	81,92	25	2	96,6
Metrics	**U_QM1**	**U_UM2**	**U_QM3**	
RO_PUF1	49,05	0,92	0,558	
RO_PUF2	49,55	0,94	0,631	
Robustness Analysis	**Time per bit μs**	**# of Meas.**	**Conf. Int.**	**Conf. Level**
RO_PUF1	81,92	1000	0,1	99,9
RO_PUF2	81,92	1000	0,1	99,9
Metrics	**R_QM1 at NOC**	**R_QM1 at VT**	**R_QM2**	**R_QM3**
RO_PUF1	0,89	2,63	3,9	1,4
RO_PUF2	1,31	3,65	3,9	2,4
Metrics	**R_QM4 at NOC**	**R_QM4 at VT**	**R_QM5 at NOC**	**R_QM5 at VT**
RO_PUF1	0,77	2,55	92,18	85,15
RO_PUF2	1,17	3,62	92,96	88,06

TABLE II
Relation between the number of circuits to be identified, PUF length and security level.

U_QM3	**PUF length**	**Security Level (min. HD)**	**Max. IC to identify**
RO_PUF1	128	0,2	912938
RO_PUF2	128	0,2	5809570
RO_PUF1	128	0,3	357
RO_PUF2	128	0,3	776
U_QM3	**# of IC to identify**	**Security Level (min. HD)**	**Required PUF length**
RO_PUF1	10000000	0,2	149
RO_PUF2	10000000	0,2	132
RO_PUF1	10000000	0,3	351
RO_PUF2	10000000	0,3	310

References

[1] R. S. Pappu, "Physical one-way functions," *Ph.D. dissertation*, 2001.

[2] B. Gassend, D. Clarke, M. V. Dijk, and S. Devadas, "Silicon pysical random functions," in *Proceedings of the Computer and Communications Security Conference*, 2002.

[3] B. Gassend, D. Clarke, M. Dijk, and S. Devadas, "Controlled physical random functions," in *18th Annual Computer Security Applications Conference*, 2002.

[4] D. Lim, J. Lee, B. Gasend, G.E.Suh, M. V. Dijk, and S. Devadas, "Extracting secret keys from integrated circuits," *IEEE Transactions on VLSI Systems*, 2005.

[5] B. Gassend, D. Clarke, M. V. Dijk, and S. Devadas, "Delay-based circuit authentication and applications," in *ACM Symposium on Applied Computing*, 2003.

[6] B. Gassend, "Physical random functions," *Master Thesis*, 2003.

[7] Y. Hori, T. Yoshida, A. Satoh, and T. Katashita, "Quantitative and statistical performance evaluation of arbiter physical unclonable functions on fpgas," *Reconfigurable Computing and FPGAs (ReConFig)*, pp. 298–303, 2010.

[8] A. Maiti, P.Schaumont, and V. Gunreddy, "A systematic method to evaluate and compare the performance of physical unclonable functions," *IACR ePrint*, vol. 657, 2011.

[9] D.E.Lazich and M.Wuensche, "Protection of sensitive security parameters in integrated circuits," *LNCS*, no. 393, pp. 157–178, 2008.

[10] D.Suzuki and K.Shimizu, "The glitch puf: A new delay-puf architecture exploiting glitch shapes," in *CHESS*, 2010.

[11] M. Majzoobi and F. Koushanfar, "Techniques for design and implementation of secure reconfigurable pufs," *ACM Transactions on Reconfigurable Technology and Systems*, vol. 2, no. 1, 2009.

[12] M. Abramowitz and I. A. Stegun, in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York: Dover, 1972, p. 11.