Quality Metric Evaluation of a Physical Unclonable Function Derived from an IC's Power Distribution System

Ryan Helinski
University of New Mexico
Albuquerque, NM
helinski@unm.edu

Dhruva Acharyya
Verigy Inc.
Cupertino, CA
dhruva.acharyya@verigy.com

Jim Plusquellic
University of New Mexico
Albuquerque, NM
jimp@ece.unm.edu

ABSTRACT

The level of security provided by digital rights management functions and cryptographic protocols depend heavily on the security of an embedded secret key. The current practice of embedding the key as digital data in the integrated circuit (IC) weakens these security protocols because the keys can be learned through attacks. Physical unclonable functions (PUFs) are a recent alternative to storing digital keys on the IC. A PUF leverages the inherent manufacturing variations of an IC to define a random function. Given environmental variations such as temperature and supply noise, PUF quality criteria such as reproducibility and the level of randomness in the responses may be difficult to achieve for a given PUF circuit architecture. In this paper, we evaluate a PUF derived from the power distribution system of an IC with regard to a set of quality metrics including single-bit and collision probability and entropy. The analysis is carried out using data obtained from 36 chips fabricated in IBM's 65 nm SOI technology.

Categories and Subject Descriptors

K.6.5 [**Management of Computing and Information Systems**]: Security and Protection -- *Authentication.*

General Terms

Security

Keywords

Hardware security, unique identifier, process variations

1.INTRODUCTION

Secret keys define the basis of many hardware security protocols. The traditional approach to giving ICs a unique identifier is to embed a predefined key in a eFUSE or ROM immediately after the IC is manufactured. This method works well for chip IDs, since every key is controlled and is therefore truly unique. However, ROM-based keys can be learned through attacks, and therefore, are not well suited for security protocols.

Physical Unclonable Functions (PUF) address this problem by exploiting process variations in order to define a function $R = f(C)$, which maps a challenge C to a response R [1]. Since the exact mapping of C to R is driven by the random nature of process variations, such a function can be used to generate a set of unique identifiers for each IC. PUFs are realized by special on-chip circuits that are intentionally designed to be sensitive to process variations. For example, an arbiter PUF is proposed that *races* two edges along nearly identical paths, and defines the response bit as a function of the relative path delays [2]. Unfortunately, such designs also show significant sensitivity to environmental variations such as temperature, supply noise, coupling effects, etc. and their specific implementations may not produce responses that are truly random. Therefore, it is necessary to characterize the responses of a PUF in order to determine its overall quality.

In this paper, we evaluate a previously proposed PUF that is defined using the resistance variations in the power distribution system of an IC [3]. We evaluate several quality criteria of the PUF, on a set of 36 chips fabricated in IBM's 65 nm SOI process technology, including the randomness of the responses and their reproducibility. Randomness relates to the uniqueness of the function and specifies the level of probability that the function will have the same mapping on different ICs. Reproducibility relates to the integrity of the function under different environmental variations, such as temperature and supply voltage variations.

The organization of this paper is as follows. A brief background is presented in Section 2. The PUF is described in Section 3. Sections 4 through 6 provides a quality metric analysis of the hardware derived PUF's responses. Section 7 concludes.

2.BACKGROUND

PUFs have been proposed to replace ROM-based storage of secret keys. Most proposed PUF implementations fall into the following categories: SRAM power-on patterns [4], mis-matched delay lines [2] or Ring Oscillators (RO) [5,6], MOS drive current mismatch [7], and leakage currents [8]. All PUFs are subject to environmental variations, but some implementations are more sensitive than others.

Two PUFs based on variations in the Power Distribution System (PDS) and the corresponding experimental design and setup are described in [3]. The PUF is implemented using Stimulus/Measure Circuit (SMC) hardware primitives, that enable measurements of PDS resistance variations. It consists of a shorting inverter which creates a ~1 mA short on the power grid, a voltage observe transistor which enables the voltage drop to be measured at the shorting location, and a scan chain for control. The Voltage Drop (VDrop) PUF is defined as simply the voltage drop between the power supply and the observe voltage V_{obs}, i.e., $V_{\mathrm{drop}} = V_{\mathrm{PWR}} - V_{\mathrm{obs}}$, with $V_{\mathrm{PWR}} = 0.9$ V. The Equivalent Resistance (ER) PUF is defined as the voltage drop divided by the shorting current, $R_{\mathrm{eq}} = (V_{\mathrm{PWR}} - V_{\mathrm{obs}})/I_{\mathrm{short}}$. In this implementation, the challenge specifies which shorting inverter is enabled and the response is the measured

DAC'09, July 26-31, 2009, San Francisco, California, USA
Copyright 2009 ACM 978-1-60558-497-3/09/07....5.00

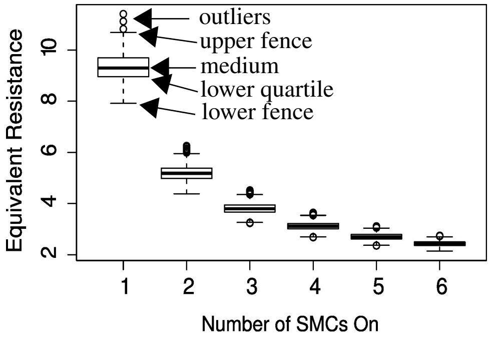


Figure 1. Box-plots of 1-on through 6-on (x-axis) ER values (y-axis) measured from 36 chips.

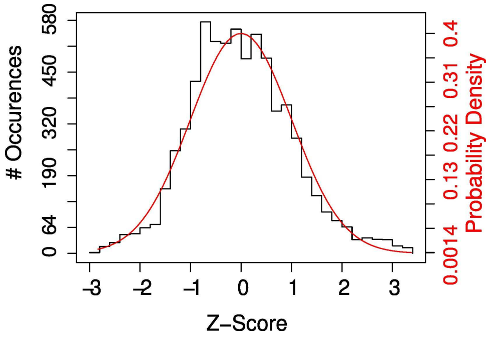


Figure 2. Histogram and Gaussian fit of standardized ERs from 192 responses and 36 chips.

VDrop or ER. Our design included 6 SMCs, and therefore, we obtain 6 Challenge/Response Pairs (CRPs) that constitute the VDrop or ER PUF of each IC, respectively. We kindly refer the reader to [3] for further details.

The ER PUF proposed in [3] is based exclusively on resistance variations in the PDS, which is a passive component, and therefore is less susceptible to environmental variations than PUFs based on transistor variations (active components). Another significant advantage of using the power grid as a PUF is that it is an existing, distributed resource in every design. Therefore, the overhead is limited to the challenge/response circuitry which is approximately 0.025% for a 25 mm^2 chip with 121 (50 um^2) SMCs.

3.MULTIPLE-SHORTING SCENARIOS

In this paper, we investigate an extension to our PUF where additional challenges are introduced by allowing more than one of the shorting inverters to be enabled at a time. For example, if the shorting inverters from two SMCs are enabled simultaneously, then two responses can be obtained by measuring the VDrop at each SMC location separately. We refer to these configurations as **x-on scenarios**, to distinguish them from the 1-on scenario described in [3]. A corresponding set of ERs can be computed by dividing each of the VDrops by I_{short}, the sum of the shorting currents from the set of enabled SMCs. With a total of six SMCs in our test chips, it is possible to obtain a total of 192 response bits by enabling different combinations of SMCs. For example, there are a total of 15 configurations in which two SMCs are enabled (**2-on** scenario), with each configuration generating 2 responses, for a total of 30 responses. For the **3-on** scenario, there are 20 configurations and 60 response bits. The closed form expression for the number of possible response values for n SMCs is given by Eq. 1.

$$\sum_{i=1}^{n} i\binom{n}{i} = n2^{n-1} \qquad \textbf{Eq. 1.}$$

Although the VDrops under the x-on scenarios remain relatively constant, the mean ERs decrease by a factor proportional to the number of enabled SMCs because of the increasing magnitude of the accumulating stimulus currents. Figure 1 gives a box plot analysis of the ERs computed from our 36 chips split into 6 groups along the x-axis, where each group is a different **x-on** scenario, 1-on, 2-on, etc. The distribution is summarized by 5 values in each box plot: the medium, the upper and lower fence limits (for largest and smallest observations, respectively), upper and lower quartiles, and outliers (see 1-on scenario in Figure 1). From the figure, it is clear that the variation among the ERs also decreases for higher-order x-on scenarios, i.e., 2-on through 6-on. Unlike the ERs, the VDrop responses (not shown) increase linearly over a small range from 7 mV to 11 mV for higher-order x-on scenarios, and the variation remains approximately constant.

An important consequence to the decreasing mean and variation in the higher-order x-on scenarios shown in Figure 1 is a firm upper limit on the number of enabled SMCs. Our analysis below shows this limit to be approximately 6, so enabling more than 6 SMCs simultaneously does not produce useful response bits in practice. Therefore, the number of challenges is linear to the number of embedded SMCs and smaller than the exponential given by Eq. 1.

In order to enable an analysis that combines all 192 responses across all x-on groups, we standardize the data by group using the standard z-score equation, $z = (x - \mu) / \sigma$. As indicated earlier, an important quality metric of a PUF is its degree of randomness. A first order measure of randomness can be obtained by constructing a histogram that bins the z-score representation of the responses. The ideal distribution with respect to randomness is a uniform distribution. Non-uniform behavior, e.g., clustering, in the responses makes the PUF susceptible to certain attacks such as the prediction attack [2]. Figure 2 gives the histogram of all 192 ER z-scores from the 36 chips. The distribution is best fit with a Gaussian curve, shown superimposed on the histogram in the figure. Although not ideal, the symmetric nature of a Gaussian is desirable and more robust to attacks in comparison to skewed distributions. A similar distribution and conclusion holds for the VDrop analysis (not shown).

4.SINGLE-BIT PROBABILITY ANALYSIS

A second, more quantitative means of evaluating randomness is through single-bit probability analysis, which evaluates the symmetry in the statistical distribution of the PUF responses. In this analysis, we first *discretize* the ER responses by computing a set of means across the 36 chips for each of the 192 response values. Each of the 192 means are used to threshold the 36 individual responses from the chips. Chip values larger than the mean are assigned ‘1’ while those below the mean are assigned ‘0’.

The level of randomness can then be easily measured by counting the number of ‘1’s and ‘0’s in each set. Sets that have equal numbers of ‘1’s and ‘0’s, i.e., 50% of each, are perfectly random. Figure 3 gives the results of the analysis using ERs. The x-axis numbers the response bit groups from 1 to 192 and the y-axis gives the probability of a ‘1’ across the 36 chips analyzed. It is clear that the individual distributions cluster around the ideal behavior of 50%, with deviations ranging from 40% to 60%. The average

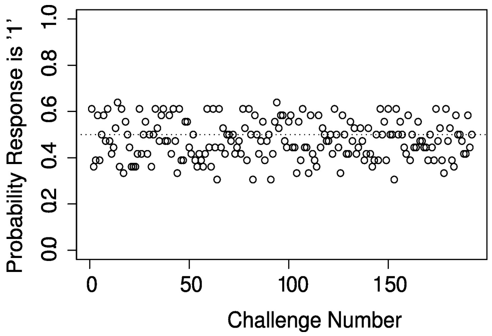


Figure 3. Single-bit probability analysis of ER PUF.

probability across all 192 groups is 47.5% for the ER analysis, and 54.5% for the VDrop analysis (not shown).

5.COLLISION PROBABILITY ANALYSIS

In this section, we analyze the probability that two chips produce the same response. Although this analysis can be performed using binary versions of the ERs, as described in the previous section, we chose to use the ERs directly because they more accurately portray the true variations in the data and allow noise to be more easily factored into the analysis. The analysis is carried out on pairings of the chip response vectors. With 36 chips, there are 630 such pairings (36 choose 2). The ERs (or VDrops) for a specific IC are arranged into a 192-dimensional vector and the Euclidean Distance (ED) between each pairing of vectors is computed.

The probability of a collision is computed by creating two histograms: one constructed using all 630 ER EDs from the 36 chips and one constructed from a set of noise samples. The noise data is obtained by repeating the entire SMC measurement process 72 times using one of the chips. The number of pairings and resulting EDs in this case is 2556 (72 choose 2). We then fit each histogram using a gamma probability density function (PDF). The probability of a collision is computed by first determining an ED value that bounds 99.73% (3 sigma) of the area under the noise PDF. The area to the left of this value in the chip PDF expresses the probability of collision [3].

n	Max Noise	Min Chip	Threshold	P(collision)
6	0.1172	0.4740	0.1092	4.27e-07
36	0.1507	0.8061	0.1247	3.29e-08
96	0.1735	1.0055	0.1431	1.55e-08
156	0.1811	1.0728	0.1539	1.22e-08
186	0.1841	1.0858	0.1583	1.18e-08
192	0.1849	1.0875	0.1591	1.19e-08

Table 1: Collision Analysis

In order to determine the impact on the probability of a collision as additional, higher-order x-on responses are added, the analysis is carried out on incrementally larger sets of response bits. Figure 4 plots the inverse probability of collision (*y*-axis) as the response vector size is increased from 6 to 192 (*x*-axis). The increasing trend associated with the curve illustrates that by adding the responses from higher-order x-on tests, the inverse probability of collision increases to a maximum that is 36 times larger than it is for the 1-on scenario. Table 1 summarizes the individual components of this analysis, including the maximum noise EDs, the minimum chip EDs and the 3-sigma noise value used in the collision analysis. The upward trend of the curve in the figure shows that the higher-order x-on responses add diversity (and value) to the overall response. It is also apparent that the increase in diversity begins to saturate with the addition of the 5-on and 6-on responses. Therefore, increasing the number of simultaneously SMCs beyond 6 is of limited value.

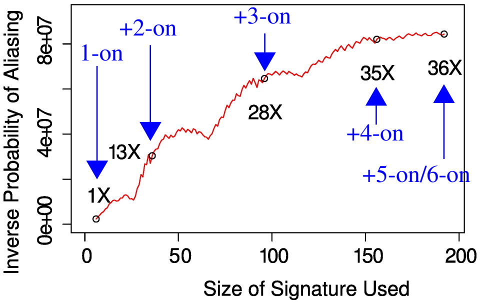
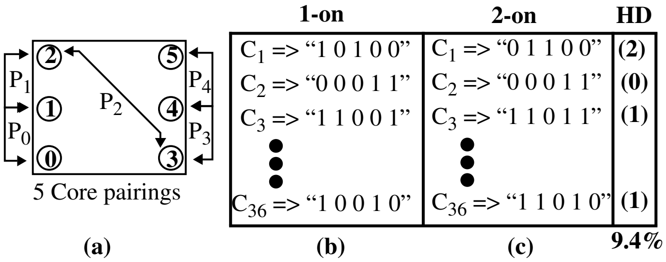


Figure 4. Collision probability using ER response vector sizes from 6 to 192.



	1-on	2-on	HD
	C_1 => "1 0 1 0 0"	C_1 => "0 1 1 0 0"	(2)
	C_2 => "0 0 0 1 1"	C_2 => "0 0 0 1 1"	(0)
	C_3 => "1 1 0 0 1"	C_3 => "1 1 0 1 1"	(1)
	⋮	⋮	
	C_{36} => "1 0 0 1 0"	C_{36} => "1 1 0 1 0"	(1)
			9.4%

Figure 5. Pairing and analysis illustration

6.ENTROPY ANALYSIS

The primary objective of this analysis is to determine the level of entropy that exists in various subsets of the ER and VDrop response vectors. The analysis is performed on the digital values computed by **comparing pairs** of ER and VDrop responses on the same chip. This models an *actual use scenario* in which a response bit is determined by the relative differences in the analog responses from two on-chip configurations of the PUF.

A response bit in our analysis is '1' if the first ER or VDrop response of a SMC pairing is larger than the second, and '0' otherwise. To determine upper and lower bounds on entropy, we consider two ways of selecting the pairs. In the first, called ***Core***, only *5* pairings of the 6 SMCs are considered, as a means of avoiding correlation (see [5]). We treat the results of this analysis as a lower bound on the available entropy. The Core analysis pairings are illustrated in Figure 5 (a) as P_0 through P_4. The second, called ***All***, includes all possible pairings of the 6 SMCs, which generates 6*5/2 = 15 bits.

As indicated earlier, it is possible to enable more than one SMC at a time. The ER response bits under the x-on scenarios can be different from the response bits from the 1-on scenario because they are affected by the total current, which is a function of multiple independent shorting currents. Changes in the relative values of the ERs on the same chip will reflect as bit-flips as shown by the example in Figure 5 (b) and (c). The response vectors under (b) portray the response bits across the 5 pairings in the 1-on Core analysis. The response vectors for each of the chips, C_x, are given

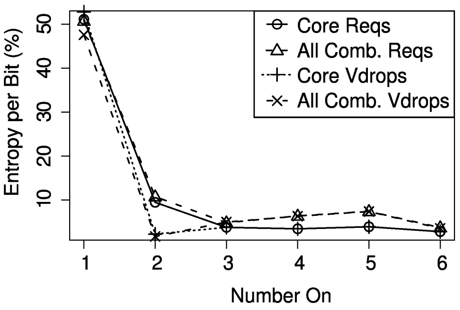


Figure 6. Entropy Analysis of VDrops and ERs.

as rows. In contrast, (c) shows the response vectors under the 2-on scenario for the same chips and pairings. The values in parenthesis on the far right are the Hamming Distances (HD) between the two vectors. For example, C_1 under (c) has two bit-flips (and an HD of 2) when compared with the vector under (b).

We use the HD to measure how much entropy is added over the 1-on base case for each of the x-on scenarios. The value of 9.4% given in the bottom right of Figure 5 (c) is computed by summing the HDs of the individual chips and dividing by the total number of bits that are compared. For the Core analysis shown in the example, the sum of the 36 chip HDs is 17. The entropy measure of 9.4% is computed as 17/180, where the denominator is computed as 36 chips * 5 bits.

The curves in Figure 6 show the average increase in entropy across the 6 *x*-on analyses as 4 curves, one each for the Core and All analyses using the VDrop and ER data sets. The 1-on base case given as the left-most data point on each of the curves is the probability of an arbitrary response being ‘1’. For the ER data curves, the probability is precisely 50%. For example, under the Core analysis, the response vector size is 5 bits for each of the 36 chips. Of the 180 bits, we observed exactly 90 ‘1’s. The result under the All analysis is 270 ‘1’s, exactly half of the 540 bits (15*36).

The remaining points on the graph each represent the average HD between the previous response vector and the vector generated using the *x*-on data identified on the *x*-axis. We refer to this change in entropy as ‘delta entropy’. For example, the ER Core analysis value for the 2-on scenario is given as 9.4% (we described this case earlier in reference to Figure 5). From the graph, the All analysis produced a similar value. Both of these values represent a relatively small increase in entropy over the 1-on base case. The VDrop values indicate very little delta entropy. This is true because the VDrop responses under the x-on scenarios cannot leverage the interaction of the SMC shorting currents used in the ER response calculation.

For the 3-on through 6-on scenarios, the delta entropies, although small, are not zero and therefore represent a positive increase in the cumulative entropy. For the 3-on through 5-on scenarios, we arbitrarily chose the locations of the additional enabled SMCs, e.g., 1 additional SMC for 3-on, 2 for 4-on, etc., beyond the two used in the pairing. The trends in delta entropy in Figure 6 support the behavior of the curve shown in Figure 4, which tends to saturate, particularly for the right-most data points representing the 5-on and 6-on scenarios.

Given these results, we can approximate the number of response bits that are truly random. As indicated earlier, the Core analysis represents a conservative bound where the number of pairing is restricted to (*n*-1) per *x*-on scenario. Therefore, a chip with *n* SMCs can produce 6*(*n*-1) unique response bits, assuming the delta entropy goes to zero for more than 6 enabled SMCs. For the optimistic All analysis, the number of meaningful response bits is given by Eq 2. For our chips, these expressions produce 30 and 255 bits, respectively with 6 SMCs.

$$N_{\text{bits}} = \binom{n}{2} + \sum_{i=0}^{n-2} \binom{n-2}{i}\binom{n}{2} = \frac{n(n-1)}{2} + (n-1)n2^{n-3} \quad \textbf{Eq.2.}$$

We also performed a pairwise HD analysis using the entire 30-bit and 255-bit response vectors from the Core and All analyses, respectively. We compute the average HD per bit by computing the HDs between all possible chip pairs, taking the average HD, and then dividing by the number of bits in the response. Ideally, each comparison should produce an HD that is exactly half of the number of bits in the response vector. The evaluation of our PUF under this metric is as follows. The average HDs per bit under the Core and All analyses using the ER data are 48.3% and 48.5%, and for the VDrop analyses, they are 48.4% and 48.5%, respectively. These values compare favorably to 46.15%, as reported in [5].

We also evaluate reproducibility by carrying out a second pairwise HD analysis using the 72 sets of ‘noise’ samples described earlier. The average HD is computed, as described above, using the 30-bit and 255-bit response vectors. The results for the Core and All analyses are 0% and 0.64%, respectively, using the ER data and 0% and 0.59%, using the VDrop data. These results also compare favorably with 0.48% obtained in [5] and provides evidence that our PUF is robust to environmental noise and ambient temperature variations.

7.CONCLUSIONS

We analyze hardware data of a PUF derived from the resistance variations in the power distribution system of an IC. Single bit and conditional probability analysis as well as entropy analysis are used to determine the quality of the power grid PUF. The results show that the PUF possesses a high degree of randomness and stability, and performs well on the quality metrics evaluated.

8.REFERENCES

[1] R. S. Pappu, B. Recht, J. Taylor and N. Gershenfeld, “Physical One-Way Functions,” *Science*, 2002, pp. 2026-2030.

[2] M. Majzoobi, F. Koushanfar, M. Potkonjak, "Testing Techniques for Hardware Security", *ITC*, 2008, pp. 185-189.

[3] R. Helinski, D. Acharyya and J. Plusquellic, “A Physical Unclonable Function Defined Using Power Distribution System Equivalent Resistance Variations”, *DAC*, 2009, pp. 676-681.

[4] S. S. Kumar and J. Guajardo and R. Maes and Geert-Jan Schrijen and P. Tuyls, “Extended Abstract: The Butterfly PUF Protecting IP on Every FPGA,” *HOST*, 2008, pp. 70-73.

[5] G. E. Suh and S. Devadas, “Physical Unclonable Functions for Device Authentication and Secret Key Generation”, *DAC*, 2007, pp. 9-14.

[6] A. Maiti, P. Schaumont, "Improving the Quality of a Physical Unclonable Function using Configurable Ring Oscillators", *Conf. on Field Prog. Logic and Apps*, 2009, pp. 703 - 707.

[7] Y. Su and J. Holleman and B. Otis, “A 1.6pJ/bit 96% Stable Chip ID Generating Circuit using Process Variations,” *Solid State Circuits Conf.*, 2007, pp. 406-407.

[8] Y. Alkabani and F. Koushanfar and N. Kiyavash and M. Potkonjak, “Trusted Integrated Circuits: A Nondestructive Hidden Characteristics Extraction Approach,” *Information Hiding*, 2008.