Effect of Aging on Linear and Nonlinear MUX PUFs by Statistical Modeling

Anoop Koyily, Satya Venkata Sandeep Avvaru, Chen Zhou, Chris H. Kim, Keshab K. Parhi [*]
University of Minnesota, Twin Cities

Abstract— This paper addresses the effect of aging on linear and non-linear MUX physical unclonable functions (PUFs). It is well known that a PUF response can be modeled in terms of the delay difference of MUX stages. In this paper, we show that the aging effects can be modeled in terms of variations in delay-difference and arbiter delay. Specifically, with aging, the *percent delay-difference variation* of each MUX stage can be modeled as a ratio of two correlated Gaussian random variables. This ratio distribution is shown to be approximately Gaussian with zero mean and variance increasing with time. In case of the arbiter, the ratio distribution is modeled as a Gaussian with positive mean. The paper makes three contributions: modeling the effect of aging in terms of percent variations in delay-difference of the MUX stages and arbiter delay, analysis of authentication accuracy with aging, and approaches to increase the PUF's lifetime by either recalibrating it to obtain new delay-difference parameters, or by tuning a threshold based on the *total delay-difference*. A general approach for selecting the threshold values is described in the paper. It is shown that the authentication accuracy of a PUF is significantly affected due to aging effects of the arbiter itself. Therefore, under the assumption that the variations in arbiter delay are considerably more than in delay-differences, the performance degradation in the case of aging alone is prominent compared to noise alone. We show that the authentication accuracy of a feed-forward PUF is more degraded compared to linear or modified feed-forward PUF. Metrics like Jenson-Shannon and Henze-Penrose divergence are also used to analyze the effect of aging.

Index Terms*—*Physical Unclonable Function, Aging, Statistical Modeling, Noise, Divergence metrics.

I. Introduction

Physical unclonable functions (PUFs) are hardware circuits that can exploit unique signatures due to random process variations during manufacturing. In this paper, we consider multiplexer (MUX) PUFs whose properties are based on delay variations [1], [2]. Three PUF configurations are considered, namely linear, feed-forward and modified feed-forward [3]. The inputs to the PUFs are N-bit challenges and outputs are 1-bit responses. It is well known that the behavior of challenge-response pairs (CRPs) of MUX PUFs can be modeled using various adaptive learning techniques [4], [5], [6]. In this paper, we are particularly interested in a least mean square (LMS) based approach which has been used to estimate the delay-difference of MUX stages [4]. The estimated model parameters can then be stored in a server database and used for authentication. While similar approaches have been proposed before [5], [7], this method proposes to store the physical parameters of the model as opposed to those of an artificial neural network (ANN) or other non-linear models [6]. Furthermore, the parameters of ANN or other models do not correspond to the intrinsic physical parameters of the chip. The proposed method in [4] can be considered canonic as it is based on the physical parameters that correspond to N multiplexer stages and arbiter(s).

For applications in authentication, we desire PUFs to perform reliably over time. But various uncontrollable factors like environmental noise and aging degrade the authentication accuracy of a PUF. Environmental noise mainly occurs due to variations in temperature and supply voltage. Another source of noise in PUFs is due to the metastability condition of the arbiter. A PUF goes into metastable state when the inputs to the arbiter change at about the same time, or in other words, when the *total delay-difference* at the arbiter input is smaller than arbiter's setup or hold time [8], [9]. In such cases, the response of the arbiter becomes independent of the delay-difference of MUX stages and is determined by its initial condition or random noise [10], [11]. More secure configurations like feed-forward show a higher degree of metastability compared to linear or modified feed-forward configurations [11].

In contrast to environmental noise, aging affects the reliability of a PUF over a longer period of time. Aging is mainly caused due to undesirable changes in hardware structure such as negative bias temperature instability (NBTI), hot carrier injection (HCI) and time dependent dielectric breakdown (TDDB) [12], [13]. NBTI and HCI, in particular, are known to induce progressive slowdown in hardware [13], [14]. This results in a gradual increase in the delays of hardware structures like multiplexers. In proposed aging model, these increases in delays result in a slowdown effect in terms of the delays of multiplexers and arbiter(s). It is known that the multiplexer delays of a PUF are distributed randomly [15], [16]. We assume that, due to aging, the mean and variance of these delays gradually increase. This has been shown to be a valid assumption in prior aging works on delay-based PUFs [17]. However, our model considers the delay-difference rather than the delays for each MUX stage. We assume that the delay-differences are distributed with zero mean and variance which increases gradually with aging. In addition to the delay stages, the arbiter (which is typically an SR latch) also forms a key

[*] This research has been supported by the National Science Foundation under grant number CNS-1441639 and the semiconductor research corporation under contract number 2014-TS-2560.

978-1-5090-0602-1/18/$31.00 ©2018 IEEE

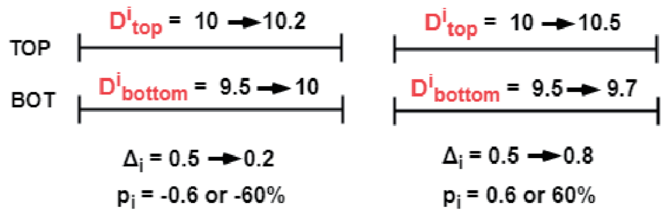

Fig. 1. An example showing two scenarios of how the delay-difference of a MUX stage varies with aging. D^i corresponds to delays of top and bottom multiplexers, Δ^i to the delay difference and p_i to the percentage change in delay-difference. p_i can be both positive or negative with aging.

component in the functioning of a PUF. In [9], [18], it is shown that for a latch/flip-flop, various timing related factors like setup time, hold time, clock-to-output and data-to-output generally increase with aging. This paper considers the effect of arbiter in terms of its propagation delay (or clock-to-output). In [17], it is shown that the arbiter in a delay-based PUF forms an Achilles' heel due to its "asymmetric" aging. Therefore, variations in the arbiter due to aging would be much more significant compared to the stage delay-differences.

This paper proposes an aging and noise based model for analyzing MUX-based PUFs. The model simulates conditions close to that for a real chip. The role of such a statistical simulation framework is important in the sense that both existing and new PUF structures can be characterized with respect to aging and noise effects without fabricating chips. Using the model, we investigate how various PUF configurations are affected by aging. Such comparisons of several PUF structures with respect to aging have never been presented before. One may be led to believe that the delay difference of a MUX stage would not be affected by aging if both the delay paths age by same amount. However, this is not true as the top and bottom path-delays both increase but by different amounts even after applying the same stress condition for the same stress duration [19], [20]. If the top path-delay increases more (or less) than the bottom path-delay, the total delay-difference increases (or decreases) and thus, the final response bit can flip. Intuitively, this explains why the delay difference variation can be modeled as a zero-mean Gaussian random variable. An example is shown in Fig. 1. In this work, we show that the variations in delay-difference due to aging can be modeled in terms of a ratio between two correlated Gaussian distributions and then, approximated as a zero mean Gaussian distribution with increasing variance. The arbiter delay, on the other hand, is modeled in the same way except the ratio distribution has a positive mean.

For a PUF, the challenge is applied multiple times to obtain the response in terms of a probability. This is done because there is a possibility of variation in the response bits due to environmental noise. Such a response is called a *soft-response* [21]. The soft-response of each challenge corresponds to the probability of the response being '1' (or '0'). The soft-response can be converted to a hard-response based on a threshold on the probability. We chose a threshold of 0.1-0.9. This means that if the soft-response is greater than or equal to 0.9, the output bit is '1'; if it is less than or equal to 0.1, the output bit is '0'; otherwise it is *unstable*. For authentication purposes, it is preferred to use *stable* challenges so that the response bits are more reliable. In our aging analysis, we focus on the authentication accuracy of the (initially) stable CRPs.

A prior aging work on linear PUFs [22] concluded that the effect of aging on PUFs is permanent and, therefore, the unstable (or unreliable) challenges need to discarded. In our work, we argue that, despite irreversible changes in the PUF structure due to aging, it can still be used for authentication by recalibrating the model parameters (i.e., the delay differences and the arbiter delay). The recalibration can be done by using the LMS method described in [4]. However, recalibrating hundreds of devices is not a feasible solution. Another way for improving the authentication performance is by choosing appropriate thresholds on *total delay-difference* to discard challenges that are unreliable. This approach is investigated in our paper.

The paper is organized as follows. Section II briefly discusses some background on PUFs and then presents the proposed aging model. Section III discusses the setup, results and the proposed techniques such as recalibration and the threshold selection process for improving reliability. Section IV presents the implications of the approach and Section V concludes the paper.

II. Background and Aging Model

A. Aging model for the delay chain

A linear MUX PUF has N multiplexer stages and an arbiter at the end. Control bits for each MUX stage are obtained from the input challenge. For MUX stage i, the delay-difference, $\Delta^i = D^i_{top} - D^i_{bottom}$, where D^i_{top} and D^i_{bottom} are the delays associated with top and bottom multiplexers of the i^{th} stage. We will assume multiplexer delays, D^i_{top} and D^i_{bottom}, to be Gaussian distributed as $N(\mu, \sigma^2)$ [15] (this can be attributed to the manufacturing process variations which tend to be random in nature). Therefore, delay-difference of the i^{th} stage, Δ^i, will be distributed as $N(0, 2\sigma^2)$.

Due to aging, the multiplexer delays of each stage gradually increase, which corresponds to an increase of mean and variance to (μ',σ_a^2) [17]. Therefore, the new delay-difference of the i^{th} stage, Δ^i_{aged}, will be distributed as $N(0, 2\sigma_a^2)$ (where $\sigma_a > \sigma$). This is validated in prior aging work [17], where it is shown that the standard deviation of delay-difference, $\sqrt{2}\sigma$, increases with aging. The new delay-difference of the i^{th} stage for an aged PUF can be expressed as:

$$\Delta^i_{aged} = \Delta^i \left(1 + \frac{\Delta^i_{aged} - \Delta^i}{\Delta^i}\right) = \Delta^i \left(1 + p_i\right) \qquad (1)$$

where p_i is the *percent delay-difference variation* in the i^{th} stage. p_i is basically the ratio of two correlated Gaussian distributions, $N(0, 2(\sigma^2 + \sigma_a^2 + 2\rho\sigma\sigma_a))$ and $N(0, 2\sigma^2)$ (where ρ is the correlation coefficient between Δ^i and Δ^i_{aged}). As the variance of Δ^i increases with aging, σ_a^2 term in the variance of

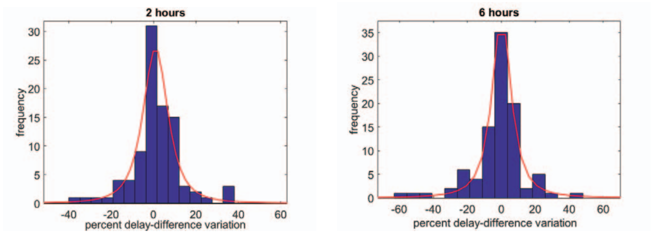

Fig. 2. Ratio distribution with *t-distribution* fit for data collected from 3 chips for 2 hours (left) and 6 hours (right) of aging. X axis is the percent delay-difference variation.

p_i will start to dominate. Therefore, the variance of these approximate distributions is to an extent proportional to σ_a^2. Prior work in [23] suggests some approximate distributions for correlated ratio distributions in terms of Gaussian, t-distribution etc. A ratio of two Gaussians also has been used to model true random number generators [24].

Using aging data collected from test chips for upto 10 hours, we observe that the ratio distributions have a good fit with *t-distribution*. Fig. 2 shows ratio distributions for recovery times of 2 and 6 hours using a voltage based stress test, where the voltage was increased from a nominal value of 0.9V to 1.8V at 25°C. However, the data collected was only for 3 chips and corresponds to 3x32=96 samples for 32-stage MUX PUFs. This is insufficient in order to obtain a precise distribution of p_i.

For our model, we adopt a Gaussian approximation for the ratio distribution (as t-distribution is a good approximation for Gaussian when dealing with small sample sizes). An example is ratio distribution, p_i, with standard deviation=0.05 (or 5%). In prior work [17], a 5% standard deviation in delay-difference roughly corresponds to 2 years of aging.

By using the model in (1) with a set of initial Δ^i (sampled from a Gaussian distribution with standard deviation, $\sqrt{2}\sigma$) and p_i (sampled from a ratio distribution with a given standard deviation), we can generate various instances of delay-difference, Δ^i_{aged}, for MUX stage i. Note that σ is obtained using the values of Δ^i estimated [4] from the test chip.

B. Aging model for arbiter

The arbiter, placed at the end of the delay chain, is modeled in a slightly different manner. We model it in terms of its delay, Δ^{arb}, which accounts for its propagation delay. However, being a delay element, it takes only positive values unlike delay-difference and, therefore, has a positive mean.

An aged Δ^{arb} can be expressed similar to (1) as:

$$\Delta^{arb}_{aged} = \Delta^{arb}\left(1+q\right) \tag{2}$$

where q is the percent change in arbiter delay. The only difference is that q is Gaussian distributed with a positive mean. Similar to the delay-difference, the mean and variance of q increase with aging. Here for simplicity, we assume that mean and variance are related as μ_q^2=$3\sigma_q^2$ (assuming q to be a uniform random variable between 0 and $2\mu_q$).

C. Total delay-difference and Noise model

Depending on the input challenge, the difference between the two paths traversed is termed as *total delay-difference*. It is denoted by r_N and for an N-stage linear MUX PUF can be computed as [15, 16]:

$$r_N = \sum_{i=1}^{N+1}(-1)^{C'_i}\Delta^i = \sum_{i=1}^{N}(-1)^{C'_i}\Delta^i + \Delta^{arb} \tag{3}$$

where C'_i=$\oplus_{j=i}^{N}C_j$ is the XOR^{ed} challenge bit corresponding to challenge bit C_i, Δ^i is the i^{th} stage delay-difference and Δ^{arb}=Δ^{N+1} (C'_{N+1}=0) is the bias corresponding to the arbiter delay [3].

However, the model in (3) is deterministic and does not include any variations that occur due to environmental noise. This model generates output which has zero intra-chip variation (or 100% reliability). Therefore, in order to simulate the model close to that of a real chip, we consider environmental noise in addition to the delay parameters, Δ^i and Δ^{arb}. Both the delay parameters vary depending on the amount of noise as Δ^i+n_i and Δ^{arb}+n_{arb}, where n_i and n_{arb} are the noise factors affecting the i^{th} delay stage and the arbiter. A modified model of (3) to mimic a real chip performance can be expressed as:

$$r_N = \sum_{i=1}^{N+1}(-1)^{C'_i}\Delta^i + \sum_{i=1}^{N+1} n_i \tag{4}$$

where n_i is the environmental noise contributed by each stage.

Final response bit, *R*, of the PUF is decided based on r_N as [25]:

$$R = sign(r_N) = \begin{cases} 1, & r_N \geq 0 \\ 0, & r_N < 0 \end{cases} \tag{5}$$

In case of feed-forward and modified feed-forward configurations, similar expressions for total delay-difference, r_N, and final response bit, R, can be obtained. Independent of the PUF configuration, an un-aged N-stage MUX PUF has total delay-difference, r_N, distributed as $N\left(\mu_{arb}, 2N\sigma^2 + \sigma_{arb}^2 + \left(N+1\right)\sigma_n^2\right)$, where $N(0, 2N\sigma^2)$ is due to the multiplexer stages, $N(\mu_{arb}, \sigma_{arb}^2)$ is due to the arbiter delay and $N\left(0, \left(N+1\right)\sigma_n^2\right)$ is due to environmental noise.

Fig. 3 shows the distribution of total delay-difference, r_N, with 10,000 random challenges for the three PUF configurations. Observe that the feed-forward configuration has a higher spread of r_N for unstable challenges (blue color). This is due to an increased metastability [11]. For our chips, variance of r_N for unstable challenges in case of feed-forward configuration is atleast 10 times more than the other two configurations. For linear and modified feed-forward configurations, total delay-difference, r_N, of unstable challenges are much more closer to 0. As discussed before, the definition of unstable CRPs comes from a thresholding (0.1-0.9) based on the *soft-response* [21]. Fig. 4 shows the probability distribution of r_N corresponding to stable 0 and stable 1 response bits for a linear PUF. As can be observed from the figure, there is a

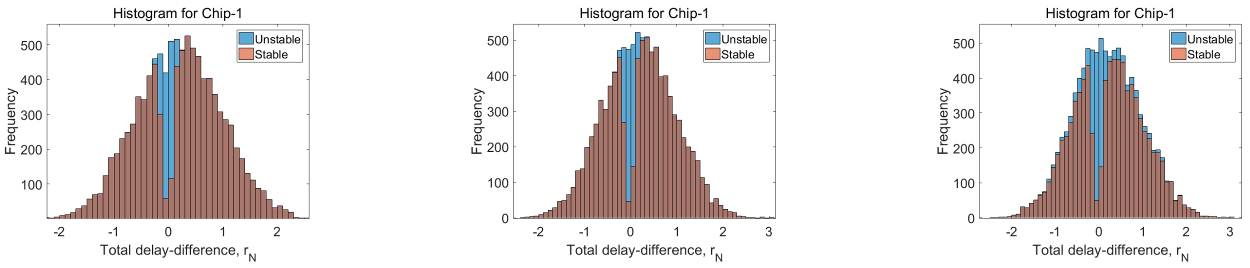

Fig. 3. Histogram of total delay-difference, r_N, for un-aged (from left to right) linear, modified feed-forward and feed-forward configurations (using ground truth from the chips).

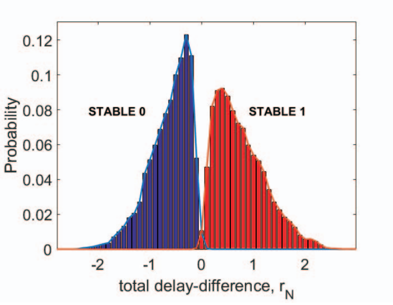

Fig. 4. Probability distributions of r_N corresponding to stable 0 and stable 1 response bits for an unaged linear PUF with noise std=5% of std of Δ^i.

small degree of overlap between the distributions. This corresponds to the environmental noise added to the model in (4) and represents the error present in the proposed model.

We now summarize the assumptions made for the proposed model:

- For a fixed set of conditions, the effect of environmental noise on the PUF performance is static. This means that for a given environmental condition, the value of σ_n can be assumed to be constant. This is true even when the PUF is undergoing a gradual change due to the aging process.
- Variance of the delay parameters like delay-difference, $2\sigma^2$, and arbiter delay, σ_{arb}^2, increase gradually with aging. The model considers them indirectly in the form of variance of percentage distributions, p_i and q.
- The variance of the percentage change in arbiter delay, q, increases much more rapidly than that of delay-difference, p_i. This means for a given amount of aging, $\sigma_{arb}>\sigma$.

III. Setup and Results

A. PUF Setup for Aging Simulation

We analyze the effects of aging on 6 32-bit MUX PUFs fabricated across 2 chips which use IBM 32nm HKMG technology [21]. Aging measurements were made for upto 10 hours at a voltage of 1.8V (from a nominal value of 0.9V) and temperature 25°C. From the silicon chip, *soft* responses are collected for 10,000 random challenges. Each challenge-response pair (CRP) is measured 102,400 times and converted to a soft response using an on-chip 1/1024 divider. The soft-response is, then, converted to a hard response bit using 0.1-0.9 threshold. These hard responses are used to estimate the delay-differences using LMS technique [4].

For a given amount of aging, we generate 1000 PUF instances using Monte-Carlo simulation with percentage delay paramaters, p_i and q, sampled from Gaussian distributions with certain standard deviation. The (initial) delay-differences of individual stages, Δ^i, and arbiter delay, Δ^{arb}, are taken from the models obtained from the actual chip. Gaussian noise with a fixed standard deviation, σ_n, is also added. To simulate aging over a period of time, standard deviations of p_i and q are varied while keeping the standard deviation of noise fixed.

B. Authentication Performance with Noise

The model in (4)-(5) includes the effect of environmental noise to the final response bit. For a given environmental condition (like fixed temperature, voltage supply etc), the noise is modeled as a Gaussian distribution with fixed variance. Unlike aging, noise is static and, therefore, the degradation in PUF performance is fixed. Fig. 4 shows the probability distributions of r_N for stable 0 and stable 1 response bits in the presence of noise with standard deviation=5% std. of Δ^i.

We can quantify the overlap between the distributions in terms of metrics like Jensen-Shannon (JS) [26] or Henze-Penrose (HP) divergence [27]. Jenson-Shannon (or JS) divergence is a symmetric form of Kullback-leibler (or KL) divergence [28]. JS divergence between two distributions P and Q is defined as:

$$JS(P||Q) = \frac{1}{2}(KL(P||R) + KL(Q||R)), \qquad where\ R = \frac{1}{2}(P+Q) \tag{6}$$

where *KL(.)* corresponds to KL divergence. We found KL divergence to be sensitive to probabilities close to 0 which deemed it unsuitable. JS divergence takes values between 0 to 1, whereas HP divergence takes values between 0.5 to 1.

For a good PUF reliability, we desire a higher divergence between the stable 0 and stable 1 probability distributions. In Fig. 5 (**dashed lines**), we show how the two metrics vary for different amounts of noise. The x-axis is the standard deviation of noise (σ_n) represented as a percentage of the standard deviation ($\sqrt{2}\sigma$) of delay-difference, Δ^i. We observe that as

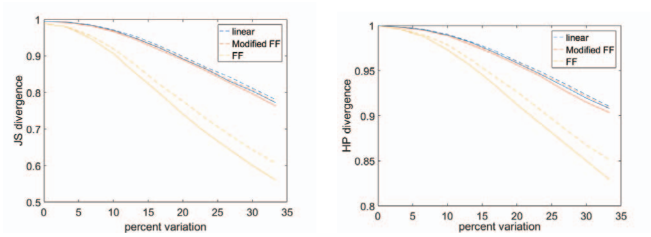

Fig. 5. Divergence metric comparisons in case of noise alone (dashed line) and aging alone (solid line) scenarios with (left) Jensen-Shannon divergence (right) Henze-Penrose divergence.

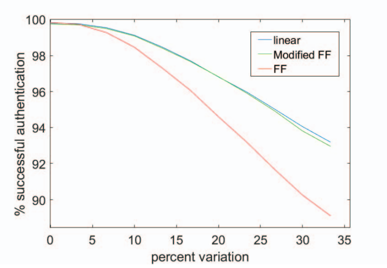

Fig. 6. Percentage of successful authentications by assuming equal variance for the delay chain and arbiter due to aging and in presence of 5% std of noise.

the amount of noise increases, the overlap between the distributions also increases and, therefore, the divergence values decrease. Out of the three configurations, feed-forward is affected the most by environmental noise.

C. Authentication Performance with Aging

As mentioned before, we assume that the model parameters, i.e., delay-differences and the arbiter delay, have been estimated and are stored in the server database. These stored parameters were estimated for an un-aged PUF instance. But with time as the PUF starts to age, the delay-differences start to vary gradually and therefore, the stored model parameters (or even a CRP look-up table or other adaptive parameters) become outdated. This occurs due to flipping of response bits for these challenges. The result is a decrease in the percentage of successful authentications, as shown in Fig. 6. An authentication is considered successful if the responses to the challenges match with their expected values (which is obtained from the stored model parameters or from a CRP look-up table).

The percentage of successful authentication at 0% standard deviation, i.e., for an un-aged PUF, depends upon the amount of environmental noise added to the model. Otherwise, the prediction accuracy is very close to 100%. Fig. 6 shows the percentage of successful authentications under equal aging scenario for the delay chain and arbiter, i.e., under the assumption of equal variation for both. Note that unless otherwise mentioned, the standard deviations for percentage variation, p_i and q will be assumed to be equal. However, as discussed before we expect a higher variance for arbiter than the delay chain. Fig. 7 show the authentication accuracies for aging effects considered separately on the delay chain and arbiter. We can observe that the performance degradation due to aging in arbiter is significant. Tables I, II show the percentage of successful authentications under equal and unequal aging conditions of the arbiter and delay-difference. From Table II, we observe that under the assumption that the variation in arbiter delay, q, is considerably more (i.e., 20%) than in delay-difference, p_i, the performance degradation in the case of aging alone is prominent than noise alone.

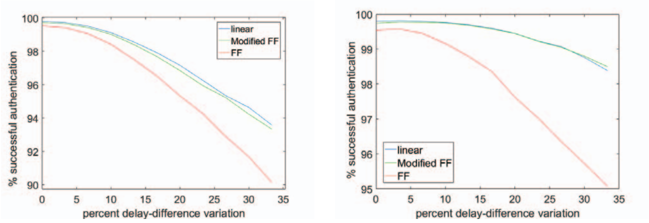

Fig. 7. Percentage of successful authentications with aging effect considered only on delay chain (left) and only arbiter (right) in presence of 5% noise std.

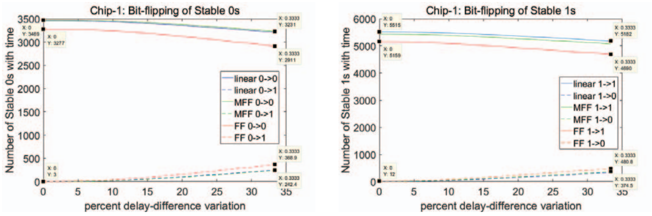

Fig. 8. Number of bit flips 0→1 in Stable-0 (left) and 1→0 in Stable-1 (right) with equal delay chain and arbiter aging.

TABLE I
PERCENTAGE SUCCESSFUL AUTHENTICATION UNDER EQUAL AGING SCENARIO; $STD(q)$=$STD(p_i)$

% STD		No Noise	Noise STD=5%	Noise STD=10%	Noise STD=20%
Linear	Original	0.9993	0.9980	0.9930	0.9729
	5%	0.9981	0.9967	0.9917	0.9714
	10%	0.9927	0.9911	0.9860	0.9674
	20%	0.9697	0.9683	0.9639	0.9487
MFF	Original	0.9985	0.9974	0.9921	0.9710
	5%	0.9977	0.9963	0.9911	0.9698
	10%	0.9923	0.9906	0.9854	0.9661
	20%	0.9690	0.9675	0.9629	0.9486
FF	Original	0.9982	0.9954	0.9863	0.9528
	5%	0.9955	0.9927	0.9837	0.9523
	10%	0.9842	0.9817	0.9728	0.9450
	20%	0.9463	0.9442	0.9387	0.9187

With aging, we observe that a feed-forward PUF is more prone to bit flips than the other two configurations. This means that among the three configurations, feed-forward is most liable to have an authentication failure for a fixed amount of tolerance in error. This is because any bit-flip in the intermediate response bit (or the internal challenge bit) affects the final response bit much more significantly than in the case of modified feed-forward or linear. In case of modified feed-forward, the interconnection between consecutive stages (almost) negates the effect of internal challenge bits on the final response.

Fig. 8 shows the number of bit flips for Stable-0 and Stable-1 response bits with aging. The top curves (solid line) in each figure show a decrease in the number of stable 0s (or stable

TABLE II
PERCENTAGE SUCCESSFUL AUTHENTICATION UNDER UNEQUAL AGING SCENARIO; $STD(q)$=$STD(p_i)$+20%

$\%STD(p,q)$		No Noise	Noise STD=5%	Noise STD=10%	Noise STD=20%
Linear	(0,20)	0.9961	0.9941	0.9892	0.9704
	(5,25)	0.9914	0.9898	0.9843	0.9657
	(10,30)	0.9825	0.9805	0.9761	0.9594
	(20,40)	0.9568	0.9556	0.9526	0.9409
MFF	(0,20)	0.9960	0.9944	0.9891	0.9694
	(5,25)	0.9912	0.9896	0.9848	0.9663
	(10,30)	0.9827	0.9815	0.9761	0.9592
	(20,40)	0.9566	0.9561	0.9528	0.9391
FF	(0,20)	0.9795	0.9773	0.9700	0.9441
	(5,25)	0.9672	0.9654	0.9591	0.9354
	(10,30)	0.9506	0.9494	0.9433	0.9248
	(20,40)	0.9136	0.9124	0.9096	0.8937

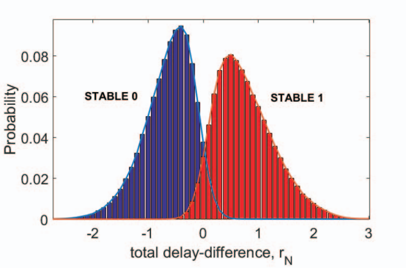

Fig. 9. Probability distributions of stable 0 and stable 1 response bits with aging with 30% delay variation for a linear PUF

1s) with aging. The bottom curves (dashed line) show an increase in the number of flipped bits corresponding to stable 0s (or stable 1s). For a fixed level of aging, we observe that the number of bit-flips is the highest for the feed-forward configuration. For example, for 33% standard deviation of delay-difference variation, the percentage of Stable-0 bit-flips for linear/modified feed-forward is roughly $\frac{242}{3469}$=6.97% and for feed-forward is $\frac{369}{3277}$=11.3%. Similarly, for a 33% standard deviation of percent delay-difference variation, the percentage of Stable-1 bit-flips for linear/modified feed-forward is roughly $\frac{374}{5515}$=6.78% and for feed-forward is $\frac{481}{5159}$=9.32%. We can observe that the number of bit-flips Stable-0$\rightarrow$1 is more than Stable-1$\rightarrow$0. This indicates that the mean of total delay-difference, r_N, (i.e., μ_{arb}) increases with aging. Also, note that the response bits are slightly skewed towards bit '1'. This is because the overall mean of distributions in Fig. 3 is slightly greater than 0 (= μ_{arb}).

Similar to the case of noise, the overlap between the stable-0 and stable-1 distributions would increase with aging (as shown in Fig. 9). Fig. 5 (**solid lines**) shows how both the divergences (JSD and HPD) change with aging. We observe that both aging and noise affect the authentication performance in a similar manner. Even so, for the same amount of percent variation, the performance is slightly worse for aging. The difference is mainly due to the way the arbiter ages with time.

D. Temporal Properties of unstable CRPs

In this section, we discuss how aging affects the (initial) unstable CRPs. Note that these CRPs are unstable in an un-aged PUF due to the effects of metastability or environmental noise. Here, we are interested in studying the variations in the total delay-difference, r_N, for these CRPs due to aging. Fig. 10 shows how the variance of r_N for unstable challenges varies with time. As observed, the variance for all the three PUF configurations increases with aging. From our earlier discussions on un-aged PUFs, we know that feed-forward has a higher variance of r_N for unstable challenges. With aging, this variance increases further. Hence, it is logical to think that some of these unstable challenges would become stable. To validate this, we choose a threshold for r_N (say, α=0.4σ, where σ is the standard deviation of total delay-difference, r_N, for *un-aged* PUF shown in Fig. 3). We then observe how the percentage of unstable challenges with $|r_N|>\alpha$ varies with aging. This is shown in Fig. 10.

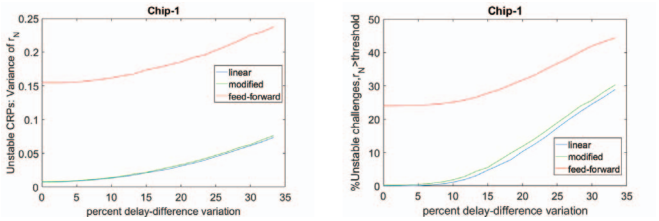

Fig. 10. (Left) Variance of r_N for unstable challenges with aging, (right) Percentage of unstable challenges with $|r_N|>\alpha$, where α=0.4σ.

We can observe that the percentage of unstable challenges with $|r_N|>\alpha$ increases with time (or aging). For linear and modified feed-forward configurations, the percentage remains near 0 till about 7% standard deviation of delay-difference variation (p_i or q). This means that, in the initial periods of aging, there is hardly any increase in the number of unstable challenges with $|r_N|>\alpha$. This is due to the choice of α (=0.4σ), which essentially guarantees the stability of these CRPs, i.e., these challenges have soft-responses between 0.1 and 0.9. Feed-forward, on the other hand, has higher ($\sim$25%) percentage of unstable challenges with $|r_N|>\alpha$, even in the un-aged case. This indicates that for feed-forward, α=0.4σ is not an optimal value for guaranteeing stability. Empirically, we choose a value of α=2.4σ for this case.

Furthermore, from Fig. 10 we can say that for linear and modified feed-forward configuration, about 25% of (initially) unstable CRPs have become stable for a 33% standard deviation of percent delay-difference. Note that 25% of unstable CRPs is roughly $\frac{25}{100}$x10%$\approx$2.5% of total CRPs. In case of feed-forward, such claims can only be made at higher values of α. This is because for feed-forward, α=0.4σ does not guarantee stability. However at a higher threshold (α=2.4σ), we observe that only a very small percent (<0.1%) of unstable CRPs becomes stable. Therefore, unstable CRPs with $|r_N|>\alpha$ can be later used for authentication purposes in an aged PUF.

E. Recalibration and Threshold Tuning

Previously in Fig. 6, we had discussed about the authentication failures occurring due to aging. This happens as the (aged) delay-differences start to vary from the ones stored in the server database. A simple solution for this problem is to re-estimate the new delay-differences using the LMS technique described in [4]. For the LMS estimation, a sufficient number of CRPs (<2000) will give us a prediction accuracy close to 100% in case of linear MUX PUFs. However, a disadvantage of the approach is that with increasing number of PUF devices in the market, the number of devices needing recalibration will be very high. Therefore, this approach will prove to be costly and is not feasible in practise.

We propose an alternative approach where the goal is to select appropriate thresholds (say, β) on the total delay-difference, r_N, of the un-aged PUF which can improve the reliability. A desired value of β will correspond to the total delay-difference, r_N, of CRPs immune to aging related bit-flips.

Fig. 11 shows the percentage of error (or failures) in authentications against varying threshold, β, for a linear PUF. Note that, *100-(% error)* is equal to percentage of successful authentication. From the plot, it is easy to observe that CRPs with a higher total delay-difference, r_N, are more immune to bit-flips. Therefore, a logical choice for optimal β is to choose it as high as possible. For example, β=0.3 is a good threshold value to choose for up to a standard deviation of p_i=q=16.7% (equal aging scenario). It corresponds to low bit-flips (∼0.02%) for 16.7% standard deviation. From previous discussion, we observe that β=α=0.4σ≈0.3 (σ=0.75 is the standard deviation of r_N for our chip). Therefore, the same value of threshold (=0.4σ) is useful for two purposes: First, for guaranteeing the stability of CRPs (specified by α) and second, to improve the unreliability due to aging for up to a standard deviation of p_i=16.7% (specified by β). For both cases, CRPs corresponding to total delay-difference, $|r_N|$>α or β are the same and can be termed as *highly stable CRPs*.

In the case of modified feed-forward, the performance is quite similar to that of linear PUF, which is expected. Therefore the thresholds, α and β (=0.4σ) are the same as before. For the case of feed-forward, we observe that the threshold value, β, required for good reliability is much higher. In this case, α=2.4σ≈1.8=β (for up to 33% standard deviation of p_i) is a desirable choice. In general for 33% standard deviation of p_i (or q), thresholds of 0.8σ, 0.8σ and 2.4σ are good choices for linear, modified feed-forward and feed-forward configurations, respectively.

The thresholds, β, chosen in the previous case correspond to the case when the tolerance to error is very low (close to 0%). However in practical authentication scenarios, we can tolerate a certain amount of error in the responses. For example, prior work in [29] considers a tolerance of 10 bits for a 128-bit response. Fig. 11 shows thresholds, β, corresponding to a tolerance of about 1.5%. As can be observed for linear PUF, threshold $\beta^*_{16.67}$ is much less than $0.4\sigma = 0.3$, obtained for low tolerance scenario. This is true for all the three configurations. Hence, in a practical authentication scenario, lower thresholds are good enough for maintaining the required level of reliability (depending on the tolerance). The threshold *vs* percentage successful authentication (or % error) curve (Fig. 11) can be learnt using a polynomial fit of order 3 or more.

IV. Discussion

We observed in our results that the reliability (or intra-chip variation) of a PUF decreased with aging. The effect of aging on other metrics is summarized below:

- *Uniqueness*: Also called inter-chip variation, is the ability of a PUF to produce outputs that are significantly different from other PUFs. From our simulations for synthesized PUFs, we observe that for at least 70% permutations of the chips, uniqueness increases with aging. This is intuitive as aging is essentially a randomization process due to its Gaussian nature.
- *Randomness*: It is the ability of a PUF to produce unbiased '0' and '1' response bits. From Fig. 8, we observe that even for an un-aged PUF, the response is slightly skewed towards bit '1'. Furthermore, with aging we observed that the number of bit flips from Stable-0→1 is much higher. This is due to increase in the mean of total delay-difference, r_N, which is μ_{arb}.

For countering the unreliability in the stable CRPs due to aging, we suggested an approach to tune a threshold, β, based on the total delay-difference, r_N. An un-aged PUF has about 85-90% stable CRPs (Fig. 3). That is, for a 32-bit un-aged PUF, we have more than 2^{31} stable CRPs. However with aging, the number of stable CRPs decreases further (Fig. 8). The number of stable CRPs depends on our choice of threshold, β. For values of β equal to 0.8σ, 0.8σ and 2.4σ for linear, modified feed-forward and feed-forward configurations, we get about 42%, 42% and 2% stable CRPs, respectively. This corresponds to approximately 2^{31}, 2^{31} and 2^{26} number of stable CRPs. However, for a practical authentication scenario, the number of stable CRPs would be considerably higher.

V. Conclusion

This paper discusses the impact of aging on linear and non-linear PUF configurations. We observe that certain structures (like feed-forward) are much more significantly affected by aging than the others. We also observed that the arbiter can largely dictate how a MUX PUF performs with time. Approaches to improve the reliability by estimating new model parameters or by tuning a threshold based on the total delay-difference were discussed. It was further observed that the authentication accuracy of feed-forward PUFs is degraded by 3% if the number of MUX stages increases from 32 to 64 under equal aging scenario. Furthermore, the effect of arbiter aging due to asymmetry is lessened as the number of MUX stages increases. Future work will be directed towards validating the proposed models on test measurements from chips, quantifying the effect of aging due to voltage and temperate and quantifying the effect of aging on higher number of MUX stages.

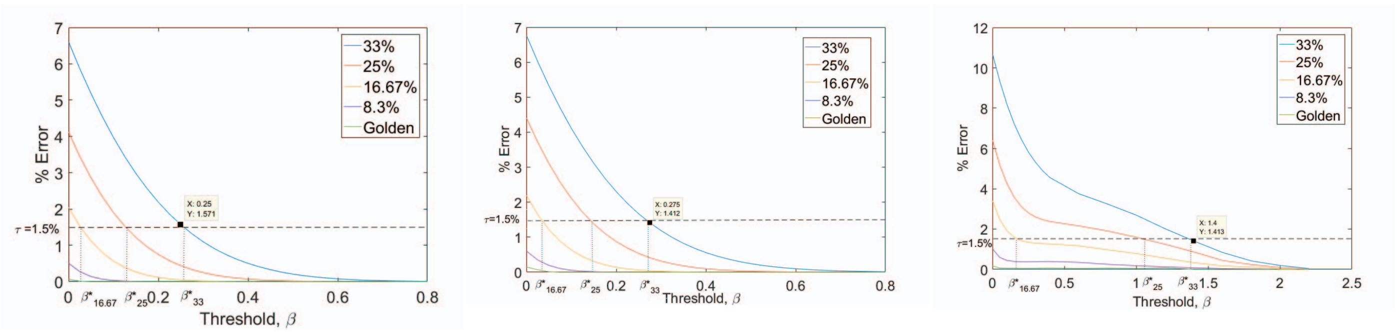

Fig. 11. Thresholds, β, for a fixed error tolerance=1.5% for linear (leftmost), modified feed-forward (middle), and feed-forward (rightmost).

References

[1] B. Gassend, D. Clarke, M. Van Dijk, and S. Devadas, “Silicon physical random functions,” in *Proceedings of the 9th ACM conference on Computer and communications security*. ACM, 2002, pp. 148–160.

[2] C. Herder, M.-D. Yu, F. Koushanfar, and S. Devadas, “Physical unclonable functions and applications: A tutorial,” *Proceedings of the IEEE*, vol. 102, no. 8, pp. 1126–1141, 2014.

[3] Y. Lao and K. K. Parhi, “Statistical analysis of MUX-based physical unclonable functions,” *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 33, no. 5, pp. 649–662, 2014.

[4] S. S. Avvaru, C. Zhou, S. Satapathy, Y. Lao, C. H. Kim, and K. K. Parhi, “Estimating delay differences of arbiter PUFs using silicon data,” in *2016 Design, Automation & Test in Europe Conference & Exhibition (DATE)*. IEEE, 2016, pp. 543–546.

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

[6] S. S. Avvaru, C. Zhou, C. H. Kim, and K. K. Parhi, “Predicting hard and soft-responses and identifying stable challenges of MUX PUFs using ANNs,” in *2011 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS)*. IEEE, 2017, pp. 934–937.

[7] Y. Gao, D. C. Ranasinghe, G. Li, S. F. Al-Sarawi, O. Kavehei, and D. Abbott, “A challenge obfuscation method for thwarting model building attacks on PUFs.” *IACR Cryptology ePrint Archive*, vol. 2015, p. 471, 2015.

[8] V. Stojanovic and V. G. Oklobdzija, “Comparative analysis of master-slave latches and flip-flops for high-performance and low-power systems,” *IEEE Journal of solid-state circuits*, vol. 34, no. 4, pp. 536–548, 1999.

[9] C. Nunes, P. F. Butzen, A. I. Reis, and R. P. Ribas, “A methodology to evaluate the aging impact on flip-flops performance,” in *2013 26th Symposium on Integrated Circuits and Systems Design (SBCCI)*. IEEE, 2013, pp. 1–6.

[10] S. Tajik, E. Dietz, S. Frohmann, J.-P. Seifert, D. Nedospasov, C. Helfmeier, C. Boit, and H. Dittrich, “Physical characterization of arbiter PUFs,” in *International Workshop on Cryptographic Hardware and Embedded Systems*. Springer, 2014, pp. 493–509.

[11] K. Markantonakis and K. Mayes, “Errata to: Secure smart embedded devices, platforms and applications,” in *Secure Smart Embedded Devices, Platforms and Applications*. Springer, 2014, pp. E3–E14.

[12] D. Ganta and L. Nazhandali, “Study of IC aging on ring oscillator physical unclonable functions,” in *Fifteenth International Symposium on Quality Electronic Design*. IEEE, 2014, pp. 461–466.

[13] J. Keane, X. Wang, D. Persaud, and C. H. Kim, “An all-in-one silicon odometer for separately monitoring HCI, BTI, and TDDB,” *IEEE Journal of Solid-State Circuits*, vol. 45, no. 4, pp. 817–829, 2010.

[14] A. Tiwari and J. Torrellas, “Facelift: Hiding and slowing down aging in multicores,” in *2008 41st IEEE/ACM International Symposium on Microarchitecture*. IEEE, 2008, pp. 129–140.

[15] Z. C. Jouini, J.-L. Danger, and L. Bossuet, “Performance evaluation of physically unclonable function by delay statistics,” in *New Circuits and Systems Conference (NEWCAS), 2011 IEEE 9th International*. IEEE, 2011, pp. 482–485.

[16] Z. Tariguliyev and B. Ors, “Reliability and security of arbiter-based physical unclonable function circuits,” *International Journal of Communication Systems*, vol. 26, no. 6, pp. 757–769, 2013.

[17] N. Karimi, J.-L. Danger, F. Lozach, and S. Guilley, “Predictive aging of reliability of two delay PUFs,” in *International Conference on Security, Privacy, and Applied Cryptography Engineering*. Springer, 2016, pp. 213–232.

[18] V. G. Rao and H. Mahmoodi, “Analysis of reliability of flip-flops under transistor aging effects in nano-scale CMOS technology,” in *2011 IEEE 29th International Conference on Computer Design (ICCD)*. IEEE, 2011, pp. 439–440.

[19] S. Pae, J. Maiz, C. Prasad, and B. Woolery, “Effect of BTI degradation on transistor variability in advanced semiconductor technologies,” *IEEE Transactions on Device and Materials Reliability*, vol. 8, no. 3, pp. 519–525, 2008.

[20] S. E. Rauch, “Review and reexamination of reliability effects related to NBTI-induced statistical variations,” *IEEE Transactions on Device and Materials Reliability*, vol. 7, no. 4, pp. 524–530, 2007.

[21] C. Zhou, S. Satapathy, Y. Lao, K. K. Parhi, and C. H. Kim, “Soft response generation and thresholding strategies for linear and feed-forward mux pufs,” in *Proceedings of the 2016 International Symposium on Low Power Electronics and Design*. ACM, 2016, pp. 124–129.

[22] X. Xu, W. Burleson, and D. E. Holcomb, “Using statistical models to improve the reliability of delay-based PUFs,” in *2016 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*. IEEE, 2016, pp. 547–552.

[23] G. Marsaglia, “Ratios of normal variables and ratios of sums of uniform variables,” *Journal of the American Statistical Association*, vol. 60, no. 309, pp. 193–204, 1965.

[24] Y. Lao, Q. Tang, C. H. Kim, and K. K. Parhi, “Beat frequency detector–based high-speed true random number generators: Statistical modeling and analysis,” *ACM Journal on Emerging Technologies in Computing Systems (JETC)*, vol. 13, no. 1, p. 9, 2016.

[25] D. Lim, J. W. Lee, B. Gassend, G. E. Suh, M. Van Dijk, and S. Devadas, “Extracting secret keys from integrated circuits,” *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 13, no. 10, pp. 1200–1205, 2005.

[26] D. M. Endres and J. E. Schindelin, “A new metric for probability distributions,” *IEEE Transactions on Information theory*, vol. 49, no. 7, pp. 1858–1860, 2003.

[27] N. Henze and M. D. Penrose, “On the multivariate runs test,” *Annals of statistics*, pp. 290–298, 1999.

[28] S. Kullback and R. A. Leibler, “On information and sufficiency,” *The annals of mathematical statistics*, vol. 22, no. 1, pp. 79–86, 1951.

[29] 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.