RON: An On-Chip Ring Oscillator Network for Hardware Trojan Detection

Xuehui Zhang and Mohammad Tehranipoor
ECE Department, University of Connecticut
{Xuehui.Zhang, tehrani}@engr.uconn.edu

Abstract

Integrated circuits (ICs) are becoming increasingly vulnerable to malicious alterations, referred to as hardware Trojans. Detection of these inclusions is of utmost importance, as they may potentially be inserted into ICs bound for military, financial, or other critical applications. A novel on-chip structure including a ring oscillator network (RON), distributed across the entire chip, is proposed to verify whether the chip is Trojan-free. This structure effectively eliminates the issue of measurement noise, localizes the measurement of dynamic power, and additionally compensates for the impact of process variations. Combined with statistical data analysis, the separation of process variations from the Trojan contribution to the circuit's transient power is made possible. Simulation results featuring Trojans inserted into a benchmark circuit using 90nm technology and experimental results on Xilinx Spartan-3E FPGA demonstrate the efficiency and scalability of the RON architecture for Trojan detection.

I. Introduction

Due to the globalization of the semiconductor design and fabrication processes, integrated circuits (ICs) are becoming increasingly vulnerable to malicious inclusions and alterations (hardware Trojans) [1]. These inclusions may disable an IC at a target time in the future, or potentially leak confidential information to an adversary. An adversary may embed a hardware Trojan featuring different physical, activation, or functional characteristics [2] into the unused spaces of an IC. In general, Trojan detection is difficult for several reasons: (*i*) Trojans may be activated under very specific conditions. (*ii*) hardware Trojans may be decomposed into different categories based on: structure, function, distribution, parameters, and size, however, it is impossible to model all possibilities, then utilize these models to identify Trojans in a design by comparison. (*iii*) tests used to detect manufactured faults such as stuck-at faults and delay faults cannot guarantee Trojan detection, since the flow utilizes nets in a circuit rather than the circuit's function.

A. Previous Work

The topic of IC trust has gained considerable attention in the past few years, yielding several approaches proposed for detection of hardware Trojans. Generally, the detection methods are classified into three categories: side-channel signal analysis, Trojan activation, and monitoring architectures. Side-channel signal analysis has been utilized to detect hardware Trojans by measuring circuit parameters. Examples of this include: power-based analysis [3] [4] [8], current analysis [5], and delay-based analysis [6] [7]. The authors in [3] were the first to use power signatures, according to the survey [8], to measure the power contribution of Trojans by applying random patterns, and observing the power consumption. Side-channel analysis methods are effective for Trojans that have a significant effect on power, current, and delay. However, there are many variables which affect these parameters, such as measurement noise, process variations, and environmental variations, and may mask Trojan's contribution to the side-channel signals.

Several strategies are presented to fully activate hardware Trojans and then detect them [9] [10] [11]. The disadvantage of the Trojan activation methods is in the difficulty of activating Trojans that are designed to be enabled under specific conditions and the inability in detecting many of the non-functional Trojans listed in the taxonomy developed in [2]. A hardware threat modeling concept is suggested in [12]. Given that for every IC, there are an exponential number of different configurations for a Trojan, it will be impossible to model every variation with the intent of comparison for Trojan identification.

Monitoring structures have been proposed to prevent the damages caused by Trojans. A system-on-chip (SoC) design with design-for-enabling security logic is suggested in [13] to monitor the most significant signals in the system. In [14], the Trojan-resistant SoC bus architecture can prevent untrusted access to the secure memory or the data contained within. Once the bus has detected malicious data, it will block the attacking packet and report it to the system, which will reset and initialize necessary registers.

B. Contributions and Paper Organization

The power signature of an IC with a Trojan will be different from that of a Trojan-free IC. In certain situations, the power fingerprint may be too vague to be detected by previously proposed Trojan detection methods [8] due to measurement noise, process variations, and less sensitivity to smaller Trojans. In this paper, we propose a new structure, called ring oscillator network (RON), featuring the ability to detect Trojans that cause power fluctuations, thereby uncovering the malicious inclusion. A number of ring oscillators (ROs) acting as *power monitors*, distributed across the entire IC, constitute the RON, which takes into account the noise caused by the Trojan gates and those caused by both inter-die and intra-die process variations. The output of each ring oscillator represents one part of the power signature of the entire IC. With N_{RO} ring oscillators in the IC, a series of power signatures can be generated by the RON. An off-chip test equipment would be able to select which ring oscillator should be used to generate the signature and could disable the RON when IC operates in functional mode. The number of ring oscillators, N_{RO}, could be adjusted according to the size of the IC and sensitivity to Trojans, thereby scaling the network and optimizing Trojan detection. Simulation and FPGA implementation results demonstrate that the RON combined with statistical data analysis effectively distinguishes the power differences caused by Trojans from those of process variations, and identifies inserted hardware Trojans in the IC. RON presents a small area overhead and is resilient to removal, tampering, and modeling attacks.

The rest of the paper is organized as follows: Section II analyzes the impact of power supply noise on ring oscillators. Section III presents the RON architecture and statistical data analysis flow is described in Section IV. Simulation results as well as FPGA implementation results are presented in Section V. Finally, and concluding remarks are given in Section VI.

II. Analyzing Impact of Power Supply Noise on Ring Oscillators

Two simple five-stage ring oscillators are shown in Figure 1: the ring oscillator in Figure 1(a) consists of inverters and the ring

978-3-9810801-7-9/DATE11/©2011 EDAA.

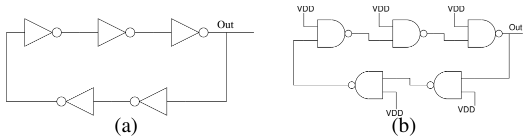


Fig. 1. Five-stage ring oscillators

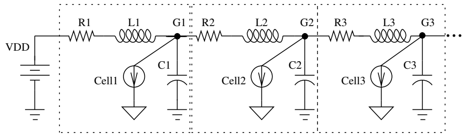


Fig. 2. The RLC model of a simple power line in a power distribution network

oscillator in Figure 1(b) is composed of NAND gate. The second ring oscillator has a higher sensitivity to supply noise since one of its inputs is connected to power supply but offers larger area overhead. In this work, we only use the first ring oscillator as power monitor due to its easier analytical analysis. The frequency of this ring oscillator is determined by the total delay of all the inverters, in the presence of supply voltage and process variations. Assume that each stage in the ring oscillator provides a delay of t_d. The delay of the n-stage ring oscillator is approximately $2*n*t_d$ and the oscillation frequency will be:

$$f = \frac{1}{2*n*t_d} \tag{1}$$

The delay of each inverter varies according to parameters such as temperature, supply voltage (VDD), load capacitance (CL), threshold voltage (Vth), channel length (L), oxide thickness (Tox), and transistor channel width (W). Since all ICs can be tested under the same temperature, the environmental variation will not be considered in this work. All the remaining parameters are susceptible to process variations and power supply noise.

Power supply noise (also known as voltage drop) impacts the delay of the logic gates. When the voltage drops, the delay of the gates increases. Thus, a change in the supply voltage of any inverter in a ring oscillator impacts the delay of all associated gates, and therefore impacts the oscillation frequency. Concerning today's tightly designed power supply distribution networks, transitions in some gates can impact the power supply of other gates within close proximity [15]. Figure 2 shows a simple power line model in which VDD supplies one row in standard cell design. The indicated VDD represents the point where a via connects the power rail to the upper metal layer in a power distribution network. Nodes G1, G2, and G3 connect to adjacent cells represented as current source for Cell 1, Cell 2, and Cell 3. Here, for sake of simplicity, the power via is assumed to have zero impedance and each interconnect is modeled by a resistance, inductance, and capacitance (RLC) network. The contribution of each current source to the overall noise is described in Equation 2 where $V1$, $V2$, and $V3$ (voltage at nodes G1, G2, and G3) are the power supply noise spectrum, $Vii = Z_{ii} * I_{ii}(i = 1,2,3)$ (Z_{ii} is the impedance of node i and I_{ii} is the current) is the power noise, $\rho_{ij}(i, j = 1,2,3)$ is voltage division coefficient, and ω is the frequency of the circuit. From the equation, we can see that $V1$, $V2$, and $V3$ are related to the neighboring gates, demonstrating that a gate's transition has effect on neighboring gates connected to the same VDD line.

$$\begin{aligned} V1 &= V11 + \rho_{21}(\omega) * V22 + \rho_{31}(\omega) * V33 \\ V2 &= \rho_{12}(\omega) * V11 + V22 + \rho_{32}(\omega) * V33 \\ V3 &= \rho_{13}(\omega) * V11 + \rho_{23}(\omega) * V22 + V33 \end{aligned} \tag{2}$$

For Trojan-inserted ICs, the switching gates in the Trojan would cause small voltage drop on the VDD line and ground bounce on the VSS line. Thus, with the same input patterns, the power supply noise affecting the Trojan-free IC and Trojan-inserted IC will differ. In order to verify the impact of the Trojan on the frequency of the ring oscillator, we implemented a 5-stage ring oscillator (shown in Figure 1(a)) in 90nm technology for simulation.

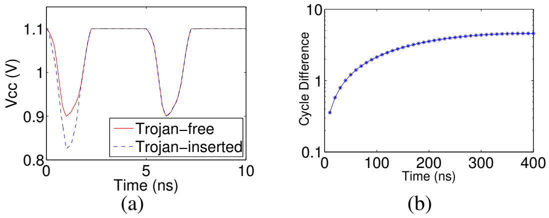


Fig. 3. (a) Power supply variations for Trojan-free and Trojan-inserted circuits; (b) Cycle difference caused by Trojan gates' switching.

In Figure 3(a), assume that the dashed line denotes the dynamic power in the presence of a Trojan and the solid line denotes the Trojan-free power (assuming $VDD = 1.1V$). As can be observed, the two supply voltages only differ during the first 2ns. These two power waveforms are applied to the ring oscillator for 400ns. Figure 3(b) shows the cycle count difference due to the extra noise caused by the Trojan. At time 0, the two ring oscillators denoting *with* and *without* an inserted Trojan have the same period. However, with the presence of power supply noise, the difference will grow steadily as the measurement duration increases.

III. Ring Oscillator Network

As mentioned earlier, Trojan gates switching impacts the frequency of a ring oscillator due to injected power supply noise. Process variations can impact the threshold voltage, channel length, and oxide thickness in circuit gates which, in turn, impacts power supply noise distribution in an IC. Since these effects may be localized, one ring oscillator may not have enough sensitivity to distinguish the effect of Trojans and process variations. A ring oscillator placed in one corner of an IC may not be able to capture noise effects which occur due to a Trojan placed in another corner of the IC. A ring oscillator network however can improve the sensitivity to Trojan noise, and increase the accuracy in determining Trojan's contributions using relative values.

Our RON is composed of N_{RO} ring oscillators distributed across the entire IC. For different ICs, the number of ring oscillators can be adjusted accordingly depending on the sensitivity of the ring oscillators to the gate switching in a pre-determined proximity. The output of RON in Trojan-free ICs generates a power signature. Similar to previous methods [3] [4] [5] [6], in this work, we assume that a number of golden ICs can be identified via a thorough test process. If the output of an IC under authentication is not compatible with the expected signature, the IC may contain a Trojan. We acknowledge that our proposed architecture can also use power signatures generated during simulation for Trojan detection eliminating the need for the golden IC. Demonstration of this fact is part of our future work and is outside of the scope of this paper.

The oscillation cycle count generated from the ring oscillators in the RON is used to generate the IC's signature. For i_{th} ring oscillator, the total accumulated cycles, C_i, in the measurement time T is:

$$C_i = \int_0^T \frac{1}{2*n*t_{di}(t)} dt \tag{3}$$

where $t_{di}(t)$ is the inverter delay which will vary with time as the input patterns change. Let $\Delta t_{dti}(t)$ represent the change in inverter delay of i_{th} ring oscillator caused by Trojan effects and C_{TFi} and C_{TIi} denote the total cycle count for Trojan-free and Trojan-inserted

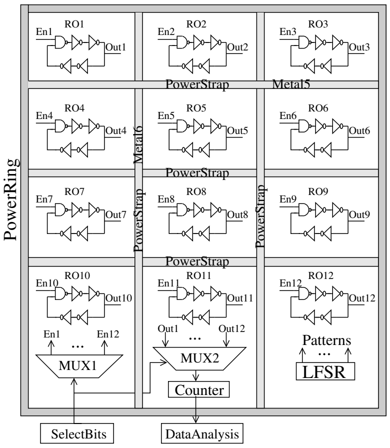


Fig. 4. A RON with N_{RO}=12 ring oscillators distributed in the circuit layout.

ICs, respectively. The effect a Trojan has on i_{th} ring oscillator (ΔC_i) is presented by Equation 4. The value of ΔC_i is related to the number of stages in a ring oscillator (n), the measurement time (T), and the Trojan's impact on inverter delay ($\Delta t_{dti}(t)$). The Trojan's impact on a ring oscillator is determined by the size of the Trojan, switching activity of the Trojan, and the distance between the Trojan and the ring oscillator.

$$\Delta C_i = C_{TIi} - C_{TFi} = -\int_0^T \frac{\Delta t_{dti}(t)}{2*n*t_{di}(t)*(t_{di}(t)+\Delta t_{dti}(t))}\,dt \tag{4}$$

Figure 4 shows the proposed ring oscillator network with N_{RO}=12 oscillators inserted into the ISCAS'89 s9234 benchmark circuit according to power straps in the layout. One RO is inserted into each grid surrounded by power straps. In addition, one multiplexer is used to select a ring oscillator in the network to be enabled during the authentication and another multiplexer chooses the same ring oscillator to be recorded. When the IC is in functional mode, the RON would be disabled and will have no impact on circuit dynamic power. The counter in the architecture will calculate the oscillation cycles occurring in the selected ring oscillator. Since N_{RO} ring oscillators are used to generate the signature, the same pattern set generated by linear feedback shift register (LFSR) must be applied to the IC N_{RO} times.

The RON architecture has a small area overhead, mainly caused by the counter and LFSR. For instance, the overhead is 10.8% for the smaller benchmark circuit, s9234 (two vertical power straps and three horizontal power straps, $N_{RO} = 12$), 3.6% for s35932 benchmark circuit (three vertical power straps and three horizontal power straps, $N_{RO} = 16$), and 0.9% for DES circuit (five vertical power straps and five horizontal power straps, $N_{RO} = 30$). We believe that the area overhead will be negligible for larger circuits even if N_{RO} increases considerably based on power planning, since the counter size does not increase linearly with N_{RO}. Also, LFSR is commonly used for built-in self-test (BIST) in modern designs.

The RON is resilient to removal and tampering attacks. It is inherently difficult for an attacker to remove the ring oscillator network, due to (*i*) its distributed placement throughout the entire IC and (*ii*) the expected measurement results from each ring oscillator, i.e., the designer relies on the ability to capture RON data from each embedded ring oscillator. If a specific ring oscillator is not reporting data, the designer should assume the design has been attacked. On the other hand, ring oscillator is sensitive to its stage count and inverter type. For the RON inserted by the designer, the frequency falls in a certain range considering variations. If one of them is not within the range, it must be tampered with. In addition, similar to ring oscillator based physical unclonable functions (PUFs), the RON architecture is also resilient to modeling and reverse engineering attacks.

```
Step 1: Measurement
Collect data from $N_{TF}$ Trojan-free ICs with $N_{RO}$ ring oscillators
for ($i = 1, i <= N_{RO}, i++$) {                         //select $i^{th}RO$
  for ($j = 1, j <= N_{RO}, j++$) ($j \neq i$) {          //select $j^{th}RO$
    for ($k = 1, k <= N_{TF}, k++$) {           //select $k^{th}$ Trojan-free IC
      $x_{ki}=(\sum_{m=1}^{N_{RO}} C_{km} - C_{ki})/C_{ki}$;
      $y_{kj}=(\sum_{m=1}^{N_{RO}} C_{km} - C_{ki})/C_{kj}$;
      plot($x_{ki}$, $y_{kj}$);
    }
    The power signature, $PS_{ij}$, is created from all $N_{TF}$ ICs.
  }     //$x_{ki}$ is named as the FirstVector
}     //$y_{kj}$ is named as the SecondVector
Note: $C_{km}$, $C_{ki}$, $C_{kj}$: Oscillation cycle count of $RO_m$, $RO_i$, $RO_j$ in $k^{th}$ IC

Step 2: Authentication
For each IC under authentication:
Collect data from $N_{RO}$ ring oscillators ($C_{RON} = \sum_{i=1}^{N_{RO}} C_i$).
for ($i = 1, i <= N_{RO}, i++$)
  { $x = (C_{RON} - C_i)/C_i$;
  for ($j = 1, j <= N_{RO}, j++$) ($j \neq i$);
    { $y = (C_{RON} - C_i)/C_j$;
    plot($x$,$y$);
    if (($x$,$y$) is outside of the power signature $PS_{ij}$
      {The IC is Trojan-inserted; Break; }
    else go on;
  }
}
Note: $C_i$, $C_j$: Oscillation cycle count of $RO_i$, $RO_j$
```

Fig. 5. Advanced outlier analysis procedure.

IV. Statistical Analysis

When the Trojan is small or widely distributed, distinguishing between noise generated by Trojan gates and process variations may be exceedingly difficult. Therefore, as an extension, a signature must be generated by recording all ring oscillators cycle count from a large number of ICs of the same design. Since the ICs will all be subject to different process variations, this signature can be statistically more tolerant to errors. In order to separate the effect of process variations and Trojans, a data analysis flow is suggested in this work, which includes three methods namely: (i) Simple Outlier Analysis, (ii) Principal Component Analysis (PCA), and (iii) Advanced Outlier Analysis. Simple outlier analysis offers least complexity compared with the other two data analysis methods.

Simple outlier analysis is based on the oscillation cycle distribution of each ring oscillator in the RON. For each ring oscillator, the oscillation cycle is within a certain range for Trojan-free ICs. If the oscillation cycle of one ring oscillator in the IC under authentication is outside of the range, this IC is considered suspicious and might contain a Trojan. This method uses the information from individual ring oscillators but not the relationship between them in the RON. Usually, this method can identify a small number of Trojan-inserted ICs but not most based on our results. If oscillation cycle count of all ring oscillators in an IC under authentication is within each Trojan-free IC's signature, the data collected from this IC will be processed by PCA and advanced outlier analysis.

The concept of principal component analysis [16] is used to account for the N_{RO} variables (one variable represents one ring oscillator). The relationship between the data from the N_{RO} ring oscillators is considered by PCA when it transforms the N_{RO} variables into uncorrelated variables. For example, noting similarities in oscillation readings between two adjacent ring oscillators, would imply a correlation in the data. The oscillation cycle count of N_{RO} ring oscillators in the Trojan-free ICs will be analyzed by PCA and convex hull [17] is constructed with the first three components. If the output of RON is beyond the convex, a Trojan must exist in the IC under authentication. However, if the output is inside the convex, advanced outlier is used for futher analysis and validation.

Advanced outlier analysis is developed to identify the ICs with Trojan that cannot be detected by simple outlier analysis and PCA. It considers the relationship between ring oscillators in the RON. The pseudo-code is shown in Figure 5, which consists of two steps. The first step, Measurement, generates N_{RO}*(N_{RO}-1) power signatures from N_{TF} Trojan-free ICs. For each Trojan-free IC, the total oscillation cycle count from the RON is $C_{RON} = \sum_{m=1}^{N_{RO}} C_m$. Then, the data from the RO_i (C_i) and RO_j ($j \neq i$) (C_j) are selected to calculate $x_i = (C_{RON} - C_i)/C_i$ and $y_j = (C_{RON} - C_i)/C_j$. Finally, ($x_i$, y_j) from all the Trojan-free ICs would be plotted to generate PS_{ij} power signature. Thus, N_{RO}*(N_{RO}-1) power signatures can be generated. The second step, Authentication, deals with the IC under authentication using the same process. If one of the IC's signatures is beyond the N_{RO}*(N_{RO}-1) power signatures, then it is assumed to contain a Trojan.

V. Results and Analysis

In order to verify the effectiveness of the RON architecture, we implemented $N_{RO} = 12$ ring oscillators with 5-stage inverters in (i) s9234 benchmark using 90nm technology, including 2 vertical and 3 horizontal power straps, for IC simulation and (ii) AES circuit on Xilinx Spartan-3E FPGA for hardware validation. For IC simulation, six Trojans (T_1 through T_6) with different sizes, distributions, and switching activities are inserted into s9234 benchmark. s9234, which is a small benchmark with 145 flip-flops and 420 gates, is selected for simulation rather than AES (6,089 flip-flops and 18,103 gates) to be able to run the very slow process of Monte Carlo simulations. Few of the Trojans can change the output of the original circuits when they are enabled. The location of the ring oscillators and Trojans are shown in Figure 6. The dark-colored circles in the figure represent the corresponding regions used in the actual layout by the Trojans. Four of the Trojans (T_1, T_2, T_4, and T_5) are placed around the ring oscillator RO8. Gates in Trojans (T_3 and T_6) are distributed at different regions within close proximity to RO5, RO7, RO8, and RO9. All Trojans have passed our validation test suite including 100,000 random functional patterns as well as structural patterns generated using automatic test pattern generation (ATPG) tool. During simulation, the same input patterns generated by LFSR are applied to all ICs, including those which are Trojan-free, to ensure the variable $t_d(t)$ in Equation 4 is identical. The fast Spice simulation tool Nanosim from Synopsys is used to conduct the power analysis and collect the oscillation cycle count in presence of process variations.

A. Trojan Distribution Analysis

As previously mentioned, six Trojans with different distributions (see Figure 6) are inserted into the benchmark to verify the Trojans' distribution impact on the RON. The counter results without process variations are shown in Table I when simulating for 1μs and applying 100 patterns. Only RO1, RO5, RO8, and RO12 are selected to show detailed results. ΔC shows the difference in oscillation cycle count between the Trojan-inserted (C_{TI}) and Trojan-free (C_{TF}) ICs. From the table, we can see that all the ΔC entries are negative. This occurs as a result of the Trojan gates' impact on VDD noise, thereby increasing the delay of RO gates.

Table I shows that T1, T2, T4, and T5 have a larger impact on the oscillation frequency of RO8 than the other ring oscillators. This is because the power supply voltage is related to the voltage division coefficient, which is partially determined by the distance between two gates. The smaller this distance, the greater impact the Trojan gates have on the ring oscillators. Contrarily, for T3 and T6, there is a larger impact on RO5 and RO8 than RO1 and RO12.

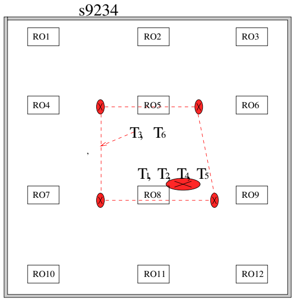


Fig. 6. s9234 with 12 ROs and 6 Trojans. One Trojan at a time is inserted into the circuit.

Thus, for a distributed Trojan, the combined effect on multiple ROs can be used for detection.

B. Trojan Size Analysis

The six inserted Trojans are designed with varying sizes to analyze the impact they would have on the RON architecture. T1, T2, and T3, are composed of 8 inverters, 12 inverters, and 25 inverters, respectively. 8 combinational gates consisting of AND, INV, and OR constitute T4, while T5 and T6 are comprised of 25 and 22 combinational gates, respectively. We observed that in T1, T2, and T3, the oscillation cycle count difference of RO8 increased with Trojan size from -31 (for T1) to -59 (for T3). This occurred due to the greater power supply noise imparted from the Trojan gates. As the power supply voltage is lowered, the speed of the ring oscillator is dropped. For T4, T5, and T6, we observed similar results. In general, the greater the size of the Trojan, the larger impact it can have on the power supply network and consequently the greater impact on the ring oscillators.

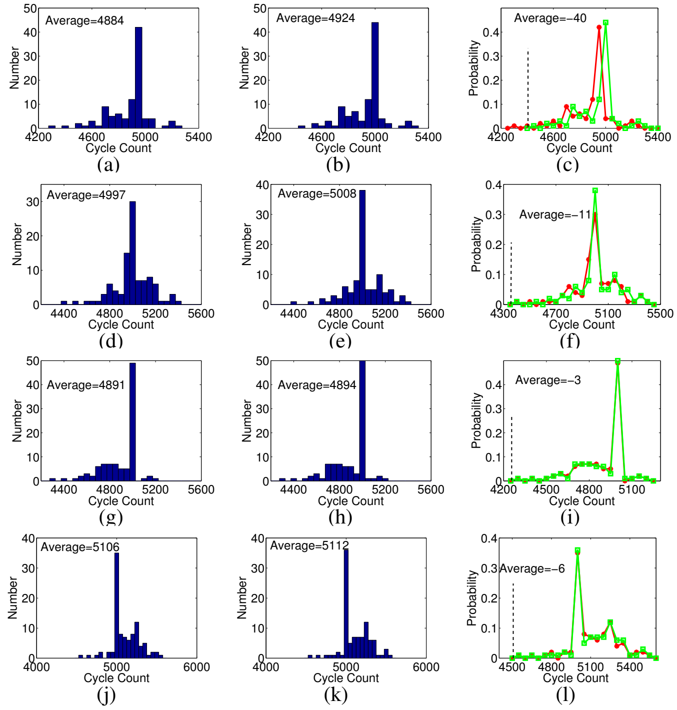


Fig. 7. Oscillation cycle distribution of RON with 100 Monte Carlo simulations when T5 is inserted in s9234. (a) RO8 with Trojan; (b) RO8 w/o Trojan; (c) Cycle count distribution of RO8; (d) RO5 with Trojan; (e) RO5 w/o Trojan; (f) Cycle count distribution of RO5; (g) RO1 with Trojan; (h) RO1 w/o Trojan; (i) Cycle count distribution of RO1; (j) RO12 with Trojan; (k) RO12 w/o Trojan; (l) Cycle count distribution of RO12.

C. Trojan Switching Activity Analysis

Trojan size is not the only parameter impacting the frequency of the ring oscillators.The Trojan switching activity plays an important role as well. In the interest of simulation running time, we designed few Trojans featuring frequent switching activities; e.g., T1, T2, and T3 switch 760 times, 1140 times, and 2375 times

TABLE I

OSCILLATION CYCLE COUNT OF RING OSCILLATORS IN PRESENCE OF TROJAN GATES SWITCHING WITHOUT PROCESS VARIATIONS

RO	T1			T2			T3			T4			T5			T6		
	C_{TI}	C_{TF}	ΔC	C_{TI}	C_{TF}	ΔC	C_{TI}	C_{TF}	ΔC	C_{TI}	C_{TF}	ΔC	C_{TI}	C_{TF}	ΔC	C_{TI}	C_{TF}	ΔC
RO1	4933	4939	-6	4985	4989	-4	4944	4965	-21	4999	4999	0	4976	4985	-9	5000	5000	0
RO5	4735	4744	-9	4740	4749	-9	4908	4948	-40	4906	4925	-19	4989	4994	-5	4792	4819	-27
RO8	4714	4545	-31	4932	4974	-42	4796	4855	-59	4604	4635	-31	4936	4981	-45	4925	4974	-49
RO12	5279	5282	-3	4999	4999	0	4943	4966	-23	5027	5031	-4	5054	5062	-8	5242	5250	-8

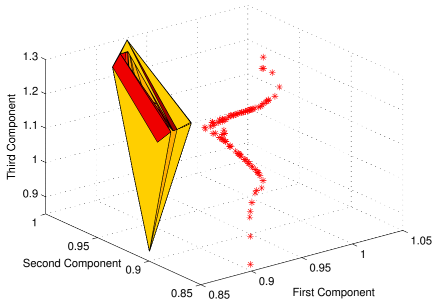


Fig. 8. Power signature using PCA for Trojan-free ICs and Trojan-inserted ICs with T5.

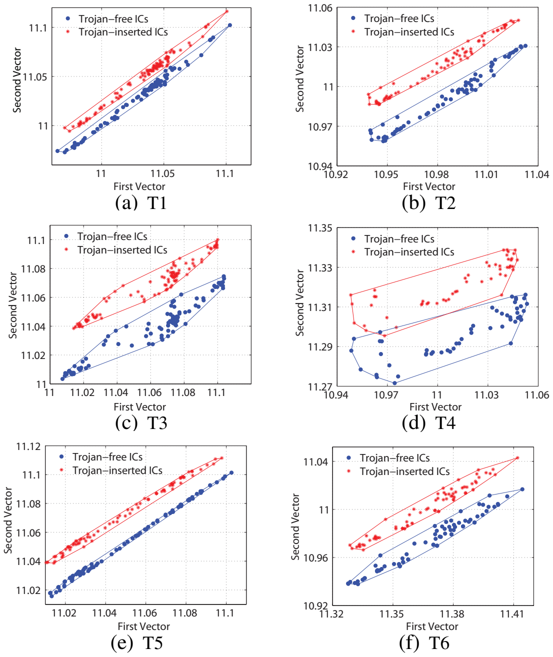
respectively during the pattern application period. T4, T5, and T6 switch 665 times, 2090 times and 1850 times during the simulation. From the Table I, one can notice the trend: the more frequently the Trojan switches, the greater the voltage drop imparted on the ring oscillator gates, which in turn, impacts oscillation cycle count reported by the ring oscillator.

D. Process Variations Analysis

Random process variations, consisting of 10% voltage threshold (8% inter-die and 2% intra-die), 3% oxide thickness (2% inter-die and 1% intra-die), and 10% channel length (8% inter-die and 2% intra-die) in 90nm technology library, are used in the following simulations. All the simulations are done under temperature 25 °C. 100 Trojan-free ICs and 600 Trojan-inserted ICs (100 per Trojan) are generated by Monte Carlo simulations. The statistical data analysis flow proposed in the previous section processed the data collected from these ICs. T5 is used to show the detailed results of the data analysis flow.

Simple outlier analysis is first applied to distinguish the effect of Trojan and process variations. Histograms obtained from RO1, RO5, RO8, and RO12 are shown in Figure 7, each showing the distribution of oscillation cycle count plotted from the data obtained in the presence of process variations with T5. Figure 7(a) displays the histogram of the cycle count of oscillations reported by RO8 with the Trojan inserted and Figure 7(b) shows the same result without (w/o) the Trojan. The distribution of the two sets of oscillation cycle count are plotted in Figure 7(c). The remaining figures (7(d)-7(l)) show the data distribution collected from RO5, RO1, and RO12, respectively. We do not notice a significant change in RO5, RO1, and RO12. However, due to the presence of T5, RO8's distribution shifts toward left considerably. For RO8, the oscillation cycle range is 4400 – 5350 in Trojan-free ICs and the boundary is marked by the black dashed line in Figure 7(c). 3 ICs out of the 100 ICs under authentication fell outside of the range, which are identified to contain Trojan.

For the remaining 97 ICs, PCA is done to analyze the data. Figure 8 shows the power signature comparison using PCA for Trojan detection. The convex is drawn from the first three principal components with Trojan-free ICs. The asterisks denote data obtained from ICs with the inserted Trojan, which are shown to be separate from the convex hull. Thus, with the RON architecture and statistical analysis, T5 can be detected with 100% accuracy. However, limited by the statistical methods and the increasingly larger process variations of nano-scale technologies, *smaller Trojans* may not necessarily be detected with such accuracy.

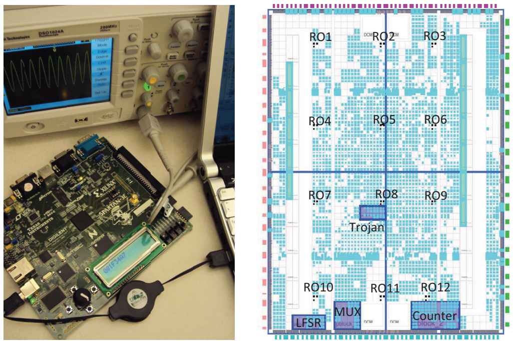


(a) T1 (b) T2 (c) T3 (d) T4 (e) T5 (f) T6

Fig. 9. Power signatures with advanced outlier data analysis from IC simulation.

Thus, advanced outlier analysis shown in Figure 5 is also used to identify Trojan-inserted ICs. There are a total of 12*11=132 power signatures generated by the Trojan-free ICs. In the following, for advanced outlier analysis results, only the power signature that can detect the most Trojan-inserted ICs is shown. As an example, for T5, Figure 9(e) shows the advanced outlier analysis result. The blue dots represent Trojan-free ICs and the red asterisks denote Trojan-inserted ICs. We can see that all of the Trojan-inserted ICs are outside of the Trojan-free ICs. Thus, the detection rate with T5 using advanced outlier analysis is 100%.

Similarly, the remaining 5 Trojans (T1, T2, T3, T4, and T6) with 100 Trojan-free ICs and 100 Trojan-inserted ICs are also simulated and the data analysis flow is applied for every Trojan. By simple outlier analysis, one Trojan-inserted IC is detected with T1, T2, and T4 and two Trojan-inserted ICs are identified with T3 and T6. Using PCA, Trojan-inserted ICs detected with T1, T2, T3, T4, and T6 are 16, 17, 8, 10, and 29, respectively. The remaining Trojan-inserted ICs are analyzed by advanced outlier analysis, shown in figures 9(a) - 9(f). In order to show the effectiveness of RON when using our advanced outlier analysis, the Trojan-inserted ICs detected by simple outlier analysis and PCA are also plotted in these figures. Combined simple outlier analysis, PCA, and advanced outlier analysis, the Trojan detection rates for T2, T3, and T6 are 100%. For smaller Trojan T1, the detection rate is 100%, even though the Trojan-inserted ICs are so close to the Trojan-free ICs. For T4, 98% Trojan-inserted ICs are detected. Note that the detection rates presented above are all only from the best distributions selected from 132 power signatures. When we analyze all power signatures, the detection rate for all Trojans including T4 is 100%. We acknowledge that further analysis is needed for very small Trojans, different pattern sets, and Trojans that switch rarely or even do not switch. Note that the more effective the pattern set is in generating switching in the Trojan circuit, the more effective the RON will be.



(a) (b)

Fig. 10. (a) Xilinx Spartan-3E FPGA board and (b) AES layout after placement.

E. Validation on Spartan-3E FPGA

The same RON architecture is applied to AES implemented on Xilinx Spartan-3E FPGA (shown in Figure 10(a)). Three Trojans (T7, T8, and T9) are inserted into the benchmark. T9 consists of 80 gates. The area overhead of T7 is 0.17%, T8 is 0.25%, and T9 is 0.33%. 24 Trojan-free FPGAs and 24 Trojan-inserted FPGAs are used. The oscillation cycle count from different FPGAs represent inter-die process variations and the oscillation cycle count from the same FPGA but different ring oscillators denote intra-die process variations.

The layout of FPGA after the placement and routing is shown in Figure 10(b). 12 ring oscillators with five inverters constitute RON while Trojans are placed near RO8. LFSR module generates patterns during authentication process. Multiplexer module selects which ring oscillator would be enabled and recorded. The implementation temperature is 25°C. Several measurements are done for each ring oscillator in every FPGA in order to eliminate the measurement noise, and the average value is used to perform data analysis.

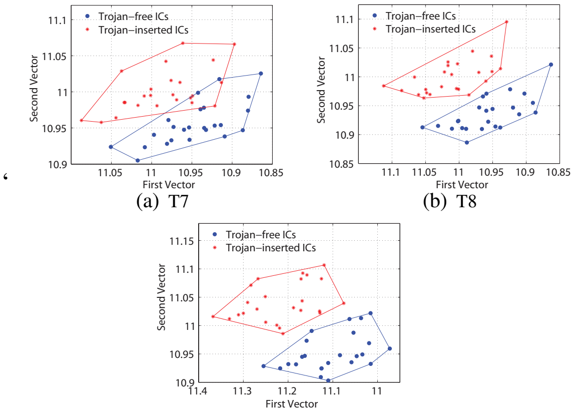
One Trojan-inserted FPGA is detected by simple outlier analysis for each Trojan. PCA detects 9 Trojan-inserted FPGAs with T7, 10 Trojan-inserted FPGAs with T8, and 16 Trojan-inserted FPGAs with T9. The remaining Trojan-inserted FPGAs are analyzed by advanced outlier analysis (shown in Figure 11). In order to show all the detected Trojan-inserted FPGAs, the FPGAs detected by simple outlier analysis and PCA are also plotted in these figures. From combined simple outlier analysis, PCA, and advanced outlier analysis, 100% Trojan-inserted FPGAs are detected for T8 and T9 but 80% for T7 from the best selected power signature. Performing similar analysis for all power signatures, we are able to increase the Trojan detection rate to 92% for T7. In addition, we also implemented Trojans of smaller size (T10=30 gates and T11=20 gates) to verify the sensitivity of RON. Trojan detection rate is 92% for T10 but 100% for T11 since it experiences more switching activity.

VI. Conclusion

We have presented an effective structure to detect hardware Trojans inserted into an IC. The RON architecture generates a power supply fingerprint, used to identify malicious alterations. Statistical analysis distinguishes the effects of hardware Trojans from process variations. The experimental results demonstrate that this approach is very effective in identifying Trojan-inserted ICs.

VII. Acknowledgment

This work was supported in part by National Science Foundation (NSF) grants CNS-0844995 and CNS-0716535.



(a) T7 (b) T8

(c) T9

Fig. 11. Advanced outlier analysis results from FPGA implementation.

References

[1] “Report of the Defense Science Board Task Force on High Performance Microchip Supply,” Defense Science Board, US DoD, *http://www.acq.osd.mil/dsb/reports/2005-02-HPMSi_Report_Final.pdf*, Feb, 2005.

[2] X. Wang, M. Tehranipoor, and J. Plusquellic, “Detecting Malicious Inclusions in Secure Hardware: Challenges and Solutions,” *IEEE Int. Hardware-Oriented Security and Trust (HOST)*, 2008.

[3] D. Agrawal, S. Baktir, D. Karakoyunlu, P. Rohatgi, and B. Sunar, “Trojan Detection using IC Fingerprinting,” in *IEEE Symposium on Security and Privacy (SP)*, pp. 296–310, 2007.

[4] R. Rad, J. Plusquellic, and M. Tehranipoor, “Sensitivity Analysis to Hardware Trojans using Power Supply Transient Signals,” *IEEE International Workshop on Hardware-Oriented Security and Trust*, pp. 3-7, June, 2008.

[5] X. Wang, H. Salmani, M. Tehranipoor, and J. Plusquellic, “Hardware Trojan Detection and Isolation using Current Integration and Localized Current Analysis,” in *IEEE International Symposium on Defect and Fault Tolerance of VLSI Systems (DFTVS08)*, pp. 87-95, 2008.

[6] Y. Jin and Y. Makris, “Hardware Trojan Detection using Path Delay Fingerprint,” *IEEE International Workshop on Hardware-Oriented Security and Trust*, pp. 51-57, 2008.

[7] J. Li and J. Lach, “At-Speed Delay Characterization for IC Authentication and Trojan Horse Detection,” *IEEE Int. Hardware-Oriented Security and Trust*, 2008.

[8] M. Tehranipoor and F. Koushanfar, “A Survey of Hardware Trojan Taxonomy and Detection,” *IEEE Design and Test of Computers*, pp. 10-25, 2010.

[9] S. Jha and S. K. Jha, “Randomization Based Probabilistic Approach to Detect Trojan Circuits,” *Proc. 11th IEEE High Assurance System Engineering Symposium*, pp. 117-124, 2008.

[10] M. Banga and M. Hsiao, “A Region based Approach for the Identification of Hardware Trojans,” *IEEE Int.Workshop on Hardware-Oriented Security and Trust (HOST’08)*, pp. 40-47, 2008.

[11] F. Wolff, C. Papachristou, S. Bhunia, and R. S. Chakraborty, “Towards Trojan-free Trusted ICs: Problem Analysis and Detection Scheme” in *Design, Automation and Test in Europe (DATE)*, 2008.

[12] J. Di and S. Smith, “A Hardware Threat Modeling Concept for Trustable Integarted Circuits,” *IEEE Region 5 Technical Conference*, 2007.

[13] M. Abramovici and P. Bradley, “Integrated Circuit Security: new Threats and Solutions,” in *5th Annual Workshop on Cyber Security and information intelligence Research : Cyber Security and information intelligence Challenges and Strategies*, pp. 13 - 15, April. 2009.

[14] L. Kim, J. Villasenor, and C. K. Koc, “A Trojan-resistant system-on-chip Bus Architecture,” *Proceedings of IEEE Military Communication (MILCOM)*, Boston, Oct. 2009.

[15] S. Zhao, K. Roy, and C. Koh, “Frequency Domain Analysis of Switching Noise on Power Supply Network,” *Tecnical Reports*, Purdue, 2000.

[16] I. T. Jolliffe, “Principal Component Analysis (2ed Edition),” Springer, pp. 150-165, 2002.

[17] F. P. Preparata and S. J. Hong, “Convex Hulls of Finite Sets of Points in Two and Three Dimensions,” *Commun. ACM*, vol. 20, no. 2, pp. 8793, 1977.