Real-Time Trust Evaluation in Integrated Circuits

Yier Jin and Dean Sullivan
Department of Electrical Engineering and Computer Science, University of Central Florida
{yier.jin@eecs.ucf.edu, dean.sullivan@knights.ucf.edu}

***Abstract*—The use of side-channel measurements and fingerprinting, in conjunction with statistical analysis, has proven to be the most effective method for accurately detecting hardware Trojans in fabricated integrated circuits. However, these post-fabrication trust evaluation methods overlook the capabilities of advanced design skills that attackers can use in designing sophisticated Trojans. To this end, we have designed a Trojan using power-gating techniques and demonstrate that it can be masked from advanced side-channel fingerprinting detection while dormant. We then propose a real-time trust evaluation framework that continuously monitors the on-board global power consumption to monitor chip trustworthiness. The measurements obtained corroborate our frameworks effectiveness for detecting Trojans. Finally, the results presented are experimentally verified by performing measurements on fabricated Trojan-free and Trojan-infected variants of a reconfigurable linear feedback shift register (LFSR) array.**

I. Introduction

Malicious modifications to integrated circuits (ICs), commonly referred to as hardware Trojans, have been the subject of intense study in recent years. Such modifications, which are done without the knowledge of the designer or end-user of a chip, provide additional functionality that can be exploited by a perpetrator to cause erroneous results, steal sensitive information or incapacitate a chip. The impact of hardware Trojan modifications can be catastrophic given the range of application where ICs are deployed. Currently, state-of-the-art EDA tools contribute little to the task of hardware Trojan detection, and only destructive, high-cost reverse engineering techniques are potentially effective in determining whether manufactured chips are genuine. However, it is clear that reverse engineering can only be used on a sample of chips with no guarantee that the remaining untested chips are Trojan-free [1]. Accordingly, various Trojan detection methods have been proposed to date among which the most effective rely on side-channel measurements and fingerprints. Even though a hardware Trojan can easily evade functional testing [2] or enhanced functional testing [3], [4], it has to alter the parametric profile of a chip [5]. Therefore, researchers rely on a fingerprint constructed from side-channel parameters such as global power consumption [1], path delays [6], or currents on power grids [7], [8], along with a trusted region which is statistically learned from genuine circuits (golden models), to differentiate Trojan-infected from Trojan-free chips.

Since the aforementioned methods are typically applied prior to chip deployment, a possible attack strategy to evade them is to design hardware Trojans that are dormant at test time and are only activated later in the field of operation. However, attackers confront a dilemma that the inserted Trojan should be small enough to evade power consumption based fingerprinting detection methods and still be sophisticated enough to cause erroneous results or leak internal information. Recently, a Trojan design was presented in [9], [10] that was both small and sophisticated. They leveraged on-chip resources to construct a Trojan-channel on top of a legal ultrawide band (UWB) channel by altering the digital portion of the target cryptographic IC. However, these types of designs are rare because it is not always possible to exploit hardware and still comply with all of its functional specifications.

It seems that the task is challenging for general integrated circuits with limited on-chip resources, but power saving techniques show promise for complicating the effectiveness of current side-channel methods of Trojan detection and prevention. Supported by power gating techniques [11], attackers can insert sophisticated Trojan logic of large size without worrying about disturbing transient/dynamic power consumption; so that if the power supply of Trojan logic is power-gated, then it will be effectively transparent to side-channel methods when dormant. This is true even when the Trojan is large compared to the on-chip resources. In order to counter this kind of Trojan, a life-time trust evaluation framework becomes a necessity, such that any dormant Trojans, in case they evaded the pre-deployment detection methods, will be identified if activated during operation. Similar work has been done in [10] where a post-deployment trust evaluation structure is proposed. However, the trusted evaluation process will only be triggered externally through primary inputs halting the normal operation and, even worse, leaving plenty of time for attackers to trigger and mute the Trojans during the testing intervals.

To address the request of trust evaluation at the post-deployment stage and overcome the shortage of the method of detection in [10], a real-time trust evaluation structure is proposed that can constantly monitor the operational status of the target circuit and report circuit abnormalities instantly. The contribution of the paper includes:

- A new hardware Trojan design is developed that relies on power-gating to evade functional and side-channel fingerprinting Trojan detection methods. Experimental results are presented showing that Trojans traces cannot be isolated from the global and local power consumption, even helped by advanced data analysis methods;
- A real time post-deployment framework for trust evaluation is proposed that can constantly monitor the chip status during operation and raise an alarm whenever an

978-3-9815370-2-4/DATE14/©2014 EDAA

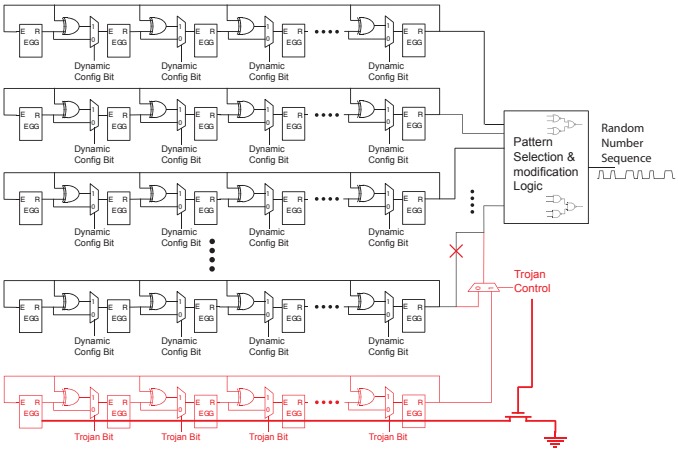


Fig. 1. Diagram of the Trojan-infected LFSR Array (The red part represents the inserted Trojan logic)

abnormal signal is detected. The proposed framework has been applied to a cryptographic circuit with experimental results demonstrating its effectiveness;

- Measurement from fabricated cryptographic chips are collected to solidify the simulation results, and provide concrete data showing an on-line trust evaluation framework is a necessity for critical systems.

The rest of the paper is organized in the following way: in section II, we outline the power-gating enhanced Trojan design and report on the simulated results. In section III, we introduce the real-time trust evaluation framework that is used for continuous Trojan detection. In section IV, chip fabrication and measurements are presented that corroborate the effectiveness of the proposed framework.

II. Power-Gating Enhanced Trojan Design

Hardware Trojans are of various types and can be embedded into different locations of the circuit [2]. However, most of the previously proposed Trojan structures are constructed to evade conventional functional/structural testing so that Trojan-infected designs can pass commercial testing and be delivered to end-users. That is, ample research has been conducted in order to enhance Trojan detection methods during the IC supply chain, but this research has focused primarily on enhancing the robustness of detection against Trojan designs that show a specific power profile. However, little work has been done to enhance Trojan designs to avoid detection by side-channel analytical methods.

The assumption that attackers can only design hardware Trojans to evade traditional testing methods weakens researcher's efforts to counter hardware Trojan attacks because it over-simplifies the task of trusted circuit design. In fact, smart attackers can always utilize cutting-edge design techniques to facilitate them in designing hard-to-detect Trojans. Among all of these design techniques, power-gating serves as a prime example. This is because many Trojan detection methods rely on global/local power consumption fingerprints. The power-gating technique has been widely used in low-power systems to shut down either part of the circuit, or the whole circuit for power saving purposes, and the muted logic will be "woken" up through certain triggers. The similarity between muted circuit logic and inactivated Trojan logic makes it possible for attackers to enhance their Trojan designs to minimize the power trace of the Trojan logic, thereby enabling large Trojan logic to be maliciously inserted without Trojan power consumption being increased. Effectively, power-gating isolates the size of the Trojan logic from its power profile leaving attackers more freedom to insert sophisticated logic in target designs without worrying that the power consumption will be significantly disturbed. In the rest of the section, we present a power-gating enhanced Trojan design and validate how the Trojan can evade current Trojan detection methods.

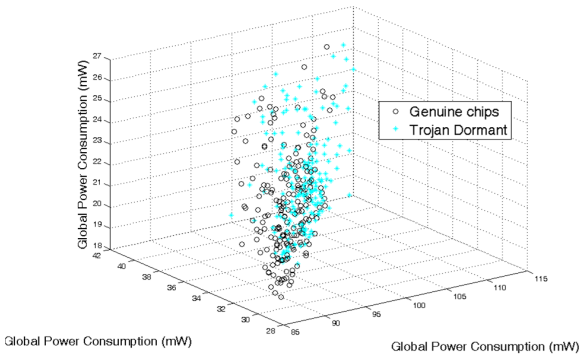
A. Experimental Vehicle

The experimental platform which we used to insert malicious logic and to test the effectiveness of the previously proposed Trojan design is a reconfigurable Linear Feedback Shift Register (LFSR) array. LFSRs have been widely used in cryptographic designs as pseudo-random number generators because of the low area cost and high entropy in generating random numbers. A reconfigurable LFSR array combines the flexibility and security to make it widely used in the cryptographic domain. Figure 1 shows the schematic of the genuine reconfigurable LFSR array, where each LFSR row contains 16-bit data with all polynomial settings configurable through the controlling 'dynamic configuration bit'. There are eight 16-bit LFSR rows in the design. Before propagating to the final output, an extra pattern modification stage is inserted to adjust the randomness of the output sequence and reshape the probabilities of '1's and '0's for special applications. To increase the testability of the target circuit for traditional testing and Trojan detection, 8 extra current nodes are added into the circuit from which local power consumption data can be measured after the chip is packaged.

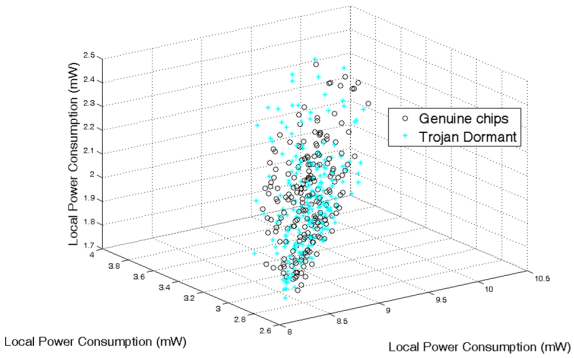
A power-gating enhanced Trojan circuit is then designed and inserted into the reconfigurable LFSR array, as shown in red in Figure 1. The Trojan is in the format of another 16-bit LFSR and the configuration polynomial is pre-configured by the attackers. When activated, the output of the Trojan LFSR will bypass one of the internal LFSRs and deteriorate the security level of the LFSR array. A gated NMOS transistor is also added between the global power supply and the VDD pin of Trojan logic. Therefore, if the Trojan Control bit is low, the overall power supply will be cut off so that the dynamic power consumption of the Trojan logic is minimized. The size of the power-gating NMOS is adjusted to minimize the leakage current and lower the impact of its transient current profile.

B. Simulation Results

The total area of Trojan logic, as can be seen from Figure 1, is non-trivial compared to the whole circuit design and is well above the detection threshold demonstrated in [1], [6], [8]. Therefore, assuming that a large population of genuine circuits exists, the genuine versus Trojan-infected circuits should be easily detected via comparison of side-channel fingerprints.

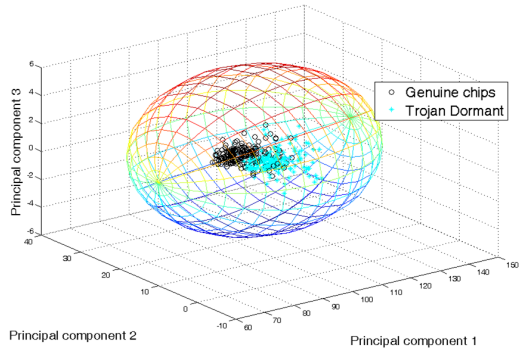


(a)

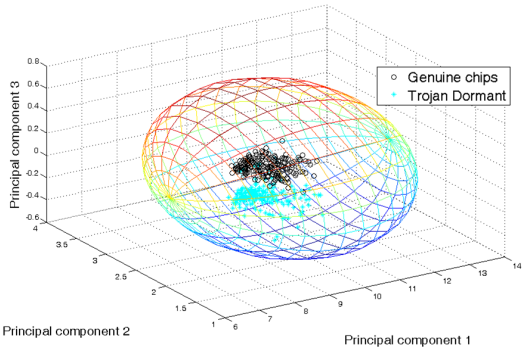


(b)

Fig. 2. (a) Global Power Consumption of Genuine and Trojan-Dormant Circuits; (b) Local Power consumption of Genuine and Trojan-Dormant Circuits



(a)



(b)

Fig. 3. (a) Trusted Operation Zones after PCA of Global Power Consumption; (b) Trusted Operation Zones after PCA for Local Power consumption

Statistical data analysis methods can also increase the probability of detection by uncovering the Trojan-related structure from the noise-affected side-channel measurements. To check whether the power-gating enhanced Trojan will indeed be detected or not, a side-channel detection experimentation was set up following the standard procedure for side-channel based Trojan detection.

1) Dataset Generation: Using Spice-level Monte-Carlo simulation with $\pm 15\%$ process variation on all circuit parameters, we generate 200 chip instances of genuine LFSR circuit as well as 200 chip instances infected by the power-gating enhanced Trojan. For each of the circuits, measurements of the global power consumption average is collected at the global power supply, and the local power consumption average is similarly collected at each of the 8 local power nodes. Moreover, 6 testing conditions are applied to each chip by varying operating frequency, reset intervals, etc. so that we generate a 9×6 power consumption table for each of the chips under test (both genuine chips and Trojan-infected chips).

2) Observations: In order to solidify our claim that power-gating enhanced hardware Trojan designs will be undetectable to current side-channel analysis methods when dormant, we randomly pick three testing conditions and plotted the corresponding measurements in Figures 2(a) and 2(b). It is apparent that in both figures we cannot differentiate the power-profiles of the genuine and Trojan-dormant chips, even though the inserted Trojan is of relatively large size. Because principal component analysis (PCA) is effective in identifying hardware Trojan traces from noise, we applied it to the collected simulation data. The PCA was run on both the global power consumption and local power consumption measurements and we have plotted the most significant principal components in Figures 3(a) and 3(b). The 3-σ trusted zones are also generated to outline the trusted spaces indicating that if the chips are operating within the trusted zones, then we can trust the chip under test. Helped by the PCA data analysis method, we are able to wipe out most of the noise caused by process variation and only collect the structure of the power consumption profile. PCA methods have helped researchers detect trivial-size Trojans and have proven to be one of the most effective data analysis methods thus far [9], [6]. With that in mind, the use of the power-gating technique has complicated our dormant Trojan detection efforts such that the Trojan related power traces are even smaller than the minimum threshold allowable. As shown in Figures 3(a) and 3(b), the effectiveness of the enhanced power-gating Trojan design has been validated because the whole Trojan circuit is powered off when the Trojan Control bit is at a low voltage. In this state, the Trojan-dormant power fingerprints are all located inside the trust boundary and are mixed with the genuine population. This is the case for both the global power traces (Figure 3(a)) and for the local power traces (Figure 3(b)). Although we have only shown one local power trace in this section, similar results were derived for all of the eight local power traces measured

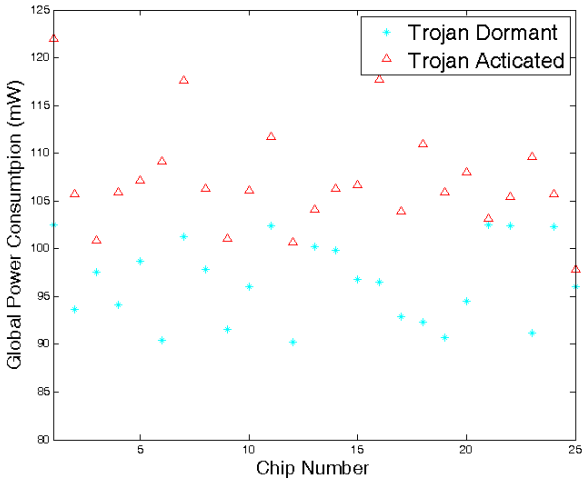


Fig. 4. Global Power Consumption Comparisons of 25 Trojan-infected Chips

from the eight local power nodes.

Although the power-gating technique can help the hardware Trojan evade side-channel based Trojan detection methods, and allow the Trojan-dormant chip to work inside the trust boundary, the power profile of Trojan-infected circuit would be disturbed significantly if the Trojan is activated. We randomly pick 25 Trojan-infected circuits and compare the average global power consumption when the Trojan is dormant and when the Trojan is activated. The comparison results are shown in Figure 4 where we have two findings. First, because of process variation, the total power consumption varies among different chips. The global power consumption of a Trojan activated chip may consume less power than another chip with the Trojan dormant. Second, for a specific chip, where the process variation parameters are fixed, there will be a power consumption gap when the Trojan is turned on or off.

III. Real-Time Trust Evaluation Framework

The simulation results demonstrate that advanced circuit design techniques, such as the power-gating method, can be utilized by attackers to design sophisticated hardware Trojans. The enhanced Trojan designs make the Trojan detection task more challenging and invalidate many of the previously proposed Trojan detection methods. However, our testing results also showed that for any specific chip, if we can collect the power traces for both Trojan on and Trojan off modes, we can easily identify the existence of hardware Trojans. Therefore, continuing to evaluate trustworthiness after chip deployment becomes even more important than pre-deployment trust evaluation. Little work has been done in this area, with the only exception being the proposed trust evaluation structure in wireless cryptographic ICs [10], which relied on on-chip neural network to analyze side-channel parameters and detect the hardware Trojan at the post-deployment stage. However, the trust evaluation framework proposed in [10] suffers from various problems. First, the framework is explicitly controlled by primary inputs, a loophole which can be easily leveraged by attackers to mute the inserted Trojans during the trust

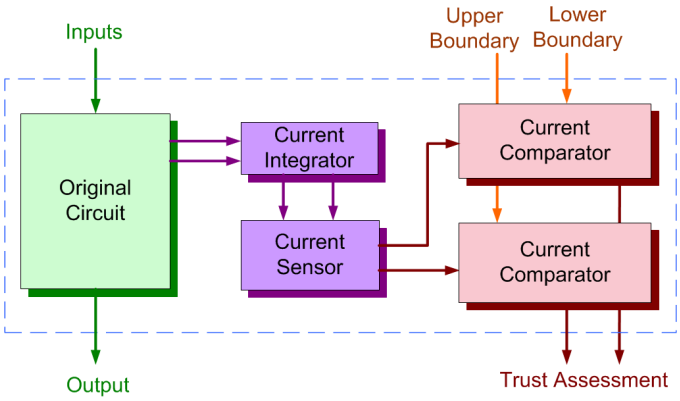


Fig. 5. Real-Time Post-Deployment Trust Evaluation Framework

evaluation process. Second, when the trust evaluation signal is ON, the whole circuit switches into the evaluation mode by suspending the normal operation. Therefore, a tradeoff needs to be considered between security level (trust evaluation frequency) and performance overhead (normal operation period).

To enhance the post-deployment trust evaluation structure and to overcome the shortages within the previously proposed method, a new real-time trust evaluation framework is developed that can continuously monitor the circuit status and trigger an alarm if an abnormality is detected. Furthermore, the operation of the trust evaluation framework can work in parallel with the circuit's normal operation so that no performance downgrading would occur. Figure 5 shows the structure of the newly proposed real-time trust evaluation framework. The new framework simplifies the on-chip trust evaluation components with low area overhead and achieves high detection capability. The basic working procedure of the trust evaluation framework is straightforward. That is, no modifications have been made to the original circuit so the throughput of the original circuit is not altered. Meanwhile, the total power consumption is measured by the on-chip current sensor (Note that in order to collect the average power consumption information, a current integrator is inserted in front of the current sensor). The measured current will then be compared to two trusted boundaries to check whether the chip is running in normal mode. If the inserted Trojan is activated, the extra power consumption would push the total power consumption outside the trusted boundary and trigger the security alarm for further diagnosis. We want to emphasize that the reference trust boundary, including upper boundary and lower boundary, are not hard wired on the chip but from off-chip resources for two reasons: 1) The trust boundaries vary from chip to chip due to process variation so that these values will only be available after the chip is fabricated and measured during the post-fabrication testing stage; 2) While the attackers may be able to understand the details of the trust evaluation framework and know what parameters are being measured, they are not privy to the actual trust boundary for each chip and, therefore, are very difficult to bypass in the proposed evaluation framework. The measurement of trust boundaries for each fabricated chip is of no cost because routine testing

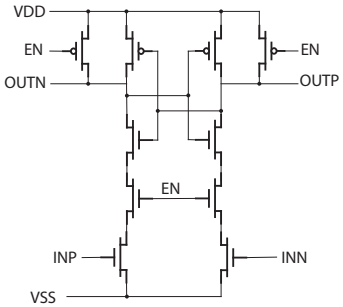


Fig. 6. The Schematic of the Voltage Comparator

stage already covers power consumption measurements of corner cases for all chips.

Once again, we set up a testing platform to demonstrate the effectiveness of the proposed real-time trust evaluation framework where the reconfigurable LFSR array serves as the target circuit, a current integrator and a current sensor are inserted at the power supply side for global power consumption acquisition, and two current comparators are used to compare power measurements with reference trusted boundaries. The current comparators will first convert the input current into voltage and then rely on voltage comparators to generate the trust assessment output. Figure 6 shows the schematic view of the designed dynamic voltage comparator where the INP and INN are the two inputs connected to the on-chip measurement and the off-chip reference. The two differential outputs, OUTP and OUTN, indicate the comparison results and will be charged every cycle when the EN controlling signal is low. The frequency of EN signal is adjustable according to the design request, but, while in the proposed real-time detection framework, the frequency of EN is set close to the clock signal of the target design. Thus, the trust evaluation would be processed almost every clock cycle to detect any Trojans being activated for more than two clock cycles[1]. The experimental results of the proposed real-time trust evaluation framework are shown in Figure 7. The `T_control` signal is the Trojan trigger indicating whether the inserted Trojan is on or off. The `OUTP<0>` and `OUTP<1>` are the positive output from the upper boundary comparator and lower boundary comparator, respectively. For the Trojan-free or Trojan-dormant circuit, the power consumption is within the trusted boundary so that the `OUTP<0>` should always be high while the `OUTP<1>` should oscillate at the same pace as the EN signal. If the Trojan is triggered to make the whole circuit consume more power, the upper boundary is violated causing both `OUTP<0>` and `OUTP<1>` to oscillate and raise the alarm signal. As shown in Figure 7, when the `T_control` is low, `OUTP<0>` is stuck at a high voltage whereas when the `T_control` is turned on, both `OUTP<0>` and `OUTP<1>` switch. The results in Figure 7 clearly demonstrates the effectiveness of the proposed framework in detecting dormant Trojans at the post-deployment stage.

[1]We point out that it will be very difficult to design hardware Trojans that will only be activated for less than two clock cycles to cause harm to the original designs.

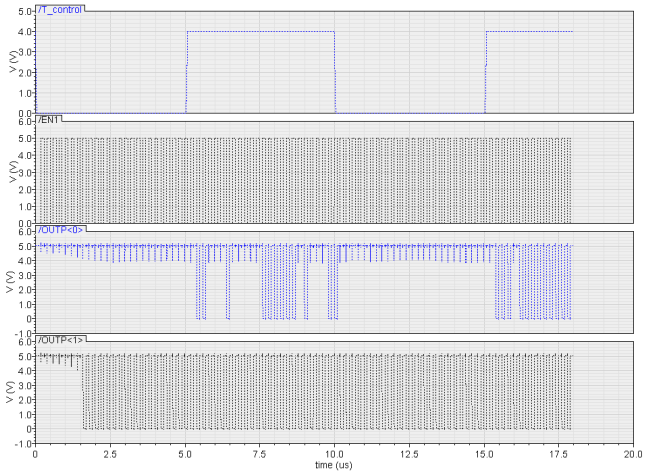


Fig. 7. The Simulation Results of the Trust Evaluation Framework

IV. Chip Fabrication and Measurements

To solidify our findings that a power-gating enhanced Trojan can evade most of the power/current based side-channel fingerprinting Trojan detection methods (Section II) and that only the on-chip trust evaluation framework can help to detect the dormant Trojans (Section III), we fabricated two versions of the designed reconfigurable LFSR array chips, both with and without Trojans, under the SEMI C5N process and collected the power measurements from the taped-out chips. In total, 18 genuine chips are fabricated which are assigned No. 1 $\sim$ No. 18 and 18 Trojan-infected chips are fabricated from No. 19 $\sim$ No. 38. As we will show shortly, the measurements taken from the fabricated chips are consistent with our findings from the simulation results in section II, proof that a post-deployment trust evaluation framework is a necessity when designing trusted systems for critical infrastructure.

A. Experimental Setup

Both the Trojan-free and Trojan-infected circuits are fabricated through SEMI C5N process and packaged with DIP28 (note that we connect the Trojan control signals and control output to the unused pins of the genuine chip so that both versions of chips use the same package type). The micrograph of the fabricated Trojan-infected chip is showed in Figure 8. An arbitrary function generator (Tektronix AFG320) is used to provide the power supply and the clock signal. The global power consumption of each chip is measured through a current probe connecting to a Tektronix MSO4104 oscillator. Considering the fact that the average power consumption would vary under different operating conditions, as well as the fact that measurement noise will be added to the measurements that differs from the simulation results where the average power consumption of each circuit is represented by one point, the measurements from the fabricated chips are presented over a range. Three sets of average power consumption measurements will be collected from genuine chips, Trojan-dormant chips and Trojan-activated chips under various operation conditions.

B. Experimental Results of Fabricated Chips

Similar to our simulation stage in Section II, two levels of comparison will be performed among the three sets

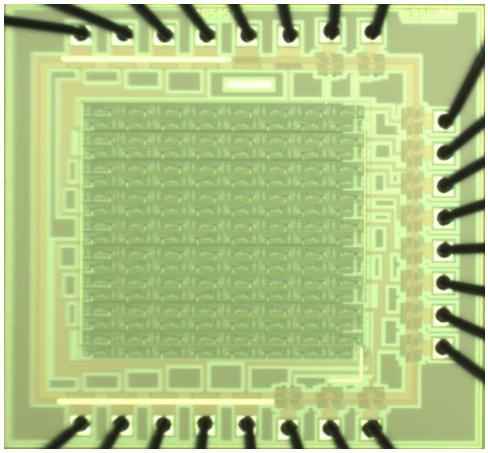


Fig. 8. Micrograph of the Fabricated Trojan-Infected LFSR Chip

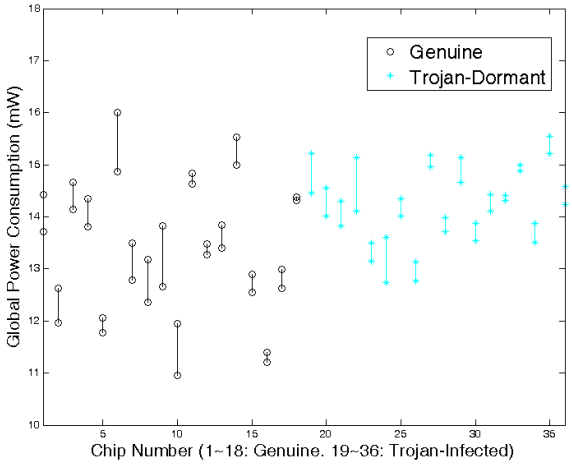


Fig. 9. Global Power Consumption of Fabricated Genuine and Trojan-Infected Chips

of collected power data. First, we will compare the power consumption data between genuine chips and Trojan-dormant chips with the results shown in Figure 9, where the x-axis lists all 36 fabricated chips with No. 1-18 indicating genuine chips and No. 19-36 indicating the Trojan-infected chips. The y-axis represents the range of total power consumption for each fabricated chip. As we can find from the figure, when both genuine and Trojan-dormant chips are tested under the same operation conditions, the average power consumption of Trojan-dormant chips is fully overlapped with that of genuine chips. However, Figure 10 shows that 16 out of the total 18 Trojan-infected chips have measurable gaps when the inserted Trojan is turned on or muted, which gives the measurement noise and circuit status differences that will be used for calibration. The delineation of global power consumption between Trojan-dormant and Trojan-activated chips is apparent and sets a clear upper limit to be used during real-time evaluation. The experimental results in Figure 10 verify our previous claim that a post-deployment stage Trojan detection method is an effective way to ensure the trustworthiness of deployed chips, even when attacked by power-gating enhanced Trojans.

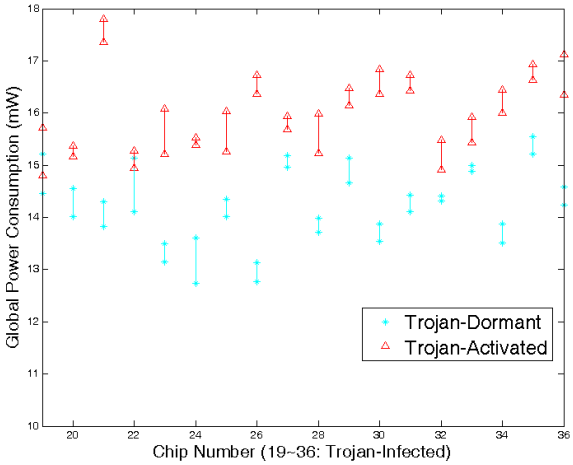


Fig. 10. Global Power Consumption of Fabricated Trojan-Infected Chips

V. Conclusions

To address the limitation of previously proposed Trojan detection methods, we designed a new Trojan using power-gating techniques and demonstrated that it can be masked from advanced side-channel fingerprinting detection while dormant. A real-time trust evaluation framework is proposed that continuously monitors the on-board global power consumption and runs trust evaluation constantly so that it can catch any activated Trojans. Both Trojan-free and Trojan-infected chips are fabricated where the experimental results uphold our claim that the proposed real-time trust evaluation framework can provide a solution for trusted system design.

References

[1] D. Agrawal, S. Baktir, D. Karakoyunlu, P. Rohatgi, and B. Sunar, "Trojan detection using IC fingerprinting," in *IEEE Symposium on Security and Privacy*, 2007, pp. 296–310.

[2] Y. Jin, N. Kupp, and Y. Makris, "Experiences in hardware Trojan design and implementation," in *IEEE International Workshop on Hardware-Oriented Security and Trust*, 2009, pp. 50–57.

[3] Y. Jin and Y. Makris, "Is single trojan detection scheme enough?" in *Proceedings of the IEEE International Conference on Computer Design (ICCD)*, 2011, pp. 305–308.

[4] J. Rajendran, V. Jyothi, and R. Karri, "Blue team red team approach to hardware trust assessment," in *Computer Design (ICCD), IEEE 29th International Conference on*, 2011, pp. 285–288.

[5] M. Tehranipoor and F. Koushanfar, "A survey of hardware Trojan taxonomy and detection," *Design Test of Computers, IEEE*, vol. 27, pp. 10–25, 2010.

[6] Y. Jin and Y. Makris, "Hardware Trojan detection using path delay fingerprint," in *IEEE International Workshop on Hardware-Oriented Security and Trust*, 2008, pp. 51–57.

[7] R. M. Rad, X. Wang, M. Tehranipoor, and J. Plusquellic, "Power supply signal calibration techniques for improving detection resolution to hardware Trojans," in *IEEE/ACM International Conference on Computer-Aided Design*, 2008, pp. 632–639.

[8] R. Rad, J. Plusquellic, and M. Tehranipoor, "Sensitivity analysis to hardware Trojans using power supply transient signals," in *IEEE International Workshop on Hardware-Oriented Security and Trust*, 2008, pp. 3–7.

[9] Y. Jin and Y. Makris, "Hardware Trojans in wireless cryptographic ICs," *IEEE Design and Test of Computers*, vol. 27, pp. 26–35, 2010.

[10] Y. Jin, D. Maliuk, and Y. Makris, "Post-deployment trust evaluation in wireless cryptographic ICs," in *Design, Automation Test in Europe Conference Exhibition (DATE), 2012*, 2012, pp. 965–970.

[11] K. Agarwal, K. Nowka, H. Deogun, and D. Sylvester, "Power gating with multiple sleep modes," in *Proceedings of the 7th International Symposium on Quality Electronic Design*, ser. ISQED '06. IEEE Computer Society, 2006, pp. 633–637.