Proceedings of the Sixth International Conference on Research in Intelligent and Computing pp. 53–56
DOI: 10.15439/2021R26
ISSN 2300-5963 ACSIS, Vol. 27

Hardware Trojan Detection Based on Side-Channel Analysis Using Power Traces and Machine Learning

Van-Phuc Hoang
Le Quy Don Technical University, 236 Hoang Quoc Viet Str., Hanoi, Vietnam
Email: phuchv@lqdtu.edu.vn

***Abstract*—With the continuous development of the Integrated Circuit (IC) manufacturing where international outsourcing is one of the main trends, hardware Trojan (HT) has been considered as a serious problem for hardware security in modern electronic systems. This paper presents a novel HT detection method based on the side-channel analysis with power traces and the machine learning (ML) technique. Side-channel information of the AES encryption core was acquired by the power consumption measurement equipment and then classified with Softmax regression. The ML technique was applied to classify and detect the HT. The experimental results have clarified the efficiency of the proposed method.**

***IndexTerms*—Hardware Trojan, power traces, SCA, machine learning**

I. Introduction

RECENTLY, the issues of cybersecurity and hardware oriented security become very critical [1], [2], [3]. Specifically, in the field of semiconductor, almost Integrated Circuit (IC) vendors aim to outsource different steps in the chip production cycle to different companies from different countries so that the production cost and time can be reduced. On the other hand, this business model of semiconductor industry also leads to the threat of hardware security including hardware Trojan (HT).

By definition, a HT is a malicious hardware module inserted in the ICs during any step of the IC design or fabrication cycle [4]. An HT consists of two components: Trigger and Payload, as shown in Fig. 1. The Trigger is the condition (such as the value of S_2S_1 in this example) so that the HT becomes active from the inactive state. On the other hand, the Payload performs the function of the HT. Once inserted in an IC, the HT can execute number of dangerous operations such as Denial of Service (DoS), extracting the secret information (for example, private cipher key) or changing the circuit behavior, etc. HT designs are often difficult for detecting either accidentally through production testing or deliberately using specially designed tests which can activate and detect HTs. The advanced HT insertion methods also allow the resistance to popular HT detection techniques using high-resolution side-channel signals such as power dissipation data, electromagnetic emission (EM), computation latency and temperature. As shown in Fig. 2, the attacker can insert HTs in some steps of the IC fabrication cycle so that different HT detection methods are required, respectively.

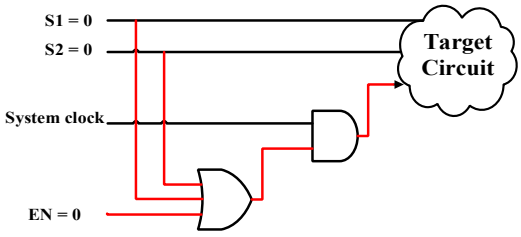

Fig. 1. A minimalist HT example.

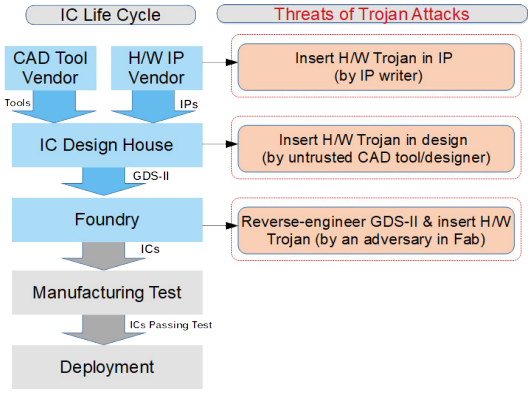

Fig. 2. The threat of HT at different steps in the IC fabrication cycle, adapted from [4].

Basically, the IC manufacturers also perform the chip testing when the production cycle is completed. However, the HT is often designed with too small size and activated by very specific conditions. Hence, the normal testing methods can not detect the HT effectively, this leads to the requirement of efficient HT detection methods. There are number of papers in literature mentioning HT detection techniques. However, there are very few papers concerning the use of machine learning (ML) techniques for HT detection. On the other hand, the recent advancements of ML techniques have inspired the researchers to apply these techniques in a broad range of applications. Therefore, this paper targets the feasibility study and experiments of ML assisted techniques for HT detection.

The remainder of this paper is organized as follows. In Section 2, the existing HT detection techniques will be introduced briefly. Then, the proposed HT detection based on

power traces and machine learning technique is described in Section 3, together with the implementation results. Finally, Section 4 concludes the paper and proposes some ideas for our future work.

II. EXISTING TECHNIQUES FOR HARDWARE TROJAN DETECTION

Currently, detection and prevention are two main categories to protect the embedded systems from the risk of HTs [4]. Prevention consists of modifying the original circuit during the conception phase to provide a secure design (against HT), to support one particular HT detection method or to make a trusted IC/chip production chain. On the other hand, detection includes techniques to clarify the presence of HTs in the design. The summary of the existing techniques to detect the HT is shown in Fig. 3. In the destructive methods, the reverse engineering is often employed to extract the circuit netlist or layout. The reconstruction of circuit layers is performed with the methods using chemical or optical principles. Hence, the destructive techniques can detect HT in the circuit with very high accuracy. However, the disadvantages of these destructive techniques are the high cost, long time and that the requirement of destroyed tested circuits so that these circuits cannot be re-used. Therefore, the non-destructive techniques are attracting most of researchers in literature. Consequently, two types of test-time and run-time techniques are often applied for non-destructive techniques.

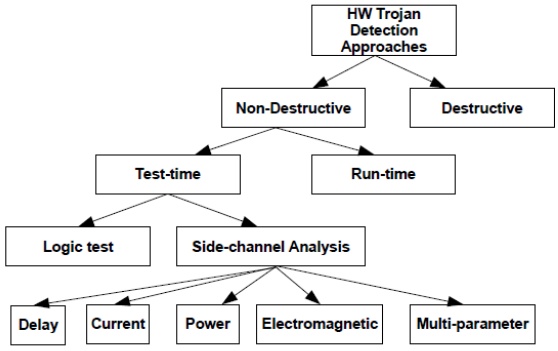

Fig. 3. Classification existing techniques for HT detection.

Side-channel analysis (SCA) is a non-destructive technique providing many potentials to improve the HT detection performance. SCA techniques use the side-channel information from the HT infected circuit and compare with side-channel information from the HT free circuits (or reference circuits). Side-channel information is divided into two groups: Energy (current, power consumption, emission energy, etc.) and signal path delay. The main advantage of the SCA approach is its ability to detect HTs even when HTs are not activated. The larger the HT, the more effective this method is and the simpler testing process [4-5]. Currently, SCA is considered among the most efficient techniques for HT detection [5]. Since the changes of the IC design will also lead to corresponding changes in the physical parameters, SCA techniques can detect many types of HTs with different sizes and structures. Moreover, recently, the research on the application of different artificial intelligence (AI) techniques including ML and deep learning (DL) in hardware security has also shown promising results [6-7]. Hence, in this work, we aim to propose a new HT detection technique with power traces and ML.

III. PROPOSED HT DETECTION METHOD AND RESULTS

In this paper, the main design is the 128-bit AES encryption core (AES_128) with the block diagram as shown in Fig. 4. The AES core uses the secret key to encrypt the plain text (Msg) and provides the cipher text after 10 rounds [8]. The HT based on the well-known Trust-Hub library [9] is inserted in this 128-bit AES encryption design.

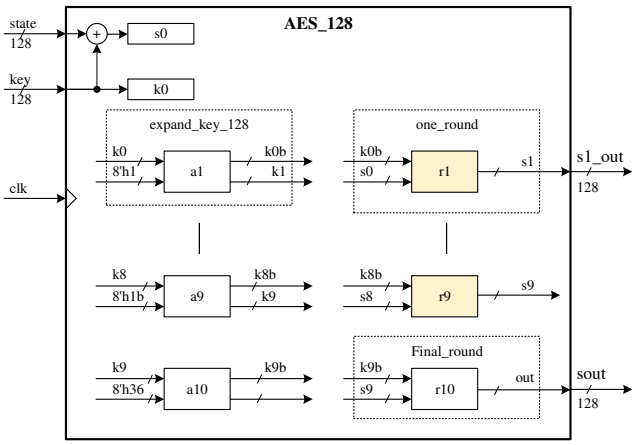

Fig. 4. AES-128 block diagram.

Table 1 presents the power consumption analysis results with the Xilinx XPower Analyzer tool for the 128-bit AES encryption core in FPGA hardware. It can be seen that the HT-infected circuit has a power consumption higher about 1% than the HT-free circuit. Based on this analysis, we propose the use of power traces of the AES core and advanced ML based signal processing techniques to detect the HT.

TABLE 1. POWER CONSUMPTION (MW) OF THE CIRCUITS USING XILINX XPOWER ANALYZER TOOL.

Power consumption type	Without HT	With HT
Dynamic power	25.84	26.54
Static power	64.69	64.71
Total power	90.53	91.26

In this work, a hardware security evaluation board (SAKURA-G) was used to implement the AES encryption core in Xilinx Spartan-6 FPGA device. This board is embedded with a specific design for SCA, especially with the power consumption data. Figure 5 presents the power traces acquisition process in our experiments. On the other hand, the measurement and experimental configuration setup for this work is shown in Fig. 6. The digital oscilloscope can capture the power traces and save in CSV files which can be displayed as in Fig. 7 for an example of the 10-round AES encryption

operation. The effect of HT in the AES operation is presented in Fig. 8 where the cipher output was wrong due the activation of the inserted HT.

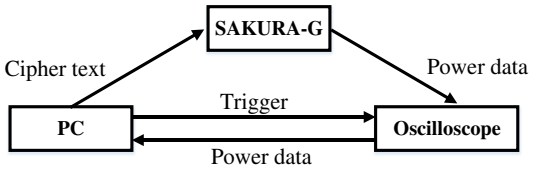

Fig. 5. Power consumption acquisition process for HT detection experiments.

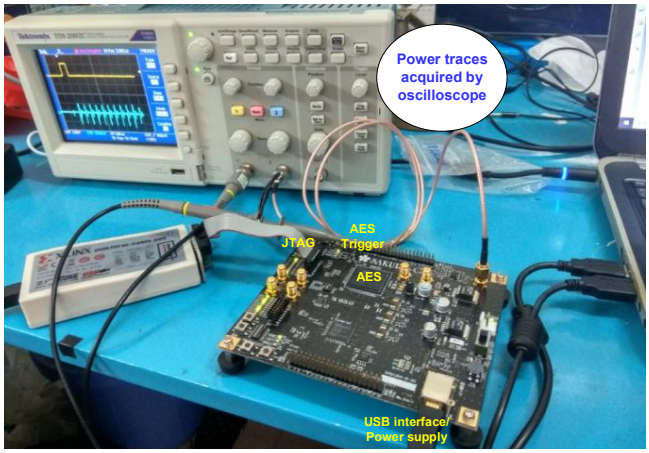

Fig. 6. Measurement setup and experimental configuration for the proposed HT detection method.

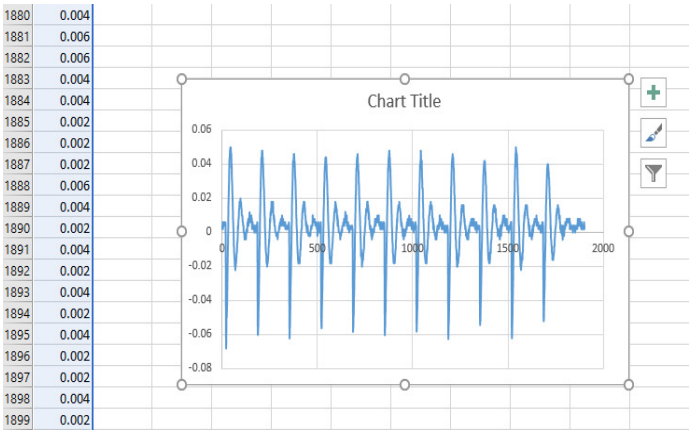

Fig. 7. The plot from CSV file captured by oscilloscope in the 10-round AES encryption process.

In this preliminary work, four different values of the AES encryption key were used to collect the power consumption data (power traces). For each key, we perform a power sampling by 30 times. In this experiment, the data with forth key is used as the reference design. The high resolution oscilloscope (Tektronix TDS2002C) was used with the support of NI LABVIEW software to build the dataset of power traces. After removing the unwanted data, we obtained 1837 power traces. From this dataset, Softmax Regression algorithm was used to classify the circuits with and without HT.

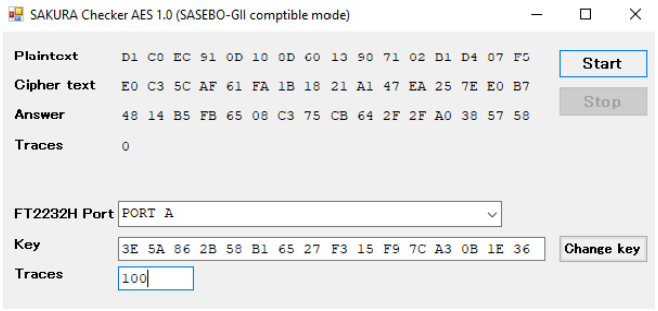

Fig. 8. The effect of HT in the wrong AES operation.

The results in Fig. 9-11 show that for the fourth key (running in an HT circuit) corresponding to the class 3, the Receiver Operation Characteristic (ROC) value is significantly different from the remaining three keys (corresponding to the class 0, 1, 2) running in the AES circuit without HT, thereby showing the ability to detect HT. Be noted that, the labels L1, L2, L3 and L4 in Figs. 9 - 10 correspond to the classes 0, 1, 2 and 3 in Fig. 11. Figure 9 is also the results of data visualization by the 2-dimension principal component analysis (PCA). With the result of HT detection accuracy of up to 97%, we can see the feasibility of applying ML techniques in classifying the power signals obtained from IC designs to detect HT in the chip (an FPGA device in this work).

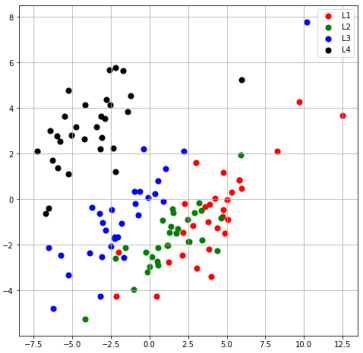

Fig. 9. Data visualization by 2-dimension PCA.

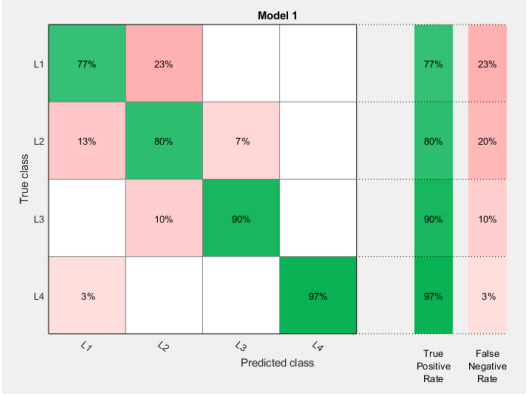

Fig. 10. Confusion Matrix of the proposed ML based HT detection method with two cases of True Positive and False Negative.

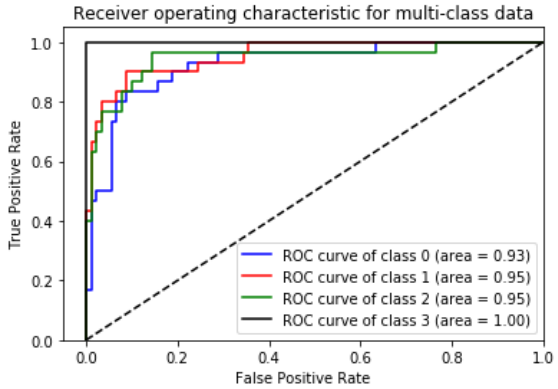

Fig. 11. ROC curve for the proposed HT detection technique.

IV. CONCLUSION

This paper has proposed a new method to detect HT using the SCA using the power traces and ML techniques. The implementation and analysis results with SAKURA-G hardware security board have shown the feasibility of the proposed solution. Moreover, with the detection accuracy of up to 97%, the proposed ML based HT detection method will have a high potential for practical applications. Furthermore, we will investigate, evaluate the proposed technique with more data, add effects of technology variations and complete the implementations on ASIC designs to provide more efficient HT detection methods.

ACKNOWLEDGMENT

This work is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 102.02-2020.14.

REFERENCES

[1] W. Ou, J. Zeng, Z. Guo, W. Yan, D. Liu, S. Fuentes, "A Homomorphic-encryption-based Vertical Federated Learning Scheme for Rick Management," Computer Science and Information Systems, vol. 17, no. 3,pp. 819–834, 2020.

[2] Faisal Alotaibi, Alexei Lisitsa, "Matrix profile for DDoS attacks detection," Proceedings of the 16th Conference on Computer Science and Intelligence Systems, Annals of Computer Science and Information Systems, vol. 25, pp. 357–361, 2021.

[3] Zhifeng Hu, Feng Zhao, Lina Qin, Hongkai Lin, "Network Virus and Computer Network Security Detection Technology Optimization," Scalable Computing: Practice and Experience, vol. 22. no. 2, 2021.

[4] S. Bhunia and M. M. Tehranipoor, *The Hardware Trojan War: Attacks, Myths, and Defenses*, Springer, pp. 15-51, 2018.

[5] A. Amelian and S.E. Borujeni, "A Side-Channel Analysis for Hardware Trojan detection based on Path Delay Measurement," Journal of Circuits, Systems, and Computers Vol. 27, No. 9, (2018).

[6] Elnaggar, R. & Chakrabarty, "Machine Learning for Hardware Security-Opportunities and Risks," K. J Electron Test (2018) 34: 183.

[7] N. -T. Do, V. -P. Hoang and V. -S. Doan, "Performance Analysis of Non-Profiled Side Channel Attacks Based on Convolutional Neural Networks," *2020 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS)*, Ha Long, Vietnam, 2020, pp. 66-69.

[8] M. Dao, V. Hoang, V. Dao and X. Tran, "An Energy Efficient AES Encryption Core for Hardware Security Implementation in IoT Systems," Proc. 2018 International Conference on Advanced Technologies for Communications (ATC), Ho Chi Minh City, Vietnam, 2018, pp. 301-304.

[9] Trojan Benchmarks, Available: https://www.trust-hub.org/resource/benchmarks.