Systematically Classifying Trusthub Hardware Trojan Benchmarks

Ana Flávia Marcondes Bomfim, José Augusto M. Nacif
Universidade Federal de Viçosa, Florestal, Brazil
{ana.bomfim, jnacif}@ufv.br

***Abstract*—This paper presents a systematic classification of benchmarks for hardware Trojan detection methods. With the globalization of integrated circuits (IC), the insertion of hardware Trojans has become a security concern. However, there was a lack of standardized benchmarks to evaluate the efficiency of detection methods. In this work, we present a classification based on Trojans' effects and activation mechanisms and classify Trusthub benchmarks accordingly. By correlating the benchmarks with the detection techniques, other researchers can now select the most suitable benchmarks for evaluating their methods, leading to faster progress in IC security.**

***Index Terms*—Hardware Trojan, Benchmark, Detection methods, Integrated circuits (ICs), Taxonomy, Activation mechanisms, Detection techniques, IC security**

I. Introduction

The globalization of Integrated Circuits (ICs) manufacturing makes it difficult to know which companies are trustworthy since inserting a Trojan Hardware (HT) during production is possible. Therefore this malicious alteration of an IC will result in improper behavior. It is possible to insert the HT at any level of the production process. With the globalization of ICs, performing post-tape-out security tests has become even more difficult.

Several HT detection techniques have emerged, and with them, dozens of benchmarks have been created [1], [2]. These benchmarks help test the efficiency of a method. However, each HT detection technique targets specific HT characteristics, i.e., it will fail to detect an HT it was not designed for [3]. Thus, the possibility of knowing previously the attributes to be tested, researchers can have their time optimized if they only use the appropriate benchmarks.

In this paper, we show that, based on the effects and activation mechanisms, it is possible to classify the most suitable benchmarks to test the efficiency of specific HT detection methods. In [3], a categorization of the possible detection methods was made and, using this as a basis together with [1], it was possible to correlate the benchmarks and the techniques.

From this paper, researchers will have an easier time testing their detection methods, leading to faster results and advances in the safety of integrated circuits. Moreover, by reducing the required test cases, they will have more time to focus on improving their projects, so it is more likely that new detection techniques will be created more quickly.

The remainder of this paper is structured as follows, in section II, it is possible to find the background, the essential pre-knowledge for a better understanding of the text, and how the benchmarks were classified. In the III section, a better description of the work, together with the articles that were used as a basis. Sections IV and V present and discuss the benchmark classification. Finally, we conclude the paper and present our final remarks in Section VI.

II. Background

This Section depicts important concepts that are a basis for understanding the paper. Sections II-A and II-B present the definitions for benchmarks, HT, and integrated circuits. Section II-C shows the HT taxonomy based on [4] and its importance. Section II-D explains many detection methods, based on [3], such as side-channel signal analysis, multiple parameters, hybrid techniques, ring oscillator, and chip partition techniques.

A. Integrated Circuit and Hardware of Trojan

An IC is an electronic device containing thousands of electronic components, such as switches, logic gates, capacitors, and transistors interconnected on a single silicon wafer. ICs are manufactured using semiconductor technology, i.e., they can work as both a conductor and an insulator, which is essential for the keys to work. HT is a type of threat to integrated circuits (IC). It is a malicious component intentionally inserted during the manufacturing process of an IC or its assembly. These HTs are difficult to detect. After all, some require a trigger to activate. However, it is possible to make these detections, and some of these ways will be better explained in II-D.

B. Benchmark

For a better understanding of this paper, it is interesting to know the concept of benchmarks, which are the way to compare the performance of different systems, in this case, the HT's detection methods. With this, there is an objective base of different technologies, allowing to show advances and innovations, in addition to facilitating the researcher to find areas of improvement of his project.

C. HT's Taxonomy Evolution

Wang, Tehranipoor, and Plusquellic [4] developed the first detailed taxonomy and only considered payload triggering and activation logic. They divided the trojan into three main categories: physical, activation, and action characteristics. This taxonomy was useful for evaluating detection methods, as it is based on the fundamental characteristics of HTs.

1) Physical Characteristics: According to physical characteristics, HTs are divided between functional and parametric. The functional ones are related to modifications in logic gates and the parametric ones in wires and logic.

2) Activation features: Also very similar to the current ones, it refers to the trojan's triggering criteria so that it can carry out its malicious action.

3) Action characteristics: are the damage it does to an IC.

D. Methodologies of HT's detection

1) Side-channel signal analysis: Trojans change the parametric characteristics of the design, that is, they can change the amount of energy used in an IC or degrade the performance. With this, it is possible to analyze the circuit's delay between wires and gates and the energy consumption characteristics. The side-channel signal analysis allows visibility into the internal structure and activities within the IC, so it is possible to detect the presence of a Trojan if the chip is tested using efficient delay tests.

Power-based analysis: First of all, it is important to obtain the power consumption pattern of Trojans-free ICs. For this purpose, random patterns are applied, and energy measurement is performed. After obtaining the references, the same patterns are applied in the circuit that wants to do the analysis. If the results are different, the circuit is considered suspicious. The analysis may differ because trojans need energy to function and may have a very large or very small impact.

Timing-based analysis: This method uses a sweeping-clock-delay measurement technique to measure selected register-to-register path delays. Basically, Trojans can be detected when delays occur beyond the threshold determined by the level of process variations.

2) Multiple Parameters: Multiple parameters is a technique that evaluates multiple parameters to increase the detection rate of HT's and decrease the false positive rate. The technique listed above depends on a trojan-free integrated circuit to be used as a base parameter in the analysis, which is why this technique was developed.

Thermal and Power: The parameters used for detection are the maps of heat and energy. To improve the accuracy of the method groupings (organization of data based on similar characteristics), thresholds (establishing references, maximums, and minimums), and statistical analysis are used.

3) Hybrid Techniques:

Current and Operating Frequency: Current and Operating Frequency: Techniques were used to improve the detection sensitivity of the HT so it is easier to detect transient leakage and supply currents between gates. A relevant fact is that this method is only really effective for small trojans.

4) Ring Oscillator:

Length Optimized Ring Oscillators: A Ring Oscillator (RO) is an electronic circuit composed of delay elements (usually inverters) connected in a closed loop, forming a ring. In this case, an RO was used to detect hardware trojans. It is possible to do this based on the differences in the frequency of the RO due to the presence of a trojan.

Ring Oscillator Network: Similar to the previous method but composed of a network of oscillators. In a network of oscillators, each one works independently, but the connection between them allows them to influence other results. Finally, it can detect a trojan by the behavior of the network.

5) Chip Partition Technique (CPT):

Current: The circuit is partitioned into regions, and each region is tested separately using a corresponding sensor so that side-channel analysis can be done. If the signature, made by the sensors, of time and power differs significantly from the one used as a base, there may be the presence of a trojan.

Power: : Analyzes the behavior of the gates, that is, the chip was partitioned and, using the power ports of each region, it is possible to detect abnormal activity.

6) Run-Time Monitoring:

Temperature Tracking: From thermal sensors inserted in the design and thermal measurements it is possible to detect trojans in the design.

Redundancy: In order to determine the reliability of other suppliers IP, design rules, optimized based on constraints such as latency, area, number of operations and cost, for trojan detection were created.

III. Related Work

A. Current Trojan Taxonomy

With HTs becoming a current threat, several ways of detection and prevention were created, but standardization was still needed to evaluate such methods. With that in mind, benchmarks were developed, in [1], to be used as parameters. Therefore, with them, it is possible to verify the quality of a Trojan Hardware detection method. Due to the similarity of the trojan with a possible error, a taxonomy was elaborated which, based on the possible insertion targets and the generated effects, classifies the trojan between:

1) Insertion phase: at what level is it inserted;

2) Abstraction phase: are the parts of the circuit susceptible to attack, basically the degree of control and flexibility available to an adversary in implementing a Trojan;

3) Activation mechanism: if any trigger is needed to activate it, whether internal or external or if it is always in operation.

4) Effect: what are the damages caused, which can vary between information leakage, functionality change, reduction reability, and service damage.

5) Location: which part of the circuit is it located.

6) Physical Characteristic: whether it is at the level of logical or physical gates.

TABLE I
BENCHMARKS CLASSIFICATION

Benchmarks	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
AES		Y					Y				Y	Y	Y		Y	Y		Y	Y		Y				
B15		Y						Y							Y	Y			Y						
B19		Y					Y				Y	Y				Y			Y		Y				
BASICRSA		Y					Y						Y		Y	Y			Y	Y	Y				
ETHERNETMAC10GE		Y	Y					Y			Y	Y			Y	Y	Y		Y					Y	Y
MC8051		Y					Y					Y		Y	Y	Y			Y	Y	Y				
MULTPYRAMID			Y						Y						Y		Y	Y							
PIC16F84		Y					Y						Y		Y	Y			Y		Y				
RS232		Y					Y		Y		Y	Y	Y	Y	Y	Y			Y	Y					
S15850		Y									Y	Y			Y	Y			Y		Y				
S35932		Y									Y	Y	Y		Y	Y			Y		Y				
S38417		Y									Y	Y			Y	Y			Y		Y				
S38584		Y									Y	Y	Y		Y	Y			Y		Y				
VGALCD		Y									Y	Y			Y	Y			Y		Y				
WBCONMAX		Y					Y				Y	Y			Y	Y			Y				Y		
TRIT		Y									Y	Y				Y			Y						

TABLE II
CLASSIFICATIONS OF HT BENCHMARKS BASED ON EFFECTS AND ACTIVATION MECHANISMS

	Change in Functionality	Information leakage	reduced reliability	Denial of service
Always on	AES	AES	-	AES MULTPYRAMID
Internally triggered	AES B19 ETHERNETMAC10GE MC8051 RS232 S15850 S35932 S38417 S38584 VGALCD WBCONMAX TRIT	AES BASICRSA PICI6F84 RS232 S35932	MC8051 RS232	AES BASICRA ETHERNETMAC10 MC8051 PIC16F84 RS232 S15850 S35932 S38417
External triggered	MC8051 RS232	BASICRA RS232	MC8051 RS232	BASICRA MAC8051 RS232

B. Relating the benchmarks and important papers

In the articles used as a basis, available in [1] and [3], benchmarks and detection methods were classified, respectively, using the trojan taxonomy. With this, we created Table I with the same parameters so that a comparison could be made. It is worth mentioning that the main parameters that will be taken into account to make the recommendations are the effect and the activation. The benchmarks were grouped as AES, B15, B19, BACRISA, ETHERMAC10GE, MC8051, MULTPYRAMID, PIC16F84, RS232, S15850, S35932, S38417, S38584,VGALCD, WBCONMAX and TRIT. The taxonomy was parameterized based on [3] and [1], being divided in following:

1) Specification;
2) Design;
3) Fabrication;
4) Testing;
5) Assembly;
6) System;
7) RTL;
8) Development Environment;
9) Logic;
10) Transistor;
11) Physical;
12) Change in Functionality;
13) Information Leakage;
14) Reduced Reliability;
15) Denial of Service;
16) Functional;
17) Parametric;
18) Always On;
19) Internally Triggered;
20) External Triggered;
21) Processor;
22) Memory;
23) I/O;
24) Power Supply;
25) Clock Grid.

With these data, it was possible to create Table I, which is important to relate the benchmarks with the comparison techniques. This is necessary to eliminate possible confounding variables that could affect the results of the comparison. In other words, the objective is a practical, unbiased, and meaningful evaluation.

IV. TRUSTHUB BENCHMARK CLASSIFICATION

Table II presents that minimizing the number of tests necessary to guarantee the quality of a certain detection method is possible. The logic behind this is that knowing the limitations of a certain method, it doesn't have to be tested

in all benchmarks. After all, it is possible to know previously by analyzing Table II when identifying the trojan will not be possible.

The taxonomy has been filtered between activation mechanisms and effects. This is because by categorizing trojans based on their effects, it is possible to gain a clearer understanding of the damage they can cause. Basically, the reason a trojan is a bad thing is its effects, if they didn't exist, hardware security wouldn't be an issue.

It is also important to consider the effects because, with this, it is possible to identify patterns and trends of attacks. In this way, it is simpler to implement specific security measures to prevent this from happening. Finally, the taxonomy filtered in this way helps quickly identify the type of trojan in a circuit.

V. Discussion

The paper aimed to make it easier for researchers to test their detection techniques only on relevant benchmarks. An example of application of the work would be using detection method 11 (path delay sensors), available in [3]. With their classification of activation mechanisms and effects, respectively, always on and reduced reliability and denial of service, the recommended benchmarks for the quality test are those of the AES and MULTPYRAMIDE groups.

Another example is number 7 (Power Consumption and Delay, which is part of the Gate Level Characterization method of detection, described in II), which is also available is [3]). The effects are change in functionality, leak of information and reduced reliability and the activation mechanisms are always on and internally triggered. Therefore, the group of benchmarks recommended are AES, B19, ETHERNETMAC10GE, MC8051, RS232, S15850, S35932, S38417, s38584,BACRISA, VGALCD, WBCONMAX,TRIT and PIC16F84.

VI. Conclusion

This paper shows that not all detection methods need to be tested in all dozens of benchmarks. This is possible to do because, as benchmarks, detection methods are classifiable according to HTs taxonomy. To save researchers time and resources, we have created a table that lists groups of benchmarks with their activation mechanisms and effects so that they can verify, based on the taxonomy of the detection method, which are the appropriate benchmarks to guarantee its effectiveness.

References

[1] B. Shakya, T. He, H. Salmani, D. Forte, S. Bhunia, and M. Tehranipoor, "Benchmarking of hardware trojans and maliciously affected circuits," *Journal of Hardware and Systems Security*, vol. 1, pp. 85–102, 2017.

[2] H. Salmani, M. Tehranipoor, and R. Karri, "On design vulnerability analysis and trust benchmarks development," in *2013 IEEE 31st international conference on computer design (ICCD)*. IEEE, 2013, pp. 471–474.

[3] S. Moein, J. Subramnian, T. A. Gulliver, F. Gebali, and M. W. El-Kharashi, "Classification of hardware trojan detection techniques," in *2015 Tenth International Conference on Computer Engineering & Systems (ICCES)*. IEEE, 2015, pp. 357–362.

[4] N. Jacob, D. Merli, J. Heyszl, and G. Sigl, "Hardware trojans: current challenges and approaches," *IET Computers & Digital Techniques*, vol. 8, no. 6, pp. 264–273, 2014.