Issue ISSN: …..-…..

IJFEAT

INTERNATIONAL JOURNAL FOR ENGINEERING APPLICATIONS AND TECHNOLOGY

"A REVIEW ON DETECTION & POTENTIAL DEFENCE SOLUTION ON HT ATTACKS IN PLD"

Manash Mitra[1], Kunal R. Nawade[2], Krutika A. Mandape[3], Nisha B. Kankenwar[4]

[1] UG Student, Electronics & Telecommunications, JDIET, Maharashtra, India, ***memanash2016@gmail.com***
[2] UG Student, Electronics & Telecommunications, JDIET, Maharashtra, India, ***Kunal.nawade2497@gmail.com***
[3] UG Student, Electronics & Telecommunications, JDIET, Maharashtra, India,
Krutikaashokmandape@gmail.com
[4] UG Student, Electronics & Telecommunications, JDIET, Maharashtra, India, ***Kankenwarnisha786@gmail.com***

Abstract

Over past few decades CPLDs and FPGA are extensively using as basic building blocks in most digital systems for their various features like robustness etc. The tremendous advancement in Very Large Scale Integration (VLSI) encourages the chip designer to rely upon commercial electronic design automation tools, offshore fabrication system and outsourced Intellectual Property (IP) cores. Hence the third party involvement in the chip design cycle, integrated circuits (IC) such as programmable logic devices (PLDs) and application specific ICs are highly vulnerable to Hardware Trojan attacks. HTs are that some malicious attacks which can be inserted in any life-cycle of the PLDs. A HT can destroy the functionality, reduces its reliability, can also leak sensible information etc. Besides the HT can be programmed either as always on or triggered in some certain conditions. A Trojan is inserted into PLD lifecycle in any manufacturing stages and remains inactive unless it is trigged by a rare value or time-event, then it produces a payload error in the circuit potentially catastrophic. For example, HT attacks in safety critical applications such as nuclear power plant, space and defense will jeopardies the safety and security of the unit. It can cause serious damage of a unit and hence lead to a war between the two countries by leaking sensitive information. Because of its nature some extensive tests can only detect the Trojans in the circuit. Based on the extensive survey HTs and their countermeasure techniques, the state of art of HT prevention, detection and diagnosis techniques are techniques are recommended to the valid stages of PLDs and ASIC cycles.

Index Terms: *CPLD, FPGA, ASIC, Trojans.*

---------- *** ----------

1. Introduction

Because of the globalization of semiconductor design and fabrication process, ICs are becoming increasingly vulnerable to malicious activities and alternations. The tremendous advancement in VLSI industry encourages chip-designer to rely upon commercial electronic design automation tools, offshore fabrication services and outsourced intellectual property cores. Hence third party involvement are more in chip-design cycle. An adversary can introduce a Trojan to disable or destroy the system. A typical Hardware Trojan (HT) design consist of "HT trigger" and "HT payload logics". Upon receiving the signal from trigger logic, payload circuitry initiates HT actions. HTs can be designed with and without trigger logic to launch serious attacks such as altering or destroying the functionality or leaking sensitive data. A simple HT design maybe as simple as a paragraph, change in specifications, an extra line of source code, modification of the silicon die at the fabrication plant, or changes in the CMOS geometrics used etc. However, the effects of HT maybe unacceptable and can have severe affects.

For a Trojan circuit to be effective, from the attacker's view point, it should be triggered under [a] rare events or

conditions, [b] It should not be easily detected by regular testing, both functional and ATPG. The Trojan detection problem is different from test problem because the Trojan effects cannot be modelled directly as traditional digital circuit faults. Apparently, there will be a need for Trojan detection models, methods and techniques to alleviate this emerging trust issue of semiconductor ICs especially in critical applications. These need as provided the main motivation for this work.

2. Literature Survey

Several literature surveys have been published on HT attacks. Some of them are listed below:

- Wolff et al. proposed the first HT taxonomy using HT trigger and payload circuits [2].
- Wang et al. classified HT taxonomy based on its physical activation and action characteristics [3].
- Alkabani et al. presented the first HT model using the storage, trigger and drive components [2].
- Tehranipoor et al. exposed a in-depth discussion of HT taxonomy, detection and design for trust techniques [5].
- Rajendran et al. developed HT taxonomy using five attributes such as design, abstraction, activation, effects and locations [4].

3. Life-cycles of PLDs

The life-cycle of PLD has a constant deal with multiple third-party accesses such as vendors during chip designs and trade, foundry during manufacture, customers during configuration and field exposure during operation. To categorize potential HTs allied with various stages of PLDs, we can classify the life cycle of PLDs as

[a] Pre-customization stage
[b] Customization stage
[c] Post customization stage.

Pre-customization stage deals with the blank i.e. un-programmed devices, which includes design and manufacturing stages. Next stage is the customization stage, where the PLD is integrated into the final system and is programmed to final stage.

In details in customization phase, it starts from coding the design using HDL, synthesis, simulation, mapping to technology-specific components, place and route(PAR),generation of the configuration bit-stream and finally programming the target device.

4. HT attacks Block Diagram

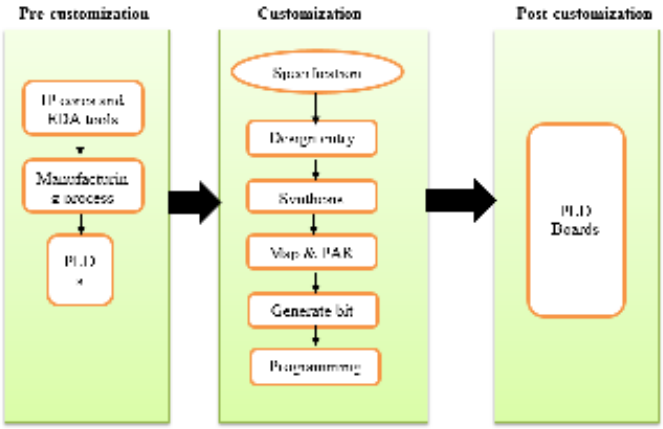

Fig 4.1: Feasible HT threats in PLD life cycles

5. Insertion of HT in PLDs

An adversary may introduce HT attacks in any of these stages either to malign the image of competitive company or to cause a malfunction in electronics used by critical system or leak secret information from the device. In most of the scenarios the IP cores, EDA tools, an intruder in chip design team and untrusted foundry are the main causes of HT attacks in the pre-customization phase. Unless the end user have a facility to investigate the chip architecture, the wide range of users trust chip vendors and directly integrates the chips with the system without any testing. Therefore, an intruder in a PLD design team can exploit this strategy to embed HTs into the device to bypass the security facilities or keep backdoors for internal access of devices used in the critical missions.

It is widely believed that the functionality of PLDs is unknown during fabrication, hence it is very difficult to insert HTs in it. However, in modern advanced PLDs have inbuilt hardcore macros like DSP, memory blocks in addition to programmable fabrics. In a research it is found that these HTs will be intelligently inserted in such a way that they left undetected during normal test process or post-manufacturing process.

6. Pre-Customization Phase Detection

To detect HTs inserted in PLDs during the pre-customization phase, both destructive and non-destructive testing can be done on the devices. As a non-destructive testing it is recommended to analyze the un-progammed blank devices for their default output values and also verify their electrical characteristics as per the datasheet provided.

To detect the HTs inserted at die-level and wire bound area, X-ray imaging is highly useful. Unfortunately the radiation can still cause damage to that specific area. In destructive approach, chip de-cap or de-packaging is carried out first. Furthermore, visual inspection/imaging techniques such as scanning electron microscopy etc. shall be used to identify HTs inserted into the die.

But these technique requires golden ICs to compare as well as they are time consuming and more complex to detect. Moreover an adversary aims to insert the HTs in such a way that they can't be detected by post-manufacturing testing. Hence these methods can also be used by chip-vendors and end users as chip-screening process that might be inserted by IP cores or EDA tools or intruder or foundry in pre-customization process.

7. Customization phase HT Detection

In customization phase design tampering at specification level is extremely difficult to detect in later stages. Hence it is good design practice to perform the functional verification of specifications by trusted outsiders. As mentioned before IP cores are the sources of black-boxes in design. So far a lot of researches are carried out either by code/structural analysis or by formal verifications techniques. In general IP cores are delivered as RTL/V HDL/Verilog codes to identify suspicious signals that may be part of an HT. Also IP cores can automatically analyzed to spot suspicious signals either using controllability values of signal or by identifying gates with low activation probability. Later Zhang explained a technique called Veritrust which marks the gates that are not driven by functional input as suspicious. That is the suspicious gates are driven by HTs, as they don't perform any computation on functional inputs. Finally the small numbers of gates are analyzed to determine if they are part of HTs.

8. Trojan Activation Methods

The Trojan detection strategies can accelerate the Trojan detection process, and in some cases have been combined with power analysis during implementation. If portion of Trojan circuitry will be activated, the Trojan circuit will consume more dynamic power, which will further help differentiate the power traces of Trojan-inserted and Trojan-free circuits. The existing Trojans in a circuit can be categorized as follows:

[a] Region-free Trojan activation.

[b] Region-aware Trojan activation.

Because the Test engineer doesn't know the Trojan type or size, both region-free & region-aware methods are necessary. If a Trojan circuit's input come from the part where they are functionally dependent i.e. part of the same logic cone the region-aware method can be effective. However if the Trojans are selected randomly from various part of the circuit, region-free could increase the probability of detection.

9. Region-free Trojan activation

These methods don't rely on the region but depend upon accidental or systematic activation of Trojans. For example Jha & Jha presented a randomization base probabilistic approach to detect Trojans. They showed that it is possible to construct a unique probabilistic signature of a circuit on the basis of specific probability for patterns applied to its inputs. They applied input patterns based on the specific probability to IUA and compare its output with the original circuit. If there are differences in the outputs, a Trojan is present. For a Trojan detection in a manufactured IC patterns can be applied only on the basis of search probability to obtain a confidence level regarding whether the original design and the fabricated chips are the same.

Wolff et al. analyzed rare net combinations designs. These rarely activated nets are used as Trojans triggers. At the same time, nets with low observability are used as payloads. Wolff et al. generated a set of vectors to activate such nets and suggested combining them with traditional ATPG test vectors to activate Trojan and to propagate its impact if the Trojan was connected to these nets.

10. Region-Aware Trojan activation

Banga and Hsiao developed a two staged test generation technique that targets magnifying the difference between the IUA and the genuine design power waveforms. In the first stage i.e. circuit partitioning, a region aware patterns helps identify the potential Trojan intersection region. To detect a Trojan circuit, the activity within a portion of the circuit is increased while the activity for the rest of the circuit is simultaneously minimized. The flip-flops in a circuit are classified into different groups, depending on structural connectivity. In the next stage i.e. activity, magnification, new test pattern concentrating on the identified region are applied to magnify the disparity between the original and Trojan inserted circuits. Regions (a set of flip-flops) exhibiting increased relative activity are identified by using the vector sequence generated in the first stage to compare the relative difference between the power profiles of the genuine and Trojan circuits. In this stage, more vectors for the specific region, marked as possible Trojan regions, are generated using the same test

generation approach as in the circuit partitioning stage.

11. Architecture level detection

Two tests for the architecture level detection includes:

1. *Liveness check:* Pseudo-random non-cached memory accesses that prevent simple prediction, delay and replay attacks.
2. *Memory protection check:* The naïve solution continuously tries to read the kernel memory with the support from RTOS.

The solutions proposed are evaluated on *SPECit 2006 benchmark* with the approximation to nearly 2.2%.

12. Power Based analysis

To obtain the power signature of Trojan-free ICs, random patterns are applied and power measurement is performed. The data belonging to each power measurement consists of several elements, including power consumption of the circuit after applying inputs that are the same in all Trojan-free ICs; measurement noise, which can be removed by several measurements; process variations, which are random and cannot be removed; and Trojan contributions to the measured power consumption. After patterns are applied, a limited number of ICs are reverse engineered to ensure they are Trojan free. Once the reference signature is obtained, the same random patterns are applied to the IC under authentication (IUA). If the IUA's power signature differs from the reference signature, the IUA is considered suspicious and that it might contain a Trojan.

13. Current Integration Method

The authors developed a multi-supply transient-current integration methodology to detect a hardware Trojan. Then, they introduced a Trojan isolation method based on localized-current analysis. They assumed the current is measured from various power ports or controlled collapse chip connections (C4s) on the die, and they applied random patterns to increase the switching in the circuit in a test-per-clock fashion. The amount of current that a Trojan can draw might be so small that it could be submerged into an envelope of noise and process variation effects, and thus be undetectable by conventional measurement equipment. However, Trojan detection capability can be greatly enhanced by measuring currents locally and from multiple power ports or pads. Figure 4 shows the current (charge) integration methodology presented by Wang et al. for detecting hardware Trojans. The die includes four power ports. The golden die can be identified using an exhaustive test for several randomly selected dies. It can also be identified via the pattern set used in the current integration method by comparing the results of all patterns in an exhaustive fashion. If the same results (within the range of variations) are obtained for all selected dies, those dies can be identified as Trojan free. The authors assumed the adversary will insert the Trojans randomly in a selected number of chips. After the golden dies are identified, the worst-case charge is obtained (dashed line in Figure 4 in response to the pattern set. The worst-case charge is based on the worst-case process variations in one of the genuine ICs. Next, the pattern set is applied to each chip, and the current is measured for each pattern locally via the power ports or C4 bumps. Figure 4 shows the current waveform of n number of patterns applied to the chips. The figure also illustrates the charge variations with time for all the current waveforms obtained after applying the patterns.

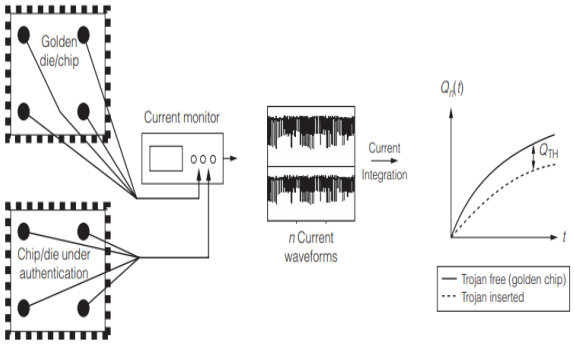

Fig 13.1: Current Integration Method

14. Design of Hardware Trust

Salmani, Tehranipoor and Plusquellic developed a methodology to increase the probability of generating a transition in functional Trojan circuits and to analyze transition generation time. They modelled transition probability using geometric distribution and estimated it on the basis of the number of clock cycles needed to generate a transition on a net. They proposed a dummy flip-flop insertion procedure in such a way that they don't

alter the design's functionality. It increase the transition probability of nets when it is lower than a specific probability threshold. Figure shows a circuit with T as Trojan gate. The transition probability at the gate output is extremely low. However after adding the dummy scan flip-flops to a net with a low transition probability, the transition probability at the Trojan output increased considerably; similarly, the average number of clock cycles per transition decreased.

To examine the effectiveness of dummy flip-flop insertion, the authors evaluated different transition probability thresholds for various Trojan circuits. They studied in detail the relationships among authentication time, the number of required transitions in the Trojan circuit, and the tester clock. These parameters can help determine a design's transition probability threshold. The transition probability threshold, in turn, provides an estimation of the area overhead induced by the insertion of dummy flip-flops.

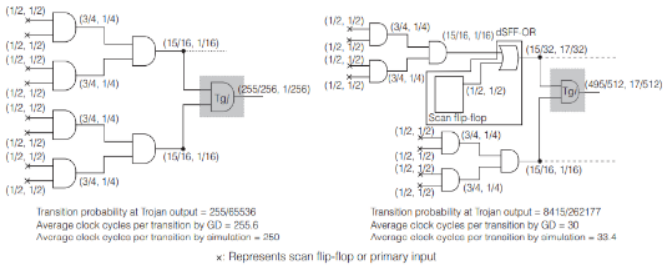

Fig 14.1: Path delay measurement Architecture

15. Voltage Inversion Technique

Banga and Hsiao proposed an inverted voltage scheme to magnify Trojan activity. Because the Trojan is assumed to be activated only under rare conditions, IC inputs could be changed so that rare combinations are created to activate the Trojan. For example, for an AND gate with four inputs, a rare condition would be when all its inputs are 1 (a probability of 1/16). The goal is to change the Trojan's functionality to remove the rare condition. Reversing a gate's power supply voltage (V_{DD}) and ground (GND) changes its function and reduces the noise margin as the output swings between V_{DD} V_{TH} and V_{TH} (where V_{TH} is the transistor voltage threshold). Thus, AND changes to NAND, and 1 at the output of a NAND Trojan is no longer a rare value, its probability becomes 15/16. However, this method must face the difficulty of switching between power supply voltage and ground for each gate on the circuit, because current power distribution networks are not designed to support an inverted voltage scheme.

To monitor an IC's system operation and detect unexpected or illegal behavior, Abramovici and Bradley suggested employing reconfigurable design-for enabling-security (Defense) logic to the functional design. When an attack is detected, the first step is to deploy countermeasures such as disabling a suspect block or forcing a safe operational mode. Figure 10 shows the architecture of a SoC with Defense inserted. Signal probe networks (SPNs) are configured to select a subset of the monitored signals and transport them to security monitors. A security monitor is a programmable transaction engine configured to implement an FSM, to check the current signals' user-specified behavior properties.

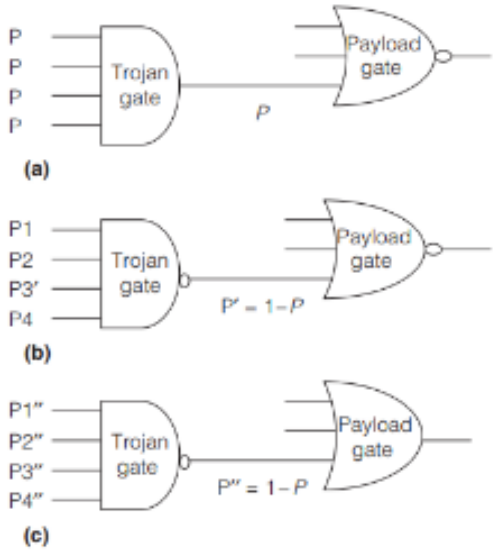

Fig 15.1: Inverted Voltage Scheme

16. Conclusion

In recent years, most of the digital designs are surfaced with hardware security threats like HTs, with effects range from a subtle degradation of service to a complete and permanent shut-down of a system. Especially, in applications such as nuclear power plant, space, and defence where safety critical systems play an important role, it is mandatory to increase system immunity for the data and design security against HT attacks. Therefore, it is utmost importance to ensure that the chip being in use performs only the intended function. As PLDs are dominant digital devices being used in these applications, we ultimately analyzed them for feasible HT attacks in pre-customization, customization, and post-customization stages. We also summarized a set of possible HT defence solutions applicable to each phase. Therefore, this work will definitely help the digital system designers/users to understand the severity of HTs associated with PLD life cycles and to incorporate a

combination of HT preventive, detective and diagnosis methods at respective stages.

References

1. G. Sumathi, L. Srivani, D. Thirugnana Murthy, K. Madhusoodaanan and S.A.V. Satya Murty, "A Review on HT attacks in PLD and ASIC Designs with Potential Defence Solutions", *IETE Technical Review, Vol 3r, No 1, 64-77*.

2. F. Wolff. C. Papachristou, S. Bhunia, and R.S Chakraborty, "Towards Trojan-free trusted ICs: problem analysis and detection scheme", *Proc. Des. Auto. Test Europe,* pp 1362-65, 2008.

3. X. Wang, M. Tehranipoor, and J. Plusquellic, "Detecting malicious inclusions in secure hardware: challenges and solutions", *Proc. IEEE Int. Workshop Hardware Oriented Security Trust,* pp 9-15, 2008.

4. J. Rajendran, E. Gavas, J. Jimenez, V.Padman, and R. Karri, "*Towards a comprehensive and systematic classification of hardware Trojans",* pp 1871-74, 2010.

5. Mohammad Tehranipoor, Farinaz Kaushanfar, "A Survey of Hardware Trojan Taxonomy and Detection", *IEEE Design & Test of Computers,* pp 10-25, 2010.

6. H. salmani, M. Terhranipoor, and J. Plusquellic, "New Design strategy For Improving Hardware Trojan Detection and Reducing Trojan Activation Time" *Proc. IEEE Workshop Hardware-Oriented Security and Trust (HOST09),* IEEE CS Press, 2009, pp 66-73.

7. M. Banga and M. Hsiao, "VITAMIN: Voltage Inversion Technique to Ascertain Malicious Insertion in ICs", *Proc. 2nd IEEE Int'l Workshop Hardware Oriented Security and Trust (Host 09),* IEEE CS Press, 2009, pp104-107.

8. S. Jha & S.K. Jha, "Randomized Based probabilistic Approach to detect Trojan Circuits", *Proc. 11th IEEE High Assurance systems engineering Symp.,* IEEE CS Press, 2008, pp. 117-124.