The Study on Hardware Security and Its Defense Measures

Yang He*

School of Communications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing, China, 21000

ABSTRACT: With the globalization of the IC design supply chain and the universal interconnection of computers, the problem of hardware security has become increasingly prominent. The increasing number of hardware security attacks in recent years shows that integrated circuit hardware has become an effective entry point for launching cross-layer and long-range attacks, and it is urgent for academia and industry to propose effective solutions to hardware security problems. This paper focuses on the threat models, defense measures and metrics of hardware attacks including Hardware Trojans, Reverse Engineering and SAT attacks (using algorithms based on satisfiability checking) by literature analysis and comparative analysis. The result shows that most of the protection against Hardware Trojans should be planned at the design stage. Camouflage, logic confusion and watermarking are good ways to defend Reverse Engineering, and for SAT attacks, it is necessary to improve the existing logical encryption technology and use an anti-SAT module.

1. INTRODUCTION

Modern computer hardware is usually in a cross-trust computing environment and shared back-to-back models by computing tasks with different security levels. The increasingly abundant interconnection features expose the key hardware resources to attackers, so it is easy for them to launch hardware attacks without physically touching hardware devices. Attackers can implant destructive Hardware Trojans into the circuit causing damage to the system. Attackers can also use reverse engineering technology on IC (integrated circuits)/IP (Intellectual Property), resulting in intellectual property piracy and over-construction of integrated circuits. Recently, the algorithm based on the satisfiability test (SAT) even showed that many hardware defense measures are vulnerable. The research goal of this paper is to do research and summarize the collected information including threat models, defense measures and metrics of Hardware Trojans, reverse engineering and SAT attacks. The research purpose of this paper aims to provide people in related fields with rich knowledge of hardware security which will make a big contribution to the security of hardware development.

2. HARDWARE TROJANS

The production of diversified chips may have to be completed by many manufacturers whose security is unknown. In these processes, the chip circuit may be implanted with malicious functions or Hardware Trojans by competitors or attackers to achieve the purpose of malicious tampering and attack control of the chip, which brings great challenges to the security of the integrated circuit. If the integrated circuit is infected with Hardware Trojan, it will lead to the change of function and specification, the leakage of sensitive information and even the paralysis of the system. That could be likely to threaten the security of the national military system, transportation, finance, medical care and other fields.

2.1. Principle And Classification of Hardware Trojans

Hardware Trojan is a circuit structure that is embedded in the original circuit and has some harmfulness. It includes a Trigger and a Payload at least. The trigger is used to monitor the signals or events in IC, such as temperature, electromagnetic changes, etc. Once it detects the expected events or conditions, it will activate the payload to perform malicious behaviors, such as denial of service, information leakage, performance barriers, etc.

Hardware Trojans are divided into many types according to their characteristics. Moein et al. further improved the classification of hardware Trojans by adding three attributes including logic type, physical layout, functional characteristics, and extended the previous classification attributes of Hardware Trojans to eight categories, as shown in figure 1 [1].

*Corresponding author. Email: 3205902561@qq.com

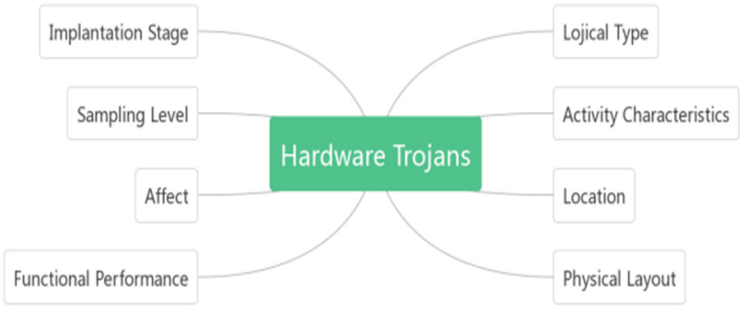

Figure 1 Types of Hardware Trojans [1]

In addition, in an effort to make the traditional method more effective and cheap, a new behavior-oriented classification is proposed by analyzing how the payload part of Trojans functions. It divides Trojans into two parts: explicit payload Trojan and implicit payload Trojan. This will be described later.

2.2. Hardware Trojan Detection Technology

Hardware Trojan detection methods mainly include destructive methods and non-destructive methods.

The destructive method destroys the chip surface package. It uses high-precision detection instruments to take pictures of the circuit layout of the chip. Then, the chip is reconstructed by reverse engineering, and compared with the chip without Trojan's called "golden circuit" to see if it contains redundant structure thus logically judging whether there is a hardware Trojan. Although the destructive method makes the Hardware Trojan invisible, it leads to too much time and high equipment cost.

There are two main types of non-destructive methods: logical test and side-channel analysis. The logical test uses the functional test to detect Hardware Trojan. Side-channel analysis checks side-channel parameters such as power supply, delay, transient and static current and maximum frequency. Because the existence of Hardware Trojan may change these parameters.

In addition, the classification of Hardware Trojans mentioned in 2.1 divides Trojans into two categories: explicit payload Trojan and implicit payload Trojan can build Trojan horse model, thus reducing test cost. Yier Jin.et al used pass delay information of the whole chip, as a result of which they proposed a new fingerprint generating method [2]. In their method, focusing the Trojan detection under manufacturing process, they found the path fingerprints which are used to characterize genuine design are more complex than power traces. That means more genuine chips. The detection rate of this method detecting explicit payload Trojan is 100%. However, it is difficult to detect hidden payload Trojans with this method.

2.3. Defend Measures

At present, most of the defense measures used are planned in advance during the design stage to deal with the problem of Hardware Trojans.

The first method aims to promote the above detection methods. Because of the concealment of Trojans, it is difficult to trigger Trojans from input and observe the influence of Trojans from the output. As a result of which, the possibility of activating Trojans can be greatly reduced because of a large number of low-controllable and low-observable nets in this design. Salmani et al. tried to insert test points in order to improve the controllability and observability of nodes [3].

Another method is a kind of preventive method which tries to prevent attackers from inserting hardware Trojans. By logic obfuscation, camouflage, and padding cells (It is very secluded for attackers to insert Trojans in circuit layout by replacing padding cells. Because it will not affect circuit parameters when removing these inactive padding cells, as a result of which, it is particularly important to find ways to fill in the blank cells). etc. Confusion, camouflage will be described in reverse engineering below.

2.4. Metrics

Metrics of hardware Trojan detection technology. (1) Detection probability: in the design, the ratio of the number of Trojans detected by this technology to the total number of Trojans. (2) Probability of wrong judgment: the ratio of the number of designs with no Trojans which are wrongly classified as Trojans to the total number of designs with no Trojans. (3) The number of clock cycles required to detect Trojans in the 3PIPs scene. [4]

3. REVERSE ENGINEERING

Reverse engineering refers to the process of identifying the structure, design and function of integrated circuits. On the one hand, reverse engineering has been successfully applied to re-record programs and relational databases, identify reusable assets, and so on. But on the other hand, reverse engineering of integrated circuits can be harmful since it can lead to collect competitive intelligence, steal the design, commercial piracy and patent infringement through attackers' malicious use.

3.1. Threat Models

The attackers must first have the tools for reverse engineering IC including a device for delaying IC, optical microscope or electron microscope for imaging and image processing tool. Then they use these tools to image the top view of each layer containing the metal wiring, contacts, vias and pins. Thus disguised standard cells and normal standard cells can be distinguished from the images of different layers. Finally, they will know what a disguised cell can do [5].

3.2. Defend Measures

One method is to apply IC camouflage to prevent reverse engineering. It means that logic gates are designed to look the same, which can lead to incorrect extraction. In order to prevent the reverse engineering of IC, any camouflage technology must provide the following guarantee. (1) The elasticity of reverse engineering: the attacker should not be able to identify the function of the camouflage door. (2) Damaged output: the output of the original netlist and the decomposed netlist should be controllable different [5].

A low-cost strategy of camouflage unit is proposed by Meng Li et al., including XOR-type strategy and STF-type (stuck-at-fault type) strategy [6]. For a camouflage unit, its cost is determined by its function in the netlist. By creating a camouflage unit with negligible overhead for a specific function, a large number of camouflage gates can be inserted into the original netlist, and the overhead of most functions can be ignored. In addition, they also proposed a strategy based on and-tree. This way achieves the purpose of camouflage by changing the circuit structure. It is worth noting that when the input pins are disguised, the and-tree structure has good resistance to the anti-camouflage attack based on SAT. An example of a camouflaged and-tree structures are shown in Figure 2.

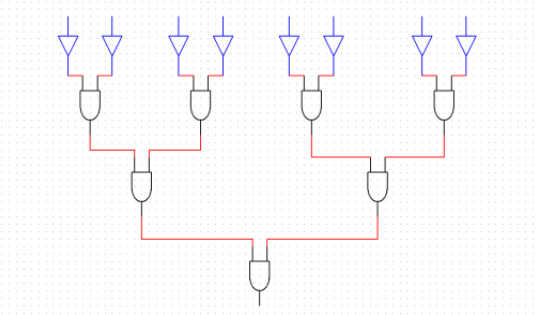

Figure 2 Example of a camouflaged and-tree structure**.**

The second method is logic confusion. It means that IC and IP can't be used immediately after manufacturing or selling through circuit logic or other parts of IC design, and can only be used normally after being unlocked by the designer. Logic confusion can protect IC and IP from theft and cloning.

Traditional logic obfuscation includes combinational logic obfuscation, integrated circuit camouflage and PC-based obfuscation. They all have the same shortcomings which are high consumption and high power. Natalia Dragan et al proposed an effective logic obfuscation having the advantage of small area, low power consumption and zero performance overhead [7]. The principle of this method is that by modifying the original netlist, the confused netlist can not be obtained by the attacker unless authorized by the designer. Therefore, the proposed confusion framework can prevent IC theft, piracy and reverse engineering.

The third method is watermarking, an active method to protect VLSI (Very Large Scale Integration) design from IP attacks. The IP core designer embeds the signature into his core to show his ownership and sells the protected IP core. If the IP core developer suspects that his core may be used for a product without authorization, he can obtain the product and check the existence of his signature to see if the third-party company illegally obtains an unauthorized copy of the protected IP core and uses the copy in their products. If this attempt is successful, and his signature provides strong proof that the product manufacturer is suspected of intellectual property fraud[8].

3.3. Metrics

Confusion: metrics of confusion include: (1) The number of violent attempts required to unlock FSM or determine key [9]. (2) Hamming distance between the output of the paste netlist and the original netlist when applying incorrect key (or configuration) [10]. (3) The number of input patterns that produce incorrect output when applying incorrect keys to the design [11].

Camouflage: metrics of camouflage include (1) The number of violent attempts needed to identify the function of the cam gate [12]. (2) The hamming distance between the original netlist and the function of the camouflaged gate assigned by the attacker [12].

Watermark: metrics of watermark include (1) the probability that the watermark algorithm generates the same solution for the signatures of different buyers. (2) The probability of attackers modifying the design in order to change one or more watermarks [4].

4. SAT ATTACK

Logic encryption was invented to resist piracy, overproduction and counterfeiting of integrated circuits. Logic encryption modifies IC design. Only when the key input is set to the correct value can it be operated correctly.

However, when Sayak Ray et al. used algorithms based on satisfiability checking (SAT) to study the security of logical encryption, they found that most logical encryption algorithms were vulnerable to this attack [13].

4.1. Threat models

Logic encryption is used in obfuscation defense. The traditional deobfuscation attack assumes that attackers can access the input and output of a chip with normal or unlocked functions. It is easy for attackers to query different inputs of the circuit and use the correct input-output pair to find the correct key value or disguised function through the oracle access to the chip [14]. However, the most strongest attacks proposed recently are based on inconsistent methods. These attacks are able to query oracle circuits according to input patterns, leading to assumptions of camouflage function or different outputs of different key values [13]. Such modes are called discriminating input patterns (DIP). The attack mentioned above is formulated as Boolean satisfiability (SAT) problem and an SAT solver is used, as a result of which, it is often called SAT attack. Being able to eliminate almost all known gate-level confusion scheme, SAT attack is now becoming a huge hidden danger threatening hardware security.

SAT attack model is considered as an untrusted foundry, and their goal is to obtain the key to unlock the circuit. Because of the layout details provided by the designer, the attacker can access the locked door-level netlist by reverse engineering a GDSII layout file in the untrusted foundry site. Then an activated functional chip is needed which can be acquired on the open market, the attackers can use the chip to evaluate a set of input modes and observe the correct output modes as a black-box model.

4.2. Defend Measures

Many scholars have expounded on the defensive measures of SAT attacks. These methods all depend on limiting the number of error keys that each DIP can exclude. Furthermore, a tree structure is needed in these methods, which outputs 1 for only one input mode and 0 for all other modes. As a result of which, these measures have very low output stability and are vulnerable to "find and remove" attacks. What's more, it is said that the defense measures recently proposed are all based on an implicit assumption that attackers need to obtain perfect accuracy when understanding the function of the circuit. For many practical situations, this may not be true [15]. For example, for the instruction decoder, the circuit must operate correctly for a possible known subset of all input modes. The following two methods are introduced.

Inspired by how the configurable and cyclic interconnection network allow modern programmable logic to realize a large set of Boolean functions with small logic elements, Kaveh Shamsi et al. proposed a low-cost SAT elastic obfuscation scheme [14]. The core idea is that if a logic circuit is generated in the circuit, the opponent can't launch the existing SAT attack by adding dummy wires and gates, because the circuit can no longer be represented as a directed acyclic graph.

Yang Xie et al. developed a relatively lightweight circuit module (called anti-SAT module), which can reduce SAT attacks by being inserted into the chip [15]. The key idea is that the total executing time to obtain the correct key and the total number of SAT attack iterations is exactly an exponential function of the key size in the Anti-SAT module. As a result of which, setting time limit of getting the right key can effectively prevent SAT attacks.

4.3. Metrics

Metrics of defense measures of SAT attack (1) High query complexity: The query complexity measure is the minimum number of queries needed to resolve the key.

High corruptibility: The effect of the key on the output is captured by the corruptibility measure of the obfuscation [14].

5. CONCLUSION

This paper introduces the threat model, defense measures and measurement standards of hardware Trojan, reverse engineering and SAT attacks. The research shows that it will be better to defense against Hardware Trojans in the design stage. Camouflage, logic confusion and watermarking are good ways to defend against Reverse Engineering. Improving the existing logical encryption technology and using the anti-SAT modules are effective methods to resist SAT attacks. The research significance of this paper lies in that it will offer a fund of knowledge of hardware security to people in related fields which shall make a big contribution to the security of hardware development. This paper is only anecdotal and informal for most of the evaluation metrics of defensive measures. There is an urgent need for a unified measurement standard in hardware protection, and only in this way can we better resist hardware attacks. At the same time, the defensive measures proposed in this paper are not specific enough and do not take every aspect into consideration. The research significance of this paper lies in providing staff in related fields with rich knowledge of hardware security which will make a big contribution to the security of hardware development.

REFERENCES

1. Huang Zhao, Wang Quan, Yang Pengfei. Hardware Trojan: Research progress and new trends of key issues [J]. Chinese journal of computers, 2019, 42(5)
2. Yier Jin, Yiorgos Makris. Hardware Trojan detection using path delay fingerprint. 2008 IEEE International Workshop on Hardware-Oriented Security and Trust.
3. Salmani and M. Tehranipoor. 2012. Layout-aware switching activity localization to enhance hardware Trojan detection. IEEE Transactions on Information Forensics and Security 7, 1 (Feb. 2012), 76–87.

4. M Rostami, F Koushanfar. Hardware security: Threat models and metrics 2013 IEEE/ACM International Conference on Computer-Aided Design (ICCAD).
5. J Rajendran, M Sam, O Sinanoglu, R Karri. Security analysis of integrated circuit camouflaging. Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security 2013.
6. Meng Li, Kaveh Shamsi. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 38, no. 8, pp. 1399-1412, Aug. 2019.
7. Jiliang Zhang. A Practical Logic Obfuscation Technique for Hardware Security IEEE Transactions on Very Large Scale Integration (VLSI) Systems 2015.
8. F. Leitao,Intellectual property (IP) protection using Watermarking and Fingerprinting techniques. 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), 2016, pp. 433-438.
9. Y. Alkabani and F. Koushanfar, Active hardware metering for intellectual property protection and security. USENIX Security, pp. 291–306, 2007.
10. A. Baumgarten, A. Tyagi, and J. Zambreno,. Preventing IC Piracy Using Reconfigurable Logic Barriers. IEEE Design and Test of Computers,vol. 27, no. 1, pp. 66–75, 2010.
11. R. Chakraborty and S. Bhunia, “HARPOON: An Obfuscation-Based SoC Design Methodology for Hardware Protection,” IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems, vol. 28, no.10, pp. 1493–1502, 2009.
12. J. Rajendran, M. Sam, O. Sinanoglu, and R. Karri,. Security Analysis of Integrated Circuit Camouflaging. ACM Conference on Computer Communications and Security, 2013.
13. P. Subramanyan, S. Ray and S. Malik, "Evaluating the security of logic encryption algorithms," 2015 IEEE International Symposium on Hardware Oriented Security and Trust (HOST), 2015, pp. 137-143
14. Kaveh Shamsi, Meng Li, Travis Meade, Zheng Zhao, David Z. Pan, and Yier Jin. Cyclic Obfuscation for Creating SAT-Unresolvable CircuitsGLSVLSI '17: Proceedings of the on Great Lakes Symposium on VLSI 2017May 2017.
15. K. Shamsi, M. Li, T. Meade, Z. Zhao, D. Z. Pan and Y. Jin, "AppSAT: Approximately deobfuscating integrated circuits," 2017 IEEE International Symposium on Hardware Oriented Security and Trust (HOST), 2017, pp. 95-100.