Literature Review of Hardware Trojans for Side Channel Attack and Detection Techniques

M. Mubdya Barry Sahya[1] and Amiruddin Amiruddin[2]

Sekolah Tinggi Sandi Negara
Bogor, Indonesia
m.mubdya@student.stsn-nci.ac.id[1]
amir@stsn-nci.ac.id[2]

Abstract — The theory of destruction of important infrastructure such as electricity, transportation and manufacturing facilities, has long been studied and developed. Such attacks are often called side channel attacks. One way to carry out such attacks is to use a hardware Trojan utilizing Integrated Circuit (IC) on a system to drain electricity or memory and hinder the system performance. This paper presents the result of a review of several hardware Trojan attack models and their detection methods that can be used as reference material for designing detection-resistant hardware Trojans as well as designing effective and efficient hardware Trojan detection methods.

Keywords— hardware Trojan, Side Channel attack, IC, detection

I. INTRODUCTION

Malicious codes continue to grow in line with technological developments. These malicious codes include viruses, Trojan Horses, worms, scripts, and rogue internet code. Mobile malicious codes are software programs that are developed to modify a computer system without the consent or permission of the owner [Yunos, 2003]. Computer viruses can be developed as cyber weapons, which can attack security, confidentiality, integrity and availability of data and computer systems. One type of virus that can be used to become a cyber weapon is a hardware Trojan horse virus, which generally attacks a system by utilizing weaknesses in hardware or often called side channel attacks. This attack can be done by utilizing weaknesses in the Integrated Circuit (IC) in which there is a defect or error in the design stage that can be inserted by the hardware Trojan.

In the Second World War, allied armies used the theory of the destruction of important infrastructure such as electrical installations, transportation, and manufacturing facilities [Yunos, 2003]. Until today, a similar theory using hardware Trojans is used as a weapon to cripple the network infrastructure and equipment.

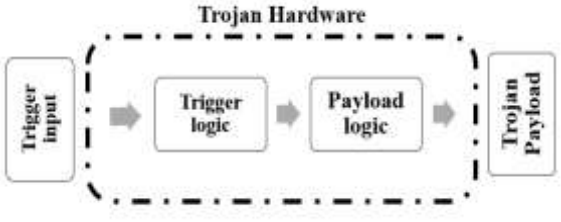

Fig. 1. General structure of Hardware Trojans [Bhunia et al., 2014]

The hardware Trojan structure in general is as shown in Figure 1 which consists of logic triggers and logic payloads or malicious programs. Triggers work to activate the payload and after the payload works, the effects of the hardware Trojan will be happening such as information leakage, changes in functions and denial of service. Triggers can be activated with input from outside or within the system itself (always active).

This paper presents the results of a review of several papers on hardware Trojan model for side-channel attacks and methods for their detection.

The structure or organization of our paper is described as follows. Section 1 describes the background of the study. The research method is explained in Section 2. Section 3 discusses several hardware Trojan models for side channel attacks and detection techniques, and Section 4 presents the comparative analysis of the hardware Trojan based on attacks and detection methods. Section 5 concludes this paper.

II. RESEARCH METHOD

This study used the traditional review method, in which several papers related to hardware Trojan were reviewed to get the details of Hardware Trojan attack models and their detection techniques. After that, we conducted a theoretical security analysis and compared the hardware Trojan model and its detection method. Hardware Trojan methods are classified according to the categories made by [Rajendran et al, 2010] which consists five classifications of Hardware Trojan, including:

1. Insertion phase in IC
2. Insertion phase at the level of abstraction
3. Activation method

ISBN 978-602-53524-0-9

4. Impact
5. Location

In the insertion phase classification, there are five steps that the Trojans can be inserted in i.e. the the phase of specification, design, fabrication, testing and packaging. In the classification of insertion at the level of abstraction, there are six classifications i.e. system level, development environment, register transfer level, gate level, transistor level, and physical level. In the activation classification, there are two classifications which are always active and triggered. For hardware Trojans that are triggered are divided into two, which are triggered internally and externally. In classifying the impact of hardware Trojans, there are four impacts that can be caused by changes in function, specification changes, information leakage, and denial of service. And finally, in the classification of locations where hardware Trojans are placed, there are five places, namely the processor, memory, place to enter and output a system, resources, and clock [Rajendran et al., 2010].

III. Review Result of Hardware Trojans

Based on several reviewed papers, hardware Trojan models and their detection methods include the following:

1. Hardware Trojan with Spread Spectrum Theory

Hardware Trojan with spread spectrum theory is a hardware Trojan that adapts the concept of Code-Division Multiple Access (CDMA). This type of hardware Trojan is intended to distribute a single bit leak for many clock cycles [Lin et al., 2009]. The leaked secret key will be modulated with PRNG which is then channeled to the leakage circuit.

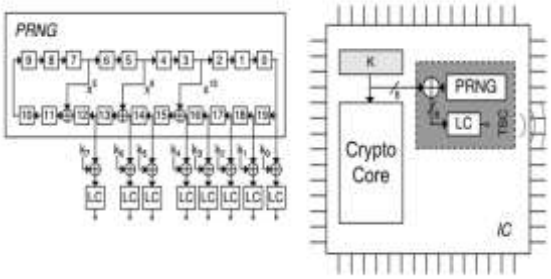

Fig. 2. Scheme of hardware Trojan with Spread-Spectrum (Lin L et al., 2009)

In Fig. 2, it is shown a hardware Trojan inserted on a cryptographic chip. The attack will be active when the flip-flop that connects between the PRNG and the key transmission line works. Therefore this trigger is included in internal triggers.

2. Hardware Trojan based on Input

This is a hardware Trojan whose trigger comes from known input values, by utilizing irregularities in key schedules that are exploited by analysis and differential attacks [Lin et al, 2009]. This hardware-based Trojan input value uses differential attacks on power. In the experiment, it was assumed that the information was leaked which was then used as input to perform differential attacks in obtaining other appropriate secret key information.

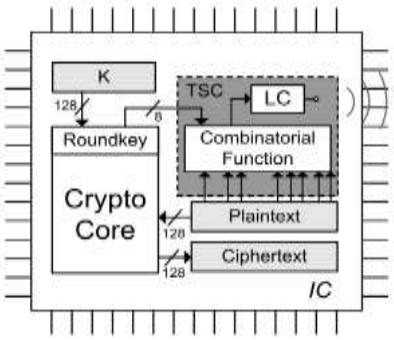

Fig. 3. Scheme of hardware Trojan based on input value

3. Hardware Trojan with Masking Scheme

Masking scheme on hardware Trojan is used to increase the difficulty of the detection of the Trojan.

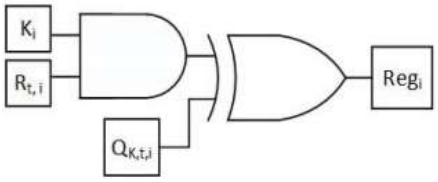

Fig. 4 Scheme of XOR Masking

As given in Fig. 4, the way the Hardware Trojan works is the same as the CDMA hardware Trojan, which is the leaked secret key modulated with PRNG. Then, in order to complicate Trojan detection, the modulated secret key can be covered with key-independence as illustrated in Fig. 4 namely $Q_{K,\,t,\,i} = F_a\,(K,\, t,\, i)$, where the F_a is some non-linear functions from K (Meng et al., 2017].

4. Hardware Trojan on wireless cryptographic IC

[Liu et al., 2013] demonstrated a hardware Trojan that could leak the secret key from a wireless cryptographic integrated circuit (IC) consisting of the core of Advance Encryption Satndard (AES) and Ultra-Wide Band (UWB) transmitters. This hardware Trojan is designed on a chip at the system level so that it can change the protocol or function on the core containing the secret key [Liu et al., 2013].

5. Hardware Trojan by [Zhang et al, 2013]

The hardware Trojan designed by [Zhang et al., 2013], is a hardware Trojan that is at the register-transfer level. In their design they only focus on the triggers of hardware Trojan detection and its implementation to be difficult to detect and more effective [Zhang et al., 2013].

ISBN 978-602-53524-0-9

6. Detect Hardware Trojan by [Aliyu et al., 2014]

The Trojan detection method carried out by [Aliyu et al., 2014] is to analyze the discrepancies in the time, power and memory of the IC. Trojans will be detected when there is a power usage mismatch by applying random patterns and calculating the power, then the memory will be checked by checking the memory usage that is not suitable for the intended use. When hit by an attack from a hardware Trojans, the time from the IC process will be affected, hence the detection of time is also done [Aliyu et al., 2014].

7. Detect hardware Trojans Based on Fingerprint Tracks

Detection of this hardware Trojan is done on critical path which is the path with the biggest time delay on the fingerprint tracks. This technique is carried out by increasing the probability of detection which reduces the excess load in the area [Nejat et al., 2014].

8. Detection and Diagnosis of hardware Trojans with Power-Based Self-Referencing Schemes

This detection method is done by comparing the power of the IC in a different time window. This method is a scheme applied to the golden chip to detect hardware Trojans sequentially [Maneesh et al., 2015].

IV. Comparative Analisis of Hardware Trojan

A. Attack Method of Hardware Trojan

Hardware Trojan analysis is done by classifying the hardware Trojan method based on the classification of Rajendran et al. Table 1 shows the results of the classification of the hardware Trojan method from the papers that have been reviewed. Five types of hardware Trojans taken from paper reviews are included in the hardware Trojan category where the insertion phase occurs in the design phase, because all hardware Trojans are made when the clock setting, flip-flop, logic gate stages are created. Types of activators in hardware Trojan models that have been classified include internal and external triggers, some are always active, and some can use external and internal triggers. Hardware Trojans with spread spectrum methods are categorized in hardware Trojans with activators that are triggered internally because this Trojan activity is active when the AES module is running and activates the flip-flop that is connected to the secret leak module. Hardware Trojan based on input values categorized in hardware Trojans triggered externally, because Trojan activity will be active when a differential attack process is performed using known plaintext input.

Table 1 Comparison of hardware Trojan attack method based on [Rajendran et al., 2010]

Method	**Insertion Phase**	**Abstrac-tion Level**	**Activa-tion by**	**Effects**
Spread Spectrum theory	Design	Physic	Internal-triggered	information leakage
Based on input value	Design	Gate	External-triggered	information leakage
Masking scheme	Design	Gate	Aalways active	Denial of service
Wireless crypto-graphic IC	Design	System	Not reported	key leakage
Zhang et al., 2013	Design	Register Transfer	Triggered	-

Hardware Trojans with masking schemes are categorized in hardware Trojans that use activators that are always active, because Trojans always work from the beginning of an infected chip being activated. In wireless cryptographic IC, the Trojan can be triggered or always be active. Hardware Trojan by [Zhang et al., 2013] categorized in a hardware Trojans that is triggered because the Trojans focuses on developing triggers that are difficult to detect.

Then, for effects caused by hardware Trojans that have been reviewed, there are hardware Trojans that leak confidential information, namely hardware Trojans with spread spectrum theory; known input values; hardware Trojan on wireless cryptographic IC; and hardware Trojans that cause denial of service, namely hardware Trojans with masking schemes that cause excess power usage. Hardware Trojans that cause leaks in the three hardware Trojan models that have been mentioned, information leakage occurs in the key used.

And lastly, for the location of the hardware Trojan that has been reviewed, two types of hardware Trojans are in the clock that is hardware Trojans with spread spectrum theory and known input values because these two hardware Trojans leak information over a flip-flop connection between the AES crypto core and the circuit leakage so long as the AES crypto core is connected the key information will leak on the leakage circuit. Furthermore, two hardware Trojans are categorized as processors, namely hardware Trojans on wireless cryptographic IC and hardware Trojans by [Zhang et al., 2013] both of which have been inserted from the beginning of the process of making a chip. And there is one hardware Trojan on the input and output of the IC i.e. the hardware Trojan with a masking scheme.

Of the five hardware Trojan methods that have been discussed, it can be seen that each hardware model has its advantages and disadvantages. Hardware Trojans with spread spectrum theory and input value-based hardware Trojans have the advantage of being resistant to conventional detection i.e. detection of physical circuits such as detection of metal layers or fingerprint circuits. The difficulty of detection is because the hardware Trojan is very small in size and juxtaposed with the core of a crypto module and is not directly connected to the input

ISBN 978-602-53524-0-9

/ output of a chip. However, the implementation of this hardware Trojan requires extra effort because the design is quite small compared to the core of a chip, thus making it difficult to be implemented.

Hardware Trojans with masking schemes have advantages similar to hardware Trojan methods with spread spectrum theory and based on known input values. The statement is based on the fact that a Trojan with this masking scheme, basically, is designed using a scheme of hardware Trojan methods with spread spectrum and value based theories that are known. A hardware Trojan is inserted at a very small level proportional to the size of the chip core being attacked. Another advantage of this Trojan is to cover secret information or leaked keys with an independent key called masking to make it difficult to detect leaks. Then, the hardware Trojan method on wireless cryptographic integrated circuit (IC) has the advantage of being able to attack on public channels by requiring only a small modification of the circuit. And, the hardware Trojan method by [Zhang et al., 2013] has the advantage of being difficult to detect with the rules made but it actually makes it difficult and does not make the performance of the hardware Trojan more efficient.

B. Detection Method of Hardware Trojan

Comparative analysis of the hardware Trojan detection method is done based on the method used in the paper of [Amin et al., 2016]. There are seven detection methods of hardware Trojan i.e. side channel analysis, time delay analysis, power analysis, reverse engineering, functional testing, Built in self-test (BIST), and logic tests [Amin et al., 2016]. The detection methods described in this study have their advantages and disadvantages. Detection methods based on fingerprint trajectory and self-reference method only focuses on the detection of anomalies time delays, in contrast to the third method proposed by [Aliyu et al., 2014] which detects hardware Trojans by looking at three aspects, anomalies in time, memory, and power.

Table 2 Comparison of hardware Trojan detection methods based on [Amin et al., 2016]

No.	Detection methods	Type of detection	Detected object
1	Fingerprint	Time delay analysis	Denial of Service
2	Self-reference based on power	Built in Self-test (BIST)	Time
3	Aliyu et al., 2014	Side channel analysis	Power, memory, and time

V. Conclusion

In this study, we reviewed several papers and obtained five types of hardware Trojan attacks and three types of hardware Trojan detection methods. The results of this review can be used as reference material in designing a detection-resistant hardware Trojan or in developing an effective and efficient hardware Trojan detection method. We also presented the results of our analysis on the strengths and weaknesses of the Trojan method and chosen an efficient detection method based on the studied methods.

Reference

[1] Aliyu, A., Bello, A., Mohammed, U. J., & Alhassan, I. H. (2014). Hardware Trojan Model for Attack and Detection Techniques, 3(3), pp. 102–105.

[2] Amin, B., Shalmani, M., Taghi, M., Ali, H., & Afshin, M. (2016). Trojan Counteraction in Hardware: A Survey and New Taxonomy, 9(May).https://doi.org/10.17485/ijst/2016/v9i18/93764

[3] Bhunia, B. S., Hsiao, M. S., Banga, M., & Narasimhan, S. (2014). Hardware Trojan Attacks: Threat Analysis and Countermeasures.

[4] Lin, L., Kasper, M., & Tim, G. (2009). Trojan Side-Channels: Lightweight Hardware Trojans through Side-Channel Engineering, 382–395.

[5] Liu, Y., Jin, Y., & Makris, Y. (2013). Hardware Trojans in Wireless Cryptographic ICs : Silicon Demonstration & Detection Method Evaluation.

[6] Meng, W., Zhu, W., Zhang, C., Liu, J., Guo, Z., & Gu, D. (2017). An Implementation of Trojan Side-Channel with a Masking Scheme.

[7] Nejat, A., Shekarian, S. M. H., & Saheb Zamani, M. (2014). A study on the efficiency of hardware Trojan detection based on path-delay fingerprinting. Microprocessors and Microsystems, 38(3), 246–252.

[8] Rajendran, J., Gavas, E., Jimenez, J., Padman, V., & Karri, R. (2010). Towards a comprehensive and systematic classification of hardware Trojans. ISCAS 2010 - 2010 IEEE International Symposium on Circuits and Systems: Nano-Bio Circuit Fabrics and Systems, 1871–1874. https://doi.org/10.1109/ISCAS.2010.5537869

[9] Maneesh P. K. and M. Nirmala Devi (2015). Power based Self-Referencing Scheme for Hardware Trojan Detection and Diagnosis. Indian Journal of Science and Technology, 8(September), 74–78.

[10] Yunos, Z., Nasir, A., Zin, M., Code, M. M., & Infrastructure, I. (2003). KOMPUTER VIRUS: FUTURE CYBER WEAPONS, (November).

[11] Zhang, J., & Xu, Q. (2013). On hardware Trojan design and implementation at register-transfer level. Proceedings of the 2013 IEEE International Symposium on Hardware-Oriented Security and Trust, HOST 2013, pp. 107–112.

ISBN 978-602-53524-0-9