IEICE Electronics Express, Vol.16, No.8, 1–4

LETTER

Fault attack hardware Trojan detection method based on ring oscillator

Qiangjia Bi[1a)], **Ning Wu**[1b)], **Fang Zhou**[1], **Jinbao Zhang**[1], **Muhammad Rehan Yahya**[1], **and Fen Ge**[1]

Abstract Recently proposed common hardware Trojan detection methods can detect wide range of Trojan types, however, there is no explicit detection method for specific Trojan types. These types of Trojan lead to low detection efficiency due to variety of categories and complex features. Keeping in view the special characteristics of Fault Attack Hardware Trojan (FAHT), a method for detection of FAHT is proposed in this letter. Proposed methodology uses ring Oscillator detection circuit that can detect extra logic gate on original cipher circuit where each byte is not required to be injected into the circuit. The experiments show that when the Trojan circuit of area 0.0495% is implanted into the original AES circuit, the detection method in this paper can detect them successfully. In addition, the proposed detection approach has high flexibility to implement in other cipher circuit.

Keywords: hardware Trojan detection, fault attack, ring oscillator
Classification: Integrated circuits

1. Introduction

With the development of the Integrated Circuit (IC), silicon chips have been used in almost every field like military, aerospace and mobile communication industry etc. In order to cope with the fierce competition in the IC market and reduce the costs, IC designers utilize third-party wafer factory to produce the chip. The fabrication process provides attacker great opportunities to insert malicious logic i.e. Hardware Trojan (HT). Generally, HT are composed of trigger circuit and payload circuit. The trigger circuit is in charge of activating the payload circuit through monitoring original circuit states. The payload circuit is responsible for performing malicious function [1].

On the other hand, fault attack is a popular method to analyze cryptographic circuit and attack the key in the field of information security. The adversary can utilize voltage glitch techniques, laser or Hardware Trojan to inject error bits while the cipher circuit is running. Then they extract the key by analyzing the correct and faulty ciphertexts [2]. Compared with other approaches, HT has good controllability and high accuracy to realize fault attack. Besides, no matter symmetric or asymmetric encryption all threatened by fault attack [2, 3, 4]. So the Fault Attack Hardware Trojan (FAHT) becomes one of the most dangerous HT for cipher circuit. It is critical and significant to do more researches on FAHT detection methods.

We have presented a brief review on related work on detection of FAHT. The previous work can be grouped into three major categories: logic test, side-channel signal analysis and trusted hardware design. First of all, it is hard to trigger the HT by using traditional logic test because of HT keep silence in most of the time. Traditional logic test is only efficient to detect small and combinatorial logic HT. For solving the problem of aforementioned logic test, in 2009, Chakraborty *et al.* proposed a statistical Trojan detection approach to activate rare value of internal nodes by generate special test patterns [5]. Lots of research on improving activation probability of HT based on statistical Trojan detection is being done [6, 7, 8, 9, 10]. But it is still hard to detect sequential logic HT. Side-channel signal analysis is another popular HT detection method. In 2007, D. Agrawal *et al.* proposed the side-channel signal analysis detection method by using power, temperature and electromagnetic profiles to reflect HT effects [11]. Improving the detection precision by changing the profiles process algorithm [12, 13, 14, 15, 16, 17, 18, 19]. In 2011, Zhang proposed an on-chip Ring Oscillator Network (RON) for HT detection as a trusted hardware design. By adding the RON and analyzing the power supply fingerprint, improving the HT detection efficient and precision [20]. Zhong use the temperature of RON to detect HT presented in [21]. In 2017, F. Pirpilidis test the effects of RO length and HT size for on AES [22].

Boneh proposed fault attack to access the key of RSA and CRT algorithm [23]. This attack method can be used in different cipher algorithms, such as DES, AES, PRESENT and SIMON and corresponding anti-fault attack method has been proposed [4, 24, 25, 26]. Debdeep and Michael implement a fault attack on AES-128 by inject error bits on the eighth round, and proposed the hack equations about ninth and tenth rounds injection [27, 28]. Different cipher algorithm has different hack equation, but the fault injection location is generally in the last few rounds [2, 3, 29, 30].

In this paper, we design a HT detection circuit based on Ring Oscillator (RO). By implementing the detection circuit on SMIC 0.18 process, we prove that our proposed detection circuit has the ability to detect small logic without considering the noise. By inserting detection circuit in AES-128 on FPGA, we verify the high detection effectiveness of proposed detection circuit.

[1]College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China
a) biqiangjia@nuaa.edu.cn
b) wunee@nuaa.edu.cn

DOI: 10.1587/elex.16.20190143
Received March 12, 2019
Accepted March 22, 2019
Publicized April 4, 2019
Copyedited April 25, 2019

2. Proposed FAHT detection circuit

A comprehensive analysis on fault attack revealed that it commonly occurs in the last few rounds [2, 3, 29, 30]. In order to defend against most existing fault attacks and improving the HT detection flexibility, we propose a FAHT detection circuit, as shown in Fig. 1.

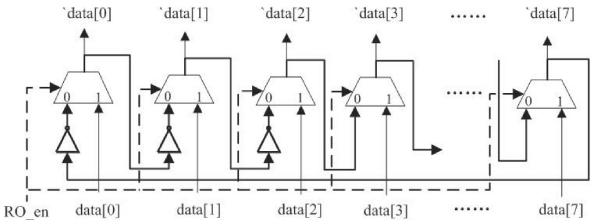

Fig. 1. FAHT detection circuit.

As the Fig. 1 shows, our FAHT detection circuit has 3 invertors that can generate oscillations and 8 multiplexers are used to transfer common data or oscillating signal. When RO_en is high, data transfer from input to output through the multiplexers. When RO_en is low, the 8 multiplexers will connected as a ring oscillator, for FAHT detection function. And the number of invertor can be changed to any odd number for different output frequency of ring oscillator. This detection circuit can be directly inserted into any cipher circuits where there is a sensitive location which can be threatened by fault attacks. This design allows designer not to modify the original circuit, hence provides higher flexibility.

3. Simulation and analysis

In order to test the performance of proposed FAHT detection circuit for series or parallel connected design with extra logic gates, we use SMIC 0.18 um process to simulate the circuit frequency in virtuoso. Every test gate is built by using two invertors in series. The test circuit is shown in Fig. 2.

As shown in Fig. 2(a) and (b), these are extra logic which may be injected by malicious third party. And the main connection of logic is (a) in series and (b) in parallel. The malicious logic may be more complex, but as long as we can detect the smallest gate in series and in parallel, we can detect the complicated logic. Actually, the location of malicious injection may be in anywhere on the ring. So the Fig. 2 just give one example.

The simulation results of gates in series are shown in Fig. 3. As shown in Fig. 3, with the increase of extra gates in series connection, the output frequencies are getting lower. It is obvious that when we add one gate the frequency is decrease from 1037.1187 MHz to 915.3156 MHz. Difference is 121.8031 MHz. So if the number of gates in series is increasing, the difference of frequency will be greater than 121.8031 MHz.

Similarly, the results of test performance for parallel connected design with extra logic gates are shown in Fig. 4.

As shown in Fig. 4, with the increase of gates in parallel connection, the output frequencies are getting lower. It is obvious that when we add one gate the frequency difference is 34.9678 MHz. So if the number of gates in parallel is increasing, the difference of frequency will be greater than 34.9678 MHz. So without considering the noise, our proposed FAHT detection circuit can distinctly reflect the injection of extra circuit.

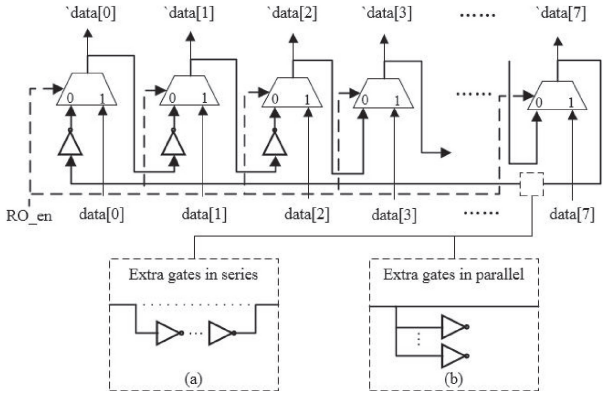

Fig. 2. Test circuit.

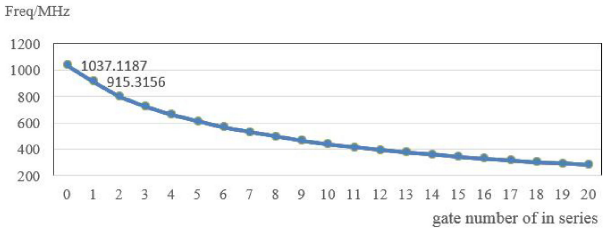

Fig. 3. Relationship between different series length and frequency.

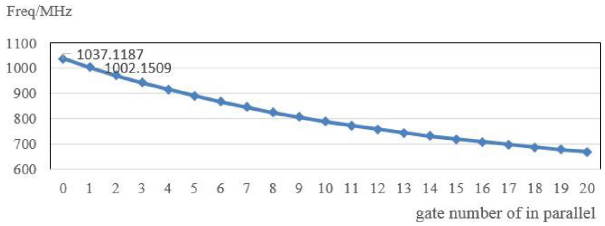

Fig. 4. Relationship between different parallel length and frequency.

4. Implementation on FPGA

In order to verify the actual detection performance of proposed detection circuit on cipher circuit, we use none-feedback mode AES-128 as the test cipher circuit. By analyzing the existing fault attacks on AES, we have concluded that those attack mainly occurred in the last three rounds [27, 28]. Based on this observation, we insert FAHT detection circuit in eighth, ninth and tenth rounds of AES. The detection location of eighth round is shown in Fig. 5.

The insert location in the last three rounds are after SUBBYTES operation. SUBBYTES are used as one of the submodules in AES [22]. Because the structure of last three rounds are almost same, except the tenth round has no MixColumn operation, we only give the eighth round structure in the Fig. 5. As shown in Fig. 5 the detection location is marked. And the data width is 128 bits, so we

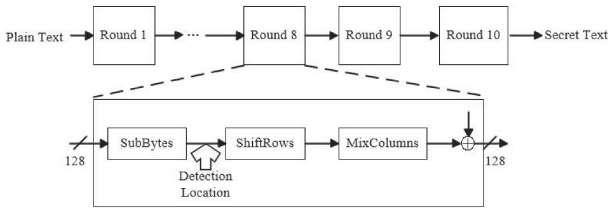

Fig. 5. Detection location of AES circuit.

commonly use state matrix to describe the data transfer process. The state matrix includes 16 bytes, every bytes has 8 bits. So the total number of bits are 128. And we extract the state matrix of the detection location in last three rounds and mark the location of detection circuits, as shown in Fig. 6.

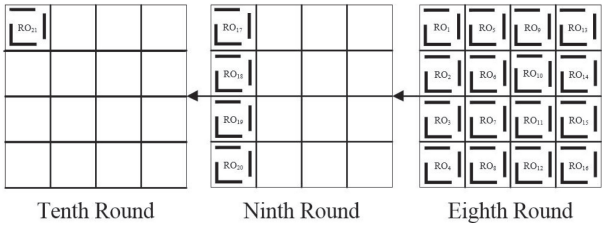

Fig. 6. Insert location of FAHT detection circuit.

As we can see in Fig. 6, RO1–RO21 are detection circuits. In order to reduce the resource consumption of the circuit, we have done some optimization at the injection, i.e. not every byte have to insert detection circuit. Because of the fault attack on tenth round must inject at least one error bit in different 16 bytes, the tenth round only needs to inject one detection circuit. In ninth round fault attack, HT has to be injected in different 4 columns to generate 4 error bits [27], so the detection circuit can insert in one of the 4 columns. By parity of reasoning in eighth round the adversary use only one bit error that can attack the key, so all 16 bytes need to be inserted in the detection circuit. The experiments are performed on Altera cyclone IV EP4CE22F17 FPGA. For easy measurement of frequency we change the number of invertor to 11.

After implementation, we use oscilloscope to measure the output frequency. As shown in Fig. 7.

It is obvious that the frequency of RO17 is lower than any other RO. Because in ninth round we inject 4 FAHT in every row to implement fault attack, RO17 detect the FAHT in first row. The FAHT injection location is shown in Fig. 8.

This experiment prove that our FAHT detection HT can detect the FAHT successfully.

5. Comparison and analysis

In order to compare the advantage of proposed method with the other HT detection method, we evaluate the hardware consumption of proposed FAHT detection circuit on FPGA. As shown in Table I.

Our original AES consume 14147 logic elements (LEs). Proposed detection circuit increase 1.93% for original AES area. And the FAHT consumes 7 LEs increasing area almost 0.0495%.

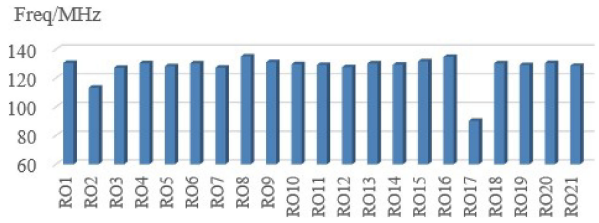

Fig. 7. Results of FAHT detection circuit on FPGA.

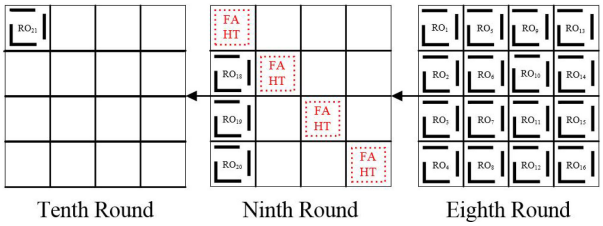

Fig. 8. FAHT insert location on state matrix.

Table I. Recourse consumption

Name	Original AES	Detection Circuit	HT	Total
LEs	14147	273	7	14427

Compared with logic test detection method [9, 10], ours do not need any time to generate the test pattern and activate the Trojan. Compared with the side-channel HT detection method [19], ours can detect the Trojan circuit of area 0.0495%. It it better than 0.6% in [19]. Compared with normal ring oscillator detection method [22], our detection method do not need any "golden" chip. Because detection circuit in [22] only based on inverter or NAND gate, the distance between HT and detection circuit impact the performance. But ours directly insert in the cipher circuit, as long as the FAHT inject it will impact the frequency.

6. Conclusion

In this letter, we proposed a FAHT detection method. Based on RO the detection circuit can detect the FAHT on original cipher circuit. The experiments show that when the Trojan circuit of area 0.0495% is implanted into the circuit, the detection method in this paper can detect them successfully. And the experiment results show that without the "golden chip" proposed detection method still can detect the HT.

Acknowledgments

This work was supported by the National Science Foundation of China (No. 61774086), the Natural Science Foundation of Jiangsu Province (BK20160806) and the Fundamental Research Funds for the Central Universities (NP2019102, NS2017023, NS2016041).

References

[1] H. Li, *et al.*: "A survey of hardware Trojan detection, diagnosis and prevention," IEEE International Conference on Computer-Aided Design and Computer Graphics (2015) (DOI: 10.1109/CADGRAPHICS.2015.41).

[2] J. Breier and W. He: "Multiple fault attack on PRESENT with a hardware Trojan implementation in FPGA," IEEE International Workshop on Secure Internet of Things (2015) (DOI: 10.1109/SIOT.2015.15).

[3] H. Marzouqi, *et al.*: "Attack countermeasure for ECC processor using One-Hot RSD encoding," 2014 21st IEEE International Conference on Electronics, Circuits and Systems (2014) (DOI: 10.1109/ICECS.2014.7049983).

[4] S. Patranabis, *et al.*: "One plus one is more than two: A practical combination of power and fault analysis attacks on PRESENT and PRESENT-like block ciphers," 2017 Workshop on Fault Diagnosis and Tolerance in Cryptography (FDTC) (2017) (DOI: 10.1109/FDTC.2017.11).

[5] R. S. Chakraborty, *et al.*: "MERO: A statistical approach for hardware Trojan detection," (2009).

[6] H. Salmani, *et al.*: "New design strategy for improving hardware Trojan detection and reducing Trojan activation time," 2009 IEEE International Workshop on Hardware-Oriented Security and Trust (2009) (DOI: 10.1109/HST.2009.5224968).

[7] I. Pomeranz and S. M. Reddy: "A measure of quality for n-detection test sets," IEEE Trans. Comput. **53** (2004) 1497 (DOI: 10.1109/TC.2004.87).

[8] S. Bhunia, *et al.*: "Hardware Trojan attacks: Threat analysis and countermeasures," Proc. IEEE **102** (2014) 1229 (DOI: 10.1109/JPROC.2014.2334493).

[9] M.-L. Flottes, *et al.*: "On the limitations of logic testing for detecting hardware Trojans horses," 2015 10th International Conference on Design & Technology of Integrated Systems in Nanoscale Era (DTIS) (2015) (DOI: 10.1109/DTIS.2015.7127362).

[10] S. Dupuis, *et al.*: "Protection against hardware Trojans with logic testing: Proposed solutions and challenges ahead," IEEE Des. Test **35** (2018) 73 (DOI: 10.1109/MDAT.2017.2766170).

[11] D. Agrawal, *et al.*: "Trojan detection using IC fingerprinting," 2007 IEEE Symposium on Security and Privacy (2007) (DOI: 10.1109/SP.2007.36).

[12] Q. Cui, *et al.*: "Hardware Trojan detection based on cluster analysis of mahalanobis distance," IEEE International Conference on Intelligent Human-machine Systems & Cybernetics (2016) (DOI: 10.1109/IHMSC.2016.65).

[13] D. Jap, *et al.*: "Supervised and unsupervised machine learning for side-channel based Trojan detection," 2016 IEEE International Conference on Application-specific Systems, Architectures and Processors (2016) (DOI: 10.1109/ASAP.2016.7760768).

[14] A. K. Marnerides, *et al.*: "Power consumption profiling using energy time-frequency distributions in smart grids," IEEE Commun. Lett. **19** (2015) 46 (DOI: 10.1109/LCOMM.2014.2371035).

[15] F. K. Lodhi, *et al.*: "Power profiling of microcontroller's instruction set for runtime hardware Trojans detection without golden circuit models," Design, Automation & Test in Europe Conference & Exhibition (DATE) (2017) (DOI: 10.23919/DATE.2017.7927002).

[16] R. Shende, *et al.*: "A side channel based power analysis technique for hardware Trojan detection using statistical learning approach," 2016 Thirteenth International Conference on Wireless and Optical Communications Networks (WOCN) (2016) (DOI: 10.1109/WOCN.2016.7759894).

[17] X. Xie, *et al.*: "Hardware Trojans classification based on controllability and observability in gate-level netlist," IEICE Electron. Express **14** (2017) 20170682 (DOI: 10.1587/elex.14.20170682).

[18] K. Hasegawa, *et al.*: "Hardware-Trojan classification method using machine learning at gate-level netlists based on Trojan features," IEICE Trans. Fundamentals **E100.A** (2017) 1427 (DOI: 10.1587/transfun.E100.A.1427).

[19] S. Jing, *et al.*: "Method on hardware Trojans detection based on side channel analysis," Netinfo Security **11** (2017) 19 (DOI: 10.3969/j.issn.1671-1122.2017.11.003).

[20] X. Zhang, *et al.*: "RON: An on-chip ring oscillator network for hardware Trojan detection," Design, Automation & Test in Europe (2011) (DOI: 10.1109/DATE.2011.5763260).

[21] J. Zhong, *et al.*: "Temperature-variation-based hardware Trojan detection through ring oscillator," Electron. Lett. **52** (2016) 1302 (DOI: 10.1049/el.2016.1411).

[22] F. Pirpilidis, *et al.*: "On the effects of ring oscillator length and hardware Trojan size on an FPGA-based implementation of AES," Microprocess. Microsyst. **54** (2017) 75 (DOI: 10.1016/j.micpro.2017.09.001).

[23] D. Boneh, *et al.*: "On the importance of checking cryptographic protocols for faults," (1997).

[24] C. H. Kim and J.-J. Quisquater: "Faults, injection methods, and fault attacks," IEEE Des. Test Comput. **24** (2007) 544 (DOI: 10.1109/MDT.2007.186).

[25] C. H. Kim: "Differential fault analysis against AES-192 and AES-256 with minimal faults," 2010 Workshop on Fault Diagnosis and Tolerance in Cryptography (2010) (DOI: 10.1109/FDTC.2010.10).

[26] J. Dofe, *et al.*: "Strengthening SIMON implementation against intelligent fault attacks," IEEE Embed. Syst. Lett. **7** (2015) 113 (DOI: 10.1109/LES.2015.2477273).

[27] D. Mukhopadhyay: "An improved fault based attack of the advanced encryption standard," International Conference on Cryptology in Africa (2009) (DOI: 10.1007/978-3-642-02384-2_26).

[28] M. Tunstall, *et al.*: "Differential fault analysis of the advanced encryption standard using a single fault," Community Ment. Health J. **49** (2011) 658 (DOI: 10.1007/s10597-012-9576-0).

[29] Y. Qin and T. Xia: "Sensitivity analysis of ring oscillator based hardware Trojan detection," IEEE International Conference on Communication Technoogy (2017) (DOI: 10.1109/ICCT.2017.8359975).

[30] W. Xiaohan, *et al.*: "AES differential fault analysis based on hardware Trojan," Comput. Eng. Appl. **53** (2017) 103 (DOI: 10.3778/j.issn.1002-8331.1507-0255).