ELectronics EXpress

LETTER

IEICE Electronics Express, Vol.10, No.17, 1–6

An intelligent classification method for Trojan detection based on side-channel analysis

Chenxu Wang[1a)], **Jinghu Li**[1], **Mingyan Yu**[1], **and Jinxiang Wang**[2]

[1] *School of Information and Electrical Engineering, Harbin Institute of Technology, Weihai, Shandong Province, China*

[2] *Harbin Institute of Technology, Harbin, Heilongjiang Province, China*

a) *wangchenxu@hit.edu.cn*

Abstract: Side-channel analysis is an important strategy for Hardware Trojan (HT) detection. Karhunen-Love (K-L) expansion can be used to improve side-channel signals analysis quality, As an auxiliary post-processing method of K-L expansion, One Class Support Vector Machine (OCSVM) is introduced to achieve ICs intelligent classification. With the OCSVM and the power traces of Genuine ICs (Genuines), a hyper sphere can be built to distinguish the Trojan ICs (Trojans) from Genuines. The effectiveness of the proposed approach is experimentally demonstrated using power simulations performed on a representative circuit with several different Trojan circuits.

Keywords: side-channel analysis, hardware Trojan, detection, One Class SVM (OCSVM)

Classification: Integrated circuits

References

[1] M. Tehranipoor and F. Koushanfar: IEEE Des. Test. Comput. **3** [1] (2010) 10.

[2] D. Agrawal, S. Baktir, D. Karakoyunlu, P. Rohatgi and B. Sunar: IEEE Symposium on Security & Privacy (2007) 296.

[3] D. Du, S. Narasimhan, R. S. Chakraborty and S. Bhunia: IEEE Workshop on Cryptographic Hardware and Embedded Systems (CHES2010) (2010) 173.

[4] Y. Jin, D. Maliuk and Y. Makris: IEEE Design, Automation & Test in Europe Conference & Exhibition (DATE2012) (2012) 965.

[5] A. Rabaoui, M. Davy, S. Rossignol, Z. Lachiri and N. Ellouze: IEEE Conf. Advanced Video and Signal Based Surveillance (2007) 117.

1 Introduction

Because of globalization of the semiconductor design and fabrication process, ICs are becoming increasingly vulnerable to malicious activities and

DOI: 10.1587/elex.10.20130602
Received July 31, 2013
Accepted August 13, 2013
Publicized August 28, 2013
Copyedited September 10, 2013

alterations, namely Hardware Trojans (HTs). Side Channel Analysis (SCA) is an important detection method to address this problem [1, 2, 3, 4]. SCA, borrowed from cryptographic analysis, was proposed by IBM in [2]. In [2], the authors modeled the IC process noise and presented the basic theoretical framework of side-channel analysis as a HT detection method. The Karhunen-Love (K-L) expansion has proved to be an effective signal processing technique for this approach in [2], however, according to the authors, we have to recognize the Trojans from the projections pictures after K-L expansion by visual inspections, which is prone to misjudgments. On the other hand, Design For Trust (DFT) is emerging for HTs real-time detection. In [4], the authors proposed a post-deployment trust evaluation architecture based on a mini neural network circuit. However, this method will depend some IC. In this paper, based on the existing flow of [2], we introduced One Class Support Vector Machine (OCSVM) a processing method after K-L expansion to implement ICs's intelligent classification. Compared with the method in [4], it is not for the specific circuit and is more universal and flexible.

2 Proposed method

In order to train a classifier to distinguish trusted ICs from untrusted ones, one should only rely on information from Trojan-free chips [4]. In other words, we are solving a One Class classification problem, where our objective is to train a classifier to envelope the region of acceptable (trusted) ICs without any data of unacceptable (untrusted) ICs. To achieve this end, we employed the One Class classification training algorithm described in [5] and we only use the Genuines data to train our classifier in our experiments. Our proposed complete HTs detection flow is shown in Figure 1.

In Step2, the data preprocessing step is as follows: (1) Sample the side-channel signals (vary with time) to get a $N \times T$ matrix $I_{N\times T}$ called side-channel signal matrix (SCSM), where N is the number of the ICsUT, and T is the number of sampling points; (2) Centralize the SCSM $I_{N\times T}$ to get the post-processing side-channel signal matrix (the post-processing SCSM).

The training and optimization of the OCSVM by GTr-data and GOTr-data mentioned in Step9 will be described in further details here: besides the training data, the structure of the minimum hyper sphere of OCSVM also depends on two basic parameters, i.e. penalty factor C and kernel function parameter factor. Here, the powerful kernel function, Gaussian RBF was taken as the OCSVM kernel function:

$$K(\mathbf{x}\cdot\mathbf{x}') = \mathbf{exp}\left(-\frac{\xi\cdot \parallel \mathbf{x}\cdot\mathbf{x}' \parallel^{\mathbf{2}}}{\mathbf{2}}\right) \tag{1}$$

Where ξ is the kernel function parameter factor [5]. The trained result of OCSVM is hardly influenced by C but very sensitive to ξ. Too large or small ξ value may result in the problem of over-fitting and under-fitting in SVM learning, and the preference should be given to search for the optimal value

DOI: 10.1587/elex.10.20130602
Received July 31, 2013
Accepted August 13, 2013
Publicized August 28, 2013
Copyedited September 10, 2013

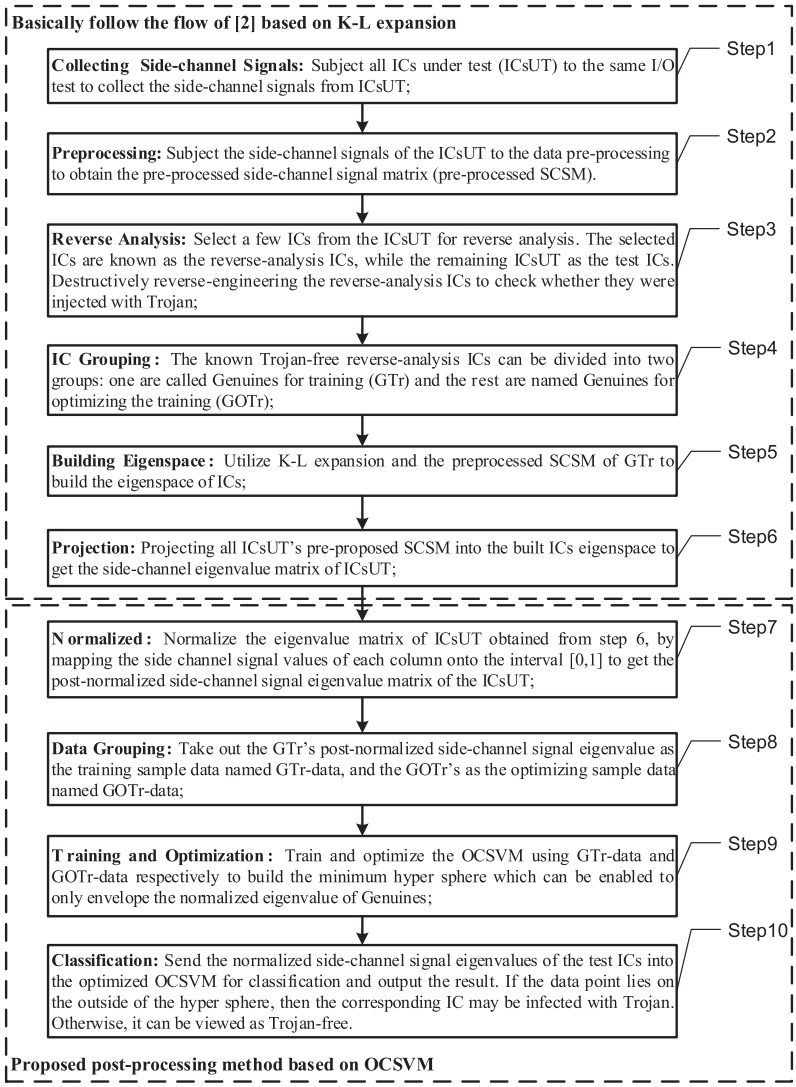

Fig. 1. Proposed Hardware Trojan Detection Flow

Table I. Description of the Algorithm to Determine ξ

step1: ξ=10; for i=1:1000 -train OCSVM using GTr-data -evaluate OCSVM using GOTr-data -get the evaluation accuracy -if the accuracy >= expection break; else ξ=ξ/2; end end	step2: for i=1:1000 ξ=ξ+Δ ξ -train OCSVM using GTr-data -evaluate OCSVM using GOTr-data -get the evaluation accuracy -if the accuracy <= expection break; end end ξ=ξ-$\Delta\xi$.

ξ to get the minimum hyper sphere by means of OCSVM training. Table I describes the optimization process of ξ in the interval $(0, 10]$ (C = 0.9).

As shown in Table I, the step $\Delta\xi$ in Step2 is determined by the value ξ obtained from Step1 of the algorithm. Take the value $\Delta\xi$ as 1/10 of order of

DOI: 10.1587/elex.10.20130602
Received July 31, 2013
Accepted August 13, 2013
Publicized August 28, 2013
Copyedited September 10, 2013

magnitude of ξ obtained from Step1. The value ξ obtained after subjected to the above algorithm processing will be regarded as the optimal value ξ, then such value and the training sample data can be utilized to complete the training of OCSVM in such a manner as to get the minimum hyper sphere.

3 Experimental setup

3.1 ICs used in our analysis

We used synthesized DES (Data Encryption Standard) circuit as the benchmark circuit in the rest of this paper. The Trojan added to this circuit ranges from a simple counter to decoder. In a counter based Trojan, the Trojan circuit counts clock cycles and output the value at every clock rising edge. In the case of a decoder based Trojan, the Trojan circuit decodes several DES ciphertext bits to output. To push the limit, we also let one 2-input NAND gate act as the Trojan, which is almost the least gate in a standard cell libray. The different types of Trojan circuit and their relative area are showed in Table III. In our experiments, the area of the synthesized DES circuit is 62945.46 μm^2.

3.2 Testbed used for circuit simulation and power trace generation

Since no two ICs are identical even if they use the same masks and go through the same fabrication process, their side-channel signals differ even for the same input. This is called process noise variation, which is the most influential misjudgment factor for Trojan detection we have to face [2]. Following the practice of the literature [2], we set up three experimental scenarios for $\pm$2%, $\pm$5%, and $\pm$7% process noise level respectively. For each scenario, the above Genuine DES and three kinds of Trojan were implemented. So, there will be 9 experimental groups which are shown in Table III to evaluate our proposed method. Similar to the method used in [2], we modeled the process noise by randomly altering the parameters of the target technology library first in the range of $\pm$2% to form 87 technology libraries, then increased process noise level up to $\pm$5%, and finally to $\pm$7%. Each different variation of parameter values represents a different physical circuit manufactured through the same process. For each experimental group, the gate level netlist of DES and the testbench are the same, and the distribution of side-channel data is shown in Table II.

Basically, we followed the power simulation method in [2]. We used Synopsys Design Compiler X-2005.09 with the Chartered 0.18 μm, 1.8 V technology library for the synthesis of the DES with and without the Trojan. We

Table II. Side-Channel Data Distribution of an Experiment

Genuines (87)	For Constructing Hyper Sphere (50)	For Training (30)
		For Training Optimization (20)
	For Test Purpose (37)	
Trojans (87)	For Test Purpose (87)	

DOI: 10.1587/elex.10.20130602
Received July 31, 2013
Accepted August 13, 2013
Publicized August 28, 2013
Copyedited September 10, 2013

also used ModelSim SE PLUS 6.0 for functional simulation and switching activity analysis and Synopsys PrimeTime PX C-2009.06 for power analysis. We ran simulations and obtained power traces at 125 MHz for different scenarios.

4 Experimental results and performance validation

After power traces acquisition for a given type Trojan circuit and process variation (e.g. ±2%), we conducted the K-L expansion to find a signal eigenspace based on known Genuines, and then Genuines and Trojans signal were projected in the eigenspace where signals from Trojans and Genuines are likely to have different characteristics. With 2-input NAND taken as an Trojan circuit, Figure 2 illustrates the above processing result (up to 87 samples) when process variation is assumed to be ±2%. Obviously, it is difficult to distinguish these curves as Trojan or genuine IC if only by observing Figure 2. However, the proposed method in this paper can avoid plotting, because it processes the ICs side-channel signal with K-L expansion and takes use of an intelligent algorithm, OCSVM, to achieve the Hardware Trojan auto-recognition.

Table III shows the recognition rate results for different scenarios based

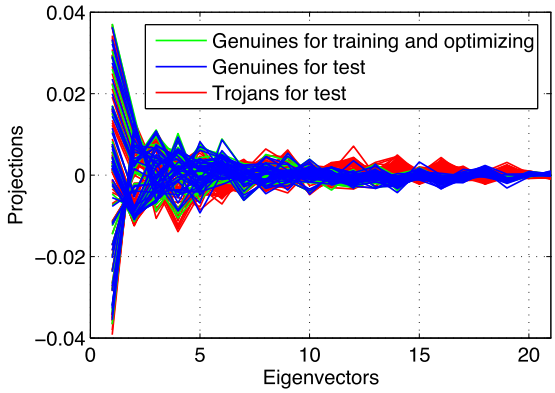

Fig. 2. Projections of Power Traces from Genuines and Trojans (2-input NAND, ±2% Process Noise Level)

Table III. Recognition Result and ξ value

Exp.	Trojan Type	Ralative Area	Process Variations	Recognition Rate			ξ value
				Genuine	Trojan	Total	
1	4-bit counter	0.68%	2%	100%	100%	100%	0.00973
2	4-bit counter	0.68%	5%	100%	100%	100%	0.00243
3	4-bit counter	0.68%	7%	100%	100%	100%	0.00153
4	2-4 decoder	0.08%	2%	100%	100%	100%	0.005859
5	2-4 decoder	0.08%	5%	100%	100%	100%	0.000152
6	2-4 decoder	0.08%	7%	100%	100%	100%	0.000087
7	2-input nand	0.015%	2%	100%	100%	100%	0.000455
8	2-input nand	0.015%	5%	90.70%	90.80%	90.77%	0.002481
9	2-input nand	0.015%	7%	90.70%	80.46%	83.85%	0.001421

DOI: 10.1587/elex.10.20130602
Received July 31, 2013
Accepted August 13, 2013
Publicized August 28, 2013
Copyedited September 10, 2013

on our proposed method. In the Table III, 'Genuine' denotes the correct recognition rate of Genuines, 'Trojan' denotes that of Trojans, and 'Total' stands for the correct recognition rate of all the ICs under test. As shown in Table III, in case of relatively large HT (e.g. the first three experiments), a very good recognition result can be achieved at different process noise levels; As the HT becomes smaller, the recognition rate can reach 100% even though the process noise level reaches ±7% (e.g. the successive three experiments). When the HT is further reduced to the extent that there is only one 2-input NAND gate (e.g. the last three experiments), the recognition rate can reach 100% only in case of relatively small process noise level (e.g. the 7th experiment). However, the recognition rate will be reduced in case of large process noise level (e.g. the 8th and 9th experiments). That's because a 2-input NAND is so little that the resulting power is completely submerged in the process noise, and the characteristic data after K-L expansion is also becoming poor. Nevertheless, we believe our proposed classification method based on OCSVM may become a good and long-term solution for Hardware Trojan detection when combined with some more effective characteristic processing technique to enlarge Trojan activity.

5 Conclusion

In this paper, our contribution is to introduce OCSVM to Hardware Trojan detection, and to propose a more sophisticated detection flow based on side-channel analysis, which can achieve intelligent classification for ICs under test, rather than depending on visual inspections upon the plotted figure. The experimental results under different scenarios show that OCSVM as a post-processing method of K-L expansion is efficient and practical for Hardware Trojan ICs recognition.

DOI: 10.1587/elex.10.20130602
Received July 31, 2013
Accepted August 13, 2013
Publicized August 28, 2013
Copyedited September 10, 2013