A Hardware Trojan Detection Method Based on Side-channel Analysis

Wang Xiaohan[1,a], LI Xiongwei[1,a], XU Lu[1,a]

[1]Department of Information Engineering, Ordnance Engineering College, Shijiazhuang, China

[a]wxh2225@126.com

Keywords: Integrated Circuit; Hardware Trojan; side-channel analysis; Maximum Margin Criterion

Abstract. For the problem of Hardware Trojan detection, we analyzed the statistical properties of the power side-channel signal, and proposed a Hardware Trojan(HT) detection method based on the power side-channel signal. We processed the power side-channel signal by the maximum margin criterion(MMC), and projected the power side-channel signal onto low-dimensional subspace. We extract signal's Characteristic vector sequence, and detect Hardware Trojan in IC chip through statistical processing and analysis. We did simulation experiments to verify by Monte Carlo, and detect Hardware Trojan in C7552 circuit. Experimental results show that at ±20% of the process noise we can detect Hardware Trojan that accounts for 0.028 percent of the circuit, and compared with the K-L transform, this method achieved a good experimental result.

Introduction

With the trend of globalization about Integrated Circuit (IC) design and manufacturing, the design process of IC chip separate from the manufacturing process, so that IC is facing a growing number of security threats. An attacker could change the original design of circuit or implant redundant circuit with malicious functions in the IC chip at any part of the IC chip manufacturing, which is called Hardware Trojan(HT)[1]. Hardware Trojan can achieve destructive function to destroy the chip or disclose confidential information inside the IC chip under certain condition, that is very hard to defend. How to detect Hardware Trojan which is hidden in the circuit to ensure the Integrated Circuit chip security has become a serious problem to be solved[2].

The detection of Hardware Trojan is very difficult. Traditional functional testing and logical testing can only detect parameter's defect or unacceptable error in device, which can not recognize additional functions that caused by the activation of Hardware Trojan, and can not meet the requirement of inspection. Nowadays, detection methods at home and abroad are mainly destructive testing, logic testing and side-channel analysis and the like. Because of its low cost and highly effectual detection side-channel analysis has become more mainstream detection method in the current[3]. For examples, the author used power signal to detect Hardware Trojan in circuit[4], the author used the thermal signal to achieve the Hardware Trojan detection[5], the author detected Hardware Trojan by detecting delay information in circuit[6], etc. There are also some studies using signal processing methods to detect Hardware Trojan , for examples , the author used Karhunen – Loève transform(referred to K-L transform) method to process the power side-channel signal, mapped the power tracks to noise's characteristic space, and built a "fingerprint" in the characteristic space to detect Hardware Trojan in [7]. The author used singular value decomposition method to construct projection subspace, found differences between reference chip("golden chip") and chip with Hardware Trojan in the subspace, and achieve detection against Hardware Trojan.

Hardware Trojan detection method described above may be effective, but there are still a large space for improvement. Such as the K-L transform in [7], its essence is to find the subspace projection that can as much as possible characterize samples as an index of the sample variance, but it may not be effective for the detection of Hardware Trojan. Therefore, we reference the idea of Hardware Trojan detection method based on K-L transform, put forward a new method of signal transform analysis to find the optimal characteristic space that can effectively distinct the "golden chip" and the IC with Hardware Trojan by Maximum Margin Criterion (MMC), and achieve the

goal to detect Hardware Trojan in circuit. We implanted a Hardware Trojan in the ISCAS85's reference circuit and detected it, verified the validity of the method, and achieved better detect results.

Power Side-channel Signal Analysis

Hardware Trojan detection based on power side-channel signal is to run "golden chip" and IC under test under the same conditions, and measure the operating current when the chip runs. Through comparing the difference between the current to determine whether the Integrated Circuit under test contain Hardware Trojan or not. Since the interaction of radiation energy between the various components, the actual measured operating current is a complex coupling time-domain signal, its composition can be roughly divided into four parts[9]:(1) Main circuit current, for all circuits are the same;(2) Measurement noise, which can be eliminated by averaging multiple measurements;(3) Process noise, which is randomly generated and can not be offset;(4) Hardware Trojan signal that may be present.

The presence of noise seriously affect the detection results of Hardware Trojan. How to model the noise is important for detecting Hardware Trojan. In fact, the noise's energy consumption of the power consumption side-channel signal at a time obey normal distribution. However, the noise of power track at two adjacent moments often changed little, there is a certain correlation between the noise of adjacent points. More nearer Two adjacent points, more relevant, and vice versa. In order to characterize the correlation between energy track, we use multivariate normal distribution to model power side-channel signal, use covariance matrix C to indicate the noise fluctuation in power side-channel signal and the correlation between adjacent points, and use mean vector m to represent the main current in circuit as well as Hardware Trojan signal[9]. Here C and m are unknown, we require to use a plurality of power side-channel signal to estimate them. The more amount of power side-channel signal, the estimated value about C and m is closer to the true value. If the power side-channel signal is enough, we can easy to know that

$$C_1 = C_2, m_1 = P_e, m_2 = P_e + P_{tr} \tag{1}$$

Wherein C_1, m_1 and C_2, m_2 are separately the covariance matrix and the mean of "gold chip" and IC with Hardware Trojan. From the view of geometric point, the power side-channel signal of "golden chip" and IC with Hardware Trojan can be seen as two super-ellipsoids which have the same size and shape but at different positions in space. Therefore, Hardware Trojan detection can be seen as a problem to find the greatest difference between two super-ellipsoids. We can detect the Hardware Trojan through this model.

Detection Method Based on Maximum Margin Criterion

The maximum distance criterion

Maximum Margin Criterion (MMC) is a linear discriminate analysis method proposed by Li et al.[10], the basic idea is similar to Fisher linear discriminate. It projected the information related to the classification characteristics in the high-dimensional data samples onto the best low-dimensional identification vector space. Sample projections in the space have the maximum distance between classes and the minimum distance within classes, so that two samples in new space have high divisibility.

Because Trojan detection can be considered as a problem to identify two categories, we introduce the Maximum Margin Criterion from the view of two categories. If we refer to the power tracks of IC without Trojan as w_1 category, and refer to the power tracks of IC with Trojan as w_2 category. The samples of two categories can be separately expressed as that $w_1=\{X_1^1,\ldots\ldots, X_{N1}^1\}$, $w_2=\{X_1^2,\ldots\ldots, X_{N2}^2\}$, in where Xi represents a D-dimensional power curve, the mean vector of category is that,

$$m_i = \frac{1}{N_i}\sum_{x_j \in w_i} x_j, i = 1, 2 \tag{2}$$

The maximum distance between classes and the minimum distance within classes are respectively that,

$$\begin{cases} S_b = \sum_{i=1}^{2} (m_i - m_0)(m_i - m_0)^T, m_0 = \dfrac{N_1 * m_1 + N_2 * m_2}{N_1 + N_2} \\ S_w = \dfrac{1}{N_1 + N_2} \sum_{i=1}^{2} \sum_{x_j \in w_i} (x_j - m_i)(x_j - m_i)^T, i = 1, 2. \end{cases} \tag{3}$$

If we project the data samples onto the new characteristic space by the projection matrix $w = \{\varphi_1, \varphi_2, \varphi_3, \cdots, \varphi_D\}$, in where φ_i is a D-dimensional unit vector, so that the projected sample changes that,

$$\mathrm{Y_i} = w^{\mathrm{T}} * \mathrm{X_i}, \mathrm{i} = 1, 2, \ldots\ldots, \mathrm{N_1} + \mathrm{N_2} \tag{4}$$

Correspondingly, The maximum distance between classes and the minimum distance within classes are respectively that,

$$\tilde{S}_b = w^T S_b w, \tilde{S}_w = w^T S_w w \tag{5}$$

In order to make separation of two projected categories possible, not only to consider the distance between two mean vectors, but also consider the dispersion of the respective categories, therefore MMC defines the Feature extraction rule formula as:

$$\max J(w) = tr(\tilde{S}_b - \tilde{S}_w) \tag{6}$$

Then Formula (6) can be reduced to

$$\max J(w) = tr(\tilde{S}_b - \tilde{S}_w) = tr(w^T (S_b - S_w) w) = \sum_{i=1}^{d} \varphi_i^T (S_b - S_w) \varphi_i \tag{7}$$

Since φ_i is an unit vector, $\varphi_i^T \varphi_i = 1$,so the optimal solution of the formula can be optimized as that,

$$\begin{cases} \max J(w) = \sum_{i=1}^{d} \varphi_i^T (S_b - S_w) \varphi_i \\ s.t. \quad \varphi_i^T \varphi_i = 1 \end{cases} \tag{8}$$

We use Lagrange method to get unconstrained objective function:

$$g(\varphi) = \sum_{i=1}^{d} \varphi_i^T (S_b - S_w) \varphi_i - \sum_{i=1}^{d} \lambda_i (\varphi_i^T \varphi_i = 1) \tag{9}$$

Then we get the partial derivative of the vector φ_i and make it zero, namely, $\dfrac{\partial g(\varphi)}{\partial \varphi_i} = 0, i = 1, \cdots, d$ .Then,

$$((S_b - S_w) - \lambda_i I)\varphi_i = 0, i = 1, \cdots, d \tag{10}$$

So that, φ_i is the characteristic vector of matrix $(S_b - S_w)$, λ_i is a corresponding characteristic value, then Formula (6) can be further into

$$J(w) = \sum_{i=1}^{d} \lambda_i. \tag{11}$$

So that, MMC transformation arranged characteristic values of matrix $(S_b - S_w)$ in descending order, select the characteristic vector Corresponding to the first d characteristic value, and constitute a new characteristic space, in which maximum the boundary between the projection of two category data sets.

Trojan Detection Program

Based on the method above, we give the detailed program about detect Hardware Trojan through Maximum Margin Criterion:

(1) Applying a fixed test vector to "golden chip" and measuring the power side-channel signal

of circuit under this vector. And getting a data set without Trojan $B=\{b(i,j)|i=1,2,\ldots,n;j=1,2,\ldots,m\}$, in which, n represents the number of power side-channel signal, m represents the sampling length of each power track. Similarly, getting power consumption data $T=\{t(i,j)|i=1,2,\ldots,n;j=1,2,\ldots,m\}$ of tested IC in the same way. By comparing the two data sets to determine whether the test chip contain a Trojan.

(2) Simultaneously processing two data sets, using Formulas above to calculate the maximum distance between classes S_b and the minimum distance within classes S_w of two data sets. Then calculating characteristic vectors(projection directions) of the matrix (S_b-S_w), and setting up the characteristic space S. Then projecting two data sets B and T to characteristic space and obtaining the projection data sets B' and T'.

(3) Comparing two projected data sets and finding differences between two projected data sets, we set projected data set's "μ±3σ" of "golden chip" to the boundary of detecting Hardware Trojan. If the projected data set of measured IC beyond the "μ±3σ" range on a certain projected direction, the test chip contains Hardware Trojan, otherwise the test chip doesn't contain Hardware Trojan.

Trojan Detection Simulation experiment

Experimental setup

In order to verify the effectiveness of the Hardware Trojan detection program we proposed, we validate the detection method through simulation in this paper. In this paper, we use HSPICE software developed by Meta-Software to complete the modification of IC's design, power simulation and other operations. We set the standard circuit c7552 in reference circuit ISCAS85 as a test circuit. In this circuit, there are 3,512 gates. We implant Hardware Trojan circuit of different sizes in the circuit, and collect power side-channel signal when the circuit works to detect.

In fact, when the power side-channel signal of the true circuit is measured, due to the effects of process noise, the form of each measured power side-channel signal is uncertain, it is more difficult to find hidden Trojan circuit. In order to better simulate the impact of process noise in real circuit, enable the simulation result as close to the actual result, in this study, we use the device parameters provided by 180nm technology library to analog circuits, and analysis the circuit by Monte Carlo. The relative change range of process noise is ±20%. Meanwhile, in order to improve simulation accuracy and get multiple samples from the circuit, the simulation time resolution is set to 1ps, and we apply fixed test vectors for each 2ns, and get 1000 energy consumption tracks for each circuit (circuit with Trojan or circuit without Trojan), Each energy consumption track contains 1000 sampling points.

Experimental result

Fig. 1 and Fig. 2 are respectively partial enlarged view of detection result for Hardware Trojan by K-L and MMC, which accounted for 0.17 percent of the size of the circuit. The abscissa represents characteristic vector sorted according to characteristic value, ordinate is the projection value on each characteristic vector, red part is the projection distribution of "golden chip", and green part is the projection distribution of IC with Hardware Trojan. Two black lines in the figure are the delineated detection boundary. From the figure, both methods can successfully detect Hardware Trojan.

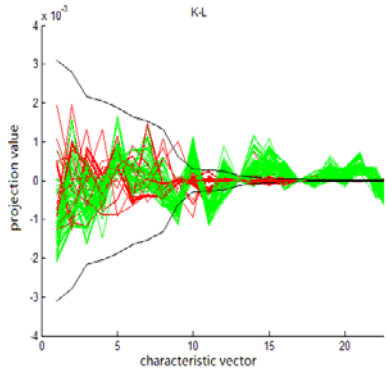

Fig 1 test result for Hardware Trojan by K-L transform

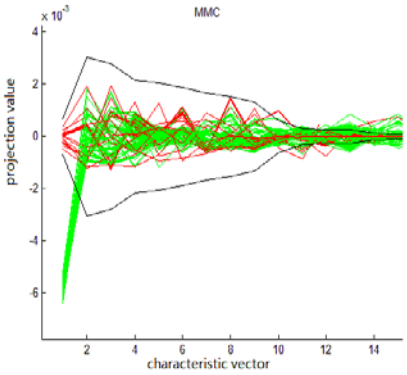

Fig 2 test result for Hardware Trojan by MMC

When the Hardware Trojan is reduced, two clusters of curve will approach each other, and begen to appear superimposing. Such as Fig 3 and Fig 4. while implanting a gate (about 0.028% of the total circuit) in the circuit, it is difficult to directly determine whether the circuit contains Trojan by K-L transform. The MMC method is still able to detect the Hardware Trojan on the first characteristic vector.

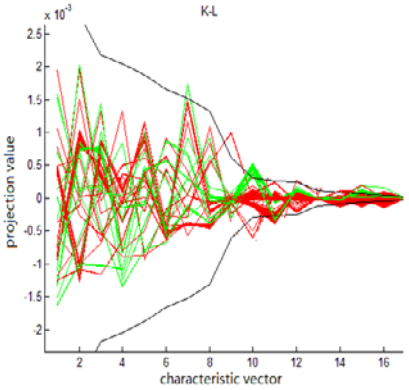

Fig 3 test result for Hardware Trojan by K-L transform

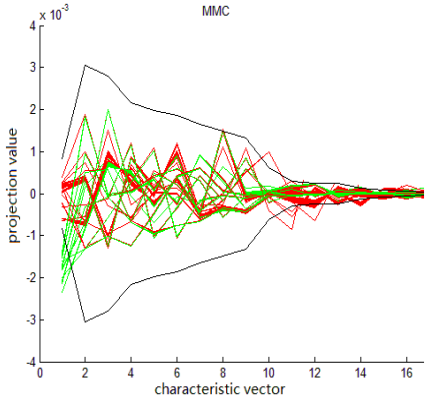

Fig 4 test result for Hardware Trojan by MMC

Summary

Hardware Trojan detection method based on MMC proposed in this paper can effectively detect Hardware Trojan circuit. We conducted simulation experiments on standard circuit c7552, and made more obvious detection results. It provides a new way to detect Hardware Trojan. In order to further improve the detect effect, it is need to future research a study of statistical decision method, and improve recognizable about the Hardware Trojan circuit on the basis of existing methods. Furthermore, the method is to detect based on the statistical characteristics of the side-channel signal, the side-channel signal is statistically non-stationary and will change over time. It is need to find analysis tools for meaningful statistical signal to detect smaller Hardware Trojan.

References

[1] Tehranipoor M, Koushanfar F. A Survey of hardware Trojan Taxonomy and Detection [J]. IEEE Design & Test of Computers, 2010, 27(1): 10-25.

[2] Wang X, Tehranipoor M, Plusquellic J. Detecting Malicious Inclusions in Secure Hardware: Challenges and Solutions [C]. in Proceedings of the IEEE International Workshop on Hardware-Oriented Security and Trust (HOST'2008), 2008: 15-19.

[3] Chakraborty R, Narasimhan S, Bhunia S. Hardware Trojan: Threats and emerging solutions [C], in proceedings of the IEEE International Workshop on High Level Design Validation and Test

Workshop, 2009: 166-171.

[4] RAD R M, WANG X X, TEHRANIPOOR M, et al. Power supply signal calibration techniques for improving detection resolution to hardware Trojan[A]. ICCAD 2008[C]. San Jose, USA, 2008. 632-639.

[5] Wei S, Meguerdichian S, Potkonjak M. Malicious circuitry using thermal conditioning[J]. Information Forensics and Security, IEEE Transactions, 2011, 6(3):1136-1145.

[6] Jin Y, Makris Y. Hardware Trojan Detection Using Path Delay Fingerprint [C], in Proceedings of IEEE International Workshop on Hardware-Oriented Trust Security, 2008: 51-57.

[7] Agrawal D, Baktir S, Karakoyunlu D, et al. Trojan Detection Using IC Fingerprinting [C], in Proceedings of the Symposium on Security and Privacy (SP'2007), 2007: 296-310.

[8] Wang liwei, Luo Hongwei, Yao Ruohe. Hardware Trojan detection method based on analysis of side-channel [J] South China University of Technology (Natural Science), 2012,40 (6): 6-10.

[9] S. Mangard, E. Oswald, T. Popp. Feng Guodeng translated. Power analysis attacks [M] Beijing: Science Press, 2010.

[10] Li H, Jiang T, Zhang K. Efficient and robust feature extraction by maximum margin criterion [J]. IEEE Trans. Neural Networks 2006, 17 (1):157-165.