PROCEEDINGS OF SCIENCE

A Hardware Trojans Detection Method by Side-channel Analysis

Lin Ni [1]
Xi'an Communications Institute
Xi'an ,710106 ,China
E-mail: nilin1008@sina.com

Kun Han[2]
Xi'an Communications Institute ;Xidian University
Xi'an ,710106 ,China
E-mail: ni_linmake@sina.com

Zhixun Zhao
National University of Defense Technology
Changsha,410073, China
E-mail: zhaozhixun1991@163.com

Pei Wei
Electronic Experiment Center of Luoyang
Luoyang, 471003, China
*E-mail:*wp3560531@sina.com

As a bomb hidden in the chips, the Hardware Trojans cannot be eliminated as the same as the software virus, which has brought about major security challenges for integrated circuit. In a variety of currently proposed methods, the method of side-channel analysis is more effective for Hardware Trojans detection. In this paper, we put forward and optimize the hardware Trojan power noise model by dissecting the power consumption of chip, utilize the method of feature projection to achieve noise optimization and feature extraction for original circuit and Trojans circuit. Then a new Hardware Trojans detection algorithm based on convergence and divergence analysis is proposed. The experiment sets up a hardware platform and designs a secret key leaked Hardware Trojans based on the sequence detection in Advanced Encryption Standard (AES) algorithm, The results show that the method can effectively detect Hardware Trojans with more robustness; and we can realize accurate detection for Hardware trojans under the actually measured data when no more than 1% area of circuit is acquired.

ISCC 2015
18-19, December, 2015
Guangzhou, China

[1]Speaker

[2]This work is supported by the "Shaanxi Province Natural Science Fund" (Project No.: 2014JM2-6097).

PoS(ISCC2015)044

1. Introduction

Given the improvement in design and manufacturing level of IC (Integrated Circuit), attacks on security chips (i.e., leaking confidential information and invalidating chips) have become simpler. In addition, driven by Moore's Law and the economic interests, division of labour of IC design, production and packaging have become more detailed and specialized; besides, more and more designs are outsourced to the third party for economic interests, which intensifies the separation of IC design and production. A large number of the third-party IP cores and EDA(Electronic Design Automation) tools are used to shorten the IC design cycle and improve market occupancy. Therefore, the integrated circuits are likely to be maliciously modified by uncontrolled third party in multiple links of IC production. This vulnerability severely threatens the security of IC, especially the encryption chips[1-3].

Hardware Trojans are small and malicious circuit. They are inserted by adversaries to change the original circuit by using redundant logic and layout of chips. In the meantime, they modify the system function, leak the confidential information, and destroy a system either unconditionally or under certain conditions. Although there is no news that Hardware Trojans were used in large-scale military confrontation, but the Western countries have already launched much exploration of Hardware Trojan attacks, which have been individual incidents. According to the front page of the *New York Times*, the US National Security agency directly implant Hardware Trojans in the USB communication protocol or USB port by spies and computer manufacturer so as to achieve the target of remote monitoring; therefore, the method of Hardware Trojans detection needs to be carried out for IC security. Up to now, the researchers have proposed many methods, in which, the side-channel analysis has received widespread attention in academic circles with the help of superior detection efficiency. The key of detection is to propose a model to optimize noise and increase precision [4,5].

2. Hardware Trojans and Power Model Optimization

The realization of Hardware Trojans is closely related to underlying hardware, they have a variety of trigger modes and functional units[6]. Fig. 1 shows the general implementation of Hardware Trojans in a single chip, which consists of two functional parts: the trigger logic activates trojan throughinput monitoring, date bus, register date, circuit work state, and time setting; as the execution unit of hardware Trojans, the payload logic implements the attacks.

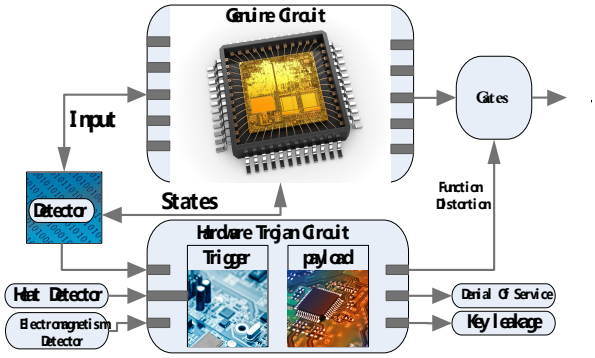


Figure 1 : Diagram of Hardware Trojans

Currently, it's difficult to perform complete detection for post-silicon chips that produced by foundries lack of controllability. In addition, a large number of reusable third-party IP cores are used for chip design;and Hardware Trojans can be flexibly hidden in IP cores. The complete examination is not realistic for all of the IP cores while the reverse engineering is destructive and exhaustive tests based on function test cannot be conducted. The method based on side-channel analysis has effective detection precision, but it can be easily affected by the process noise.

The Hardware Trojans detection based on power analysis is selected in this paper. Once the Hardware Trojans have been implanted in a genuine chip, the characteristic of chip under test varies from those of the original in terms of power and delay. Typically, the power consumption of CMOS circuit can be described as $P_{tot} = P_{dyn} + P_{short} + P_{leak}$ [7]. The experiment analyzes the global current I_{DD} in the target circuit. We obtained N block genuine chips Ci (i∈N) by changing the process parameters. We also input a constant test vector K to reiterate the calculation process at different temperatures T and supply voltages V as well as power measurements M. The I_{DD} of genuine chip C_i mainly consists of two parts in normal operation:

$$I_{DD}^{(C,K)}(\mathrm{C,K,M,t}) = I_{G}^{(C,K)}(\mathrm{C,K,t}) + \Delta I_{DD}^{(C,K)}(\mathrm{C,K,M,t}) \tag{2.1}$$

There is the genuine current of original circuit $I_{G}^{(C,K)}(\mathrm{C,K,t})$, which is measured without any noise. The current is mainly related to chip (C), calculation (K) and time (t). Similarly, the current introduced by noise $\Delta I_{DD}^{(C,K)}(\mathrm{C,K,M,t})$ is associated with PVT and M. In the chips under test that was implanted by Hardware Trojans, the additional logic that affects the total power can be seen as noise. Its contribution can be expressed as $I_{HT}^{(C,K)}(\mathrm{C,K,t})$:

$$I_{DD}^{(C,K)}(\mathrm{C,K,M,t}) = I_{G}^{(C,K)}(\mathrm{C,K,t}) + I_{HT}^{(C,K)}(\mathrm{C,K,t}) + \Delta I_{DD}^{(C,K)}(\mathrm{C,K,M,t}) \tag{2.2}$$

The measurement noise and VT noise obey N (u, σ2) in the actual sampling process. This noise is typically random and can be easily eliminated by averaging data obtained under the same input from the same IC. We input a constant test vector to reiterate the calculation process S times to get power curve containing n sampling points $\mathrm{Tra}_1, \mathrm{Tra}_2 \ldots\ldots \mathrm{Tra}_s$, of which $\mathrm{Tra}_i = (p_{i1},\ p_{i2},\ \cdots\ , p_{in})$. Each sampling point of s curve is used to do the average:

$$\overline{C_i} = \left(\overline{p_1},\ \overline{p_2},\ \cdots\ , \overline{p_n}\right) = \left(\frac{1}{s}\sum_{i=1}^{s} p_{i1},\ \frac{1}{s}\sum_{i=1}^{s} p_{i2},\ \cdots\ , \frac{1}{s}\sum_{i=1}^{s} p_{is}\right) \tag{2.3}$$

After eliminating the random noise, $I_{G}^{(C,K)}(\mathrm{C,K,t})$ can also be subtracted from all of the power traces. The power trace of genuine circuit can be expressed as $Tra_{Gc}(\mathrm{C,K,t}) = \Delta I_{P}^{(C,K)}(\mathrm{C,K,t})$, which Contains only the power variation caused by process. The power trace of Trojan circuit can be expressed as $Tra_{Hc}(\mathrm{C,K,t}) = \Delta I_{P}^{(C,K)}(\mathrm{C,K,t}) + I_{HT}^{(C,K)}(\mathrm{C,K,t})$, which contains the power variation caused by process and Trojan logic. The key of detection is to reduce the noise and extract Trojan current from I_{DD}; hence, the Hardware Trojans detection is to validate $H_0 : Tra_{Gc} \cap Tra_{Hc} = Tra_{Gc} = Tra_{Hc}$.

3.Convergence And Divergence Analysis Model Based on Feature Projection

3.1 Algorithm of Characteristic Projection

In the process of Hardware Trojans detection by using power, eliminating and optimizing the noise from the environment, the voltage jitter, the temperature fluctuations and the measurement error will determine the detection reliability. we can eliminate the effect of random noise by averaging a large number of samples; but the process variation has greater influence in the test and it is difficult to be eliminated. Hardware Trojans is activated by different degrees, the variance of P_{chip} vary in size due to influence of the Hardware Trojans, and this deviation is not available for process variation; therefore, the manifestation of difference of power variance can manifest the effect of Hardware Trojans. The noise problem such as PVT is ascribed to manifest variance of power matrix[8]. For N chips, we can get the power matrix of the original circuit by observing n sampling points of each sample.

$$X=\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \triangleq (X_1, X_2, \cdots, X_n) \tag{3.1}$$

$X_1, X_2, \ldots\ldots, X_n$ are column vectors of X, their linear combination is:

$$F = a_1X_1 + a_2X_2 + \cdots + a_nX_n \triangleq a'X \tag{3.2}$$

We define $a=(a_1,a_2,\cdots,a_n)'$ and $X=(X_1,X_2,\cdots,X_n)'$. To seek the principal component is to seek the linear function of X which makes variance as large as possible. The characteristic root of covariance matrix of original power is $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p > 0$, corresponding unit eigenvector is $u_1, u_2, \cdots, u_p$. We can prove and get maximum for $Var(a'X) = a'\Sigma a$ when $a = u_1$, and $Var(u_1'X) = u_1'\Sigma u_1 = \lambda_1$. Similarly, we get $Var(u_i'X) = \lambda_i$, $Cov(u_i'X, u_j'X) = \sum_{a=1}^{p} \lambda_a (u_i'u_a)(u_a'u_j) = 0, i \neq j$.

Above derivation shows that the main component of $X_1, X_2, \cdots, X_p$ is a linear combination whose coefficient is eigenvector of $X_1, X_2, \cdots, X_p$ and their variance is the characteristic root of Σ. As the independence between various dimensions after projection, it can avoid the overlapping of data space, realize reduction of the dimensionality and simplify the data structure so that the influence of process deviation is optimized.

PCA theory is based on feature transformation. We deal with the data matrix of the difference between the measurement power and the average power. Firstly, M chips are centrally processed. The mean value of the same sampling points of the power curves after the static de-noising is obtained, and the M samples curve can be obtained.

$$C_{TM} = \left(\frac{1}{m}\sum_{k=1}^{m}\overline{p_{k1}}, \quad \frac{1}{m}\sum_{k=1}^{m}\overline{p_{k2}}, \quad \cdots \quad , \frac{1}{m}\sum_{k=1}^{m}\overline{p_{kn}}, \right) \tag{3.3}$$

In the first analysis step, we calculate the offset value based on the mean value of power in each dimension and get the power matrix S which is to be analyzed. The specific feature projection algorithm is shown in Algorithm 1.

$$S = \begin{pmatrix} \overline{C_1} \\ \overline{C_2} \\ \vdots \\ \overline{C_m} \end{pmatrix} - \begin{pmatrix} C_{TM} \\ C_{TM} \\ \vdots \\ C_{TM} \end{pmatrix} \tag{3.4}$$

Algorithm1. Method of Removing Noise based on PCA

Inputs: chips number; UTC Power matrix; Golden power matrix

Outputs: date of noise removal: S

Step1: calculation: the centralized power matrix S;

Step2: covariance matrix P of S: $P = S^T S/(n-1)$ $(P \in R^{n\times n})$;

Step3: decompose eigenvalue λ_i and eigenvectors of P;

Step4: sort the eigenvalue λ_i of P;

Step5: $\eta(k) = \sum_{i=1}^{k}\lambda_i \Big/ \sum_{i=1}^{m}\lambda_i$ for Π>85% choose the most bigger k eigenvalue λ_i;

Step6: structure characteristic matrix K with eigenvectors;

Step7: matrix transformation $S_T = SK$ $(S_T \in R^{n\times k})$;

Step8: output the date results;

S_T is covariance matrix which is mapped by the sample matrix S. As the reference chips (GOLDEN) do the same power processing, Trojan power vector S_G is achieved.

3.2 Convergence and Divergence Analysis Model of Power

S_G and S_T are m*n matrix upon the feature transformation. We can get the spatial distribution of S_G and S_T with each row as a point in n-dimensional space. Define the distance from each point $(x_1, x_2, ..., x_n)$ to the origin as the property which is distance ralative to the power zero.

$$X_{Ti} = \sqrt{x_{Ti1}{}^2 + x_{Ti2}{}^2 + \cdots + x_{Tim}{}^2} \tag{3.5}$$

$$X_{Gi} = \sqrt{x_{Gi1}{}^2 + x_{Gi2}{}^2 + \cdots + x_{Gim}{}^2} \tag{3.6}$$

Calculate the property which is between the relative power zero and the distribution points for matrix S_G and S_T, then we can get two power property vectors which are made up of m power properties $X_G = (x_{G1}, x_{G2}, ..., x_{Gm})$ and $X_T = (x_{T1}, x_{T2}, ..., x_{Tm})$. The power distribution can be seen by putting the power vector X_G and X_T of the same figure. As well we can still use the following formula to calculate mathematical features of power.

$$E(X_T) = \frac{1}{n}\sum_{i=1}^{n} X_{Ti},\ \delta^2(X_T) = \frac{1}{n}\sum_{i=1}^{n}\left(E(X_T) - X_{Ti}\right)^2 \tag{3.7}$$

$$E(X_G) = \frac{1}{n}\sum_{i=1}^{n} X_{Gi},\ \delta^2(X_G) = \frac{1}{n}\sum_{i=1}^{n}(E(X_G) - X_{Gi})^2 \tag{3.8}$$

For convergence analysis of power property, using the power property vector $X_G = (x_{G1}, x_{G2}, ..., x_{Gm})$ and $X_T = (x_{T1}, x_{T2}, ..., x_{Tm})$ of the original circuit and the test circuit, we can get m discrete values $|(x_{Gi} - E(X_G))|$ and $|(x_{Hi} - E(X_H))|$ making the power property minus expectation, then we sort them and make them to satisfy Formula (3.9):

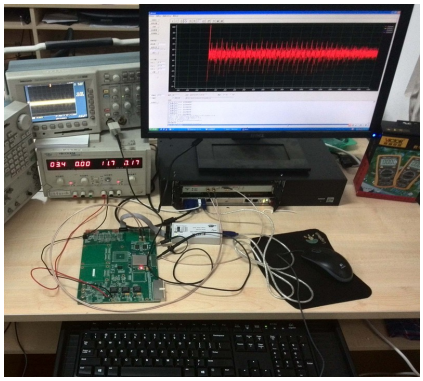
$$|(x_{Hi} - E(X_H))| \geq |(x_{Hj} - E(X_H))| \geq |(x_{Hk} - E(X_H))| \tag{3.9}$$

Among them i, j, k = 1, 2… m, and i≠j≠k. By comparing the convergence and divergence of the original circuit and Trojan circuit, we can see whether the circuit has been modified.

4. FPGA Platform Building and Experiment Analysis

4.1 Realization of Hardware Trojans in FPGA

The key of verifying robustness of Hardware Trojans detection based on the side-channel analysis has the following two points. Firstly, the detection process has transformation from the simulation platform to the measured platform; secondly, the Hardware Trojans are realized in the attack carrier that covers the general architecture. Considering the Field Programmable Gate Array(FPGA) vendor's market share and its own function, the article uses Xilinx Kintex 7 to set up the power acquisition platform. We adopt AES encryption algorithm as the attack carrier which covers the general microprocessor architecture. It has been internationally accepted. Besides, high-precision signal acquisition board, oscilloscope and PC are used to build the power platform, as shown in Fig. 2:



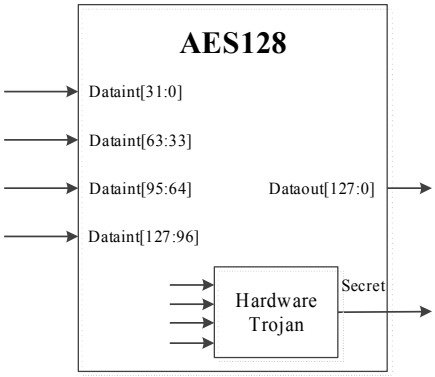


Figure 2: Power Acquisition Platform **Figure 3:** Hardware Trojans Circuit

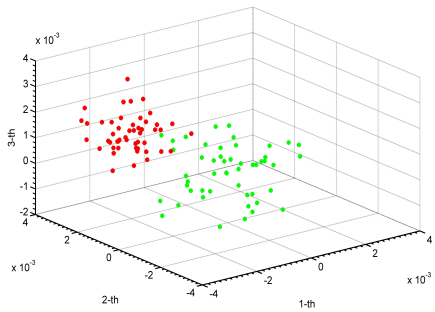
AES algorithm circuit has more practical attack value. The experiment designs a secret key leaked Hardware Trojans based on sequence detection and implements Hardware Trojans in AES circuit, including Hardware Trojans circuit equivalent gates size occupies 1.01%. As shown in Fig. 3, the secret key of 128bit is used to encrypt the input data in normal work. The trigger logic of Hardware Trojans monitors the input data to activate the circuit. After detecting a specific sequence of input data, the Trojan leaks the secret key of AES-128 through FPGA vacancy pin.

4.2 Power Analysis of Hardware Trojans

The process variation of the chip is random in the same batch, but it has the same direction in different batches. To reflect the influence of the actual process noise on the Hardware Trojans detection as far as possible, the experiment captures 50 groups of power curves reflecting different process characteristics for analysis, so a 50*500 power matrix is achieved.

In order to eliminate the interference of random noise, the test chip is sampled under the same condition as repeated; then power curve Tra_1 is achieved after averaging the data. Do as this, we get all the sample chip power curve *TRA*= (Tra1, tra2... Tra50). Then we handle the power curve *TRA* with the feature projection algorithm in 3.1, the subspace matrix of the original chip and under test chip are obtained, as described as S_G and S_T. Construct a subspace

from the first three-dimension of the power matrix S_G and S_T, the subspace character distribution is shown in Fig. 4：



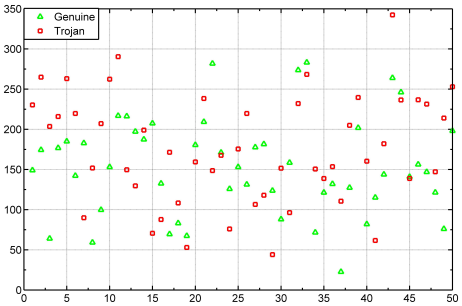


Figure 4 : Subspace Character Distribution **Figure 5 :** Projective Distribution of Power Data

The distribution areas almost separated in addition to individual point. According to the process of Hardware Trojans analysis, the distance between Subspace matrix and the zero point, the distribution of power property is shown in Fig. 5.

In Fig. 5, the red dots show the power property that Hardware Trojans is implanted and the green dots show the power property of the original AES chip. The results show that the projected power distribution displays certain differences. The power of the more than 82% under test chip is higher than that of the original AES chip and the power property difference varies with different projection points. Thus, the power property diversity based on data projection can be used as the symbol of recognition whether the Hardware Trojans is implanted.

Upon mathematical analysis of the power projection data of the original AES, we can find that the power of the under test chip is obviously higher than the original. From the perspective of the matrix transformation, the power matrixes of the original chip and the chip under test multiply the same projection matrix *K*, the projection process maintained their relative power size; thus the projection power characteristics can also reflect the relative size. From distribution of the variance of power attribute, the degree of the original chip and the chip under test deviating from its expectation varies from characteristics of the Hardware Trojans. In order to observe the convergence degree, the power attribute is used as the sequence whose expectation value is convergent.

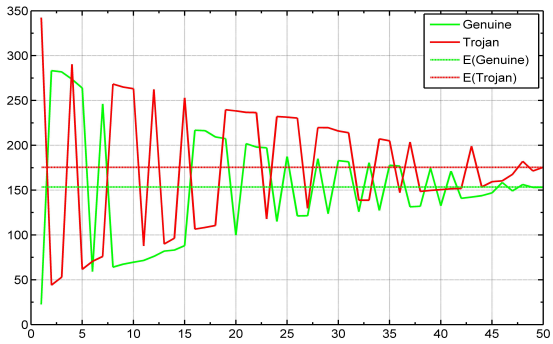


Figure 6: Distribution Trends of Data Convergence

From the data convergence distribution in Fig. 6, the power projection points converge to their expectations. From the process of convergence and divergence, we can see that the convergence speed of the chip under test varies with original AES chip; at meanwhile, in the power projection distribution, the fluctuation around the expectations of the chip under test is

smaller. From the whole process of feature extraction, the matrix feature transformation manifest the feature of power variance. Because of the existence of the Trojan, the power projection distribution of the chip under test is concentrated, and the convergence is better than the original circuit. Through convergence diversity we can well distinguish the difference between the Trojan chip and the original chip to achieve the purpose of Hardware Trojans detection.

5. Conclusion

The process noise as the key factor that affects the Hardware Trojans detection based on the side-channel information, has great influence on the sensitivity of Hardware Trojans detection. Aiming at the problem of great process influence for Hardware Trojans detection, the paper optimizes the power model of Hardware Trojans, eliminates the process noise by the method of feature extraction, and proposes a new Hardware Trojans detection model and algorithm based on convergence and divergence analysis. Finally, the experiment sets up FPGA test platform and designs a secret key leaked Hardware Trojans based on sequence detection in AES algorithm with the size of 1.01%. The results show that the method can effectively optimize the process noise and detect Hardware Trojans on the scale of landscape level 1%. It provides a feasible and effective solution for noise elimination and algorithm research.

References

[1] L. Ni, S. Q. Li, R. C. Ma, P. Wei. *Hardware Trojans Detection and Protection*[J].Digital Communication. 41(1) :59-63(2014) (In Chinese)

[2] Q. Sui. *Hardware Trojan detection based on side channel signal analysis*[D].Changsha: National University of Defense Technology(2012) (In Chinese)

[3] D. Agrawal, S. Baktir, D. Karakoyunlu, et al. *Trojan detection using IC fingerprinting*[C]. Proceeding of the 2007 IEEE Symposium Security and Privacy. Oakland:CA, USA. pp, 296-310(2007)

[4] C. L. Liu , Y. Q. Zhao, Y. F. Shi, Z. Z. Feng. *Hardware Trojan Detection Method Based on Correlation Analysis*[J].Computer Engineering. 39(9):182~186(2013) (In Chinese)

[5] R. S. Chakraborty, S. Narasimhan, S. Bhunia. *Hardware trojan: Threats and emerging solutions*[C]. Proceedings of High Level Design Validation and Test Workshop. San Francisco:CA. pp,166~171(2009)

[6] L. W. Wang, H. Xie, H. W. Luo. *Malicious circuitry detection using transient power analysis for IC security*[C].Proceedings of Quality, Reliability, Risk, Maintenance, and Safety Engineering, pp,1164-1167(2013)

[7] Neil H.E.Weste, David Money Harris. *CMOS VLSI Design*[M].Publishing House of Electronics Industry, pp,145-170(2012)

[8] Z. X. Zhao, L. Ni, S. Q. Li,Y. B. Shi. *A Feature Extraction Method for Hardware Trojans Detection*[C]. International Conference on Automation, Mechanical Control and Computational Engineering. Atlantis Press, Paris:CA. pp,1726-1731(2015)