A Side-Channel Hardware Trojan Detection Method Based on Fuzzy C-Means Clustering and Fusion Distance Algorithms

Chunhua He, Dengyun Lei, Heng Wu, Lianglun Cheng, Guizhen Yan, and Qinwen Huang

***Abstract*—With the wide application of the Internet of Things technology, the hardware security has attracted more and more attention from users around the world. Hardware Trojan (HT) of integrated circuit (IC) has become a main security threat gradually. Therefore, HT detection is very significant. In this article, a HT automatic test system used for side-channel test combined logic test is constructed with a high-performance oscilloscope, FPGA chips, a NI digital acquisition card and LabVIEW software. Besides, the test flowchart and data processing method are depicted in detail. Spectral feature analysis combined principal component analysis is proposed for feature extraction. Fuzzy C-means clustering combined spectral energy analysis is put forward to distinguish the Trojan category from the golden category. Then Fusion distance (i.e., Mahalanobis distance combined Euclidean distance) is presented for the real-time HT recognition. A 128-bit AES cipher circuit and a 2-bit counter are applied as a golden circuit and a Trojan circuit, respectively. Experimental results demonstrate that the detection accuracy is 100% and the proposed detection method can easily achieve 0.1% HT detection sensitivity, which verifies that the detection method is feasible and effective.**

***Index Terms*—Euclidean distance (ED) and Mahalanobis distance (MD), fuzzy C-means clustering (FCC), hardware Trojan (HT), principal component analysis (PCA), spectral feature analysis (SFA).**

I. Introduction

With the rapid development of technology, Internet of Things (IOT) products have entered thousands of households. In the new IOT era, hardware security is becoming more and more important [1]. Hardware Trojan (HT) threatens the information security seriously nowadays as the wide application of semiconductor technology goes on [2]. Presently, the third-party services, such as CAD tools, IP core, integrated circuit (IC) design, fabrication, test, and package, are uncontrollable and vulnerable. These untrustworthy designs may be tampered with by the opponents, resulting in serious security risks. Nowadays, ICs are mostly designed independently, but they are outsourced for manufacturing. Thus, a hacker may tamper with the gate level netlist, rather than RTL code, during the manufacturing process. Intentional tampering by the manufacturing factory may lead to disclosure of important information, change of the functions, shortening of the lifetime, refusing of the service, or even system damage [3], [4]. In general, Trojans are usually composed of combinational or sequential circuits, some are triggered by ultralow probability events, and some are triggered after a preset time. Generally, most Trojans contain triggers and payloads. According to the performance of the payload after triggering, Trojans can be divided into explicit payload Trojans and implicit payload Trojans [5]. The explicit payload Trojan, when triggered, will affect the internal signal, causing the chip to perform incorrect function or propagate secret information to the output pins. The triggering principle of an implicit payload Trojan is similar to that of an explicit payload Trojan, but after being triggered, it does not affect the internal signals, but rather leaks confidential information by transmitting radio-frequency signals or damages chips through implicit ways. Relatively speaking, running exhaustive input patterns through traditional logic testing may detect explicit payload Trojans, but cannot detect implicit payload Trojans.

Generally speaking, HT is often too small to be discovered, and it is only triggered under some rare circumstances, which makes the detection so difficult. Because it is hard to find the malicious circuits using traditional testing schemes, to study some effective detection techniques is extremely urgent. So far, more and more institutes at home and abroad focus on a variety of HT detection methods. Among them, logic test, reverse engineering test and side-channel test are three main methods adopted for HT detection. The logic test method can recognize the HT circuit by analyzing the output response of the target circuit stimulated by the effective test patterns [6]. Dynamic flip-flop conversion structures and dummy scan flip-flop can be inserted to decrease the transition generation

Manuscript received 1 September 2023; revised 14 October 2023; accepted 29 November 2023. Date of publication 5 December 2023; date of current version 9 April 2024. This work was supported in part by the National Natural Science Foundation of China under Grant U22A2012, Grant 62104047, and Grant 62173098; in part by the Natural Science Foundation of Guangdong Province under Grant 2023A1515010291; and in part by the Basic and Applied Basic Research Project of Guangzhou Basic Research Program under Grant 2023A04J1707. *(Corresponding authors: Heng Wu; Qinwen Huang.)*

Chunhua He and Lianglun Cheng are with the School of Computer, Guangdong University of Technology, Guangzhou 510000, China (e-mail: hechunhua@pku.edu.cn;llcheng@gdut.edu.cn).

Dengyun Lei is with the School of Integrated Circuits, Guangdong University of Technology, Guangzhou 510000, China (e-mail: leidy@gdut.edu.cn).

Heng Wu is with the School of Automation, Guangdong University of Technology, Guangzhou 510000, China (e-mail: heng.wu@foxmail.com).

Guizhen Yan is with the Institute of Microelectronics, Peking University, Beijing 100871, China (e-mail: gzyan@pku.edu.cn).

Qinwen Huang is with the Science and Technology on Reliability Physics and Application of Electronic Component Laboratory, No. 5 Electronics Research Institute, Ministry of Industry and Information Technology, Guangzhou 510000, China (e-mail: 971230012@163.com).

Digital Object Identifier 10.1109/JIOT.2023.3339488

time and increase the activity of HT [7], [8]. The scalable statistical test generation method and automatic test pattern generation (ATPG) approach can be applied to generate high-quality test set for creating high-relative activity in any Trojan instances [9], [10]. Even though the process variation is large, it is still effective for detecting a small HT circuit, whereas, to obtain the effective test patterns is very difficult, especially for a very large-scale circuit [11].

Reverse engineering test is the most reliable HT detection method. Considering the technical and human factors, the complexity of reverse engineering is hard to be estimated, which is still an unsolved issue [12]. In order to simplify the processing and advance the test efficiency, K-means clustering, backscattering side-channel clustering, path retrace algorithm, and histogram of oriented gradients are applied to assist the reverse engineering test [13], [14], [15], [16]. The effects are obvious, nevertheless, the operations are very time-consuming [17].

Side-channel test, used to measure the tiny differences of power dissipation, delay, maximum operating frequency, emission of light, electromagnetic interference, and thermal map, is another prevalent method applied to detect HT circuits. According to the path delay of ring oscillator networks, HT detection can be achieved [18], [19]. However, it is at the expense of many areas to ensure a high-detection sensitivity. Besides, clock scanning technique and symmetric path delay technique are useful to detect some kinds of HTs [20], [21], but confirming the critical path is so time-consuming. Path delay fingerprint technique is proved effective for detecting the explicit payload Trojan (such as a combinational comparator), instead of the implicit payload Trojan (such as a sequential counter) [5]. Multiple currents analysis has been proved to be effective for HT detection, and the detection sensitivity can be improved greatly [22], [23]. Power dissipation and maximum operating frequency can be used to achieve the HT detection [24], [25], [26], but this detection approach is so complicated since it needs the effective test patterns. The electromagnetic (EM) radiation method is presented to recognize HT [27], [28]. However, it depends so much on the resolution of EM probe. The tiny difference resulted from a HT in the emission map is apt to be affected by process variation and measurement error. Likewise, detection methods based on emission of light [29], [30], or power and thermal maps [31], [32], [33], also rely on the relevant instruments' resolution and the HT circuit's area. In addition, short-term aging effects in FinFET transistors and circuit overclocking may induce bit errors at the circuit outputs, which can be applied for HT detection [34]. A simplified equivalent circuit model and frequency spectrum analysis can be also adopted for HT detection [35], [36], although the detection accuracy is high, the analysis algorithms can be further improved.

Apart from the detection system, the mathematical analysis algorithms are also very essential for HT detection, where feature extraction and pattern recognition are the most essential. In terms of feature extraction algorithms, piecewise average filtering (PAF) [36], singular value decomposition [37] and principal component analysis (PCA) [38], [39] are widely used for the dimension reduction, then the Euclidean or Mahalanobis distance (MD) algorithm is applied to distinguish the Trojan chips from the golden chips [40], [41]. These algorithms are generally effective, but they can be further improved. In terms of pattern recognition algorithms, self-organizing map neural network [32], local outlier factor [42], clustering-based local outlier factor [43], hybrid clustering ensemble method [44], random forest classifier [45], and density-based clustering method [46], [47] are often used as the unsupervised detection algorithms, while support vector machine [48], back-propagation (BP) neural network [27], deep neural network (DNN) [49], and convolutional neural network (CNN) [50], [51] are adopted as the supervised detection algorithms. Despite that the single algorithm is useful for the HT detection, in order to further decrease the amount of computation and advance the detection accuracy, it is usually necessary to fuse multiple algorithms.

Given that outsourced fabrication of IC is uncontrolled and untrusted, the tiny tampering that occurred at this stage is difficult to discover, hence HT detection after fabrication is indispensable. In addition, considering that the detection of implicit payload Trojans is more difficult than that of explicit payload Trojans, thus this article will aim at the study of the detection for implicit payload Trojans. Considering the merits and disadvantages of the methods above, in this work, a novel HT detection approach combining side-channel test and logic test will be applied. Besides, spectral feature analysis (SFA) combined PCA will be proposed for feature extraction. Fuzzy C-means clustering (FCC) combined spectral energy analysis (SEA) will be put forward to identify the golden category and Trojan category, as well as their clustering centers. Finally, MD combined Euclidean distance (ED) will be presented for HT recognition.

II. System Design and Detection Flow

A. System Design

The architecture diagram of an automatic test system applied for side-channel test combined logic test is illustrated in Fig. 1, which is composed of a NI high-speed digital acquisition card, a personal computer (PC), a power supply, a test circuit and a high-performance oscilloscope. ISE software is utilized to program the Xilinx FPGA chip through the JTAG interface. LabVIEW software is developed to stimulate the device under test (DUT), that is FPGA chip, and obtain the response with the RS232 cable or NI digital acquisition card. Meanwhile, it can set and control the oscilloscope, and acquire the side- channel information with the USB cable. There are 32 channels in the NI digital acquisition card and the maximum operating frequency is 200 MHz. Besides, the size of the memory of each channel is 64 Mbit, which is large enough. In this work, Xilinx Spartan-3E XC3S500E FPGA chip, fabricated in 90-nm CMOS process and packaged with PQ208, is applied to simulate ASIC scenario and carry out the Trojan or golden chip. Considering the process variation (about 6%), dozens of FPGA chips are used for the test. In the PCB, a chip socket is adopted to fix the DUT, which makes it replaceable.

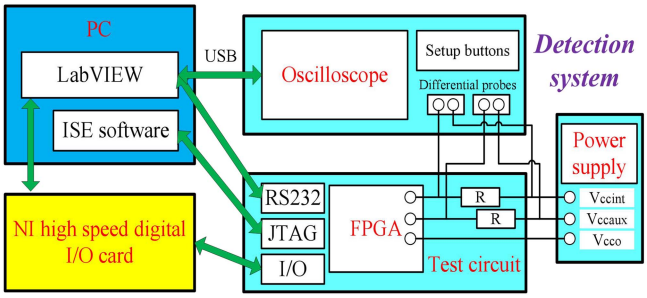


Fig. 1. Architecture diagram of an automatic test system applied for side-channel test combined logic test.

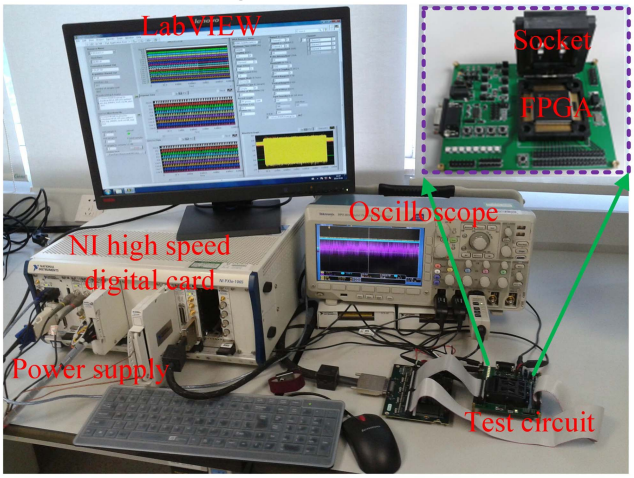


Fig. 2. Automatic test system is designed for HT detection.

Here, the high-performance oscilloscope is Tektronix DPO 3034. Its bandwidth is 350 MHz, maximum sampling frequency is 2.5 Gs/s, and memory of each channel is 5 Mbit. Differential probe is TDP0500, whose maximum sampling frequency is 500 MHz. They are utilized to measure the differential voltages of two precise resistances (0.5 Ω) located at the power cords of the auxiliary power supply (V_{ccaux}) and the internal core power supply (V_{ccint}). All the internal logic modules, such as block, CLBs, multipliers, and RAM, are powered by V_{ccint}, while the auxiliary modules, such as differential drivers, digital clock managers, dedicated configuration pins, and JTAG interface, are powered by V_{ccaux}. Therefore, these two power supplies may be influenced by the HT circuit, and ought to be monitored. Despite that side-channel information includes the maximum operating frequency, power dissipation, delay, electromagnetic interference, and so on, here only power dissipation is applied to perform the HT detection. Thus, the automatic test system can be designed and set up as Fig. 2.

B. Detection Flow

In terms of real-time HT detection, in this article, SFA combined PCA is used for feature extraction, while MD combined ED is adopted for HT recognition. However, before real-time HT detection, the clustering centers and principal eigenmatrices of the golden and Trojan categories, as well as the eigen-subspace, should be confirmed beforehand. The flowchart is shown in Fig. 3, and it mainly includes three steps.

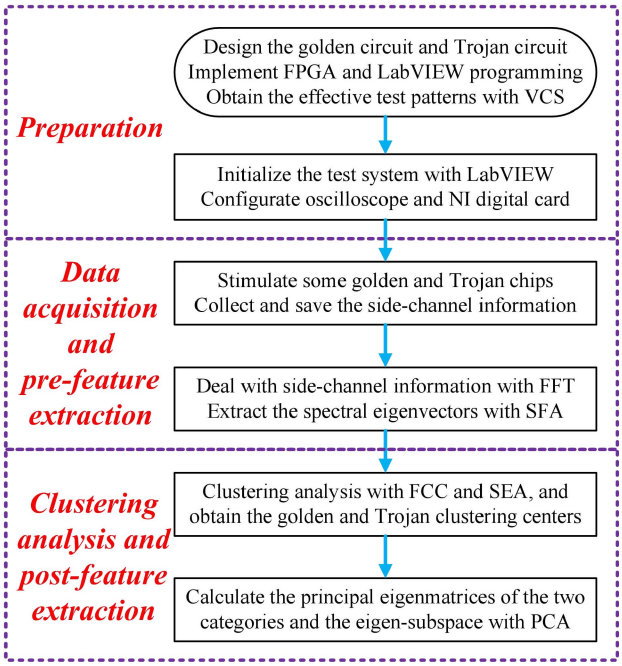


Fig. 3. Flowchart for extracting the clustering centers and principal eigenmatrices of the two categories, as well as the eigen-subspace.

1) *Preparation:* First, design a Trojan circuit and a golden circuit, then achieve HT implantation based on the activation mechanisms and typical HT taxonomy, then accomplish them in FPGA chips to imitate ASIC scenario. Automatic test software is programmed with LabVIEW, which is utilized to control oscilloscope and NI digital card, and collect the output response and side-channel information simultaneously. VCS software is applied to conduct the simulation and obtain the most effective test patterns for the logic test. Thus, the stimulus signals corresponding to the most effective test patterns will be used to stimulate the DUT, and the response outputs obtained by simulation will be save as the expected outputs compared with the real outputs of the following tests. After that, the test system, including oscilloscope and NI digital card, is initialized by the LabVIEW software.
2) *Data Acquisition and Prefeature Extraction:* The stimulus signals are applied to stimulate the golden or Trojan chip, and the side-channel information (i.e., power trace of V_{ccaux} or V_{ccint}) is collected and saved by the oscilloscope and LabVIEW. For each chip, every pattern is applied to tests for 1000 times, and their side-channel signals are averaged to further filter the measurement and temporal noise. Afterward, the side-channel information is processed with FFT, and the spectral eigenvector is extracted with SFA. For simulating the real ASIC scenario, dozens of DUTs are applied to perform the identical test. Here, 24 Trojan chips and 24 golden chips are adopted for the tests.
3) *Clustering Analysis and Post-Feature Extraction:* Based on all the tested spectral eigenvectors extracted with

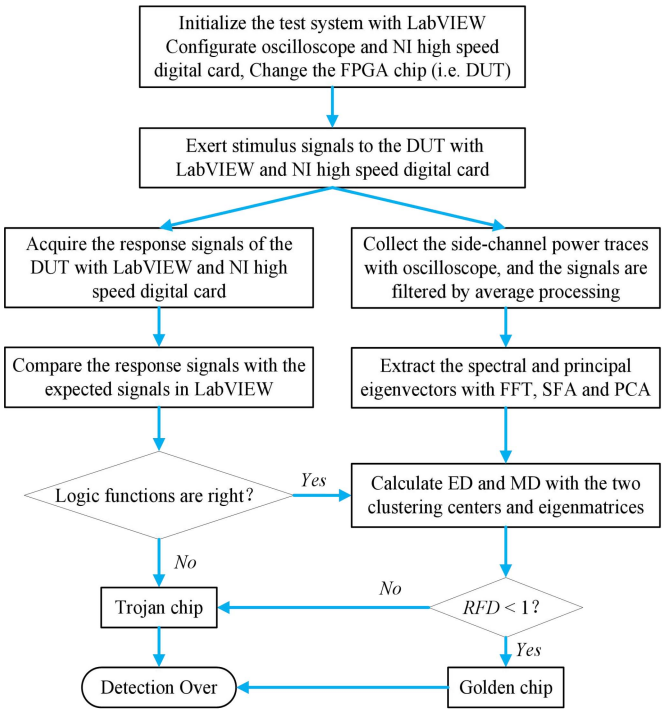


Fig. 4. Flowchart for the real-time HT detection.

SFA, the chips can be divided into two categories with FCC processing, and the Trojan category can be distinguished from the golden category based on SEA. Thus, their corresponding clustering centers and spectral eigenmatrices can be confirmed. In order to reduce the data dimension and further extract the feature, PCA is applied to deal with the spectral eigenmatrices. Finally, the principal eigenmatrices of the two categories, as well as the eigen-subspace, are obtained accordingly.

Once the clustering centers and principal eigenmatrices of the golden and Trojan categories, as well as the eigen-subspace, have been confirmed, the real-time HT detection can be performed, and the test flowchart is depicted in Fig. 4. It mainly includes five steps, as depicted as follows.

1) *Initialization:* First, the NI digital acquisition card, high-performance oscilloscope and test circuit are initialized by LabVIEW software. Then, change the FPGA chip (i.e., DUT) and launch a new test. In order to testify the proposed detection method, dozens of DUTs are applied to conduct the identical test. Here, 12 Trojan chips and 12 golden chips are applied for validation.
2) *Side-Channel Test and Logic Test:* The DUT is stimulated by LabVIEW software, then the response outputs are acquired by the NI high-speed digital card. Meanwhile, the side-channel power traces of V_{ccaux} and V_{ccint} are measured by the high- precision oscilloscope.
3) *Data Preprocessing:* If the response outputs are not equal to the expected outputs obtained by simulation, then some of the logic functions are incorrect, thus the DUT is judged to be infected by HT and the detection is over. Otherwise, the side- channel signals will be processed by the average filtering to eliminate the temporal and measurement noise, and then be stored for the following post-processing.
4) *Feature Extraction:* The side-channel signal is processed with FFT, and the spectral eigenvector is extracted with SFA. Afterward, the principal eigenvector is extracted with the spectral eigenvector and eigen-subspace.
5) *Pattern Recognition:* ED is calculated based on the clustering center and the spectral eigenvector, while MD is calculated based on the principal eigenvector and principal eigenmatrix. Thus, the fusion distance (FD) can be computed with ED and MD. For the relevant processing of the golden category, the corresponding FD is named as FDG, while that of the Trojan category is named as FDT. Then the ratio (RFD) of FDG to FDT can be computed. If RFD is less than 1, the DUT is judged as a golden chip; otherwise, it is judged as a Trojan chip. After pattern recognition, the detection is over.

III. Theoretical Analysis Algorithms

The algorithms for feature extraction and pattern recognition mentioned above are essential for HT detection, which will be described in detail in this section.

A. Feature Extraction With SFA

Since the differences of the power traces between the Trojan chips and the golden chips are so tiny that HT detection is very difficult. Hence, here frequency spectrum analysis is adopted. First, for each chip, every pattern is applied to tests for 1000 times, and their side-channel signals are averaged to filter the measurement noise. After the average filtering, the raw time-domain side-channel information can be obtained as d. In this work, the sampling frequency f_s of the high-performance oscilloscope is set as 2.5 GHz, while the bandwidth and the sampling number are set to 350 MHz and 1 M/channel, respectively. Thus, the resolution in the frequency domain is 2500 Hz, and the size of d is $1\times100\,0000$. After FFT processing, D is obtained, as shown in

$$D[k]=\sum_{i=1}^{N} d[i]\ e^{-j2\pi ki/N},\ 1\le k\le N \tag{1}$$

where N is the size of the input vector d, namely, 100 0000. Then, function *fftshift*() is applied to rearrange D by moving the zero frequency component to the center of the vector. Thus, the valid data vector Y after FFT processing can be derived as

$$\begin{aligned} D &= fftshift(D)\\ Y &= D(500001{:}1000000). \end{aligned} \tag{2}$$

Hence, the size of Y is $1\times500\,000$, and the frequency range is from 0 to 1.25 GHz. Considering that some combinational and sequential HT circuits are indirectly or directly connected to the clock tree and the power tree, hence the frequency response characteristics of the power traces are inevitably affected by the HT circuits. In particular, the spectral feature located at the multiplier or divider of the clock frequency may be different between the golden chip and Trojan chip. Thereby, based on this principle, further feature extraction can be carried out, as shown in

$$\mathrm{EV}[k]=\begin{cases} Y\left[\mathrm{round}\left(\frac{500000\times10\times10^6}{1.25\times10^9\times(66-k)}\right)\right], & 1\le k\le 65 \\ Y\left[\mathrm{round}\left(\frac{500000\times10\times10^6\times(k-64)}{1.25\times10^9}\right)\right], & 66\le k\le 129 \end{cases} \tag{3}$$

where $\boldsymbol{EV}$ is the spectral eigenvector extracted with SFA, whose size is 1×129. It means that the data compression ratio is about 129 ppm. *round*() is the rounding function. In this article, the clock frequency of the chip is set to 10 MHz.

B. Pattern Recognition With FCC and SEA

The FCC algorithm was proposed by Jim Bezdek in 1981, which soon became one of the most important unsupervised learning methods in pattern recognition. It can extract the intrinsic characteristic from a great deal of data, and reveal the internal rules, so it is widely used in clustering analysis and system modeling. It assumes that every sample is fuzzy affiliated with a certain center, which means every sample may be affiliated with multiple centers. It randomly selects some clustering centers at the beginning, and assigns the fuzzy memberships between the samples and clustering centers. The fuzzy memberships and clustering centers are adjusted by the iterative operation until the optimization target is achieved. Finally, the optimal fuzzy memberships and clustering centers, rather than a fuzzy inference system, are obtained.

Assume that there are c categories and n samples. ω_j and θ_j (j=1, 2,. . . , c) stand for the jth category and parameter of membership function, respectively. $\boldsymbol{\theta}=(\theta_1,\theta_2,\ldots,\theta_c)^T$ is a parameter vector. X_k (k=1, 2,. . . , n) represents the kth sample. Assume that $P(\omega_j)$ is the prior probability, and $p(X_k|\omega_j,\ \theta_j)$ is the conditional probability density, thus the generating probability $p(X_k|\theta)$ can be expressed as

$$p(X_k|\theta)=\sum_{j=1}^{c}p\left(X_k|\omega_j,\theta_j\right)P\left(\omega_j\right). \tag{4}$$

Introducing the posterior probability, the probability that X_k is affiliated with ω_j is depicted as

$$P\left(\omega_j|X_k,\theta\right)=\frac{p\left(X_k|\omega_j,\theta_j\right)P\left(\omega_j\right)}{p(X_k|\theta)}. \tag{5}$$

According to the Bayes theorem, the maximum likelihood estimation of posterior probability can be derived as

$$\begin{aligned}\widehat{P}\left(\omega_j|X_k,\widehat{\theta}\right)&=\frac{P\left(X_k|\omega_j,\widehat{\theta_i}\right)\widehat{P}\left(\omega_j\right)}{\sum_{j=1}^{c}P\left(X_k|\omega_j,\widehat{\theta_j}\right)\widehat{P}\left(\omega_j\right)}\\&=\frac{\left|\widehat{\Sigma}_j\right|^{-1/2}\exp\left[-\frac{1}{2}\left(X_k-\widehat{\mu}_j\right)^T\widehat{\Sigma}_j^{-1}\left(X_k-\widehat{\mu}_j\right)\right]\widehat{P}\left(\omega_j\right)}{\sum_{j=1}^{c}\left|\widehat{\Sigma}_j\right|^{-1/2}\exp\left[-\frac{1}{2}\left(X_k-\widehat{\mu}_j\right)^T\widehat{\Sigma}_j^{-1}\left(X_k-\widehat{\mu}_j\right)\right]\widehat{P}\left(\omega_j\right)}\end{aligned} \tag{6}$$

where the maximum likelihood estimations of covariance matrix, clustering center, and prior probability are shown as

$$\begin{aligned}\widehat{\Sigma}_j&=\frac{\sum_{k=1}^{n}\widehat{P}\left(\omega_j|X_k,\widehat{\theta}\right)\left(X_k-\widehat{\mu}_j\right)\left(X_k-\widehat{\mu}_j\right)^T}{\sum_{k=1}^{n}\widehat{P}\left(\omega_j|X_k,\widehat{\theta}\right)}\\\widehat{\mu}_j&=\frac{\sum_{k=1}^{n}\widehat{P}\left(\omega_j|X_k,\widehat{\theta}\right)X_k}{\sum_{k=1}^{n}\widehat{P}\left(\omega_j|X_k,\widehat{\theta}\right)}\\\widehat{P}\left(\omega_j\right)&=\frac{1}{n}\sum_{k=1}^{n}\widehat{P}\left(\omega_j|X_k,\widehat{\theta}\right).\end{aligned} \tag{7}$$

Here, (6) figures out that $\widehat{P}(\omega_j|X_k,\widehat{\theta})$ increases as $(X_k-\widehat{\mu}_j)^T\widehat{\Sigma}_j^{-1}(X_k-\widehat{\mu}_j)$ decreases. In order to simplify the operation and accelerate the convergence rate, the square of ED $\|X_k-\mu_j\|^2$can be applied to replace the square of MD $(X_k-\widehat{\mu}_j)^T\widehat{\Sigma}_j^{-1}(X_k-\widehat{\mu}_j)$. Thus, the optimization goal can be constructed as

$$J_{\mathrm{fuz}}=\sum_{j=1}^{c}\sum_{k=1}^{n}\left[\widehat{P}\left(\omega_j|X_k,\widehat{\theta}\right)\right]^b\left\|X_k-\mu_j\right\|^2 \tag{8}$$

where $\boldsymbol{\mu}_j$ is the jth clustering center and $\widehat{P}(\omega_j|X_k,\widehat{\theta})$ is the fuzzy membership grade. b is a control parameter, which is often set to 0.5. Every sample is affiliated with only one cluster when b is set to zero. However, every sample can be affiliated with multiple clusters when b is larger than zero. Due to normalization, the sum of $\widehat{P}(\omega_j|X_k,\widehat{\theta})$ is 1, as depicted in

$$\sum_{j=1}^{c}\widehat{P}\left(\omega_j|X_k,\widehat{\theta}\right)=\sum_{j=1}^{c}\widehat{P}\left(\omega_j|X_k\right)=1,\ \ k=1,2,\ldots,n. \tag{9}$$

Assume that $(\partial J_{\mathrm{fuz}}/\partial\mu_j)=0$ and $(\partial J_{\mathrm{fuz}}/\partial\widehat{P}_k)=0$, thus optimization goal can be achieved, and the optimal solutions of clustering center and fuzzy membership grade can be deducted as

$$\begin{aligned}\mu_j&=\frac{\sum_{k=1}^{n}\left[\widehat{P}\left(\omega_j|X_k\right)\right]^bX_k}{\sum_{k=1}^{n}\left[\widehat{P}\left(\omega_j|X_k\right)\right]^b}\\\widehat{P}\left(\omega_j|X_k\right)&=\frac{\left(1/\left\|X_k-\mu_j\right\|^2\right)^{1/(b-1)}}{\sum_{j=1}^{c}\left(1/\left\|X_k-\mu_j\right\|^2\right)^{1/(b-1)}}.\end{aligned} \tag{10}$$

After defuzzification, the fuzzy membership grade (posterior probability) can be obtained as

$$P\left(\omega_j|X_k\right)=\begin{cases}1,\ \left\|X_k-\mu_j\right\|<\left\|X_k-\mu_{j'}\right\|,j'\ne j\\0,\ \text{other.}\end{cases} \tag{11}$$

Therefore, after FCC, the clustering centers μ_j (j = 1, 2, . . ., c) are confirmed, thus ED can be applied to conduct the cluster judgment. In this article, MATLAB function fcm() can be adopted for FCC processing, as shown in

$$\left[\mu,U,\mathrm{objF}\right]=\mathrm{fcm}\left(X,c,\left[\mathrm{opt1},\mathrm{opt2},\mathrm{opt3},\mathrm{opt4}\right]\right) \tag{12}$$

where the number of clusters c is set to 2. The exponent *opt1* for the partition matrix U controls the amount of fuzzy overlap between clusters, and it is set to 2. The larger it is, the greater the degree of overlap is. The maximum number of iterations *opt2* is set to 1000. The minimum improvement *opt3* in objective function between two consecutive iterations is set to 1e-6. *opt4* is utilized to control whether to display the objective function value after every iteration, and it is set to true. μ is the clustering center matrix composed of c clustering centers. The number of columns equals that of $\boldsymbol{X}$. Element of fuzzy partition matrix $\boldsymbol{U}[i,j]$ indicates the degree of membership of the jth data point in the ith cluster. *objF* is the objective function values for every iteration.

$\boldsymbol{X}$ is the data set to be clustered, and it is a matrix composed of a series of $\boldsymbol{EV}$. Here, there are n samples. As mentioned above, 24 Trojan chips and 24 golden chips are applied for

the tests and FCC analysis. That is, n is set to 48. Given that the spectral eigenvector $\boldsymbol{EV}$ of each chip can be extracted with SFA, a total spectral eigenmatrix $\boldsymbol{X}$ of 48 chips can be obtained, whose size is 48×129. After FCC processing, $\boldsymbol{X}$ can be divided into two spectral eigenmatrices, namely, $\boldsymbol{EX1}$ and $\boldsymbol{EX2}$, whose sizes are both 24×129. Although two clusters are obtained with FCC processing, it is still uncertain which one is the Trojan cluster. In order to distinguish them, SEA should be adopted.

Considering that HT circuits also need a power supply, its existence will inevitably result in the increase of the spectral amplitude of the power trace. Thereby, the integral of the spectral amplitude (i.e., total energy) of the Trojan chip should be greater than that of the golden chip. Therefore, the cluster with larger energy will be judged as the Trojan cluster, while the other will be judged as the golden cluster, as shown in (13). Thus, the cluster type can be confirmed with SEA. Finally, the Trojan clustering center and spectral eigenmatrix are recorded and marked as $\boldsymbol{\mu T}$ and $\boldsymbol{EXT}$, whose sizes are 1 × 129 and 24 × 129, respectively. Similarly, the golden clustering center and spectral eigenmatrix are recorded and marked as $\boldsymbol{\mu G}$ and $\boldsymbol{EXG}$, whose sizes are also 1×129 and 24×129, respectively

$$\text{Type(EX1)} = \begin{cases} \text{Trojan}, & \sum \text{EX1} \geq \sum \text{EX2} \\ \text{Golden}, & \sum \text{EX1} < \sum \text{EX2}. \end{cases} \tag{13}$$

C. Feature Extraction With PCA

PCA is an effective and prevalent statistical analysis method to find a matrix to explain the variance of the data. With this matrix, the data can be mapped to a low-dimensional space from a high-dimensional space. In the process of dimension reduction, the correlation of the data is removed, and the main features of the data are extracted. Hence, PCA is adopted to further extract the features of spectral eigenvector and spectral eigenmatrix. MATLAB function *pca*() is adopted for PCA in this article, as shown in

$$[\text{coeff, sc, lat, tsq, epl}] = \text{pca}(X, \text{“NumComponents”, kn}) \tag{14}$$

where *NumComponents* is the number of components requested, which is set as kn ($0 < kn \leq p$). p is the number of variables in $\boldsymbol{X}$. $\boldsymbol{epl}$ outputs the percentage of the total variance explained by each principal component. $\boldsymbol{tsq}$ is Hotelling's T-Squared Statistic. $\boldsymbol{lat}$ is a vector composed of the principal component variances, that is the eigenvalues of the covariance matrix of $\boldsymbol{X}$. The elements of $\boldsymbol{lat}$ are arranged in descending order. In general, the sum of the percentages of the total variances explained by the first kn principal components is greater than 85%. Hence, they are the most important components selected as the representative. Simultaneously, the first kn columns of the principal component coefficients and scores are returned, which are named as $\boldsymbol{coeff}$ and $\boldsymbol{sc}$, respectively. $\boldsymbol{coeff}$ is the eigen-subspace composed of the eigenvectors corresponding to the first kn principal eigenvalues. Besides, it is a matrix with the size of p*kn, and its each column contains coefficients for one principal component. $\boldsymbol{sc}$ is a matrix with the size of n*kn, which is equal to $\boldsymbol{X}$ times $\boldsymbol{coeff}$. That is, $\boldsymbol{sc}$ is the orthogonal projection of $\boldsymbol{X}$ in the eigen-subspace $\boldsymbol{coeff}$. This projection represents the main feature of $\boldsymbol{X}$, thereby $\boldsymbol{sc}$ is the principal eigenmatrix. In this article, p is set to 129, while kn is often set to 10. The comparison between $\boldsymbol{sc}$ and $\boldsymbol{X}$ figures out that the data compression ratio is 7.75%.

Given that $\boldsymbol{sc}$ is the principal eigenmatrix with the size of 48×10, and the golden cluster has been distinguished from the Trojan cluster with FCC processing, so $\boldsymbol{sc}$ can be divided into two sub matrices, the golden one is marked as $\boldsymbol{scg}$, while the Trojan one is marked as $\boldsymbol{sct}$, as shown in (15). The sizes of $\boldsymbol{scg}$ and $\boldsymbol{sct}$ are both 24×10

$$\begin{bmatrix} scg \\ sct \end{bmatrix} = X \times \text{coeff} = \begin{bmatrix} EXG \\ EXT \end{bmatrix} \times \text{coeff}. \tag{15}$$

It is clear that before data dimension reduction, the number of rows in $\boldsymbol{X}$ is greater than the number of columns, but after data dimension reduction, the number of rows in $\boldsymbol{scg}$ or $\boldsymbol{sct}$ is greater than the number of columns, which is conducive to the subsequent calculation of MD. Thus, after PCA processing, the principal eigenmatrices of the golden and Trojan categories, as well as the eigen-subspace, can be confirmed.

D. Pattern Recognition With ED and MD

In terms of a certain DUT, the spectral eigenvector $\boldsymbol{EV}$ is extracted with FFT and SFA. Then the principal eigenvector $\boldsymbol{PEV}$ can be extracted with $\boldsymbol{EV}$ and $\boldsymbol{coeff}$, as shown in

$$\text{PEV} = \text{EV} \times \text{coeff}. \tag{16}$$

That is, $\boldsymbol{PEV}$ is the principal component score, which is the orthogonal projection of $\boldsymbol{EV}$ in the eigen-subspace $\boldsymbol{coeff}$. Here, the size of $\boldsymbol{PEV}$ is 1×10. Here, for the real-time HT detection, distance analysis can be adopted. Based on $\boldsymbol{EV}$, $\boldsymbol{\mu T}$ and $\boldsymbol{\mu G}$, the ED EDG between the spectral eigenvector and golden clustering center, as well as the ED EDT between the spectral eigenvector and Trojan clustering center, can be calculated, as shown in

$$\begin{aligned} \text{EDG} &= \sqrt{\sum_{i=1}^{129} (\text{EV}[i] - \mu G[i])^2} \\ \text{EDT} &= \sqrt{\sum_{i=1}^{129} (\text{EV}[i] - \mu T[i])^2}. \end{aligned} \tag{17}$$

In general, if EDT is larger than EDG, it will be judged as a golden chip, else it is a Trojan chip. However, this judgment may fail to work when the noise is large. Because there is no further feature extraction dealt with PCA, some minor elements in the spectral eigenvector may affect the distance result. In addition, ED treats all the elements in the vector equally and cannot eliminate the interference of the correlation between the elements. Hence, MD should be combined to advance the HT detection accuracy.

The advantages of MD are as follows: 1) it is not affected by dimensions, and MD is independent of the measurement units of the raw data; 2) MDs calculated with the normalized data and the centralized data are the same; and 3) MD can eliminate the interference of the correlation between the elements. Thereby, MD is adopted to compute the distance

between the principal eigenvector and the golden or Trojan principal eigenmatrix, as shown in

$$\begin{aligned} \mathrm{MDG} &= \sqrt{\mathrm{mahal}(\mathrm{PEV}, \mathrm{scg})} \\ \mathrm{MDT} &= \sqrt{\mathrm{mahal}(\mathrm{PEV}, \mathrm{sct})} \end{aligned} \tag{18}$$

where *mahal*() is a MATLAB function used to calculate the squared MD of ***PEV*** to the reference samples in ***scg*** or ***sct***. It should be noted that ***PEV***, ***scg*** and ***sct*** must have the same number of columns, but can have different numbers of rows. ***scg*** and ***sct*** must have more rows than columns, otherwise, the inverse of the covariance matrix does not exist.

For improving the detection accuracy, here, fusion distance analysis is presented, and the formulas are depicted in

$$\begin{aligned} \mathrm{FDG} &= a_1 \times \mathrm{EDG} + a_2 \times \mathrm{MDG} \\ \mathrm{FDT} &= a_1 \times \mathrm{EDT} + a_2 \times \mathrm{MDT} \\ \mathrm{RFD} &= \mathrm{FDG}/\mathrm{FDT} \end{aligned} \tag{19}$$

where FDG and FDT are the fusion distances from the new DUT to the golden cluster and Trojan cluster, respectively. a_1 and a_2 are the weights of ED and MD, respectively. Thus, the ratio (RFD) of FDG to FDT will be utilized to judge whether it is a Trojan chip, and the rule is depicted as

$$\mathrm{Type}(\mathrm{EV}) = \begin{cases} \text{golden}, & \mathrm{RFD} < 1 \\ \text{Trojan}, & \mathrm{RFD} \geq 1. \end{cases} \tag{20}$$

That is, if RFD is less than 1, the DUT is judged as a golden chip; Otherwise, it is judged as a Trojan chip. After pattern recognition with distance analysis, the detection is over.

To sum up, the novelty of this HT detection method include the following.

1) SFA combined PCA is applied for feature extraction. At the physical level, SFA can amplify the difference of the side- channel information between the golden chip and Trojan chip, especially the difference of the spectral feature located at the multiplier or divider of the clock frequency. In addition, at the mathematical level, PCA can further perform feature extraction and dimension reduction to facilitate the implementation of MD analysis, because MD calculation requires that the matrices should have more rows than columns.
2) FCC combined SEA is utilized to distinguish the golden category from the Trojan category. The FCC unsupervised learning method can be used for fast clustering, rather than category judgment. However, SEA can be adopted for category judgment since the integral of the spectral amplitude of the Trojan chip should be greater than that of the golden chip.
3) FD (i.e., MD combined ED) is used for the real-time HT recognition. In the final step, distance analysis is effective to judge whether the DUT is a Trojan chip or not.

However, ED treats all the elements in the vector equally and cannot eliminate the interference of the correlation between the elements. Hence, MD should be combined together to advance the HT detection accuracy. Obviously, the combination of these algorithms is not simple, but deeply complementary. Their combination is established on the basis of analyzing the defects of a single algorithm, making the proposed detection method maximize the advantages of each algorithm.

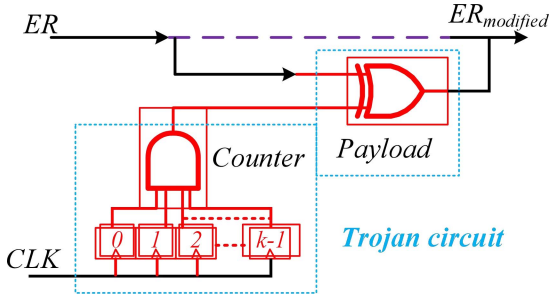


Fig. 5. Schematic of a synchronous counter Trojan circuit.

IV. Experimental Results

A. Experimental Preparation

AES encryption is one of the most important data encryption algorithms, which is widely used in most of IOT products [52]. Therefore, in this work, a 128-bit AES cipher circuit is applied as a golden circuit used for hardware security analysis. For RSA or other encryption circuit, the analysis is similar. Besides, considering that the detection of implicit payload Trojans is more difficult than that of explicit payload Trojans, thus this article will aim at the study of the detection for implicit payload Trojans. Running exhaustive input patterns through traditional logic testing can hardly detect the implicit payload Trojans, especially for the always-on HTs. Hence, in order to evaluate the effectiveness of the proposed detection method, here a tiny always-on sequential circuit, instead of a combinational circuit, can be used as the Trojan circuit. Thus, after analysis the Trojan benchmarks reported in [53] and [54], a simple 2-bit counter is chosen as a Trojan circuit, as depicted in Fig. 5. This HT does no harm to the golden circuit, it only leads to additional power consumption and increase of some path delays since it may be used to disclose secret information to hackers, such as the secret keys of the golden chip.

The Trojan circuit and golden circuit are realized in the FPGA chip, and their equivalent area ratio is about 0.1%. Considering that outsourced fabrication is uncontrollable, a hacker may tamper with the gate-level netlist of the design, hence in this work, the HT will be inserted in the gate-level netlist, rather than RTL code, to imitate the real case. In order to simulate the real ASIC scenario, a total of 72 FPGA chips are used for the tests and validation. Among them, 36 FPGA chips are utilized to implement the golden chips, while the other 36 FPGA chips are applied to implement the Trojan chips.

Before experimental tests, the most effective test patterns for the logic test should be confirmed. Here, pseudorandom test pattern generation technique is adopted to obtain the optimal and effective test patterns. In order to amplify the influence induced by the HT, the node toggle times of the golden chip should be as small as possible. Here, 10 million test patterns are pseudo-random generated for the simulations

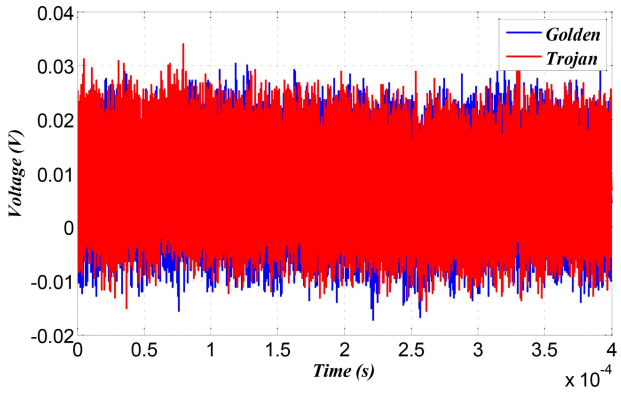


Fig. 6. Time-domain side-channel power traces of the internal core power supply in the golden chip and Trojan chip.

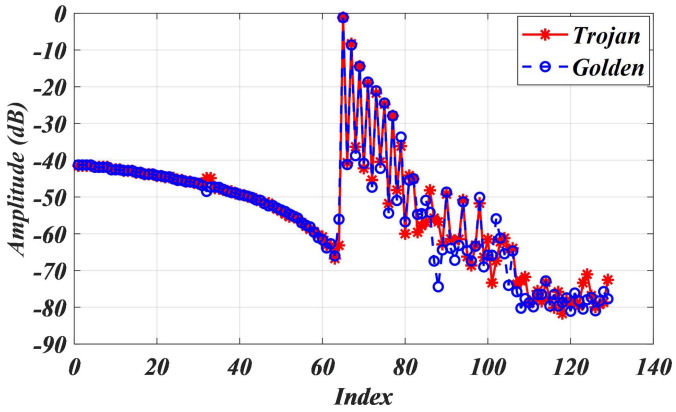


Fig. 7. Spectral eigenvectors of the Trojan chip and golden chip after feature extraction with FFT processing and SFA.

with VCS software. After statistical analysis, the toggle times can be obtained and sorted from small to large, the patterns corresponding to the top-10 toggle times are the optimization test patterns which will be applied to the logic test and side-channel test. Then, the corresponding response signals are saved as the expected signals for functional comparison. The main clock frequency of the tested chip is set to 10 MHz.

B. Experimental Results of Feature Extraction

Under the stimulus by the effective test patterns above, the time-domain side-channel power traces of the internal core power supply in the Trojan chip and golden chip are acquired by the proposed test system, as shown in Fig. 6. It figures out that the two signals are so similar that it is difficult to distinguish them. In order to advance the detection accuracy and reduce the amount of calculation, feature extraction is very important. The spectral eigenvectors (*EVs*) of the Trojan chip and golden chip after feature extraction with FFT processing and SFA are shown in Fig. 7.

In Fig. 7, the abscissa is the index, while the ordinate is the spectrum amplitude. It is clear that the two curves are a little different, and the integral of the spectral amplitude (i.e., total energy) of the Trojan chip is indeed greater than that of the golden chip. That is, the difference gradually appears after feature extraction with SFA. In order to further amplify the difference, PCA is necessary.

C. Experimental Results of Pattern Recognition

In the unsupervised learning stage, 24 Trojan chips and 24 golden chips are applied for the tests and FCC analysis. After FCC processing and SEA, the clustering centers $\boldsymbol{\mu T}$ and $\boldsymbol{\mu G}$ can be obtained. Subsequently, PCA is adopted to obtain the principal eigenmatrices ***scg*** and ***sct***, as well as the eigensubspace ***coeff***. To visualize the effect of FCC, the first three projections of ***scg*** and ***sct*** are depicted in a 3-D diagram, as shown in Fig. 8. It is obvious that the Trojan chips have been distinguished from the golden chip effectively. In addition, the energy corresponding to the three projections of the Trojan chip is also greater than that of the golden chip. In fact, in

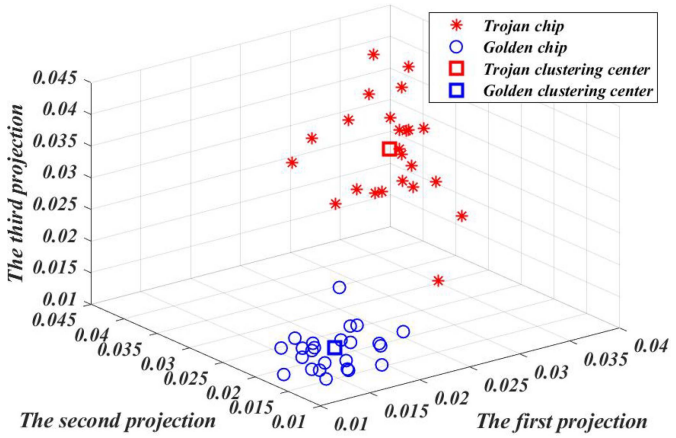


Fig. 8. First three projections of ***scg*** and ***sct*** are depicted in a 3-D diagram.

order to further recognize the golden chips reliably, at least one chip from the golden category should be used for reverse engineering analysis.

In the validation stage, the rest 12 Trojan chips and 12 golden chips are utilized for the tests and real-time pattern recognition. Since the principal eigenvector ***PEV*** is extracted with ***EV*** and ***coeff***, FDG and FDT can be obtained. Based on (20), if RFD is less than 1, the DUT is judged as a golden chip; Otherwise, it is judged as a Trojan chip. The relationships between FDG and FDT are illustrated in Fig. 9. It demonstrates that all the golden and Trojan chips can be recognized accurately, and the chips located above the boundary (i.e., $y{=}x$) are judged as golden chips, otherwise, they are Trojan chips. The true-negative rate and false-positive rate are both zero, which indicates that the proposed detection method is effective.

V. Conclusion

The contributions of this work contain: 1) SFA combined PCA is applied for feature extraction; 2) FCC combined SEA is utilized to distinguish the golden category from the Trojan category; and 3) FD (i.e., MD combined ED) is used for the real-time HT recognition. Besides, a HT automatic test system

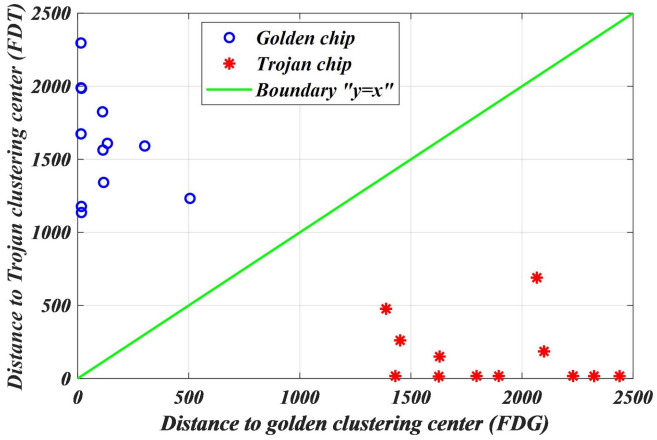


Fig. 9. Golden and Trojan chips can be recognized accurately with fusion distance analysis.

used for side-channel test combined logic test is constructed with a high-performance oscilloscope, FPGA chips, a NI digital acquisition card and LabVIEW software. The test flowchart and data processing method are depicted in detail. Considering that the detection of implicit payload Trojans is more difficult than that of explicit payload Trojans, this article aim at the study of the detection for implicit payload Trojans. Thus, a 128-bit AES cipher circuit and a 2-bit counter are applied as a golden circuit and a Trojan circuit, respectively. Experimental results demonstrate that the detection accuracy is 100%, and the proposed detection method can achieve 0.1% HT detection sensitivity, which verifies that the detection method is effective and feasible. The proposed method is easy to be realized, and it can be used for the security detection of some IOT chips.

From the references cited in this article, the reported detection sensitivities range from 0.5% to 0.01%. Generally speaking, the detection methods and sensitivities of different-type HTs are different, and it is difficult to propose a widely applicable and high-precision detection method. Given that HTs are unknown and uncertain, the reliable detection method in the future will be a combination of multiple detection methods. Considering the variety of HT circuits and their implantation approaches, the applicability and effect of the proposed method still need to be further verified. For example, for the other HT or golden circuits, more tests should be conducted to evaluate whether the proposed method is still feasible and superior. These will be our future work.

REFERENCES

[1] M. H. P. Rizi and S. A. H. Seno, "A systematic review of technologies and solutions to improve security and privacy protection of citizens in the smart city," *Internet Things*, vol. 20, Nov. 2022, Art. no. 100584.

[2] S. Bhunia and M. Tehranipoor, "Chapter 1—Introduction to hardware security," in *Hardware Security*. Burlington, MA, USA: Morgan Kaufmann, 2019, pp. 1–20.

[3] C. Higgins, L. McDonald, M. I. U. Haq, and S. Hakak, "IoT hardware-based security: A generalized review of threats and countermeasures," in *Security and Privacy in the Internet of Things: Architectures, Techniques, and Applications*. Hoboken, NJ, USA: Wiley-IEEE Press, 2022, pp. 261–296.

[4] Y. Fazea, F. Mohammed, N. H. Al-Kumaim, and M. S. Sajat, "Automatic hardware Trojan generation platforms in integrated circuits: A critical review," in *Proc. ICTSA*, 2021, pp. 1–7.

[5] Y. Jin and Y. Makris, "Hardware Trojan detection using path delay fingerprint," in *Proc. IEEE HOST*, 2008, pp. 51–57.

[6] S. Dupuis, M. L. Flottes, G. Di Natale, and B. Rouzeyre, "Protection against hardware Trojans with logic testing: Proposed solutions and challenges ahead," *IEEE Design Test*, vol. 35, no. 2, pp. 73–90, Apr. 2018.

[7] F. Khormizi, A. Shabani, and B. Alizadeh, "Hardware patching methodology for neutralizing timing hardware Trojans using vulnerability analysis and time borrowing scheme," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 69, no. 6, pp. 2937–2941, Jun. 2022.

[8] M. Rithesh, G. Harish and S. Yellampalli, "Detection and analysis of hardware Trojan using dummy scan flip-flop," in *Proc. ICSTM*, 2015, pp. 439–442.

[9] Y. Huang, S. Bhunia, and P. Mishra, "Scalable test generation for Trojan detection using side channel analysis," *IEEE Trans. Inf. Forensics Security*, vol. 13, pp. 2746–2760, 2018.

[10] T. Xu, C. Wang, S. Zhao, Z. Zhou, M. Luo, and X. Wang, "A novel ATPG method to increase activation probability of hardware Trojan," in *Proc. IEEE PACRIM*, 2019, pp. 1–5.

[11] M. Flottes, S. Dupuis, P. Ba, and B. Rouzeyre, "On the limitations of logic testing for detecting hardware Trojans horses," in *Proc. DTIS*, 2015, pp. 1–5.

[12] M. Fyrbiak et al., "Hardware reverse engineering: Overview and open challenges," in *Proc. IEEE IVSW*, 2017, pp. 88–94.

[13] C. Bao, D. Forte, and A. Srivastava, "On reverse engineering-based hardware Trojan detection," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 35, no. 1, pp. 49–57, Jan. 2016.

[14] L. N. Nguyen, B. B. Yilmaz, M. Prvulovic, and A. Zajic, "A novel golden-chip-free clustering technique using backscattering side channel for hardware Trojan detection," in *Proc. IEEE HOST*, 2020, pp. 1–12.

[15] S. Rajendran and M. L. Regeena, "A novel algorithm for hardware Trojan detection through reverse engineering," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 41, no. 4, pp. 1154–1166, Apr. 2022.

[16] A. A. Nasr and M. Z. Abdulmageed, "An efficient reverse engineering hardware Trojan detector using histogram of oriented gradient," *J. Electron. Test.*, vol. 33, no. 1, pp. 93–105, 2017.

[17] T. Zhang, J. Wang, and Z. Chen, "A reverse engineering-based framework assisting hardware Trojan detection for encrypted IPs," in *Proc. IMCCC*, 2018, pp. 1649–1652.

[18] M. R. A. Kouhanjani and A. H. Jahangir, "Improving hardware Trojan detection using scan chain based ring oscillators," *Microprocess. Microsyst.*, Vol. 63, Nov. 2018, pp. 55–65.

[19] Y. Xiao and S. Feng, "Detecting hardware Trojans by monitoring power supply noise based on ring oscillator network in FPGA," in *Proc. AEMCSE*, 2020, pp. 410–413.

[20] N. Yoshimizu, "Hardware Trojan detection by symmetry breaking in path delays," in *Proc. IEEE HOST*, 2014, pp. 107–111.

[21] M. M. Hasan, S. Baul, M. B. Hashem, and H. Rahman, "Hardware Trojan detection using slope of path delay trend: Combination of clock and DC sweep," in *Proc. ICECE*, 2020, pp. 145–148.

[22] X. Wang, H. Salmani, M. Tehranipoor, and J. Plusquellic, "Hardware Trojan detection and isolation using current integration and localized current analysis," in *Proc. IEEE Int. Symp. Defect Fault Tolerance VLSI Syst.*, 2008, pp. 87–95.

[23] J. Aarestad, D. Acharyya, R. Rad, and J. Plusquellic, "Detecting Trojans through leakage current analysis using multiple supply pad IDDQs," *IEEE Trans. Inf. Forensics Security*, vol. 5, pp. 893–904, 2010.

[24] S. Narasimhan et al., "Hardware Trojan detection by multiple-parameter side-channel analysis," *IEEE Trans. Comput.*, vol. 62, no. 11, pp. 2183–2195, Nov. 2013.

[25] S. Narasimhan et al., "Multiple-parameter side-channel analysis: A non-invasive hardware Trojan detection approach," in *Proc. IEEE HOST*, 2010, pp. 13–18.

[26] R. Gayatri, Y. Gayatri, C. Mitra, S. Mekala, and M. Priyatharishini, "System level hardware Trojan detection using side-channel power analysis and machine learning," in *Proc. ICCES*, 2020, pp. 650–654.

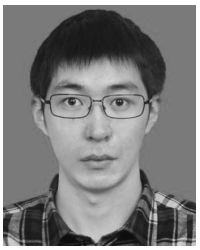
[27] Z. Chen, S. Guo, J. Wang, Y. Li, and Z. Lu, "Toward FPGA security in IoT: A new detection technique for hardware Trojans," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 7061–7068, Aug. 2019.

[28] S. Sun, H. Zhang, X. Cui, L. Dong, and X. Fang, "Electromagnetic side-channel hardware Trojan detection based on transfer learning," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 69, no. 3, pp. 1742–1746, Mar. 2022.

[29] F. Stellari, P. Song, and H. A. Ainspan, "Functional block extraction for hardware security detection using time-integrated and time-resolved emission measurements," in *Proc. IEEE VTS*, 2014, pp. 1–6.
[30] A. Stern, J. Vosatka, S. Tajik, F. Farahmandi, and M. Tehranipoor, "Trust assessment for electronic components using laser and emission-based microscopy," in *Proc. IEEE RAPID*, 2020, pp. 1–2.
[31] Y. Wen and W. Yu, "Combining thermal maps with inception neural networks for hardware Trojan detection," *IEEE Embedded Syst. Lett.*, vol. 13, no. 2, pp. 45–48, Jun. 2021.
[32] S. Guo, J. Wang, Z. Chen, Y. Li, and Z. Lu, "Securing IoT space via hardware Trojan detection," *IEEE Internet Things J.*, vol. 7, no. 11, pp. 11115–11122, Nov. 2020.
[33] D. Forte, C. Bao, and A. Srivastava, "Temperature tracking: An innovative run-time approach for hardware Trojan detection," in *Proc. IEEE/ACM ICCAD*, 2013, pp. 532–539.
[34] V. R. Surabhi, P. Krishnamurthy, H. Amrouch, J. Henkel, R. Karri, and F. Khorrami, "Trojan detection in embedded systems with FinFET technology," *IEEE Trans. Comput.*, vol. 71, no. 11, pp. 3061–3071, Nov. 2022.
[35] E. Jedari and R. Rashidzadeh, "A hardware Trojan detection method for IoT sensors using side-channel activity magnifier," *IEEE Internet Things J.*, vol. 9, no. 6, pp. 4507–4517, Mar. 2022.
[36] C. He, B. Hou, L. Wang, Y. En, and S. Xie, "A failure physics model for hardware Trojan detection based on frequency spectrum analysis," in *Proc. IEEE IRPS*, 2015, pp. PR.1.1–PR.1.4.
[37] L. W. Wang and H. W. Luo, "A power analysis based approach to detect Trojan circuits," in *Proc. QR2MSE*, 2011, pp. 380–384.
[38] R. Shende and D. D. Ambawade, "A side channel based power analysis technique for hardware Trojan detection using statistical learning approach," in *Proc. WOCN*, 2016, pp. 1–4.
[39] L. Zhang, Y. Dong, J. Wang, C. Xiao, and D. Ding, "A hardware Trojan detection method based on the electromagnetic leakage," *China Commun.*, vol. 16, no. 12, pp. 100–110, Dec. 2019.
[40] N. Tang, W. Zhou, L. Li, J. Yang, R. Li, and Y. He, "Hardware Trojan detection method based on the frequency domain characteristics of power consumption," in *Proc. ISCID*, 2020, pp. 410–413.
[41] W. Jianxin, W. Boren, Q. Ming, Z. Lei, and X. Chaoen, "Hardware Trojan detection based on the distance discrimination," in *Proc. IEEE ICCCI*, 2016, pp. 404–407.
[42] C. Dong, Y. Liu, J. Chen, X. Liu, W. Guo, and Y. Chen, "An unsupervised detection approach for hardware Trojans," *IEEE Access*, vol. 8, pp. 158169–158183, 2020.
[43] S. Meenakshi and M. N. Devi, "Performance enhancement of unsupervised hardware Trojan detection algorithm using clustering-based local outlier factor technique for design security," in *Proc. IEEE ITC*, 2022, pp. 1–8.
[44] R. Bian, M. Xue, and J. Wang, "Building trusted golden models-free hardware Trojan detection framework against untrustworthy testing parties using a novel clustering ensemble technique," in *Proc. IEEE TrustCom/BigDataSE*, 2018, pp. 1458–1463.
[45] Y. Xiang, L. Li, and W. Zhou, "Random forest classifier for hardware Trojan detection," in *Proc. ISCID*, 2019, pp. 134–137.
[46] R. Lu, H. Shen, Z. Feng, H. Li, W. Zhao, and X. Li, "HTDet: A clustering method using information entropy for hardware Trojan detection," *Tsinghua Sci. Technol.*, vol. 26, no. 1, pp. 48–61, 2021.
[47] P. Zhao and Q. Liu, "Density-based clustering method for hardware Trojan detection based on gate-level structural features," in *Proc. AsianHOST*, 2019, pp. 1–4.
[48] N. Shang, A. Wang, Y. L. Ding, K. Gai, L. Zhu, and G. Zhang, "A machine learning based golden-free detection method for command-activated hardware Trojan," *Inf. Sci.*, vol. 540, Nov. 2020, pp. 292–307.
[49] S. Mittal, H. Gupta, and S. Srivastava, "A survey on hardware security of DNN models and accelerators," *J. Syst. Archit.*, vol. 117, Aug. 2021, Art. no. 102163.
[50] R. Sharma, G. K. Sharma, and M. Pattanaik, "A few shot learning based approach for hardware Trojan detection using deep siamese CNN," in *Proc. VLSID*, 2021, pp. 163–168.
[51] N. Muralidhar, A. Zubair, N. Weidler, R. Gerdes, and N. Ramakrishnan, "Contrastive graph convolutional networks for hardware Trojan detection in third party IP cores," in *Proc. IEEE HOST*, 2021, pp. 181–191.
[52] N. Das et al., "A Trojan framework in AES core to evade state-of-the-art HT detection schemes," *Microelectron. J.*, vol. 111, May 2021, Art. no. 105023.
[53] H. Salmani, M. Tehranipoor, and R. Karri, "On design vulnerability analysis and trust benchmarks development," in *Proc.. IEEE ICCD*, 2013, pp. 471–474.
[54] "Benchmarks." Accessed: Sep. 1, 2022. [Online]. Available: https://trust-hub.org/

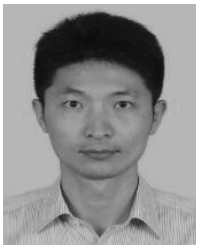
Chunhua He received the B.S. degree in microelectronics from Sun Yat-sen University, Guangzhou, China, in 2010, and the M.S. and Ph.D. degrees in microelectronics and solid-state electronics from Peking University, Beijing, China, in 2013 and 2018, respectively.

From 2013 to 2017, he was an Engineer with the No.5 Electronics Research Institute, Ministry of Industry and Information Technology, Guangzhou. From 2017 to 2019, he was an Engineer with Midea Group, Foshan, China. From 2019 to 2021, he was a Senior Engineer with Guangzhou 37 Degree Smart Home Company Ltd., Guangzhou. He was with the School of Computer, Guangdong University of Technology, Guangzhou, in 2021, where he is currently an Associate Professor. His current research interests include the design and application of MEMS sensors, IC security, and artificial intelligence.



Dengyun Lei received the B.S. degree in microelectronics from Xidian University, Xi'an, China, in 2010, and the Ph.D. degree in microelectronics from Peking University, Beijing, China, in 2015.

From 2015 to 2022, he severed as a Senior Engineer of Science and Technology with Reliability Physics and Application of Electronic Component Laboratory, No.5 Electronics Research Institute, Ministry of Industry and Information Technology, Guangzhou, China. He was with the School of Integrated Circuits, Guangdong University of Technology, Guangzhou, in 2022, where he is currently an Associate Professor. His main research interests include the reliability and security of integrated circuit.



Heng Wu received the B.S. degree in electronic and information engineering from Wuhan University of Science and Technology, Wuhan, China, in 2009, and the M.S. and Ph.D. degrees in optics and mechanical manufacture and automation from South China University of Technology, Guangzhou, China, in 2012 and 2017, respectively.

He is currently an Associate Professor with the School of Automation, Guangdong University of Technology, Guangzhou. His research interests are in the fields of optical imaging, optical measurement, machine vision, and image processing.

Lianglun Cheng received the B.S. and M.S. degrees in automatic control engineering from Huazhong University of Science and Technology, Wuhan, China, in 1988 and 1992, respectively, and the Ph.D. degree in automatic control engineering from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 1999.

He was with the School of Computer, Guangdong University of Technology, Guangzhou, China, in 1992, where he is currently a Professor and a Dean. His current research interests include Internet of Things and information physical integration system, interconnection and fusion of heterogeneous network, big data of industrial process, and high-performance computing.

Prof. Cheng is currently an expert with Special Subsidy from the Government of the State Council, the Director of the National and Local Joint Engineering Research Center for Integrated Technology of Intelligent Manufacturing Information Physics Fusion System, the Director of the Key Laboratory of Guangdong Information Physics Fusion System, and the Deputy Director of Qian Xuesen Innovation Committee of CASS, the Director of China Automation Society, the member of China Computer Society, the Vice Director of Guangdong Automation Society, and the Executive Vice President of Guangdong Robotics Society.

Guizhen Yan received the B.S. degree in semiconductor from Peking University, Beijing, China, in 1974.

From 1999 to 2001, she is a Senior Visiting Scholar with The Hong Kong University of Science and Technology, Hong Kong. She was with the National Key Laboratory of Science and Technology on Micro/Nano Fabrication, Institute of Microelectronics, Peking University, in 1974, where she is currently a Professor. She is the Founder of China's MEMS process technique and an expert in MEMS inertial sensors. She is proficient in the techniques of microelectronic devices and IC manufacturing, and has developed several sets of CMOS and MEMS process techniques. In the 1970s, she participated in the development of China's first VLSI. In the 1980s, she developed a complete set of process technique for polysilicon emitter-stage ultra-high-speed two-stage integrated circuits. She has developed the single-chip integration techniques for bulk silicon MEMS and CMOS. She has more than 40 invention patents and has published more than 200 papers in magazines and conferences. Her current research interests include the design and application of integrated circuit, MEMS, and IC process.

Qinwen Huang received the B.S. and M.S. degrees in microelectronics technology and microelectronics and solid state electronics from South China University of Technology, Guangzhou, China, in 2001 and 2004, respectively, and the Ph.D. degree in microelectronics and solid state electronics from the Institute of Microelectronics, Chinese Academy of Sciences, Beijing, China, in 2007.

From 2007 to 2009, he was a Lecturer of Institute of Mechanics, Nanjing University of Science and Technology, Nanjing, China. He was with Science and Technology on Reliability Physics and Application of Electronic Component Laboratory, No.5 Electronics Research Institute, Ministry of Industry and Information Technology, Guangzhou, China, in 2009, where he is currently a Professor of Engineering. His research interests are in the fields of MEMS sensors, MEMS reliability, and IC process and security.