重回帰分析を用いたサイドチャネル攻撃の高精度化

金　用大† 菅原　健† 林　優一† 本間 尚文† 青木 孝文†
佐藤　証‡

† 東北大学大学院情報科学研究科
980-8579 宮城県仙台市青葉区荒巻字青葉 6-6-05

{kimyd, sugawara, homma}@aoki.ecei.tohoku.ac.jp, yu-ichi@main.tains.tohoku.ac.jp, aoki@ecei.tohoku.ac.jp

‡ 独立行政法人産業技術総合研究所 情報セキュリティ研究センター
101-0021 東京都千代田区外神田 1-18-13

akashi.satoh@aist.go.jp

あらまし 本稿では，重回帰分析法を用いたサイドチャネル攻撃の高精度化について述べる．提案手法は，波形中から有用なサイドチャネル情報を含む箇所を効率的に取得するため，従来手法では困難な SN 比が非常に小さいサイドチャネル信号を取り扱う場合に有効である．また，従来のテンプレート攻撃や Stochastic モデル攻撃と比較して，プロファイリングに要する波形数を削減することが可能である．本稿では，サイドチャネル攻撃標準評価ボード SASEBO に実装した AES 回路を対象とした実験により，提案手法の有効性を評価する．特に，鍵推定率及びプロファイリング精度の向上について考察する．

Improved Side Channel Attack using Multivariate Regression Analysis

Yongdae Kim† Takeshi Sugawara† Yu-ichi Hayashi† Naofumi Homma†
Takafumi Aoki† Akashi Satoh‡

†Graduate School of Information Science, Tohoku University.
6-6-05 Aramaki Aza Aoba Aoba-ku Sendai-shi 980-8579 Japan

{kimyd, sugawara, homma}@aoki.ecei.tohoku.ac.jp, yu-ichi@main.tains.tohoku.ac.jp, aoki@ecei.tohoku.ac.jp

‡National Institute of Advanced Industrial Science and Technology
1-18-13 Sotokanda Chiyoda-ku Tokyo 101-0021 Japan

akashi.satoh@aist.go.jp

Abstract This paper presents an improved side-channel attack using multivariate regression analysis. The proposed method can acquire waveform points including significant side-channel information in an effective way, which makes it possible to handle waveform signals with lower S/N (signal-to-noise ratio) as compared with conventional methods. Our method also reduces the number of waveforms for profiling with the conventional template and stochastic model attacks. We demonstrate the performance of our method through some experiments using the side-channel attack standard evaluation board. In particular, we examine the key extraction rate and the profiling efficiency.

1 Introduction

A number of side-channel attacks on cryptographic device have been published since Kocher's paper [1]. Among these attacks, Template Attack [2] and Stochastic Model Attack with profiling phase [3], that are known as profiling attack, have been introduced as a class of the most efficient attacks. Many works have been done to different cryptographic devices and algorithms using profiling attacks [4]– [6].

On the other hand, such attacks have some issues to be considered. First, we have to determine *interesting points*: the points containing data-dependent variations. Second, we need a large number of side-channel information (*i.e.* waveforms) to be acquired to build an appropriate noise model of physical leakage.

If we choose inadequate interesting points including non-data dependent points, the accuracy of the attack would be decreased due to the wrong noise model built in the profiling phase. In the case of side-channel data with low S/N ratio, we need more traces for the profiling phase. Many papers discussed the two issues for conventional profiling attacks [4], [7], [5], [6].

Addressing the two issues, this paper proposes a new type of improved profiling attacks using multivariate regression analysis. Our analysis method can acquire the significant points of waveforms to improve profiling phase. Even if non-data dependent points are selected as the interesting points, the adverse effects for the following key extraction step can be minimized. As a result, our method has achieved the highest efficiency compared with conventional CPA (Correlation Power Analysis) and profiling attacks. In our experiments, we have applied our method to four types of side-channel information with different S/N ratios which are measured from different locations [8].

2 Conventional Methods

This section describes conventional two profiling attacks: Template Attack and Stochastic Model Attack. The attacks utilize a reference device equivalent to a target device to be attacked. We assume that the reference device can be fully controlled to determine the characteristics of the target device. The attacks consist of two phases: (i) profiling phase to learn about the device and (ii) key extraction phase to detect the secret key. The two phases for each profiling attack are briefly explained in the following.

2.1 Template Attack

Assume that a power trace is represented as a vector $\boldsymbol{t} = (t_1, t_2, \cdots, t_T)$, where T denotes the length of time instants. The profiling phase, collects a large number of power traces with different a plaintext d_i and a key k_j given as

$$d_i \in \{d_1, d_2, \cdots, d_D\}, \quad (1)$$

$$k_j \in \{k_1, k_2, \cdots, k_K\}, \quad (2)$$

where D and K denote the number of plaintexts and key candidates, respectively. Then, we estimate a mean vector and a covariance matrix of the multivariate normal distribution from the power traces with (d_i, k_j). The pair of the mean vector $\boldsymbol{m}$ and the covariance $\boldsymbol{C}$ is referred as *template*, $h_{d_i,k_j} = (\boldsymbol{m}, \boldsymbol{C})_{d_i,k_j}$. Given the power trace $\boldsymbol{t}$, and a template $h_{d_i,k_j} = (\boldsymbol{m}, \boldsymbol{C})_{d_i,k_j}$, the key extraction phase computes the probability density function of the multivariate normal distribution as follows :

$$\boldsymbol{q} = \boldsymbol{t} - \boldsymbol{m}, \quad (3)$$

$$p(\boldsymbol{t}; (\boldsymbol{m}, \boldsymbol{C})_{d_i,k_j}) = \frac{exp\left(-\frac{1}{2}\boldsymbol{q}'\boldsymbol{C}^{-1}\boldsymbol{q}\right)}{\sqrt{(2\pi)^T det(\boldsymbol{C})}}, \quad (4)$$

where $det(\boldsymbol{C})$ denotes the determinant of $\boldsymbol{C}$. We get the probabilities for every templates ($p(\boldsymbol{t}; (\boldsymbol{m}, \boldsymbol{C})_{d_1,k_1}), \cdots, p(\boldsymbol{t}; (\boldsymbol{m}, \boldsymbol{C})_{d_D,k_K})$). We estimate the correct key k_{ck} using the maximum likelihood principle with the probabilities as follows :

$$k_{ck} = \underset{k_j \in k^*}{\operatorname{argmax}}\, p(\boldsymbol{t}; h_{d_i,k_j}), \quad (5)$$

where k^* is all possible key candidates.

2.2 Stochastic Model Attack

Assume that a power trace at time t with a plaintext d_i and a key k is represented as $I_t(d_i, k)$.

Stochastic model attack assumes that the power trace can be written as the sum of a deterministic part and a random part as follows :

$$I_t(d_i, k) = h_t(d_i, k) + R_t, \tag{6}$$

where $h_t(d_i, k)$ denotes the deterministic part of the power trace as far it depending on d_i and k, and R_t denotes a random part that does not depend on d_i and k. The profiling phase is divided into two steps in order to approximate the two parts. In the first step, we generate the profile of the deterministic part $\hat{h}_t(d_i, k)$ using N_1 traces. After having determined the approximators $\hat{h}_t(d_i, k)$, we use a different set of N_2 power traces to estimate the distribution of the random part. In order to approximate the distribution, we first calculate the T-dimensional random vector $\boldsymbol{R} = (R_1, R_2, \cdots, R_T)$ as follows :

$$R_t = I_t(d_i, k) - \hat{h}_t(d_i, k), \tag{7}$$

where $t = 1, 2, \cdots, T$. We assume that the random vector is normally distributed with a covariance matrix $\boldsymbol{C}$, which is computed as follows :

$$C_{i,j} = E(R_i R_j) - E(R_i)E(R_j), \tag{8}$$

where $1 \leq i, j \leq T$, and $E(X)$ denotes the expected value of the variable X.

The key extraction phase obtains another set of N_3 new power traces from the target device corresponding to known plaintext $d_i \in \{d_1, d_2, \cdots, d_{N_3}\}$ and an estimated key k_{ck}. Using the power traces, a noise vector $\boldsymbol{z}_i$ is firstly computed as follows :

$$\boldsymbol{z}_i = I_t(d_i, k_{ck}) - \hat{h}_t(d_i, k_j), \tag{9}$$

where $k_j \in \{k_1, k_2, \cdots, k_K\}$, and K denotes the number of key candidates. This vector has a multivariate normal distribution with a covariance matrix $\boldsymbol{C}$. We can compute the following probabilities :

$$p\left(\boldsymbol{z}_i; \hat{h}_t(d_i, k_j)\right) = \frac{exp\left(-\frac{1}{2}\boldsymbol{z}_i' \boldsymbol{C}^{-1} \boldsymbol{z}_i\right)}{\sqrt{(2\pi)^T det(\boldsymbol{C})}}. \tag{10}$$

If k_{ck} is a correct key, the probability density function $p(\boldsymbol{z}_i; \hat{h}_t(d_i, k_{ck}))$ is assumed to have the highest probability. The maximum likelihood principle is applied to extract a correct key k_{ck} using all the measured N_3 traces as follows :

$$k_{ck} = \underset{k_j \in k^*}{\operatorname{argmax}} \Pi_{i=1}^{N3} p\left(\boldsymbol{z}_i; \hat{h}_t(d_i, k_j)\right). \tag{11}$$

3 Proposed Method

In this section, we first introduce multivariate regression model, and then show our proposed method applied for attacking cryptographic hardware implementations.

3.1 Multivariate Regression Model

Given N observations of P independent (explanatory) variables and a dependent (response) variable as follows.

$$\begin{aligned}
\text{Observation 1} &: (x_{1,1}, x_{1,2}, \cdots, x_{1,P}, y_1) \\
\text{Observation 2} &: (x_{2,1}, x_{2,2}, \cdots, x_{2,P}, y_2) \\
&\vdots \\
\text{Observation N} &: (x_{N,1}, x_{N,2}, \cdots, x_{N,P}, y_N)
\end{aligned}$$

Here, $x_{i,j}$ and y_i denote the j-th independent variable and the dependent variable for the i-th observation, respectively. Let $\beta_0, \beta_1, \cdots, \beta_P$ denotes $P+1$ unknown parameters (regression coefficients), then the multivariate regression model can be written as follows :

$$y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \cdots + \beta_P x_{i,P} + \epsilon_i, \tag{12}$$

where ϵ_i denotes the i-th residual part that is independent and normally distributed.

Let $\hat{\beta}_0, \hat{\beta}_1, \cdots, \hat{\beta}_P$ denotes the estimators of the parameters $\beta_0, \beta_1, \cdots, \beta_P$. For the i-th observation, the predicted value $\hat{y}_i$ is

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i,1} + \hat{\beta}_2 x_{i,2} + \cdots + \hat{\beta}_P x_{i,P}. \tag{13}$$

Therefore, the i-th residual is computed as $\epsilon_i = y_i - \hat{y}_i$. The least squares chooses the values of the parameters that make the sum of the squared residuals $\sum_{i=1}^{N} \epsilon_i^2$ as small as possible.

3.2 Side-channel Attack Based on Multivariate Regression Model

Our method applies the above multivariate regression model to profiling attacks.

3.2.1 Profiling Phase

The profiling phase defines independent and dependent variables to build a multivariate regression model.

Define independent variable: We consider all the encryption/decryption computations to define the independent variable as a Hamming distance value. For example, we derive a 128-bit Hamming distance value from the 16 S-box computations in AES even though we estimate each subkey corresponding to one S-box. We referred to this value as a *simulated value*. Let $v^l_{d_i,k}$ denotes a Hamming distance value of the l-th S-box in a N plaintext $d_i (\in \{d_1, d_2, \cdots, d_N\})$ and a key k. The simulated value s_i can be computed as follows :

$$s_i = \sum_{l=1}^{16} v^l_{d_i,k} \tag{14}$$

Define dependent variables: We refer to the power trace corresponding to a plaintext d_i as $\boldsymbol{t}_i = \{t_{i,1}, t_{i,2}, \cdots, t_{i,T}\}$, where T denotes the length of the time instants. The traces are defined as the dependent variables. We select some time instants from the T-dimensional trace $\boldsymbol{t}_i$ in order to reduce the computational time and improve the classification rate because the time instants are not always correlated to the Hamming distance value. For the selection of time instants, we calculate a squared Pearson correlation between s_i and $t_{i,t}$ considering both negative and positive correlation as follows :

$$\rho_t^2 = corr(s_i, t_{i,t})^2, \tag{15}$$

where $corr(X, Y)$ denotes the Pearson correlation value between the variables X and Y. An adversary select the $M (< T)$ instants having the highest ρ_t^2. We denote these time instants as M interesting points.

Multivariate Regression Model: We compute the regression coefficients using multivariate regression analysis, and then generate the following fitted regression model,

$$\hat{s}_i = \hat{\beta}_0 + \hat{\beta}_1 t_{i,1} + \cdots + \hat{\beta}_M t_{i,M}, \tag{16}$$

where is the estimated value of s_i.

3.2.2 Key Extraction Phase

The key extraction phase measures N power traces from the target device corresponding to N plaintexts $\{d_1, d_2, \cdots, d_N\}$ and an estimated key k_{ck}. We utilize the regression model in Eq. (16) to calculate the $\hat{s}_i$. To extract secret key, the Pearson correlation value between the predicted simulated value $\hat{s}_i$ and the Hamming distance value v_{d_i,k_j} for each key candidate $k_j \in \{k_1, k_2, \cdots, k_K\}$ of a S-Box is calculated. The correct key k_{ck} can be estimated as follows :

$$k_{ck} = \underset{k_j \in k^*}{\operatorname{argmax}}\, corr(\hat{s}_i, v_{d_i,k_j}). \tag{17}$$

4 Experimental Analysis

In this section, we demonstrate the performance of the proposed attack in comparison with those of conventional attacks. The Side-channel Attack Standard Evaluation Board (SASEBO) [9] was used for both the target and the reference devices. Four sets of side-channel waveforms (*i.e.*, power and EM waveforms) with different S/N ratios were measured from an FPGA implementation of the 128-bit AES (Advanced Encryption Standard) on the board. The measurement waveforms were (i) the voltage drop across a resistor (power), (ii) the current on the attacked power cable (power cable), (iii) the current on the communication cable (RS232C cable), and (iv) the magnetic field around the power cable (antenna). The more details on the measurement conditions are described [8].

Figure 1 shows the proposed result using the four waveform sets, where the classification rate indicates the number of s-boxes that we could distinguish a correct key from all possible key candidates. If all the subkeys were obtained, the classification rate is 100. For comparison, the figure also shows the results of conventional attacks using the same sets of

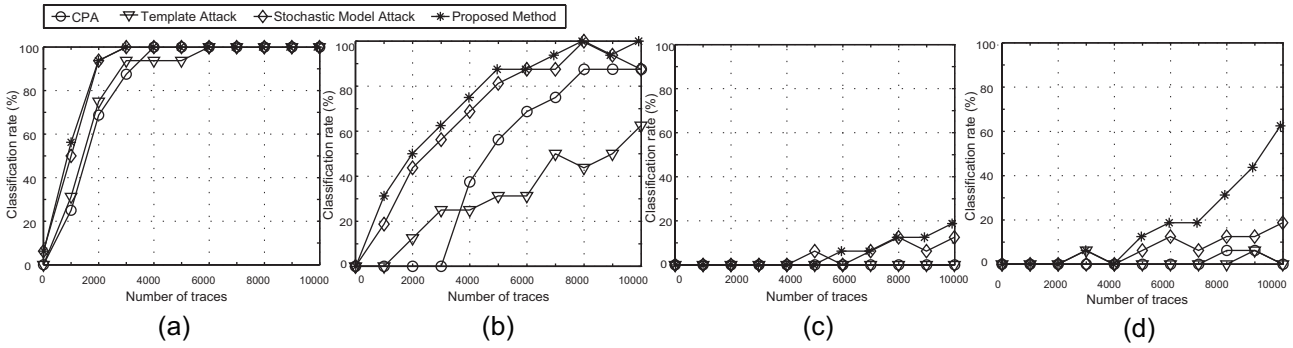

Figure 1: Experimental results of the proposed method with four types of waveforms: (a) voltage drop across a resistor (power) (b) current on an attached power cable (c) current on an attached communication cable (d) magnetic field around the power cable

waveforms. For the conventional profiling attacks and the proposed attack, 20,000 traces and 5 interesting points are used in the profiling phase and other 10,000 traces are used in key extraction phase. As a result, our method has the most least number of traces to estimate all the subkeys of AES in the measurements (i) and (ii). Even though the final classification rates of the measurements (iii) and (iv) are not 100 within the 10,000 traces, the rate of our method were higher than those of conventional attacks.

Figure 2 shows the average values of classification rates for the proposed method and the stochastic model attack, where the horizontal axis is the number of traces used in the profiling phase and the number of interesting points are fixed to 5 points. We clearly find that our method has the higher rate for all the waveform types than the stochastic model attack though the stochastic model attack is known as the most powerful attack which has the highest profiling efficiency and classification rate among the conventional profiling attacks [5]. This result indicates that our method can be more efficient than the stochastic model attack in the case that the number of cryptographic operations on the reference device is severely limited.

Figure 3 shows the average of classification rates associated with the number of interesting points, where the number of traces for profiling is 20,000. The result shows that the averages of classification rates for the stochastic model attack decrease when non-data dependent time instants are included in the interesting points as shown in Figs. 3 (a) and (b). In the proposed method, on the other hand, the classification rates do not decrease even though the number of (non-data dependent) time instants increases since such irrelevant time instants have a less significant effect on the regression model. This means that the proposed method takes less effort to determine the time instants in order to extract all keys successfully.

5 Conclusion

This paper proposed an improved side-channel attack using multivariate regression analysis. The experimental result showed that the proposed method has a significant advantage for the number of side-channel traces required to disclose secret keys as compared with the conventional attacks. We also confirmed that the profiling efficiency of the proposed method is higher than that of the stochastic model attack and the noise distribution can be successfully profiled even though the measured signals have significantly low S/N ratios.

References

[1] P. Kocher, J. Jaffe, and B. Jun, "Differential power analysis," *Lecture Notes in Computer*

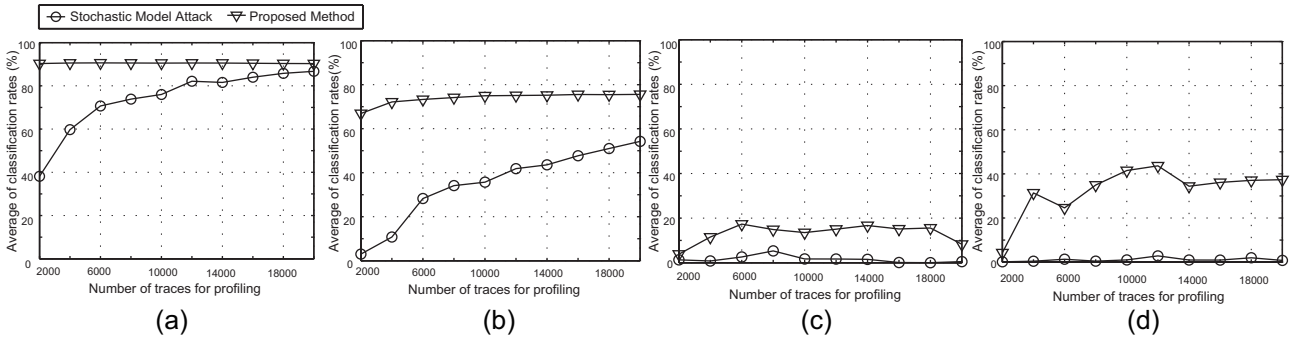

Figure 2: Average classification rate of the proposed method associated with the number of traces for profiling: (a) voltage drop across a resistor (power) (b) current on an attached power cable (c) current on an attached communication cable (d) magnetic field around the power cable

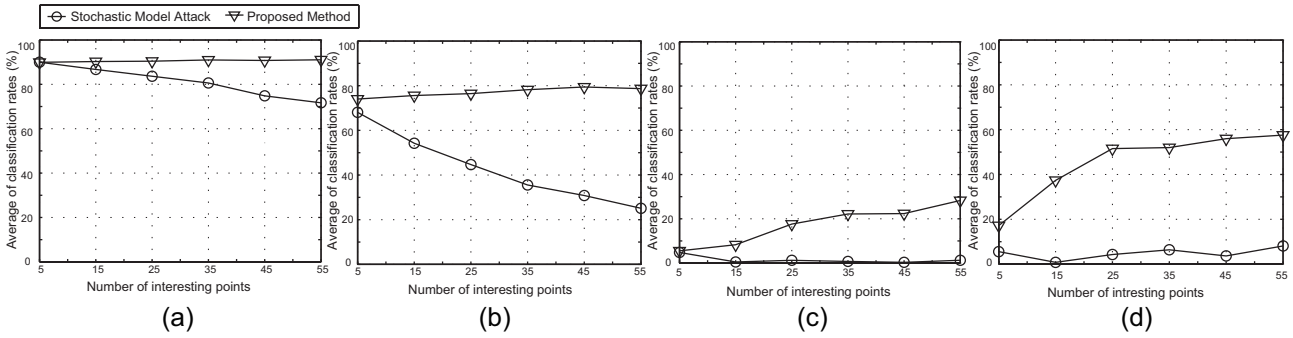

Figure 3: Average classification rate of the proposed method associated with the number of interesting points: (a) voltage drop across a resistor (power) (b) current on an attached power cable (c) current on an attached communication cable (d) magnetic field around the power cable

Science, vol. 1666, pp. 388–397, Aug. 1999.

[2] S. Chari, J. R. Rao, and P. Rohatgi, "Template attacks," *Lecture Notes in Computer Science*, vol. 2523, pp. 12–28, 2002.

[3] S. W, K. Lemke, and C. Paar, "A stochastic model for differential side channel cryptanalysis," *Lecture Notes in Computer Science*, vol. 3659, pp. 30–46, 2005.

[4] C. Rechberger and E. Oswald, "Practical template attacks," *Lecture Notes in Computer Science*, vol. 3325, pp. 440–456, 2005.

[5] B. Gierlichs, L.-R. K, and C. Paar, "Templates vs. stochastic methods," *Lecture Notes in Computer Science*, vol. 4249, pp. 15–29, 2006.

[6] F.-X. Standaert, K. F, and W. Schindler, "How to compare profiled side channel attacks," *Lecture Notes in Computer Science*, vol. 5536, pp. 485–498, 2009.

[7] C. Archambeau, E. Peeters, F.-X. Standaert, and J.-J. Quisquater, "Template attacks in principal subspaces," *Lecture Notes in Computer Science*, vol. 4249, pp. 1–14, 2006.

[8] T. Sugawara, Y. Hayashi, N. Homma, T. Mizuki, T. Aoki, H. Sone, and A. Satoh, "Mechanism behind information leakage in electromagnetic analysis of cryptographic modules," *The 10th International Workshop on Information Security Applications*, August 2009.

[9] A. Research Center for Information Security, "Side-channel Attack Standard Evaluation BOard (SASEBO)," http://www.rcis.aist.go.jp/special/SASEBO.