IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received August 28, 2018, accepted October 22, 2018, date of publication October 29, 2018, date of current version December 3, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2878421

Multi-Byte Power Analysis: A Generic Approach Based on Linear Regression

SHAN FU[1,2], ZONGYUE WANG[3], GUOAI XU[1], FANXING WEI[2], AN WANG[4], JUAN PAN[2], YUGUANG LI[2], AND NING ZHANG[5]

[1]School of Cyberspace Security, Beijing University of Posts and Telecommunications, Beijing 100876, China
[2]China Academy of Information and Communications Technology, Beijing 100191, China
[3]Open Security Research, Shenzhen 518052, China
[4]School of Computer Science, Beijing Institute of Technology, Beijing 100081, China
[5]Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China

Corresponding author: Guoai Xu (xga@bupt.edu.cn)

This work was supported in part by the National Key R&D Program of China under Grant 2018YFB0803605 and in part by the National Natural Science Foundation of China under Grant 61873069.

ABSTRACT Linear regression used to be known as a robust side-channel analysis (SCA) method as it makes use of independent bits leakage. This leakage assumption is more general than Hamming weight/Hamming distance model used in correlation power analysis (CPA). We find that in many common scenarios, linear regression is not only an alternative but also a more efficient tool compared with CPA. This paper proposes a generic SCA approach based on linear regression called multi-byte power analysis (MPA) that can be applied to any number of bytes instead of one single byte when performing SCA. Two typical cases are illustrated in this paper. One is recovering keys with XOR operation leakage and the other one is chosen plaintext attack on block ciphers with leakages from round output. Simulation results are given to compare with traditional CPA in both cases. MPA achieves up to 400% and 300% improvements for the corresponding case compared with CPA, respectively. Experiments with AES on SAKURA-G board also prove the efficiency of MPA in practice, where 128 key bits are recovered with 1500 traces using XOR operation leakage and one key byte is recovered with only 50 chosen-plaintext traces in the other case.

INDEX TERMS AES, linear regression, multi-byte power analysis, side-channel analysis.

I. INTRODUCTION

With the development of Internet of Things (IoT), embedded devices such as smart cards, mobile phones, RFID tags and even sensor networks are widely used in our daily lives. New cryptographic techniques are applied to secure these devices and networks. Though these devices are extensively used, the sensitive data contained in them might be easily recovered by adversaries. Side-channel attacks, especially power analysis attacks [5], [6], [10], [11], [14] provide an access to data in cryptographic implementations, which are well-known threats to these devices. The most famous example is Differential Power Analysis introduced by Kocher *et al.* in 1999 [14]. In their experiment, they monitored the power consumption of a smart card and extracted the secret key efficiently. Later, the correlation between power consumption traces and modeled values (e.g. under Hamming weight model) of handled data was taken into account.

A. RELATED WORKS

The Correlation Power Analysis (CPA) was proposed in 2004 by Brier *et al.* [5]. Subsequently, several works applied this idea to practical environments and achieved good results [7], [19]. Other attack models such as Partitioning Power Analysis (PPA) [15], Collision Attack (CA) [4], [23], Mutual Information Analysis (MIA) [3], [10] and Differential Cluster Analysis (DCA) [2] have also been studied. Linear regression side-channel analysis has been introduced by Schindler *et al.* in 2005 [22]. Initially, they describe an efficient profiling method for SCA. The attack can recover the independent bits leakage function coefficient based on the known subkey k using linear regression. With coefficient of determination R^2, Doget *et al.* [9] further developed a non-profiled key-recovery attack. The linear regression attacks can be applied in the same context as the CPA, but with a weaker assumption on the device behavior. This extension is denoted as Linear Regression Attack (LRA) in [26].

Furthermore, these results of LRA has been extended to apply against first-order masking techniques [8], which are the main SCA countermeasures. Other contributions [12], [13], [16] improve the efficiency and effectiveness of LRA when applied to real attack procedure. For leakage model [1], [5], [17], [20], even though independent bits leakage model is more general, the Hamming weight model and the Hamming distance model suit most devices well in practice, especially for leakages from registers and data bus. A general understanding is that Hamming weight model is applicable to software while Hamming distance model to hardware implementations.

B. OUR CONTRIBUTIONS

The recent analysis methods, including CPA, MIA and LRA, are always applied to one single byte of block ciphers, thus the analyzer has to perform the attack several times in order to recover the whole key. It consumes too much time and efforts which may be unaffordable in practice. In this paper, we investigate linear regression method under Hamming weight/ Hamming distance model and find that in many common scenarios linear regression method is not only an alternative but also a more efficient tool compared with CPA. Furthermore, we propose a generic SCA approach based on linear regression called Multi-byte Power Analysis (MPA) which can be applied to any number of bytes instead of one single byte when performing SCA. Two typical cases are recovering keys with XOR operation leakage and chosen plaintext attack on block ciphers in T-table software or round based hardware implementations.

In the first case, we recover key from the leakage $m \oplus k$ where m is the message and k is the whitening key. In Hamming weight model, leakage is expressed as a linear function of Hamming weight of the intermediate value $m \oplus k$. We find that in the expression of leakage, the signs of coefficients of every independent bit in m indicates the value of corresponding bits of k. Multi-byte Power Analysis is used to examine relation between several bytes of m and leakages, and recover the keys. Simulations of 8-bit, 32-bit, 64-bit and 128-bit leakages are given to make comparison between MPA and CPA. The result shows that MPA is much more efficient in the multi-byte situation. For 128-bit leakages, we achieve a 400% improvement compared with CPA.

In the second case, we focus on block cipher with leakage only in round output. This is a very common and very normal scenario. T-table based software and round-based hardware implementations are both examples. Typically, chosen plaintext method is used to decrease key-guessing space where some bytes of plaintext are kept stable. However, these stable bytes result in round output XORing to the calculated intermediate values used in CPA. As shown in our simulations, round output has great impact on CPA efficiency. We prove that MPA can overcome this impact. With MPA we achieve a 300% improvement compared with CPA.

C. ORGANIZATION

The remaining of the paper is organized as follows: Section II illustrates some preliminaries. Section III gives the attack on XOR operation, including simulation and experimental results. Section IV gives the attack method and experiment results on round output. Finally, we conclude this paper in Section V.

II. PRELIMINARIES

A. HAMMING WEIGHT AND HAMMING DISTANCE MODEL

The Hamming weight model is proposed by Kocher [14] and Messerges [18] which generally assumed that leakage through the power side-channel depends on the number of bits set in the data. Let T be the leakage value of data X and $HW(\cdot)$ be the Hamming weight function, the Hamming weight model is described as follow:

$$T = a \cdot HW(X) + c + \sigma.$$

where a is a scalar coefficient, c is a constant consumption and σ is noise.

The Hamming distance model was proposed by Eric Brier *et al.* in CHES 2004 [5] where the leakage is assumed to be depended on the number of bits switching from one state to another. The consumptions for a bit switching from 0 to 1 or from 1 to 0 are further assumed to be same. Let the current state be R and the next state be X. The number of flipping bits equals $HW(R \oplus X)$. The Hamming distance model is described as follow:

$$T = a \cdot HW(R \oplus X) + c + \sigma.$$

where a is a scalar coefficient, c is a constant consumption and σ is noise.

The Hamming weight and the Hamming distance model suit most devices well in practice, especially for leakages from registers and data buses. A general understanding is that Hamming weight model is applicable to software while Hamming distance model to hardware implementations. As the Hamming weight model and the Hamming distance model are similar, we only describe our method under Hamming weight model in the following for simplification.

B. CORRELATION POWER ANALYSIS

Since the significant work of Kocher [14] of side-channel analysis in late 1990's, a large amount of work has been devoted. As a most famous successor, Correlation Power Analysis is proposed by Brier *et al.* in 2004.

The correlation coefficient between Hamming weight of target data X and the power consumption T is described as follow:

$$\rho_{HW(X),T} = \frac{cov\left(HW\left(X\right), T\right)}{\sigma_{HW(X)}\sigma_T}.$$

where $cov(\cdot)$ is the covariance between $HW(X)$ and T. σ_X and σ_T are standard deviation for $HW(x)$ and T respectively. When implemented, the target data X should relate to some unknown key bits and the correlation coefficient is used as a distinguisher. The attacker guesses the unknown key bits and calculates $\rho_{HW(X),T}$ for every key candidate. It is the biggest $\left|\rho_{HW(X),T}\right|$ that is supposed to indicate the correct key. For detail, we refer to [5].

C. MULTIPLE LINEAR REGRESSION

In statistics, multiple regression is an approach for modeling the relationship between a scalar dependent variable y and several explanatory variables denoted $X = (x_1, x_2 \ldots, x_p)$. For Multiple Linear Regression (MLR), the relationships are modeled by linear predictor function:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p. \tag{1}$$

where $\beta = (\beta_0, \beta_1, \cdots, \beta_p)^T$ is the unknown model parameter which can be estimated by giving sample sets of y and X. Ordinary least square method is the most commonly used estimator. For given N sample sets

$$y_s = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}, \quad X_s = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ x_{21} & \cdots & x_{2p} \\ \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{Np} \end{pmatrix},$$

the ordinary least square method first generates a new matrix M as

$$M = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N1} & \cdots & x_{Np} \end{pmatrix}.$$

Then the estimation of β is

$$\hat{\beta} = (M^T M)^{-1} M^T y_s \tag{2}$$

where M^T is the transpose of M. The confidence of determination, denoted R^2, indicates how well samples fit the linear model established with $\hat{\beta}$.

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} (y_i - \bar{y}_i)^2}, \tag{3}$$

where $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \cdots + \hat{\beta}_p x_{pi}$ is the estimated y_i with the linear model and $\bar{y}_i$ is the mean of y_s. R^2 has a value in the range of [0, 1] with 1.0 being the best fit.

D. LINEAR REGRESSION DISTINGUISHER

Linear regression distinguisher is proposed by Doget *et al.* to perform Robust Side-channel Analysis [9]. Let T be the leakage measurement. Choose an n-bit target value V_k which depends on part of key bits. V_k is further denoted as $(v_k[n], v_k[n-1], \ldots, v_k[1])$ which is its binary decomposition. Instead of the correlation coefficient, the linear regression distinguisher tests the linear relationship using R^2. The process is as follows:

1) For every key candidate $\hat{k}$, calculate $(v_{\hat{k}}[n], v_{\hat{k}}[n-1], \ldots, v_{\hat{k}}[1])$ for every measurement.
2) Construct the model between T and $(v_{\hat{k}}[n], v_{\hat{k}}[n-1], \ldots, v_{\hat{k}}[1])$ as
$$T = \beta_{\hat{k},0} + \beta_{\hat{k},1} v_{\hat{k}}[1] + \cdots + \beta_{\hat{k},n} v_{\hat{k}}[n].$$
Estimate the parameter $\beta_{\hat{k}} = (\beta_{\hat{k},0}, \beta_{\hat{k},1}, \cdots, \beta_{\hat{k},n})^T$ with ordinary least square method as shown in Sec.II-C.
3) Compute and store the confidence of determination $R^2_{\hat{k}}$ for $\hat{k}$.

The key candidate $\hat{k}$ with largest $R^2_{\hat{k}}$ is considered to be the right key with the highest level of confidence.

III. MULTI-BYTE POWER ANALYSIS ON XOR OPERATION

Original linear regression attack is known as an alternative to CPA using R^2 instead of ρ to distinguish key candidates. The advantage of linear regression is that the model assumption is weaker as the weight of different bit could be differed. In this section, we illustrate a novel power analysis attack model which called MPA and show its high efficiency on attacking XOR operation.

A. LEAKAGE FROM XOR OPERATION

XOR Operation is a commonly used operation in ciphers. For example, in most block ciphers, the plaintext m is XORed with whitening key k as the first step. Attackers can use leakage from $m \oplus k$ to launch a side-channel attack.

Typically, CPA is used to recover k as follows:

1) For every key candidates $\hat{k}$, calculate $L_{\hat{k}} = m \oplus \hat{k}$.
2) Further calculate $\rho_{HW(L_{\hat{k}}),T}$ and sort key candidates according to $\left|\rho_{HW(L_{\hat{k}}),T}\right|$.
3) Output the key candidates with largest $\left|\rho_{HW(L_{\hat{k}}),T}\right|$.

In block ciphers, k is usually a 128-bit value. Considering the performing architecture, the leakage can be 8-bit, 16-bit, 32-bit, 64-bit and even 128-bit (most for hardware situation). In CPA procedure, guessing a multi-byte k value results in high computation cost which is almost impossible for 32-bit, 64-bit and 128-bit architectures. One solution is separating k into different parts and performing CPA in every part. When performing one part, leakages from other parts are considered noise. Although this solution works, the efficiency is reduced as only part of information is used.

Our method is based on insight of Hamming weight model. m and k are denoted as $(m[n], m[n-1], \ldots, m[1])$ and $(k[n], k[n-1], \ldots, k[1])$ respectively, which are their binary decomposition. According to the Hamming weight model, the leakage of target value $m \oplus k$ is expressed as:

$$\begin{aligned} T &= a \cdot HW(m \oplus k) + c + \sigma \\ &= a \cdot \sum_{j=1}^{n} (m[j] \oplus k[j]) + c + \sigma. \end{aligned}$$

As k is a stable value, we have

$$m[j] \oplus k[j] = \begin{cases} m[j] & \mathit{if}\, k[j] = 0 \\ 1 - m[j] & \mathit{if}\, k[j] = 1. \end{cases}$$

Hence,

$$\begin{aligned} T &= a \cdot \left(\sum_{k[j]=0} m[j] + \sum_{k[j]=1} (1 - m[j]) \right) + c + \sigma \\ &= a \cdot \sum_{k[j]=0} m[j] - a \cdot \sum_{k[j]=1} m[j] + a \cdot \sum_{k[j]=1} 1 + c + \sigma. \end{aligned}$$

Observation 1: In the leakage expression, the bits with value '0' and the bits with value '1' in k give opposite sign of coefficient of corresponding bits in m.

B. ATTACK PROCEDURE

According to Observation 1, the sign of coefficient of every bit in m gives a predict to corresponding bit in k. We use multiple linear regression to estimate the coefficient, taking T as the dependent variable and $m = (m[n], m[n-1], \dots, m[1])$ as the explanatory variables. In practical terms, T is not a single value but a set $T = (T_1, T_2, \cdots, T_N)$ which formed a trace. The attack procedure is as follows:

1) Random choose plaintext m, perform encryption and record leakage traces.
2) For every trace point variable T_j, construct the model between T_j and $(m[n], m[n-1], \dots, m[1])$ as

$$T_j = \beta_{j0} + \beta_{j1} m[1] + \cdots + \beta_{jn} m[n].$$

Estimate the coefficient $\beta_j = (\beta_{j0}, \beta_{j1}, \cdots, \beta_{jn})^T$ with ordinary least square method as shown in Sec.II-C. Calculate and store the corresponding confidence of determination R_j^2.
3) Find the largest R^2 among all R_j^2, denoting as R_b^2.
4) Recover every bit in k as $\hat{k}[i] = 0$ if β_{bi} is positive and $\hat{k}[i] = 1$ otherwise.
5) Test the $\hat{k}$ and $\neg\hat{k}$ where $\neg$ is inverse. Output the correct one.

As we do not guess key, our method is applicable to leakage on any architecture even if XORed a very long whitening key. In the multi-byte leakage situation, our method makes use of whole information of the leakage which brings a higher efficiency compared with CPA. Besides, even though we deduce this method from Hamming weight model, our assumption is still weaker. The weight of every bit could also differ. We only assume these weight have the same positive or negative sign.

C. SIMULATION

To compare the efficiency between our method and CPA, we do simulations on 8-bit, 32-bit, 64-bit and 128-bit k with noise $\sigma = 2$. For CPA, we applied divide and conquer technique, guessing 8-bit every time until all bits of k are recovered. 1000 parallel experiments using different number of traces are performed to estimate success rates, indicating the percentage of total experiment times that recovering all key bits. The Fig. 1 shows the simulation result.

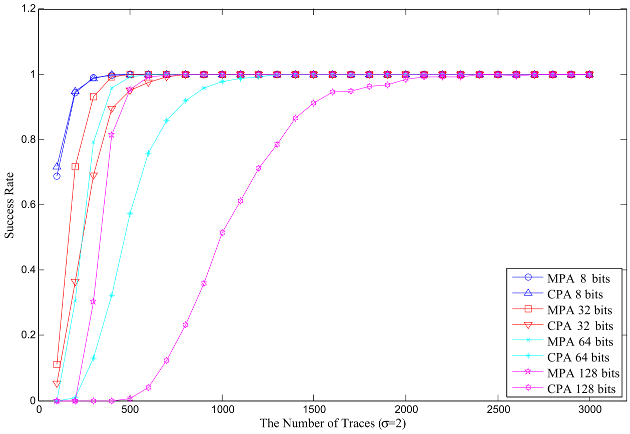


FIGURE 1. Simulation of MPA and CPA on 8-bit, 32-bit, 64-bit and 128-bit ***k***.

In the case of 8-bit, 32-bit, 64-bit and 128-bit k, the MPA method only computes β and R^2 once, while CPA computing 2^8, $4 * 2^8$, $8 * 2^8$ and $16 * 2^8$ respectively. As illustrated in Fig. 1, the efficiency of our method and CPA almost match in 8-bit situation. But in the multi-byte situation, our method is much more efficient. For 128-bit k, CPA requires more than 2000 traces to reach success rate 1 while our method needs only 500 traces which means a 400% improvement of efficiency.

D. EXPERIMENT RESULTS ON FPGA

We also test MPA in practice using SAKURA-G board, performing key XOR operation and acquiring the power consumption with an oscilloscope where key is a 128-bit target value. The result in Fig. 2 shows R^2 with 2000 traces. With the sample data of the highest R^2, all 128 bits of key are recovered.

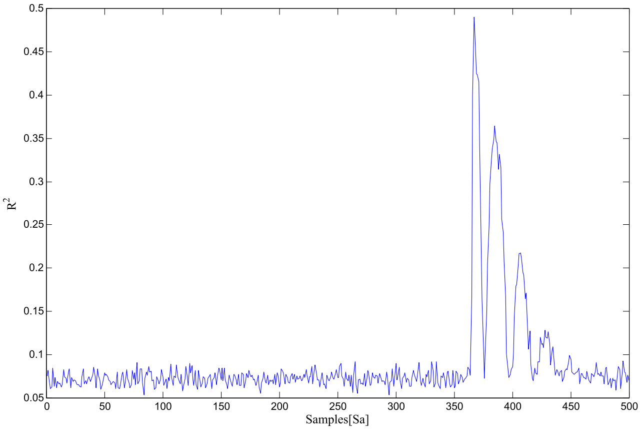


FIGURE 2. ***R*²** with 2000 traces.

Fig. 3 indicates the number of recovered bits under different number of traces. The recovered key bits would be more than 110 bits out of 128 bits in total within only 400 traces.

TABLE 1. R^2 **and** ρ **when** $HW(u)$ **changes.**

$HW(u)$	0	1	2	3	4	5	6	7	8
$R^2_{T,x}$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$\rho_{HW(x),T}$	1.000	0.746	0.495	0.245	-0.003	-0.248	-0.497	-0.747	-1.000

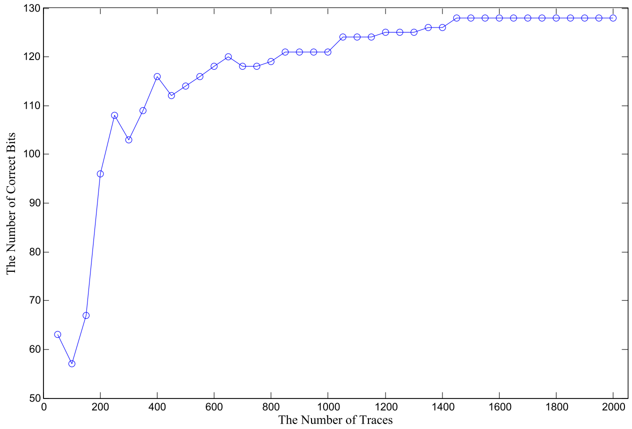


FIGURE 3. Correct bits recovered.

The experiment proves that our method has remarkable high efficiency in the multi-byte scene.

IV. MULTI-BYTE POWER ANALYSIS ON ROUND OUTPUT

When doing side-channel analysis, the attacker may concentrate on the round output in practice. Inspired from the concept of MPA, we give an efficient method to perform SCA on the round output.

A. LEAKAGE ON ROUND OUTPUT

Leakage on the round output means that leakage T is caused by $x \oplus u$ where x is an intermediate value related to some guessed key bit and u denotes the value of round output. This is a common scenario in side-channel attacks, especially in chosen plaintext cases [24]. To perform CPA, one way is to further guess u using $\rho_{HW(x \oplus u),T}$ as a distinguisher. This method is computationally infeasible when u is 32-bit or larger. In [24], Tu *et al.* suggests directly use $\rho_{HW(x),T}$ as a distinguisher as there exists some linear correlations between $HW(x)$ and $HW(x \oplus u)$. This is still not a perfect solution because $\rho_{HW(x),T}$ is highly affected by u, which may lose efficiency under some u values.

Similar as in Section III-A, in the Hamming weight model, the leakage T can be expressed in bitwise. We have

$$T = a \cdot \sum_{u[j]=0} x[j] - a \cdot \sum_{u[j]=1} x[j] + a \cdot \sum_{u[j]=1} 1 + c + \sigma.$$

Even though u changes the sign of coefficient of some bits in x, it does not affect the linear relationship between x and T. So we can perform MPA between x and T, and takes the confidence of determination $R^2_{T,x}$ as distinguisher which is not affected by u. Hence MPA with linear regression using $R^2_{T,x}$ overcomes this situation without losing efficiency.

A simulation is made to make comparison between CPA distinguiser $\rho_{HW(x),T}$ and MPA distinguisher $R^2_{T,x}$. In the simulation, u is an 8-bit value and $T = HW(x \oplus u)$ without considering the noise and constant consumption for simplicity. The result is shown in Table 1. We can see that $\rho_{HW(X),T}$ changes over $HW(u)$ while $R^2_{T,x}$ is always stable.

Observation 2: The confidence of determination R^2 in linear regression distinguisher is not affected by the round output.

B. CHOSEN PLAINTEXT MULTI-BYTE POWER ANALYSIS

Chosen plaintext [25] is a common technique when attacking block ciphers. However, this technique may face the same problem in Table 1 when performing CPA. We give an exemplified attack on AES to show how MPA with linear regression overcomes this obstacle efficiently.

AES is block cipher supporting 128-bit blocks and 128/192/256-bit keys [21]. Based on substitution permutation network (SPN) structure, AES XORs whitening key firstly and performs 10/12/14 round functions. Except the last one, every round consists of SubBytes (SB), ShiftRows (SR), MixColumns (MC) and AddRoundKeys (AK). SubBytes works on each byte of cipher state, which is the only non-linear function. In software implements, T-table is usually used for higher efficiency where SubBytes and MixColumns operations are combined outputting 32-bit states.

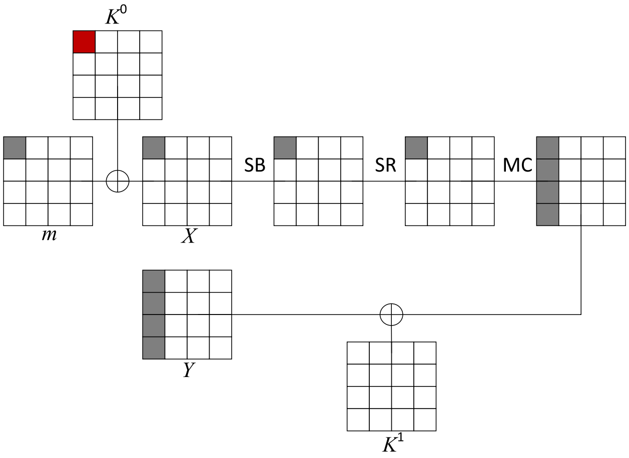


FIGURE 4. First round of AES with whitening key.

Fig. 4 illustrates the first round of AES with whitening key. In T-table software or round based hardware implementations, there is leakage in Y, the output of first round. Taking one byte of Y as target value, the adversary has to guess 5 bytes of key (4 bytes of whitening key and 1 byte of first round key) in CPA procedure which result in high computational complexity. To reduce the guessing space, chosen plaintext technique is applied. The adversary can choose plaintext varying only in one byte e.g. the first byte. Except the first column, other columns in Y are constants. As MixColumns are linear, the first column of Y is expressed as

$$Y_0 = MC\left(S\left(X_{0,0}\right), 0, 0, 0\right) \oplus c,$$

where c is the value of round output. Taking Y_0 as target value, only one byte of whitening key, $K_{0,0}$, need to be guessed. However, as we describe in Section IV-A, c has great effect on the result of CPA. According to Observation 2, we can keep the chosen plaintext attack procedure but use linear regression distinguisher to bypass the effect of c.

C. SIMULATION

We simulate chosen plaintext attack on AES using both CPA and MPA. The simulation is under 8-bit k with noise factor $\sigma = 2$. Both CPA and MPA guess 8-bit k value on AES XORs whitening key. 1000 parallel experiments with random selected key are performed to estimate success rates. As shown in Fig. 5, MPA achieves 100% success rate with about 400 traces. Because the effect of round output as described in Sec. IV-B, the highest success rate of CPA is about 60%, with 800 traces. When CPA failed, the adversary can chose another group of plaintext to change the value of round output and repeat the attack. So on average, CPA reach 100% success rate with 1333 traces. This means, our MPA attack improves more than 300% efficiency compared with CPA.

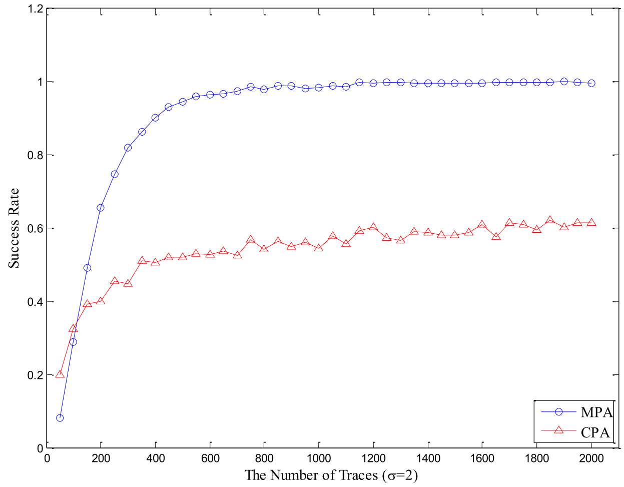


FIGURE 5. Simulation of chosen plaintext attack on AES.

D. EXPERIMENTS ON FPGA

In this section, we test our method in practice using SAKURA-G board, performing round-based AES implementation and acquiring the power consumption with an oscilloscope.

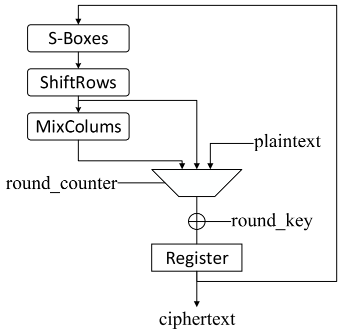


FIGURE 6. Hardware implementation of AES.

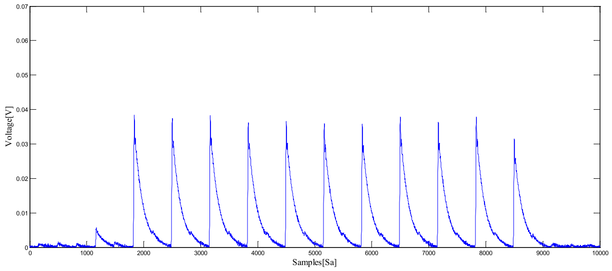


FIGURE 7. Power consumption of one AES encryption.

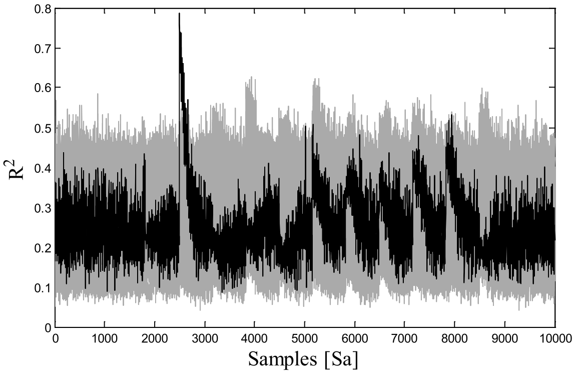


FIGURE 8. Multi-byte power analysis on AES.

1) DESCRIPTION OF AES IMPLEMENTATION

We synthesize the official code on SAKURA-G board to perform AES encryptions. This implementation is a standard paralleled hardware implementation of AES. As shown in Fig. 6, the message is first XORed with the whitening key and stored into the register. Then for every clock cycle, the chip performs one AES round including **SubBytes**, **ShiftRows**,**MixColumns** and **AddRoundKey**. The **MixColumns** step is omitted from the last round. The traces acquired by the oscilloscope is shown in Fig. 7, where we can clearly recognize the pattern of round operation.

2) EXPERIMENTAL RESULTS

We test the effectiveness of MPA with 100 traces on round output. The experimental results are shown in Fig. 8 where grey lines mean the R^2 of the wrong key, the black line means the R^2 of the correct. It is obvious that the correct key has a higher R^2 value.

To show the relation between the number of traces and R^2, the result is illustrated in Fig. 9. Grey lines mean the R^2 of the key candidates, the black line means the R^2 of the correct key. It shows that with only 50 original traces, the correct key can be recovered by MPA method.

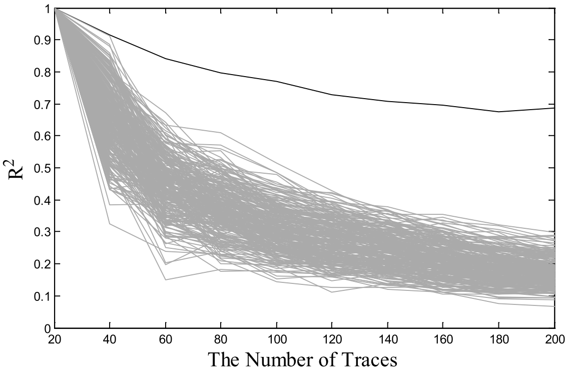


FIGURE 9. **Efficiency of multi-byte power analysis.**

V. CONCLUSION

In this paper, we give another look at linear regression under Hamming weight/ Hamming distance model and propose an innovative Multi-byte Power Analysis method which can be applied to any number of bytes instead of one single byte when performing SCA. We find that MPA with linear regression has great advantages than CPA in two typical cases, recovering keys with XOR operation leakage and chosen plaintext attack on round output of block ciphers in T-table software implementation or round based hardware implementation. For the first case we achieve as high as 400% improvement compared with CPA when recovering 128-bits keys. For the second case, we show that MPA is extremely powerful as it can overcome the problem when performing CPA on round output. Experiments on AES are also given which verify the efficiency of two typical cases in practice. We believe that this characteristic of MPA with linear regression provides a feasible attacking method which could be used in many other cases with further research.

REFERENCES

[1] M.-L. Akkar, R. Bevan, P. Dischamp, and D. Moyart, "Power analysis, what is now possible...," in *Proc. Int. Conf. Theory Appl. Cryptol. Inf. Secur.* Berlin, Germany: Springer, 2000, pp. 489–502.
[2] L. Batina, B. Gierlichs, and K. Lemke-Rust, "Differential cluster analysis," in *Cryptographic Hardware and Embedded Systems*. Berlin, Germany: Springer, 2009, pp. 112–127.
[3] L. Batina, B. Gierlichs, E. Prouff, M. Rivain, F.-X. Standaert, and N. Veyrat-Charvillon, "Mutual information analysis: A comprehensive study," *J. Cryptol.*, vol. 24, no. 2, pp. 269–291, 2011.
[4] A. Bogdanov, "Improved side-channel collision attacks on AES," in *Selected Areas in Cryptography*. Berlin, Germany: Springer, 2007, pp. 84–95.
[5] E. Brier, C. Clavier, and F. Olivier, "Correlation power analysis with a leakage model," in *Cryptographic Hardware and Embedded Systems*. Berlin, Germany: Springer, 2004, pp. 16–29.
[6] S. Chari, J. R. Rao, and P. Rohatgi, "Template attacks," in *Proc. Int. Workshop Cryptograph. Hardw. Embedded Syst.* Berlin, Germany: Springer, 2002, pp. 13–28.
[7] C. Clavier, B. Feix, G. Gagnerot, M. Roussellet, and V. Verneuil, "Improved collision-correlation power analysis on first order protected AES," in *Cryptographic Hardware and Embedded Systems*. Berlin, Germany: Springer, 2011, pp. 49–62.
[8] G. Dabosville, J. Doget, and E. Prouff, "A new second-order side channel attack based on linear regression," *IEEE Trans. Comput.*, vol. 62, no. 8, pp. 1629–1640, Aug. 2013.
[9] J. Doget, E. Prouff, M. Rivain, and F.-X. Standaert, "Univariate side channel attacks and leakage modeling," *J. Cryptogr. Eng.*, vol. 1, no. 2, pp. 123–144, 2011.
[10] B. Gierlichs, L. Batina, P. Tuyls, and B. Preneel, "Mutual information analysis," in *Cryptographic Hardware and Embedded Systems*. Berlin, Germany: Springer, 2008, pp. 426–442.
[11] B. Gierlichs, K. Lemke-Rust, and C. Paar, "Templates vs. stochastic methods," in *Proc. Int. Workshop Cryptograph. Hardw. Embedded Syst.* Berlin, Germany: Springer, 2006, pp. 15–29.
[12] A. Heuser, W. Schindler, and M. Stöttinger, "Revealing side-channel issues of complex circuits by enhanced leakage models," in *Proc. Design, Automat. Test Eur. Conf. Exhib. (DATE)*, Mar. 2012, pp. 1179–1184.
[13] M. Kasper, W. Schindler, and M. Stöttinger, "A stochastic method for security evaluation of cryptographic FPGA implementations," in *Proc. Int. Conf. Field-Program. Technol.*, Dec. 2010, pp. 146–153.
[14] P. Kocher, J. Jaffe, and B. Jun, "Differential power analysis," in *Advances in Cryptology*. Berlin, Germany: Springer, 1999, pp. 388–397.
[15] T.-H. Le, J. Clédière, C. Canovas, B. Robisson, C. Servière, and J.-L. Lacoume, "A proposition for correlation power analysis enhancement," in *Cryptographic Hardware and Embedded Systems*. Berlin, Germany: Springer, 2006, pp. 174–186.
[16] V. Lomné, E. Prouff, and T. Roche, "Behind the scene of side channel attacks," in *Advances in Cryptology—ASIACRYPT* (Lecture Notes in Computer Science), vol. 8269. Berlin, Germany: Springer, 2013, pp. 506–525.
[17] S. Mangard, N. Pramstaller, and E. Oswald, "Successfully attacking masked AES hardware implementations," in *Proc. Int. Workshop Cryptograph. Hardw. Embedded Syst.* Berlin, Germany: Springer, 2005, pp. 157–171.
[18] T. S. Messerges, E. A. Dabbish, and R. H. Sloan, "Investigations of power analysis attacks on smartcards," in *Proc. USENIX Workshop Smartcard Technol.*, 1999, p. 17.
[19] A. Moradi, O. Mischke, and T. Eisenbarth, "Correlation-enhanced power analysis collision attack," in *Cryptographic Hardware and Embedded Systems*. Berlin, Germany: Springer, 2010, pp. 125–139.
[20] E. Peeters, F. X. Standaert, and J. J. Quisquater, "Power and electromagnetic analysis: Improved model, consequences and comparisons," *Integr. VLSI J.*, vol. 40, no. 1, pp. 52–60, Jan. 2007.
[21] V. Rijmen and J. Daemen, "Advanced encryption standard," Fed. Inf. Process. Standards Publ., Nat. Inst. Standards Technol., Gaithersburg, MD, USA, Tech. Rep., 2001, pp. 19–22.
[22] W. Schindler, K. Lemke, and C. Paar, "A stochastic model for differential side channel cryptanalysis," in *Cryptographic Hardware and Embedded Systems* (Lecture Notes in Computer Science), vol. 3659. Berlin, Germany: Springer, 2005, pp. 30–46.
[23] K. Schramm, G. Leander, P. Felke, and C. Paar, "A collision-attack on AES," in *Proc. Int. Workshop Cryptograph. Hardw. Embedded Syst.* Berlin, Germany: Springer, 2004, pp. 163–175.
[24] C. Tu, L. Zhang, Z. Liu, N. Gao, and Y. Ma, "A practical chosen message power analysis approach against ciphers with the key whitening layers," in *Int. Conf. Appl. Cryptogr. Netw. Secur.* Cham, Switzerland: Springer, 2017, pp. 415–434.
[25] N. Veyrat-Charvillon and F.-X. Standaert, "Adaptive chosen-message side-channel attacks," in *Proc. Int. Conf. Appl. Cryptogr. Netw. Secur.*, 2010, pp. 186–199.
[26] C. Whitnall, E. Oswald, and F. X. Standaert, "The myth of generic DPA... and the magic of learning," in *Topics in Cryptology* (Lecture Notes in Computer Science), vol. 8366. Cham, Switzerland: Springer, 2014, pp. 183–205.

SHAN FU received the B.E. degree in automation engineering from the Beijing University of Posts and Telecommunications, Beijing, China, in 2011, and the M.S. degree in electrical engineering from Columbia University, New York, NY, USA, in 2013. She is currently pursuing the Ph.D. degree in information security with the Beijing University of Posts and Telecommunications.

Since 2013, she has been an Information Security Researcher with the China Academy of Information and Communications Technology, Beijing. Her research interests include the side channel analysis on mobile terminals, and IoT devices and their biometric scenarios.

ZONGYUE WANG received the Ph.D. degree in information security from Shandong University in 2015. From 2015 to 2017, he was a Research Engineer with the China Academy of Information and Communications Technology. He is currently at Open Security Research, Shenzhen, China. His main research interests include side channel analysis and cryptographic engineering.

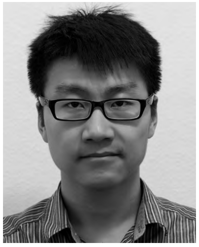
GUOAI XU received the Ph.D. degree in signal and information processing from the Beijing University of Posts and Telecommunications, China, in 2002. He was a recipient of the title of Professor in 2011. He is currently an Associate Director with the National Engineering Laboratory of Security Technology for Mobile Internet, School of Cyberspace Security, Beijing University of Posts and Telecommunications. His research interests include software security and data analysis.

FANXING WEI received the M.S. degree in computer technology from the Beijing Institute of Technology in 2017. She is currently with the China Academy of Information and Communications Technology. Her research interests include side channel analysis and fault injection.

AN WANG was born in 1983. He received the Ph.D. degree from Shandong University in 2011. From 2011 to 2015, he held a postdoctoral position at Tsinghua University. He is currently with the Beijing Institute of Technology. His main research interests include side-channel analysis, embedded systems, and cryptographic implementation.

JUAN PAN received the master's degree from the Communication School, Beijing University of Posts and Telecommunications, in 2003. She is currently the Director of information security with the Department of CTTL-Terminals, China Academy of Information and Communications Technology. Her research interests include mobile Internet security and data protection.

YUGUANG LI received the master's degree from the School of Cyber Science and Engineering, Wuhan University, China, in 2018. He is current a Research Engineer with the China Academy of Information and Communications Technology, Beijing. His research interests include cryptography and side channel analysis.



NING ZHANG received the B.E. degree from the Telecommunication School, Tongji University, in 2011, and the M.S. degree in electrical engineering from Columbia University in 2013. He is currently a Research Engineer with the Institute of Information Engineering, Chinese Academy of Sciences.