Evaluation Metrics of Physical Non-invasive Security

Huiyun Li, Keke Wu, Fengqi Yu, Hai Yuan

Evaluation Metrics of Physical Non-Invasive Security

Huiyun Li, Keke Wu, Fengqi Yu, and Hai Yuan

Department of Integrated Electronics
Shenzhen Institute of Advanced Technology
The Chinese University of Hong Kong, Chinese Academy of Sciences, China
hy.li@siat.ac.cn

Abstract. Physical non-invasive security has become crucial for cryptographic modules, which are widely used in pervasive computing. International security evaluation standards, such as U.S. Federal Information Processing Standard (FIPS) 140-3 and Common Criteria (CC) part 3 have added special requirements addressing physical non-invasive security. However, these evaluation standards lack of quantitative metrics to explicitly guide the design and measurement. This paper proposes practice-oriented quantitative evaluation metrics, in which the distinguishability between the key predictions is measured under statistical significance tests. Significant distinguishability between the most possible two key candidates suggests high success rates of the right key prediction, thus indicates a low security degree. The quantitative evaluation results provide high accountability of security performance. The accordance with FIPS 140-3 makes the proposed evaluation metrics a valuable complement to these widely adopted standards. Case studies on various smart cards demonstrate that the proposed evaluation metrics are accurate and feasible.

1 Introduction

Pervasive Computing is an emerging technology that harmonizes numerous networked devices at all scales throughout everyday life. Cryptographic modules are widely used in pervasive applications to provide security services such as confidentiality, integrity, and authentication. Modern cryptography algorithms are usually used to provide security to cryptographic modules, and are extremely robust against traditional black-box cryptanalysis attacks, such as brute force and factoring. Intelligent adversaries turn to focus their efforts on more subtle and complex attacks: physical non-invasive attacks, which exploit the correlations between the physical leakage (timing, power consumption, electromagnetic emission etc) information of the target cryptographic module and the internally used secret key. Since this correlation can be exploited with relatively cheap equipment, such as an oscilloscope and a few electromagnetic sensors, physical non-invasive attacks pose a serious threat to cryptographic modules.

So far, physical non-invasive attacks have successfully broken the hardware or software implementations of many cryptographic systems including block ciphers (such as DES, AES, Camellia, IDEA etc), stream ciphers (such as RC4, RC6, A5/1, SOBER-t32 etc), public key ciphers (such as RSA, ElGamal, ECC, XTR etc), and also the implementations of signature schemes, the MAC schemes etc [1].

Enormous research efforts have been devoted to countermeasures against physical non-invasive attacks. However, the effectiveness of these countermeasures was generally evaluated qualitatively and just contained case studies showing that the proposed countermeasures really increased the number of samples, compared to the version without any countermeasures.

There have only been a few attempts to quantitatively evaluate the physical non-invasive security. An example is paper [2], which defines the notion of physical computer that is the combination of an abstract computer (i.e. a Turing machine) and a leakage function. Another example is paper [3], which takes information theoretic conditional entropy into account. However, the models in [2, 3] are too general to be applied to specific practice. An open question is to meaningfully restrict the models to realistic adversaries or evaluators.

Paper [4] brought forward a tentative quantitative approach to evaluate the countermeasures by estimating the needed number of samples. The number of samples was deduced from signal-to-noise ratio (SNR). However, noise is hard to measure separately from the side-channel information on real products. Thus the approach in [4] is of limited use within estimation through simulation in stead of actual evaluation.

Some guiding standards and good experiments of security metrics exist, such as FIPS 140-32007FIPS [5] and Common Criteria (CC) [6]. The CC has seven levels of assurance: EAL1: Functionally Tested; EAL2: Structurally Tested; EAL3: Methodically Tested and Checked; EAL4: Methodically designed, Tested, and Revised; EAL5: Semi-formally Designed and Tested; EAL6: Semi-formally Verified Design and Tested; and EAL7: Formally Verified Design and Tested. Nevertheless, the level of trust of various methodologies is a qualitative indicator by nature. There are no mathematical formulas to be applied to obtain the level of trust as a value of such an indicator [7].

The Federal Information Processing Standard FIPS 140-3 [5, 8] specifies five increasing levels of security requirements, including the requirements on physical non-invasive security, as shown in Fig. 1. Security Level 1 requires minimum physical protection. Level 2 requires the addition of tamper-evident mechanisms such as a seal or enclosure. Level 3 specifies stronger detection and response mechanisms, and requires mechanisms against timing analysis attacks. Level 4 requires highly rigorous design processes and mechanisms against simple power analysis (SPA) and differential power analysis (DPA) attacks. Finally, Level 5 mandates mechanisms with environmental failure protection and electromagnetic emission analysis (EMA) attack countermeasures. FIPS 140-3 superseded FIPS 140-2 with emphasis on physical non-invasive security. However, there is

still lack of quantifiable metrics to explicitly guide the design and evaluation of cryptographic modules.

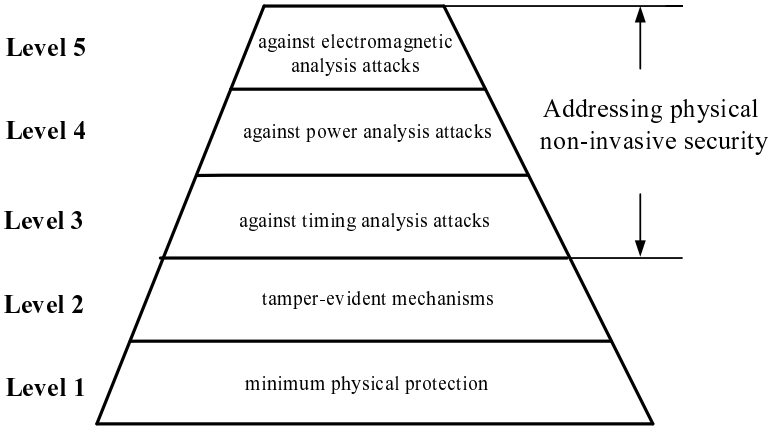

Fig. 1. Physical non-invasive security addressed by FIPS 140-3

In this paper, we for the first time, propose practice-oriented evaluation metrics to assess the physical non-invasive security. We classify the cryptographic modules into three levels, in accordance with the widely adopted FIPS 140-3 standard. In each level, the distinguishability between the key predictions is assessed under significance test. For FIPS 140-3 Level 3 timing analysis security, and Level 4 SPA security, the shape similarity of timing\power patterns at processing binary bits "0" or "1") are assessed. While for DPA security, the power magnitude confusion is measured. For FIPS 140-3 Level 5, the electromagnetic emission patterns are assessed, where simple electromagnetic analysis (SEMA) security and differential electromagnetic analysis (DEMA) security are assessed similarly as SPA and DPA respectively.

2 Understanding Physical non-invasive Attacks

Non-invasive attacks refer to attacks that exploit the implementation of target devices and identify properties of the implementation without physically damaging the target devices. These attacks can be performed relatively quickly and easily, while leaving no evidence of tampering, hence they are of particular concern to the security field. There are many forms of non-invasive attacks such as timing attacks, fault induction techniques, power and electromagnetic analysis based attacks, and so on. The following sections provide a brief introduction.

2.1 Timing Analysis

Timing Analysis attacks rely on precisely measuring the time at performing specific mathematical operations associated with a cryptographic algorithm or process. The collected timing information (often via power consumption) is analyzed to determine the relationship between the inputs and the cryptographic keys used by the underlying algorithms or processes. The analysis of the relationship may be used to exploit the timing measurements to reveal the cryptographic key [9].

Making all computations take exactly the same amount of time would eliminate the attack, but few programs operate in exactly constant time. Writing constant-time code (particularly in high-level languages) can be difficult [8]. The effectiveness of the time analysis security should be evaluated with quantitative metrics.

2.2 Power Analysis

Attacks based on the analysis of power consumption can be divided into two general categories, Simple Power Analysis (SPA) and Differential Power Analysis (DPA). SPA involves a direct (primarily visual) analysis of electrical power consumption patterns and/or timings derived from the execution of individual instructions carried out by a cryptographic module during a cryptographic process. The patterns are obtained through monitoring the variations in electrical power consumption for the purpose of revealing the features and implementations of cryptographic algorithms and subsequently values of cryptographic keys [9]. SPA and timing analysis intersects when timing information leaked via power consumption is examined. SPA also involves the situations where power consumption samples of various operations have constant time but different amplitude patterns. DPA utilizes statistical techniques to analyze the variations of the electrical power consumption of a cryptographic module.

Since SPA and DPA attacks exploit more advanced analysis techniques than timing analysis, power analysis security is often regarded as a higher level requirement in design and measurement of cryptographic modules.

Countermeasures against power analysis attacks attempt to keep operation-independent and data-independent in terms of power consumption, e.g., through software, by removing conditional branches [10] and/or adding random mask [11] etc. The countermeasures are also viable through hardware, by adding random delay or exploit balanced logic styles [12, 13] etc. The effectiveness of these countermeasures varies with realistic implementations. Quantifiable metrics are demanded in terms of security evaluation.

2.3 Electromagnetic Analysis

The cryptographic module under attack emits different amounts of electromagnetic (EM) emission depending on the instructions and data being executed. EM energy is closely correlated to power consumption but may be localized into

a smaller area. If the global current is like a river, the EM emission is then produced by streams that flow into the river.

EM emissions can be categorized into two types: direct emissions and modulated emissions [14, 15]. Direct emissions are caused directly by current flow with sharp rising/falling edges. To measure direct emissions from a signal source isolated from interference from other signal sources, one uses tiny field probes positioned very close to the signal source and special filters to minimize interference. Modulated emissions occur when a data signal modulates carrier signals which then generate EM emissions propagating into the space. A strong source of carrier signals are the harmonic-rich square-wave signals such as a clock, which may then be modulated in amplitude, phase or some other manner. The recovery of the data signals requires a receiver tuned to the carrier frequency with a corresponding demodulator.

In some cases when the global power measurement becomes useless, local EM emission may convey important information [16]. Therefore, EM analysis security is regarded as a higher level than timing analysis and power analysis security.

2.4 Fault Induction

Fault induction attacks utilize external forces such as microwaves, temperature extremes, and voltage manipulation to cause processing errors in a predictable and useful way for attackers. External glitches inserted on the power or clock line are examples of non-invasive fault induction attacks. There are other fault induction attacks that cause some damage to the chip, falling into the category of invasive or semi-invasive attacks. Many chips nowadays are designed to resist fault induction attacks by having voltage/temperature fluctuation sensors, and the effectiveness is straightforward to evaluate. Thus fault induction security evaluation is not considered in this paper.

3 Quantitative Evaluation Metrics of non-invasive Security

As discussed in previous sections, quantitative evaluation metrics for the physical non-invasive security are demanded to fill the gap between evaluation and attacks/countermeasures technology. We propose practice-oriented evaluation metrics in this section to quantitatively evaluate the physical non-invasive security.

Fig.2 demonstrates the flow chart of the evaluation procedure. First, we classify the target cryptographic modules into three levels, in accordance with FIPS 140-3 standard. The higher level should cover evaluation contents of the lower level(s). For example, level 3 covers timing analysis security evaluation, and level 4 covers timing analysis and power analysis security evaluation, while level 5 will cover timing analysis, power analysis and EM analysis security evaluation.

Second, we define the prediction function $f_{prediction}$ for the key guesses. Different cryptographic primitives correspond to different prediction functions, such as correlation coefficients, operation time mean values etc. The details to choose prediction functions are elaborated later on. After applying the prediction function on the target, we obtain the key guess vector$g = [g_1, g_2, ..., g_S]$, which denotes the key candidates sorted accordingly to their likelihood, where S denotes the key space. The highest possible candidate is ranked first.

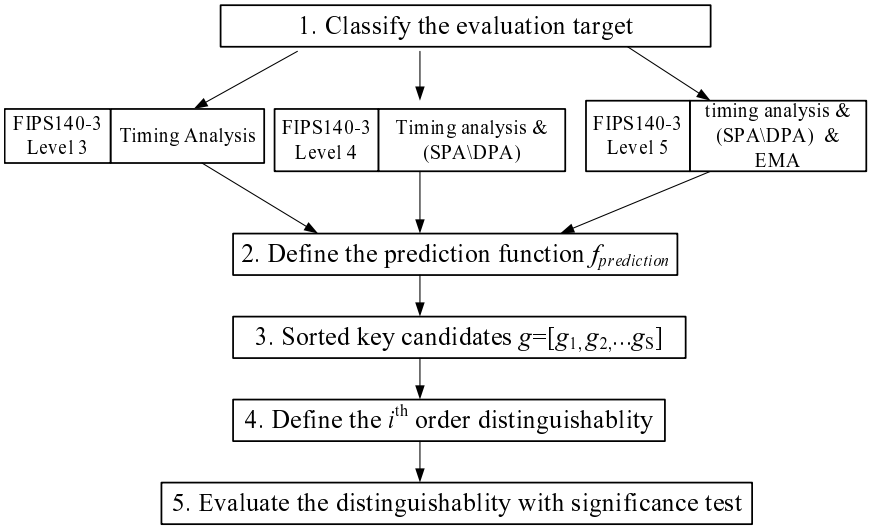

Fig. 2. Quantitative evaluation procedure for physical non-invasive security

Next, we define the ith order distinguishability $Dist^i(g_1, g_2, ...g_{i+1})$ between the first i+1 key candidates. If not specified, a first order distinguishablity between the first two key candidates is assumed. Finally, the distinguishablity is obtained through statistical analysis, and has to be evaluated with significance test under a chosen confidence degree.

Many significance tests exist and can be applied into our evaluation metrics, such as the distance of means, goodness of fit and sum of ranks [17, 18]. For the sake of simplicity, we only demonstrate the distance of means tests in this paper. However, other custom procedures are also applicable. The evaluators can select specific form suitable to the experimental environment.

3.1 Assessing the timing analysis security

The prediction functions in the proposed evaluation methodology are different according to the types of physical non-invasive leakage. In evaluating timing analysis security, we assess the distinguishability of physical non-invasive information between processing bit “0” and processing bit “1”.

For timing information leakage vector of n experimental samples $T_{bit\_0} = (t_{bit0\_1}, t_{bit0\_2}, ..., t_{bit0\_n})$ obtained when the device under test is processing binary bit "0", and m experimental samples: $T_{bit\_1} = (t_{bit1\_1}, t_{bit1\_2}, ..., t_{bit1\_m})$ obtained when the device is processing binary bit "1", we examine the operation time vectors T_{bit_0} and T_{bit_1}. If n and m are sufficiently large, by virtue of the central limit theorem, the probability distribution of the two variables (t_{bit0_i}, t_{bit1_i}), are approximately Gaussian. The mean values can be chosen as the prediction function for timing analysis security evaluation. The smaller value corresponds to bit "0", and the larger value for bit "1" due to the conditional branch operations. The distinguishability $Dist(bit0, bit1)$ between the two mean values should be evaluated with significance test.

The significance test of difference between the mean values can thus be calculated as equation (1) [18].

$$\varepsilon = \frac{|\mu_0 - \mu_1|}{\sqrt{\frac{{s_0}^2}{n} - \frac{{s_1}^2}{m}}} \tag{1}$$

where μ_0, μ_1denote the expectations of t_{bit0_i}, t_{bit1_i}, s_0, s_1denote of the standard deviation of t_{bit0_i}, t_{bit1_i}; n and m denote the number of samples of t_{bit0_i}, t_{bit1_i} respectively.

The significance test evaluates the probability that the two samples have the same mean value. If the test result turns out that the difference is significantly larger than the critical value under a certain confidence degree α(normally chosen 5%, adjustable in given situation), we judge the two groups have significant different means with the degree of reliability at 1-α.

The difference of means thus measures the distinguishability of timing information between processing bit "0" and processing bit "1". If the difference between is larger than a critical value, the larger the difference, the easier to discern the key bit and therefore less secure of the cryptographic device. If the difference between is less than a critical value, then the difference is statistically insignificant. The less the difference is, the more difficult to discern the right guess, and therefore more secure of the cryptographic device under test.

3.2 Assessing the power analysis security

The power analysis attacks are generally divided into SPA and DPA. SPA involves pattern recognitions and DPA involves statistic analysis. The evaluation metrics are accordingly classified.

– SPA

For n power samples when the device under test is processing binary bit "0", $P_{1...n,1...T|bit0}$, as shown in Fig.3 (a), where T is the number of points that are recorded per trace, and n power samples when the device under test is processing binary bit "1", $P_{1...n,1...T|bit1}$, as shown in Fig.3(b) we examine the shape similarity between the two vectors. There are numerous shape similarity models

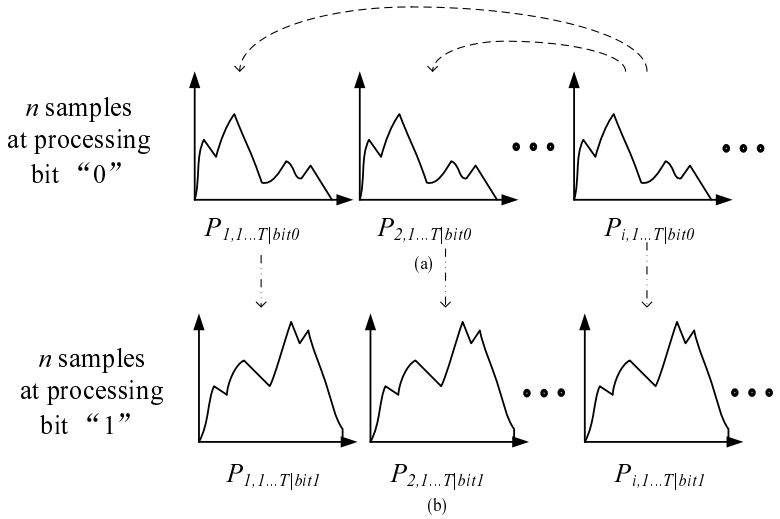

Fig. 3. Compare the similarity between the power patterns of processing bit "0" and bit "1"

based on different feature factors, such as Euclidean distance, area, circularity, major axis orientation, and a set of algebraic moments [19].

Taking the Euclidean distance model as the example, we calculate the distances between ith power sample and the rest n-1 samples at processing bit "0", as shown in the dash curves in Fig.3 (a). The distance between the i^{th} power sample and the x^{th} power sample $d_{i\_x|bit0}$ is calculated as in equation (2).

$$d_{i\_x|bit0} = \sum_{j=1}^{T} \left| P_{i,j|bit0} - P_{x,j|bit0} \right| \tag{2}$$

If n is sufficiently large, the probability distribution of the distances are approximately Gaussian. The mean values can be chosen as the prediction function for timing analysis security evaluation. The mean value μ_{bit0} of self-similarity for bit "0" is calculated as in equation (3).

$$\mu_{bit0} = \frac{\sum_{x=1}^{n} \sum_{j=1}^{T} \left| P_{i,j|bit0} - P_{x,j|bit0} \right|}{n-1} \tag{3}$$

Similarly we calculate the Euclidean distances between the n power samples at processing bit "0" and the n power samples at processing bit "1", as shown in the dash-dot lines in Fig.3(b), and obtain the mean value μ_{bit0-1} as shown in equation (4).

$$\mu_{bit0-1} = \frac{\sum_{x=1}^{n} \sum_{j=1}^{T} \left| P_{i,j|bit0} - P_{x,j|bit1} \right|}{n} \tag{4}$$

Then, the distinguishability $Dist(bit0, bit1)$ between the two mean values should be evaluated with significance test.

If the test result turns out that the difference is significantly larger than the critical value under a certain confidence degree α(normally chosen 5%, adjustable in given situation), we judge the two groups have significant different means with the degree of reliability at 1- α.

– DPA

In the different power analysis attacks, for $n$power traces $P_{1...n,1...T}$, the attacker hypothesizes a key and calculates the correlation factor ρ_{WH} [4] between power and intermediate data for each point at time t_i:

$$\rho_{WH} = \frac{E(W \cdot H) - E(W) \cdot E(H)}{\sqrt{D(W)} \cdot \sqrt{D(H)}} \tag{5}$$

where W denotes power consumption, and H denotes Hamming weight, $E()$ denotes the expectation, and $D()$ denotes the variance.

Thus he obtains correlation traces $\rho_{1...T|i}$ for each key guess key_i, where i is between 1 and the number of all possible keys k_{all}. Correlation coefficient $\rho_{1...T|i}$ is chosen to be the prediction function. There is a highest ρ_{ti} in each correlation trace, and the largest ρ_{max} of all ρ_{ti} indicates the correct key guess. The second highest ρ_{2nd_max} indicates the second possible key [20].

Effective countermeasures often make ρ_{max} hard to discern, i.e., the highest two correlation factors ρ_{max} and ρ_{2nd_max} out of two key guesses have similar values. Thus the distinguishability $Dist(\rho_{max}, \rho_{2nd\_max})$ is assessed with significance test.

We test the statistical significance of the difference between ρ_{max} and ρ_{2nd_max} under a certain confidence degree (normally chosen 95%). Since the sampling distribution of correlation factor ρ may not be normal, and is better transformed through Fisher's Z-Transformation given in equation (6).

$$z_\rho = \frac{1}{2} \ln \frac{1+\rho}{1-\rho} \tag{6}$$

The statistic z_ρ has an approximate normal distribution with variance ${z_\rho}^2 = \frac{1}{n-3}$, where n is the number of samples.

The procedure to assess the statistical significance of the difference between ρ_{max} and ρ_{2nd_max} is as following. The first step is to convert the highest correlation factors ρ_{max} and ρ_{2nd_max} to z_{max} and z_{2nd_max} through equation (6). Then their normalized difference is calculated as shown in equation (7):

$$\triangle z = \frac{z_{max} - z_{2nd\_max}}{\sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}}} \tag{7}$$

where n $_1$ and n $_2$ are the number of samples to get ρ_{max} and ρ_{2nd_max} respectively. In our test, ρ_{max} and ρ_{2nd_max} usually have the same sampling numbers, i.e., $n_1 = n_2$.

Once the difference between ρ_{max} and ρ_{2nd_max} under Z-transformation has been obtained, their difference can be assessed with statistical significance under a given confidence degree.

3.3 Assessing the electromagnetic analysis security

There are generally two types of emissions in EM analysis attacks: direct emissions and modulated emissions. As to each type of the emissions, the analysis procedure is further divided into Simple Electromagnetic Analysis (SEMA) and Differential Electromagnetic Analysis (DEMA), as shown in Fig.4. Their evaluation metrics are similar to those in SPA and DPA respectively.

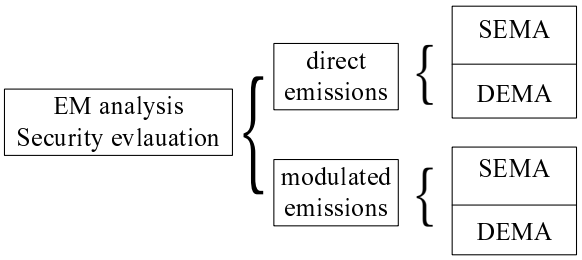

Fig. 4. EM analysis security evaluation procedure

4 Experiment Results

To verify the proposed methodology on evaluating physical non-invasive security, we performed case studies on evaluation of timing analysis security and differential power analysis security. The experimental setup is shown in Fig.5. The setup for timing analysis via power consumption is the same as that of power analysis. Electromagnetic analysis set up is also similar except the oscilloscope will collect emission through an EM probe placed near the device rather than through the resistor.

The experiment procedure is as follows: First, a PC generates random plain texts and instructs the cryptographic device under test to start cryptographic operations through the PC/Device interface. Second, the PC/Device interface, which contains a hardware trigger, will send a trigger signal to instruct a digital oscilloscope to collect the power consumption (or Electromagnetic emission) traces of the cryptographic device during the encryption operation. Third, the PC receives the sample traces from the oscilloscope along with the plain texts for each encryption operation. Finally the PC performs evaluation with the proposed quantitative metrics.

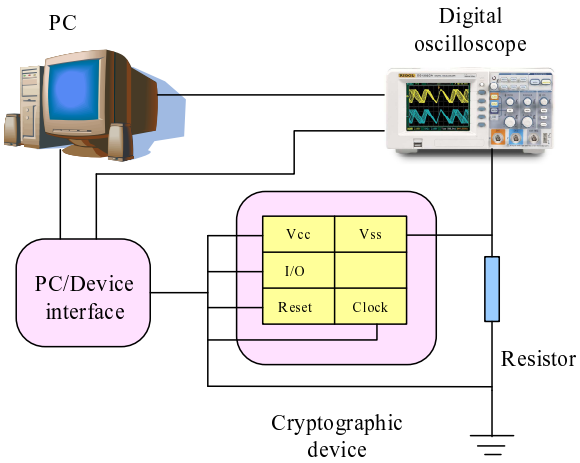

Fig. 5. Power analysis setup

Timing analysis attacks have been demonstrated to be very powerful against most straightforward implementations of public key ciphers. The modular exponentiation in RSA-type ciphers and scalar multiplication in Elliptic Curve Cryptosystems (ECC)-type ciphers are especially prone to timing analysis attacks. We perform experiments on ECC implemented cryptographic devices to illustrate the accuracy and feasibility of the proposed evaluation metrics.

An elliptic curve is a set of points (x, y) satisfying a bivariate cubic equation over a finite field F [10]. The operation of adding a point P to itself k times is called scalar multiplication by k and denoted as kP. Scalar multiplication is the basic key-involved operation for ECC, thus the main target in side-channel attacks.

The most straightforward implementation of scalar multiplication is the binary method [10] based on the binary expansion of $k = (k_{n-1}, ...k_1, k_0)_2$where $n-1$ is the most significant bit of k. Table 1 illustrates the operation of the scalar multiplication with the binary method.

Table 1. Scalar Multiplication – Binary Method [10]

Input: $\boldsymbol{k = (k_{n-1}, \ldots, k_1, k_0)_2}$, $\boldsymbol{P \in E(F_p)}$ $\boldsymbol{(k_{n-1} = 1)}$. **Output:** $\boldsymbol{kP}$.
1. $Q = P$ 2. For i from $n-2$ to 0 do 2.1 $Q = 2\ Q$ 2.2 If k_i =1 then $Q = Q + P$ 3. Return (Q)

Notice that the conditional branches containing a point addition $(Q+P)$ only happen when a bit representation of k is "1". If the sequence of field operations of point addition $(Q+P)$ has a different operation time than that of point doubling $(2Q)$, the key bit can be easily deduced through timing analysis attacks. Thus a common group of countermeasures is to make the addition and the doubling operations indistinguishable, either by means of inserting dummy instructions or operations [21], or by unifying the addition formulae [22, 23].

In our experiments, the devices under test are two smartcards running the public key cryptography ECC. Both cards are 8-bit microprocessors. One card (called device A) has been deployed with the most basic binary algorithm. Another card (called device B) has been deployed with dummy instructions insertion as the security enhancement. For each card, we collected 300 power traces when the device was running the point doubling and point addition.

4.1 Evaluation of timing analysis security of device A

Fig.6 demonstrates the power consumption patterns for device A. It is clearly shown that the operation of point addition is distinct from that of point doubling. The operations of point doubling lasts 3.25 ms, and the operations of addition lasts 3.68 ms, noticeably longer than doubling.

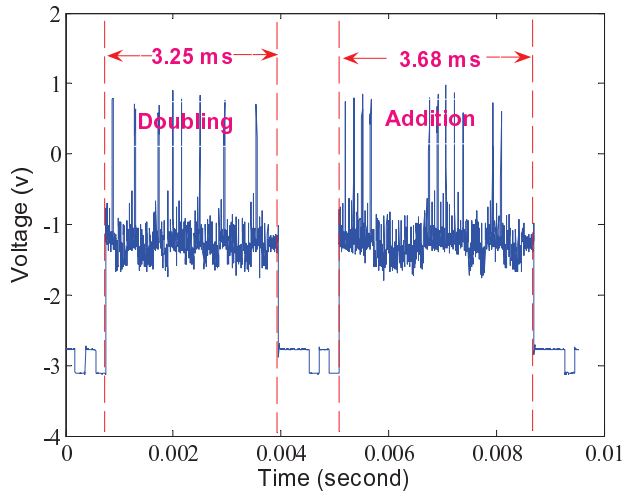

Fig. 6. Distinguishable power trace of doubling and addition on device A

Statistical experiment results are demonstrated in Fig.7. The mean value of operation time in 300 runs is chosen to be the prediction function. Fig.7 (a) shows the distribution of 300 operation time of point additions, and the superimposed red curves indicate the normal distribution fitting. The mean value is $\mu_A = 0.0036$, and standard deviation is $s_A = 1.27e^{-4}$. Fig.7 (b) shows the distribution

of 300 operation time of point doublings, with mean value $\mu_D = 0.0032$, and standard deviation $s_D = 8.47e^{-5}$.

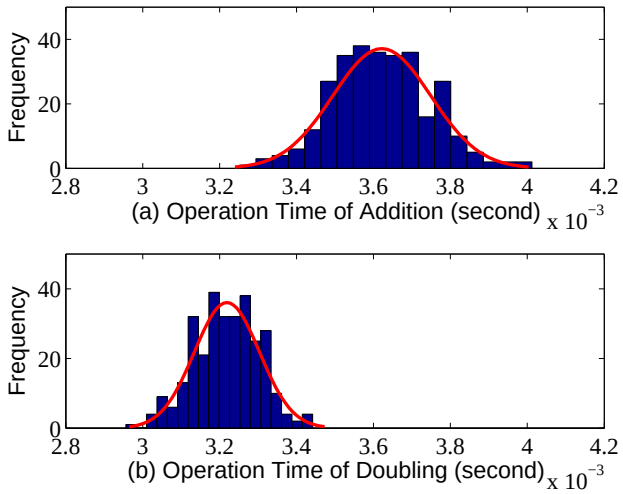

Fig. 7. The operation time distributions of device A at processing point additions and point doublings

The normalized difference of mean between μ_A and $\mu_D$is 45.6 according to equation (8), which is the final quantitative evaluation result. The result is much larger than the critical value 1.96, turning down the hypothesis that the two mean values are same. In practice, device A is vulnerable against timing analysis (as well as simple power analysis). The secret key bits "0101" can be easily read out, where "D" indicates a point doubling operation and "A" indicating a point addition. Whenever the "A" appears, key bit "1" is processed.

We continue to evaluate device B with deployment of the binary balanced method through dummy operation insertion. Fig.8 demonstrates the power consumption patterns of point addition and point doubling for device B. The two power patterns are almost indistinguishable. The operation of point doubling lasts 4.83 ms, and the operation of addition 4.86 ms.

4.2 Evaluation of timing analysis security of device B

Fig.9 demonstrates the experiment results of distribution of operation time when the device B is processing point additions and point doublings. Fig.9 (a) shows the distribution of the operation time of processing point additions, with mean value $\mu_A = 0.00479$, and standard deviation $s_A = 1.56e^{-5}$. Fig.9 (b) shows the distribution of the operation time of processing point doublings with mean value $\mu_D = 0.00477$, and standard deviation $s_D = 1.08e^{-5}$.

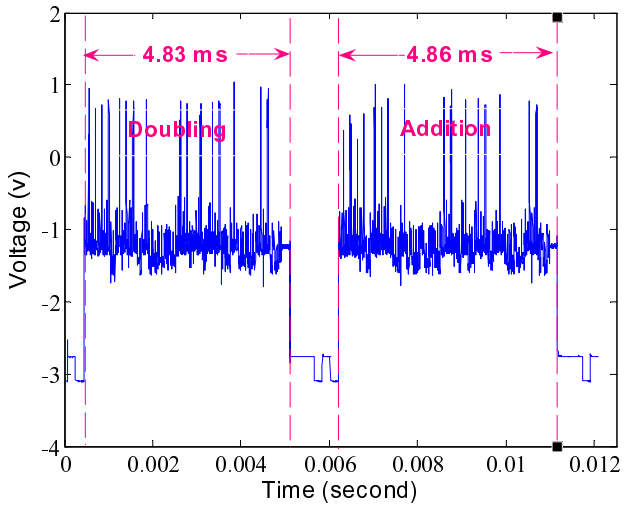

Fig. 8. Indistinguishable power trace of doubling and addition on device B

The normalized difference of mean between μ_A and μ_D is calculated as 1.62, which is the final quantitative evaluation result. The normalized difference is less than the critical value, indicating the difference between point additions and doublings is statistically insignificant. The experiment result on device B conforms to the fact that device B is secure against timing analysis attacks.

Experiment of evaluating timing analysis security verifies that the security of device A can be easily compromised in timing analysis. The normalized difference of mean between μ_A and μ_D is 45.6, demonstrating noticeable difference. While device B has much better security, in that the normalized difference of mean between μ_A and μ_D is indifferent statistically. The evaluation results conform to the fact and provide quantitative assurance.

The devices are assessed back to back, without the necessity of an extra reference card. Elimination of requirement of any references makes the evaluations applicable in objective assessment of various devices by independent designers or evaluators. The accordance with FIPS 140-3 makes the evaluation metrics a valuable complement to these widely adopted standards.

5 Conclusion

Physical non-invasive security has become crucial to pervasive computing. However, there is a noticeable gap between the attacks/countermeasure and the evaluation technology. The existing evaluation certifications and standards are usually qualitative and lack of practice-oriented guidance.

This paper, for the first time, presents a generic evaluation methodology to quantitatively evaluate physical non-invasive security in accordance with FIPS 140-3 standards. Effective quantitative evaluation metrics are further proposed,

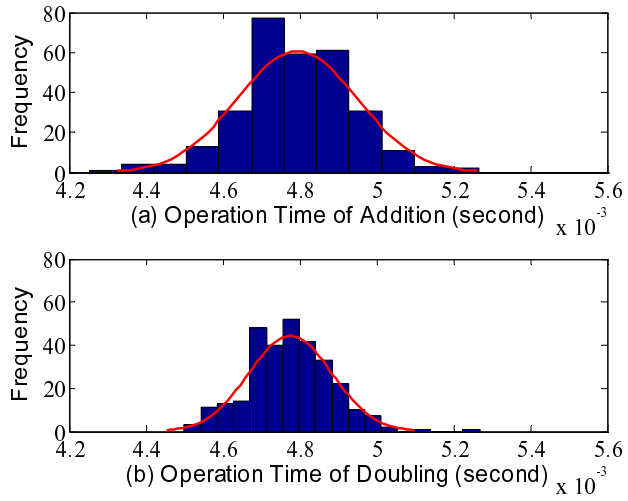

Fig. 9. The operation time distributions of device D at processing point additions and point doublings

in which the distinguishability between the key predictions is measured under statistical significance tests. The quantitative evaluation results provide high accountability of security performance and are applicable in independent evaluations. Case studies on various smart cards demonstrate that the proposed evaluation metrics are highly feasible.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (Grant No. 60901052), and the Basic Research Project of Shenzhen (Grant No. JC200903160412A).

References

1. Zhou, Y., Feng, D.: Side-channel attacks: Ten years after its publication and the impact on cryptographic module security testing. In: Information Security Seminar. (2006)
2. Micali, S., Reyzin, L.: Physically observable cryptography. In: Theory of Cryptography Conference (TCC), Springer (2004) 278–296
3. Standaert, F., Malkin, T., Yung, M.: A unified framework for the analysis of side-channel key recovery attacks. In: Cryptology ePrint Archive, Report 2006/139. (2008)
4. Mangard, S.: Hardware countermeasures against DPA - A statistical analysis of their effectiveness. In: Proceedings of the RSA Conference. (2004)
5. Security requirements for cryptographic modules, FIPS PUB 140-3 (draft), by National Institute of Standards and Technology (2007)

6. Common Criteria for information technology security evaluation, Part III: Security assurance requirements, by National Institute of Standards and Technology (1999)
7. Wang, A.: Information security models and metrics. In: Proceedings of the 43rd ACM annual Southeast regional conference. (2005)
8. Ravi, S., Raghunathan, A., Kocher, P., Hattangady, S.: Security in embedded systems: Design challenges. ACM Transactions on Embedded Computing Systems **3**(3) (2004) 461 – 491
9. Security requirements for cryptographic modules, FIPS PUB 140-2, by National Institute of Standards and Technology (2001)
10. Coron, J.: Resistance against differential power analysis for elliptic curve cryptosystems. In: Proceedings of Cryptographic Hardware and Embedded Systems - CHES99, LNCS 1717 (1999) 292–302
11. Hasan, M.: Power analysis attacks and algorithmic approaches to their countermeasures for koblitz curve cryptosystems. IEEE Transactions on Computers **50** (2001) 1071–1083
12. Tiri, K., Akmal, M., Verbauwhede, I.: A dynamic and differential cmos logic with signal independent power consumption to withstand differential power analysis on smart cards. In: IEEE 28th European Solid-state Circuit Conference (ESSCIRC). (2002)
13. Moore, S., Anderson, R., Cunningham, P., Mullins, R., Taylor, G.: Improving smart card security using self-timed circuits. In: 8th IEEE International Symposium on Asynchronous Circuits and Systems (Async), IEEE Computer Society Press (2002)
14. Agrawal, D., Archambeault, B., Rao, J., Rohatgi, P.: The em side-channel(s). In: Cryptographic Hardware and Embedded Systems (CHES). (2002) 29–45
15. Li, H., Markettos, T., Moore., S.: Security evaluation against electromagnetic analysis at design time. In: Cryptographic Hardware and Embedded Systems (CHES), LNCS 3659 (2005) 280–292
16. Gandolfi, K., Mourtel, C., Olivier, F.: Electromagnetic analysis: Concrete results. In: Proceedings of Cryptographic Hardware and Embedded Systems (CHES). (2001) 251–261
17. Coron, J.S., Kocher, P., Naccache., D.: Statistics and secret leakage. In: Financial Cryptography (FC), LNCS 1972 (2000) 157–173
18. Miller, I., Freund, J., Johnson, R.: Probability and statistics for enginners. Prentice Hill (1990)
19. Berretti, S., Bimbo, A.D., Pala, P.: Retrieval by shape similarity with perceptual distance and effective indexing. IEEE Transactions on Multimedia **2**(4) (2000) 225–239
20. Li, H., Chen, T., Wu, K., Yu, F.: Quantitative evaluation of side-channel security. In: Asia-Pacific Conference on Information Processing (APCIP). (2009)
21. Chevallier-Mames, B., Ciet, M., Joye, M.: Low-cost solutions for preventing simple side-channel analysis: Side-channel atomicity. IEEE Transactions on Computers **53**(6) (2004) 760–768
22. Brier, E., Joye, M.: Weierstrass elliptic curves and side-channel attacks. In: Public Key Cryptography (PKC), LNCS 2274 (2002)
23. Brier, E., Dechene, I., Joye., M.: Unified addition formulae for elliptic curve cryptosystems. In: Embedded Cryptographic Hardware: Methodologies and Architectures, Nova Science Publishers (2004)