Attacking Misaligned Power Tracks Using Fourth-Order Cumulant

Eng. Mustafa M. Shiple
Electronics Department
National Telecommunication Institute Cairo, Egypt

Prof. Dr. Iman S. Ashour
Electronics Department
National Telecommunication Institute Cairo, Egypt

Prof. Dr. Abdelhady A. Ammar,
Communication Department
Al Azhar University
Cairo, Egypt

***Abstract*—Side channel attacks (SCA) use the leaked confidential data to reveal the cipher key. Power consumptions, electromagnetic emissions, and operation timing of cryptographic hardware are examples of measurable parameters (analysis) effected by internal confident data. To prevent such attacks, SCA countermeasures are implemented. Misaligned power tracks is a considerable countermeasure which directly affect the effectiveness of SCA. Added to that, SCA are suffering from tremendous types of noise problems. This paper proposes Fourth-order Cumulant Analysis as preprocessing step to align power tracks dynamically and partially. Moreover, this paper illustrates that the proposed analysis can efficiently deal with Gaussian noise and misaligned tracks through comprehensive analysis of an AES 128 bit block cipher.**

Keywords—Correlation power analysis (CPA); Differential power analysis (DPA); side channel attack; FPGA; AES; cryptography; cipher; fourth-order cumulant; Gaussian noise; higher order statistics

I. INTRODUCTION

In the past decade, new threats become more and more efficient and powerful against cryptosystems. There are three categories of attacks, active invasive (e.g. fault injection), active non-invasive (e.g. tampering), and passive (power consumption, timing attack) [1].

Invasive attack bases on penetration the Device Under Attack (DUA) package and analyzes each layer to modify or monitoring the entire signals and buses. The countermeasures of this type is based on burying critical layers beneath other layers of conducting metal layers to avoid direct connection from the surface. Added to that, distributing sensors all over the chip to detect any attack trials. These sensors erase the memories and all critical registers when activated. Ffurthermore, the designers use the nonstandard cells, scrambling the bus implementation, scrambling the stored data, dummy structure to mislead the attackers to discover the design architecture.

Like invasive attacks, Semi-invasive attack requires depackaging the chip without creating contacts to the internal lines [2]. Ultraviolet ray is used to unauthorized access to the stored data meanwhile the cryptosystem designers use many of anti-fuse to prevent such attacks like Security Fuse, Program Fuse, Array Fuses, and oProbe Fuse [3]. Another technique; a transient fault during the execution of some process is injected. A fault allows bypassing security condition checks, such as PIN correctness or in some other cases reducing the cipher rounds by manipulating the cipher counter. Many researchers have mounted differential fault analysis (DFA) on symmetric key encryption algorithms, such as the triple-DES [4], Advanced Encryption Standard (AES)[5], CLEFIA [6], and ARIA [7].

On the other hand, Noninvasive attack does not depackaging the DUA since the main load of this attack derives towards exploiting the confidential data through observing the output behavior of the DUA. Noninvasive attack could be Passive, also called side-channel attacks, or Active attack. Side-channel attack does not involve any interaction with the attacked device but, usually, observation of its power consumptions and electromagnetic emissions.

Paul Kocher et al. introduced a powerful cryptanalysis technique called Differential power analysis (DPA) in 1999 [8]. This technique is based on the dependency of the processed data/the operation performed to power consumption of the device under attack (DUA). Kocher proofed that the leak information could easily reveal the confidential cipher key.

Many different countermeasures are emerged to stop SCA and secure the cryptosystems. Misaligned tracks, adding non-correlated noise, and breaking the relation between processed data and measured parameters are designers targets to protect their hardware.

A novel approach is proposed to dynamically align power traces and reducing noise signal effects in one shot. Moreover, the proposed idea shows a great improvement in processing time reaches more than 75% less than other techniques. By attacking a FPGA-based 128 bit AES block cipher, Analysis results show that the proposed idea efficiently helps SCA in harsh environment (noisy and suffering from alignment problems).

This paper is organized as follows. An overview on CPA is provided in Section II. In Section III, IV, Side channel attack Noise and Alignment techniques are presented. Then, in Section V, The background of higher order statistics is briefly introduced. Section VI provides a detailed explanation of the proposed method. Finally, we conclude with the main advantages presented in this paper.

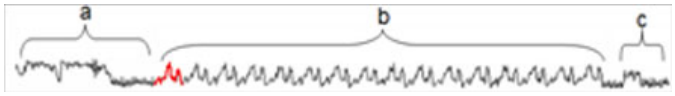

Fig. 1. SPA for DES algorithm; (a) 16 Rounds, (a) and (c) initial and final permutation respectively

II. Passive Noninvasive Attack

The types of available power analysis are Simple power analysis (SPA), Differential power Analysis (DPA), Correlation power analysis (CPA), and higher-order analysis [9].

A. Simple Power Analysis (SPA)

SPA is a side-channel attack, which involves visual examination of graphs of the power used by a device over time. Variations in power consumption occur as the device performs different operations. In short, SPA exploits the relationship between the executed operations and the power leakage. For example, Fig 1 shows the power consumption of a DES algorithm. SPA shows sixteen identical pulses for rounds and two pulses for permutations and rotations.

B. Differential Power Analysis(DPA)

DPA is statistically analyzing power consumption measurements recorded from DUA. DPA inherently reduces the noise by its averaging technique [10]. The attacker does two steps. The first step, "recording", the power consumption of cryptosystem while plaintext is processed. The second step, "comparison", the recorded traces is compared with a power model of the DUA by correlation analysis. The result of correlation will detect the correct key; high peaks means correct key no peaks means false key guessing. DPA exploits the relationship between the processed data and the power leakage. The main advantage of DPA is no details required about the structure of the attacked algorithm while SPA shows only the type of the used algorithm. High-Order Differential Power Analysis (HO-DPA) is an advanced form of DPA attack. HO-DPA enables multiple data sources and different time offsets to be incorporated in the analysis.

C. Correlation Power Analysis(CPA)

In this technique, CPA is based on how a predicted power consumption model correlates with measured power consumption of DUA.

At first glance, the adversary builds power consumption model for DUA using Hamming Weight, Hamming Distance or any other models. These models target intermediate value dependent on key and plain message $f(P_i;K_S)$, where P_i is a known non-constant data value and K_S is a small part of the key. Consequently, plain messages P_i (i=1.....N) with every possible key K_S (S=1.....256) stimulate the selected power model and the results are recorded in (Nx256) predicted power matrix M_{pp}.

In the second part of attack, the same Plain messages P_i (i=1.....N) are encrypted by DUA meanwhile, the power consumption of DUA while the chip is operating is measured and recorded in in (NxT) measurement power matrix M_{mp}, where T denotes the length of the trace.

Finally, the analytical part, the correlation coefficient is the most common way to determine linear relationships between data. The correlation coefficient is used between M_{mp} and columns of M_{pp}. These results are recorded in a matrix of estimated correlation coefficients. An efficient way to compute this linear relation is to use Pearson coefficient that can be expressed as follows;

$$\rho M_{mp(t)}, M_{pp} = \frac{E\left(M_{mp}(t).M_{pp}\right) - E\left(M_{mp}(t)\right).E(M_{pp})}{\sqrt{var\left(M_{mp}(t)\right).var(M_{pp})}} \quad (1)$$

In this expression,

E(x) denotes the mean value of matrix x.
var(x) denotes the variance of matrix x.
$M_{mp}(t)$ denotes the measured power matrix at time "t".

The next Figure 2 expresses briefly CPA steps.

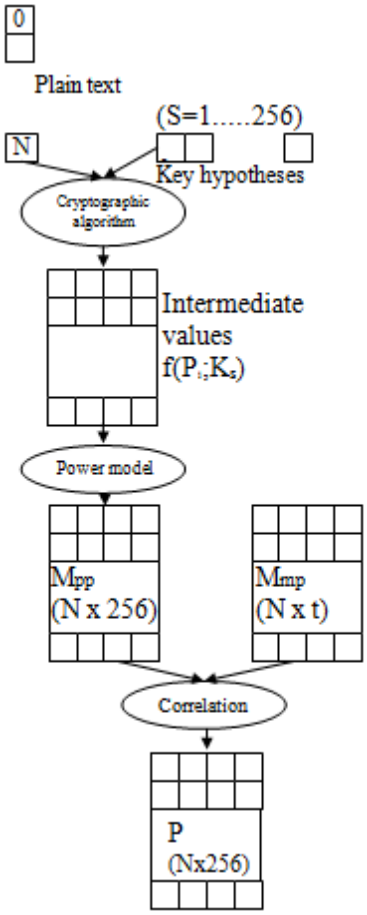

Fig. 2. Correlation Power Analysis

III. Types Of Noise Signals

Noise is an unwanted signal composed with genuine signal. According to cryptography, additive Noise to power signal have a great effect to immune side channel attack thus cryptography pursue to magnify the effect of noise on their implementation. There are two categories of noise: intentional

noise which is added by cryptography, normal noise which is any other type of noise.

When a measurement of a cryptographic device is repeated several times with constant input parameters, the resulting power traces are different. We refer to these fluctuations in the power traces as Electronic Noise [1]. Sources of this type of noise are varying from noise due to power supply, Quantization, and all other noise radiated from measurement setup components.

The device under attack consumed partial power relates to cipherkey which is important and rest of consumed power caused by cells that are not relevant for the attack as Switching Noise.

On other hand, Many cryptanalysts conduct researches and experiments to overcome the effect of the noise over attacking techniques, they could be summarized in three directions:

- Design of accurate power model to minimize Switching Noise.
- Use filters to minimize Electronic Noise.
- Conduct of Higher order statistics (HOS).

IV. Alignment Techniques

To prevent side channel attacks using power analysis techniques, cryptographers commonly implement DPA countermeasures that create misalignment in power trace sets and decrease the effectiveness of such attacks. On the other hand, Adversaries crucially work to realign the power traces.

There are two types of alignment: Static alignment and Dynamic alignment. Static misalignment is typically caused by inaccuracies in triggering the power measurements. Static alignment solves this problem by determining the duration of the timing inaccuracies, and shifting the traces accordingly [1].

In contrast, cryptographers actively use random time delays, varying clock frequencies and Random Process Interrupts (RPI). These techniques force cryptosystem sub-blocks to start operating at different and random times. Consequently, measured power consumptions are inherently misaligned. In these cases, static shifting cannot fully align the traces. Dynamic alignment is a general term for algorithms that match parts of several traces at different offsets, and perform nonlinear re-sampling of the traces.

A. Dynamic Time Warping

Dynamic Time Warping (DTW), was introduced to the data mining community by Berndt and Clifford [11]. In order to detect similar shapes with different phases, DTW method allows elastic shifting of the time axis. Also, speech processing community uses this technique for long time. The main drawback of DTW is processing time since performance on very large databases may be limited. Figure 3 demonstrates the power of DTW over traditional technique. DTW measures the distance between two sequences by elastically warping them in time.

Figure 4 shows the superiority of DTW over sliding window DPA (SW-DPA)[12]. SW-DPA is based on averaging fixed length clock cycles to restore the DPA peak in the face of random process interrupts.

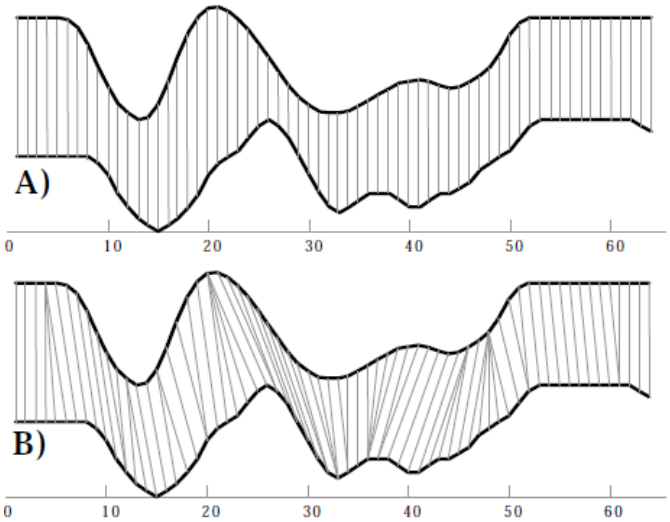

Fig. 3. Two sequences have an overall similar shape, they suffer from time misalignment, (A) i^{th} point on top sequence is aligned with the i^{th} point on the bottom that will produce a dissimilarity measure, (B) allows a more intuitive distance measure to be calculated.

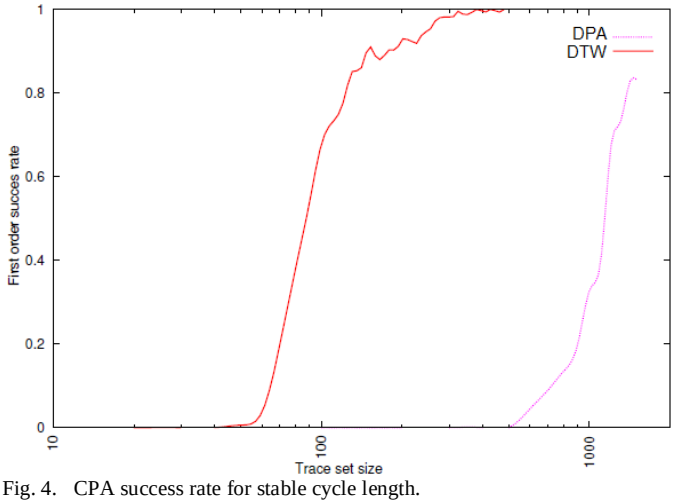

Fig. 4. CPA success rate for stable cycle length.

B. Correlation and Euclidean Alignment Technique

The main weakness of DTW is long processing time. In contrary to DTW, further techniques align power traces both partially and dynamically. Small portions of the traces (the interesting round part) are captured and aligned accordingly[13].

To save the processing time, this technique is based on extract the interesting round part from the whole power traces. This cutting concept limits later processing to shorter domain [14]. Figure 5 demonstrates the steps needed to align power traces.

Table I explains the steps of alignment. Fine tuning is considering with maximum similarity between selected round and other traces. This similarly is calculated by correlation, Euclidean [15].

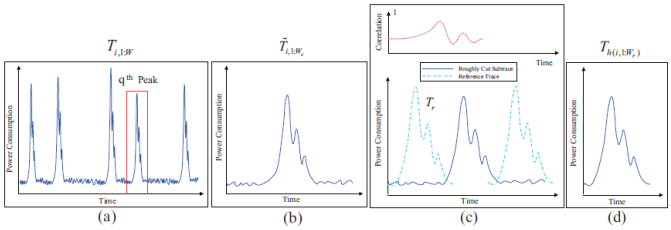

Fig. 5. Steps of aligning power traces (A)Find the qth peak, (B) Coarse Extraction, (C)Fine Tuning, and (D)Fine Tuned Extraction

TABLE I. STEPS OF ALIGNING POWER TRACES

Algorithm: Alignment Algorithm

Require: aligned power trace set

1. Find interesting round: Find the qth peak position (targeted round).
2. Coarse Extraction: Cut the selected round roughly to save all information belongs to that round.
3. Fine Tuning: Calculate maximum similarity by using Euclidean or Correlation to the selected round.
4. Fine Tuned Extraction: Cut the high similarity part in all traces.

TABLE II. MOMENTS OF DISTRIBUTION

Parameter	*Moment*	*Description*
Mean (μ)	1	Measure of central location.
Variance (σ2)	2	Measure of dispersion.
Skew	3	Measure of Asymmetry.
Kurtosis	4	Measure of peakedness.

The Euclidean distance or Euclidean metric is the distance between two vectors. The Euclidean distance is calculated by following equation:

$$d(RR,TR) = \sqrt{(RR_0 - TR_0)^2 + \cdots + (RR_n - TR_n)^2} \qquad (2)$$

Where:
RR: is the referenced vector.
TR: tested vector.

V. FOURTH ORDER CUMULANT

This research uses Higher Order Statistics (HOS) to find the moments of random signals that give the random variables some of their dominant features [16]. Moments are divided to two categories; Moments about the origin (raw moments) and Moments about the mean (central moment). Table II shows the first four distributions.

There are four ways to calculate the moments, they are:

1) Using the definition of moment.
2) Probability Generating Function [PGF].
3) Moment Generating Function [MGF].
4) Characteristic Function.

Moments are driven from the mathematical expectations of the random signal E{g(X)} [17].

The *raw moment*, is called the nth moment of X. It is denoted by n is given by

$$\mu'_n = E\{(X)^n\} = \sum_i (x_i^n) Px(x_i) \qquad (3)$$

Prime symbol is denoted to raw moment.

The central moments of random variable X are the moments of X with respect to its mean. Hence, the nth central moment of X, n, is defined as [17]

$$\mu_n = E\{(X-\mu)^n\} = \sum_i (x_i - \mu)^n Px(x_i) \qquad (4)$$

TABLE III. RAW AND CENTRAL MOMENTS

Moment	*Raw Moment*	*Central Moment*
μ_1	$E\{X\}=\mu$	0
$\mu 2$	$E\{X^2\}$	$E\{(X-\mu)^2\}=\sigma$
$\mu 3$	$E\{X^3\}$	$E\{(X-\mu)^3\}$
$\mu 4$	$E\{X^4\}$	$E\{(X-\mu)^4\}$

Table III contrasts between raw and central moments

Kurtosis measures the height and sharpness of the peak relative to the rest of the data. Higher values indicate a higher, sharper peak; lower values indicate a lower, less distinct peak. The kurtosis has no units: it's a pure number.

The normal distribution has a kurtosis of 3. Excess kurtosis is simply kurtosis-3[18].

- A normal distribution has kurtosis exactly 3 (excess kurtosis = 0). This is called mesokurtic.
- Any distribution has central peak is lower and broader, and its tails are shorter and thinner than normal distribution, its kurtosis < 3 (excess kurtosis < 0) is called platykurtic.
- Any distribution has central peak is higher and sharper, and its tails are longer and fatter than normal distribution, its kurtosis > 3 (excess kurtosis > 0) is called leptokurtic.

Kurtosis can be formally defined as a fourth population moment about the mean [19],

$$\beta_2 = \frac{E(X-\mu)^4}{(E(X-\mu)^2)^2} = \frac{\mu_4}{\sigma_4} \qquad (5)$$

Laplace was led to introduce a function known as a cumulative function, which is simply the logarithmic of the characteristic function [20]

VI. PROPOSED PREPROCESSING TECHNIQUE

The proposed idea aims to extract areas of interest using dynamic peak search [21]. Based on this technique, the 4th Cumulant method will duplicate its advantages to be used as

alignment technique. Furthermore, Cumulant shows superiority in:

- Eliminating the Gaussian noise.
- Treating misalignment.
- Reducing processing time.

We conducted two experiments to illustrate the power of the Cumulant method over the other traditional methods. The first experiment aims to compare Cumulant with other alignment techniques in absence of the noise. The second experiment highlights the advantages of Cumulant to prevent noise compared to traditional noise reduction methods with aligned power tracks.

A. *Cumulant vs. Alignment Techniques at Free Noise Environment*

In this part, free noise, misaligned measured power tracks are analyzed by the proposed idea and traditional alignment techniques. A comparison is made among these techniques to judge each one. Processing time, and differentiation between correct and false cipher keys will be the two parameters used to evaluate each technique.

Number of measured power tracks is frequently increased when alignment techniques are processed. Each time the difference between correlation results of correct and false cipherkeys is calculated and plotted. Figure 6 shows that, all traditional techniques including proposed one need the same number of power tracks to detect the correct cipherkey (approximately 81 power tracks). Over eighty one power tracks, the proposed idea and other techniques success to pickup the correct cipherkeys but in reality the traditional techniques show better results than the proposed idea.

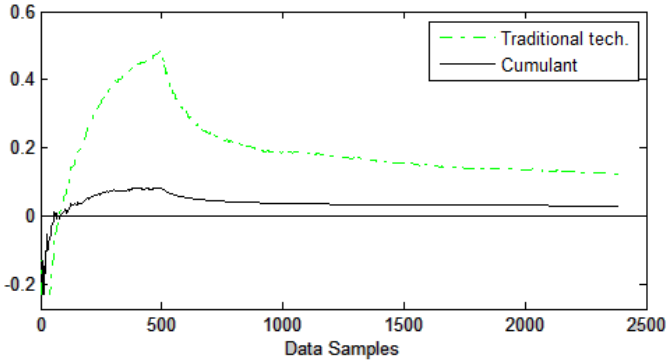

Fig. 6. Distinguishing between correct and false cipher keys.

Again, number of measured power tracks is frequently increased when alignment techniques are processed. Each time the processing time is calculated and plotted. Figure 7 shows the processing time plot for each technique.

Figure 7 illustrates that the processing time of traditional alignment techniques increases linearly with number of measured power tracks. Equations (6) and (7) express the rate of increasing of CPU processing time of traditional techniques. On other hand, It is obvious that the proposed idea keeps CPU processing time constant regardless to number of measured power tracks.

Moreover, CPU processing time of proposed idea (= $4x10^{-2}$sec) is 50% less than the lowest CPU processing time of traditional techniques.

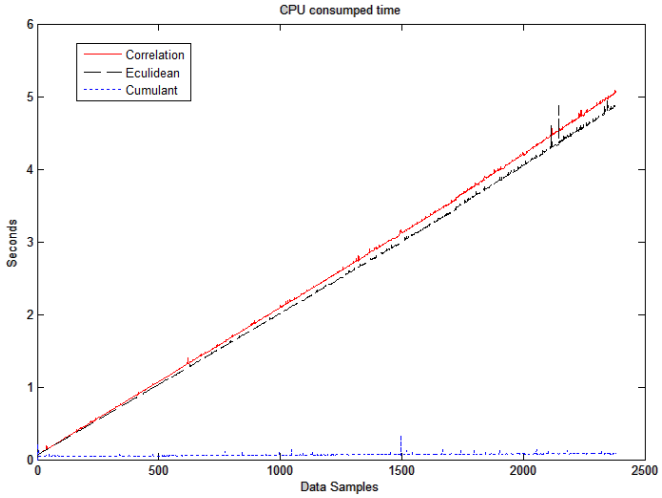

Fig. 7. The processing time of alignment techniques.

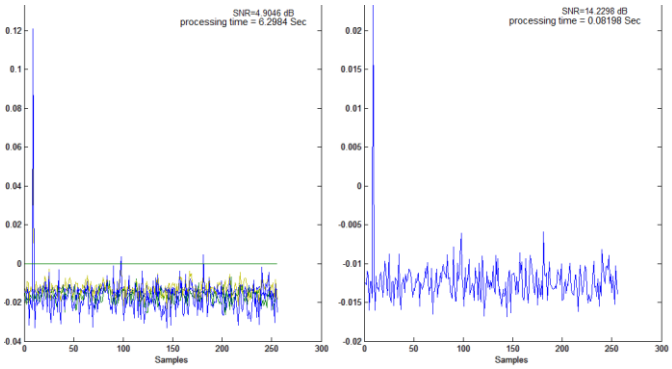

Fig. 8. A contrast between traditional alignment methods versus Cumulant.

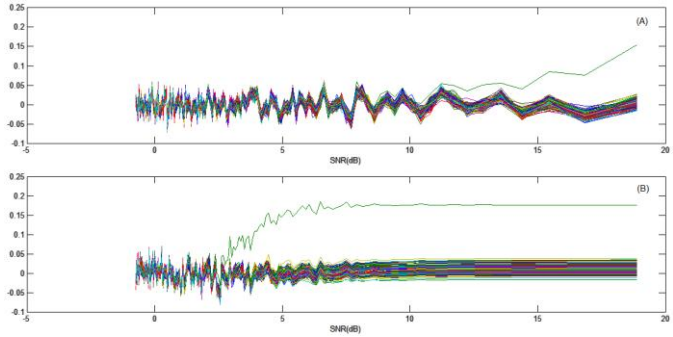

Fig. 9. RIJNDAEL SNR threshold (A)CPA behavior (B) Cumulant behavior.

$$\text{CPU time}|_{\text{Correlation}} = 2 \times 10^{-3}\text{tracks} + 7 \times 10^{-2} \quad (6)$$

$$\text{CPU time}|_{\text{Eculedian}} = 19 \times 10^{-4}\text{tracks} + 7 \times 10^{-2} \quad (7)$$

B. *Cumulant vs. Alignment Techniques at Noisy Environment*

Figure 8 shows a comparison between traditional alignment methods versus Cumulant. It's obvious that the correlation peak of true cipher key to false ones ("hint: we will denote the correlation peak of true cipher key to false ones as SNR_{corr}) of the proposed idea is enhanced three times the traditional

methods. Moreover, the speed of processing is 76 times higher than traditional methods.

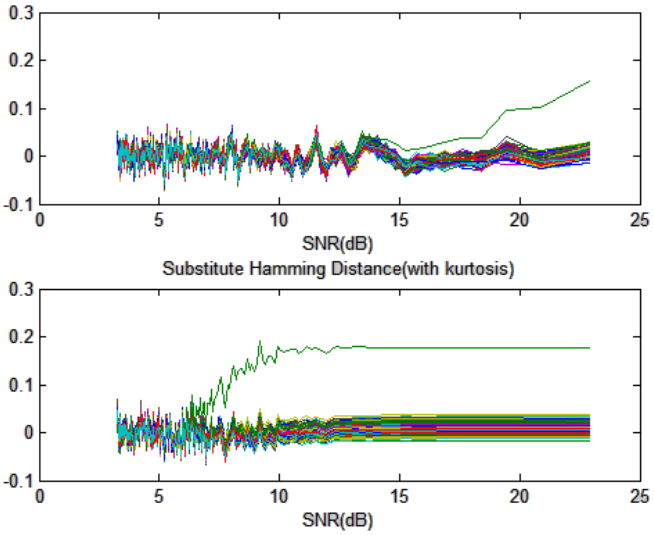

Fig. 10. Twofish SNR threshold (A)CPA behavior (B) Cumulant behavior.

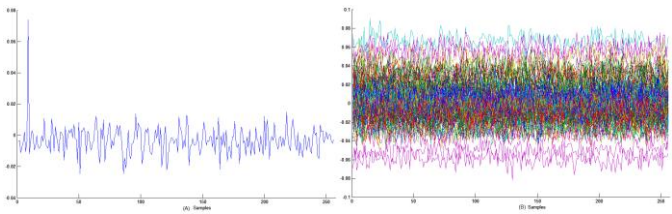

Fig. 11. Attacking RPI noisy stem by: (A)Cumulant (B) CPA.

Cumulant method demonstrates a great immunity to the noise, since it successfully attack a noisy system with SNR_{corr} = 2.5dB. This SNR_{corr} is reduced than those traditional methods by 75%.

Figure 9 shows the SNR limitation of Cumulant and traditional CPA. In Addition,Cumulant method is generic and independent of any specific cryptographic algorithm. The same results are obtained when applied to Twofish algorithm. Figure 10 shows the SNR limitation of Cumulant and traditional CPA when using Twofish algorithm.

It is obvious, that in case of a double impact system (noisy and RPI) all previous methods fails to get the cipherkey independently. As a result, combined techniques must be conducted and thus the processing time will be increased. Figure 11 shows the results of RPI noisy system exposed to attack by Cumulant and traditional methods. Figure 11 (A) demonstrates that Cumulant successfully attacks the system unlike the other.

VII. Conclusion

In this paper, we have proposed new features of fourth-order cumulant. These features include overcoming the noise added to the processed data, aligning measured tracks, speeding up the processing time to open new hopes to attack unbreakable algorithms due to time processing barriers. We have given the theoretical evaluation based on SNR criteria. The formulas to calculate these parameters have been given under a general form with flexible parameters, such as the noise level, and the number of side channel signals.

VIII. Future Work

Although the great results that are achieved by this paper, there are other noise types still affect the efficiency of SCA. These noise types need to be addressed and reduce its effects.

Extra studies are needed to highlight the benefits of the proposed algorithm over correlative noise countermeasures. Moreover, the limitation of the algorithm toward such noise would be calculated.

References

[1] S. Mangard, E. Oswald, and T. Popp, Power Analysis Attacks: Revealingthe Secrets of Smart Cards. Springer, 2007.

[2] S. P. Skorobogatov, Semi-invasive attacks a new approach to hardwaresecurity analysis. University of Cambridge Computer Laboratory,2005. [Online]. Available: http://www.cl.cam.ac.uk/TechReports/

[3] Implementation of Security in Actel Antifuse FPGAs. Alctel, ApplicationNote AC168, 2002.

[4] L. Hemme, A Differential Fault Analysis Against Early Rounds of Triple-DES. Proc. CHES, LNCS, vol. 3156, 2004.

[5] D. C. Y. K. JeaHoon Park, SangJae Moon and J. Ha, DifferentialFault Analysis for Round-Reduced AES by Fault Injection. ETRIJournal,Volume 33, Number 3, 2011.

[6] W. W. H. Chen and D. Feng, Differential Fault Analysis on CLEFIA.Proc. ICICS, LNCS, vol. 4861, 2007.

[7] D. G. W. Li and J. Li, Differential Fault Analysis on the ARIA Algorithm.Information Sciences, Elsevier, vol. 178, no. 19, 2008.

[8] J. J. P. Kocher and B. Jun, Differential Power Analysis. Proceedings of Crypto99, lecture notes in computer science, Vol. 1666, Springer-Verlag,1999.

[9] "updated time:june 2013, accessed date: Nov 2013," Wikipedia:the free encyclopedia.

[10] C. Chen, X. Li, L. Wu, and X. Zhang, "Design and implementation of a differential power analysis system for cryptographic devices," in Solid-State and Integrated Circuit Technology (ICSICT),, 2010.

[11] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," IEEE Trans. Acoustics, Speech, and Signal Processing, vol. 2, p. 143165, 1978.

[12] B. B. Jasper G. J. van Woudenberg, Marc F. Witteman, "Improving differential power analysis by elastic alignment," Cryptographers Track of the RSA Conference (CT-RSA), pp. 104–119, 2011.

[13] S. A. H. Qizhi Tian, "On clock frequency effiects in side channel attacks of symmetric block ciphers," IEEE Int. Conf. on New Technologies, Mobility and Security, 2012.

[14] M. S. Qizhi Tian, Abdulhadi Shoufan and S. A. Huss, "Power trace alignment for cryptosystems featuring random frequency countermeasures," Technical Report, Dept. of Computer Science, TU Darmstadt, 2012.

[15] S. A. H. Qizhi Tian, "On the attack of misaligned traces by power analysis methods," Technical Report, Dept. of Computer Science, TU Darmstadt, 2012.

[16] "updated time: 2012, accessed date: Nov 2013," Statistics Help.

[17] W. J. Stewart, "Probability, markov chains, queues, and simulation: The mathematical basis of performance modeling," in Princeton University Press. p. 105. ISBN 978-1-4008-3281-1, 2011.

[18] S. Brown, "Measures of Shape: Skewness and Kurtosis kernel description," http://www.tc3.edu/instruct/sbrown/stat/shape.htm, accessed: 2013-09-30.

[19] L. T. DeCarlo, "On the meaning and use of kertosis," Psychological Methods, vol. 2, pp. 292–307, 1997.

[20] R. A. F. E. A. Cornish, "Moments and cumulants in the specification of distriputions," revue de l'institute International de Statistique, vol. 5, pp. 307–320, 1937.

[21] S. A. H. Qizhi Tian, "A general approach to power trace alignment for the assessment of side-channel resistance of hardened cryptosystems," Eighth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 2012.