Improved Leakage Model Based on Genetic Algorithm

Zhenbin Zhang
zhangzb12@mails.tsinghua.edu.cn

Liji Wu
lijiwu@mail.tsinghua.edu.cn

Abstract. The classical leakage model usually exploits the power of one single S-box, which is called divide and conquer. Taking DES algorithm for example, the attack on each S-box needs to search the key space of 2^6 in a brute force way. Besides, 48-bit round key is limited to the result correctness of each single S-box. In this paper, we put forward a new leakage model based on the power consumption of multi S-box. The implementation of this method is combined with genetic algorithm. In DES algorithm, we can establish leakage model based on the Hamming distance of summing up 8 S-boxes. The genetic algorithm can search the key space of 2^{48} to complete the attack of 8 S-boxes at the same time intelligently. And we also experimentally validate the fact that the leakage model of 8 S-boxes can decrease about 60% number of traces which is needed in the classical based on one single S-box in time domain and it also decreases about 33% number of traces in frequency domain. The IC card which is used in experiment is the training card 8 provided by Riscure Company.

Keywords: side-channel analysis, leakage model, genetic algorithm, correlation power analysis.

1 Introduction

Leakage model plays an important role in power analysis attack. Attackers use it to establish the virtual power consumption of each different guess key to simulate the actual work state of cryptographic devices. With the development of power attack, the cognition of leakage model is getting more and more accurate. DPA which is used by Kocher [7,8] and formalized by Thomas Messerges et al. [10] just exploits 1 bit power information to classify the power traces. In [1,2,9,11], leakage model usually exploits the Hamming weight or Hamming distance of some bits in a S-box to depict the actual power consumption. Both approaches exhibit some limitations due to unrealistic assumptions and model imperfections. But it is obvious that the trend of leakage model is advancing towards the way which can be closer to real power consumption.

In DES algorithm, the DES key is 56-bit, which means if you want to guess the key in brute force way, you have to try all 2^{56} key. Classical DPA and CPA aim at attacking the key of one single S-box. It is a classic divide-and-conquer strategy, which suggests the key can be recovered by cracking 8 S-boxes separately. Appearance of the strategy is due to the reason that classical leakage model need to search all 2^6 key space in a brute force way. The poor ability of searching extremely limits the key space which results in the attacking target can only choose one single S-box. While in fact, the leakage model based on a single S-box is not so accurate. Because in each round of DES, it usually needs to do the parallel operation of inquiring 8 S-boxes. In time domain, this behaves as the process of inquiring 8 S-boxes is working at the same time.

However, the classical leakage model chooses a single S-box as the attacking target, while regarding the other 7 S-boxes as noise. This kind of leakage model just exploits the power information of one single S-box in essence. That means the classical leakage model which aims at one single S-box has discarded most of the power that can be available. As we know, Hamming weight or Hamming distance of a single S-box just ranges from 0 to 4. So, the problem which we face is that one single S-box may lead to low SNR, and multi S-box may bring the unacceptable searching key space.

Our Contribution. The novel contributions of this paper are as follows:

1. In this paper, we put forward a new accurate leakage model which is based on the power consumption of multi S-box. We use genetic algorithm to achieve the construction of multi S-box leakage model in which searching space can be as large as 2^{48}. Besides, it is worth mentioning that the searching ability of genetic algorithm is also efficient.
2. We find the connection between genetic algorithm and power analysis attack. It has a profound effect on raising efficiency of power analysis. Different kinds of intelligent algorithms and their parameters will affect the speed of convergence. And a better set of parameters can lead to the helpful searching path. But the optimization of the problem depends on the leakage model which can be close to the actual power consumption.
3. Our work is of high scalability. The new leakage model can also be applied for other cryptographic algorithm. And the leakage model is not limited to analysed S-box. It also can be used for the predictable operation happened concurrently which leads to the superposition of power consumption.

Organization. The paper is organized as following. Sect. 2 introduces implementation of cryptographic algorithm and mathematical foundation of genetic algorithm. Sect. 3 puts forward a new leakage model based on multi S-box. The method makes full use of the powerful searching ability of genetic algorithm. Sect. 4 confirms the efficiency of new leakage model with some experimental analysis on DES. We conclude in Sect. 5.

2 Preliminaries

2.1 Implementation of Cryptographic Algorithms and Hamming Distance Model

As we know, the implementation of cryptographic algorithms in hardware may have some predictable operation working concurrently. For the 8 S-boxes of DES is different, its implementation usually let the inquiring of 8 S-boxes complete at the same time. This can short the running time of DES and increase the safety to a certain extent.

Hamming distance model reflects the power consumption of switching state. And Hamming weight model reflects the power consumption of static state. The Hamming distance of a single S-box is defined as the Hamming weight of switching state from input to output. Usually we choose 4-bit of input and 4-bit of output to calculate the Hamming distance of a S-box.

$$\mathrm{HD}(Sbox) = \mathrm{HW}\left(S_{input\ 1-4} \oplus S_{output\ 1-4}\right)$$

2.2 Preliminaries of Genetic Algorithm

In [6], Holland puts forward the Genetic algorithm (GA) to solve optimization problem. GA is a search heuristic that mimics the process of natural selection. This heuristic is routinely used to generate useful solutions to optimization and search problems. GA uses the searching techniques inspired by natural evolution, such as selection, crossover and mutation.

In GA, a population of candidate solutions which called individuals to an optimization problem is evolved toward better solutions. Each candidate solution has a set of properties which can be altered and mutated. Traditionally, solutions are represented in binary as strings consisting of 0 and 1. A typical GA requires two key components. One is genetic representation of the solution domain. The other is fitness function to evaluate the solutions.

The evolution usually starts from a population of randomly generated individuals, and is an iterative process, with the population in each iteration called a generation. In each generation, the fitness of every individual in the population is evaluated. And the fitness is usually the value of the objective function in the optimization problem being solved. The more fit individuals are stochastically selected from the current population, and each individual's genome is modified to form a new generation. Crossover rate is denoted as p_c and mutation rate is denoted as p_m. The step of GA as below.

Algorithm 1. Genetic Algorithm

Input: Random solutions of the problem.

Output: The optimization solution.

1. Encoding. Make the strategy of mapping the solution into chromosome. Usually it is binary string.
2. Initialization. Generate random individual solutions to form an initial population.
3. Individual Evaluation. Calculate the fitness of individuals. Where fitness function is defined over the genetic representation and measures the quality of the represented solution.
4. Selection. Select the individual solutions which have high fitness value based on the ranking calculated by the fitness function.
5. Genetic Operator. Change the chromosome via crossover and mutation operation.
6. Termination. Stop the iterative process. But if the generation doesn't reach the expected number, then go to step 3.
7. Decoding. Decode the optimization chromosome into the solution of problem.

2.3 Mathematical foundation of Genetic Algorithm

In [3,4,5,15], the mathematical foundation of genetic algorithm is studied, First, we should know three definitions in GA in order to analysis the mathematical foundation.

Definition 1. *Let us consider a schema H taken from the three-letter alphabet which is the assemblage of* $\{0,1,*\}$. *And the star $*$ is a don't care or wild card symbol which matches either a 0 or a 1 at a particular position.*

Definition 2*. The order of a schema H, denoted by $O(H)$, is simply the number of fixed position (in a binary alphabet, the number of 0's and 1's) present in the template.*

Definition 3*. The defining length of a schema H, denoted by $\delta(H)$, is the distance between the first and last specific string position.*

Theorem 1 (Schema Theorem)*. Lower order, above-average schemata receive exponentially increasing trials in subsequent generation.*

$$\begin{cases} m(H,t+1) \geq m(H,t)\dfrac{f(H)}{\bar{f}}\left[1 - p_c\dfrac{\delta(H)}{l-1} - p_m O(H)\right] \\ m(H,t)\dfrac{f(H)}{\bar{f}} = (1+c)^t * m(H,0) \end{cases}$$

Where

$m(H,t)$ is there are m examples of a particular schema H containded within the population at a given time step t.

$f(H)$ is the average fitness of the strings representing schema H at time t.

$\bar{f}$ is the average fitness of the strings representing all individuals at time t.

l is the length of individual.

c is a constant that a particular schema H remains above average an amount $c\bar{f}$.

Mathematical foundation of GA is the Schema Theorem. The theorem describes a fact that a particular schema H receives an expected number of copies in the next generation under reproduction, crossover and mutation as given by the equation above.

In a word, a particular schema grows as the ratio of the average fitness of the schema to the average fitness of the population. Put another way, schemata with fitness values above the population average will receive an increasing number of samples in the next generation, while schemata with fitness values below the population average will receive a decreasing number of samples.

3 New Attack Model Based on Genetic Algorithm

3.1 Main Idea of Attack Model

Instead of attacking the Hamming distance of one single S-box, we try to analyze the Hamming distance of multi S-boxes. As for cryptographic devices which make 8 S-boxes work concurrently, the better leakage model is to sum up the hypothetical power consumption of 8 S-boxes.

Figure 1 shows the relation between correlation coefficient and leakage model based on a single S-box or multi S-boxes. In Figure 1, $\mathrm{HD}(Sbox\ 1-k)$ means the leakage model constructed by summing up the Hamming distances from S-box 1 to S-box k, where $1 \le k \le 8$. Table 1 shows the maximum correlation coefficient in different leakage model based on a single S-box or multi S-boxes.

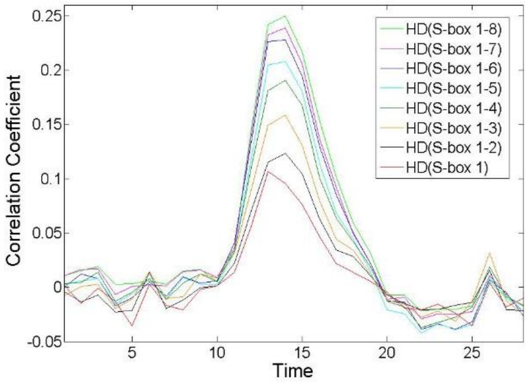


Fig.1. Relation between correlation coefficient and leakage model based on a single S-box or multi S-boxes respectively.

Table 1. Maximum correlation coefficient in different leakage model based on a single S-box or multi S-boxes.

k	1	2	3	4	5	6	7	8
Correlation	0.1064	0.1151	0.1490	0.1809	0.2048	0.2262	0.2326	0.2417

From the result of Figure 1 and Table 1, the maximum correlation coefficient which is calculated by different leakage model increases with the increasing of the number of S-boxes. That is what the inequation describes.

$$max\{corr[Trace, \sum_{i=1}^{k} \mathrm{HD}(Sbox\ i)]\} > max\{corr[Trace, \mathrm{HD}(Sbox\ i)]\}$$

Where $2 \le k \le 8$ and the Trace is the actual power consumption acquired from an IC card.

Obviously, correlation of 8 S-boxes leakage model is observably higher than that of a single S-box. That means, leakage model of multi S-box does not lose power information. Instead, it can be the better description of the actual power consumption. That is to say, we increase the SNR markedly.

So we obtain the fact is that leakage model of 8 S-boxes is more efficient and exact than leakage model of partial S-box. Actual power consumption is closer to the leakage model of 8 S-boxes. Hamming weight or Hamming distance of leakage model based on 8 S-boxes can range from 0 to 32. So the challenge of new efficient leakage model is that searching space can be as large as 2^{48}. The huge searching space of 2^{48} is extremely beyond the acceptable range in a brute force way.

3.2 New Leakage Model Based on Genetic Algorithm

Before we start the searching of 2^{48} key space, we must find the way to change the 48-bit DES round key into the solution of GA. The round key is in binary as strings of 0s and 1s, so we just make 48-bit of DES first round divided into 8 segments. And each segment has 6 bits that make each segment map the input of each S-box. Table 2 shows a mapping example.

Table 2. Mapping example between a random DES 48-bit key and solution in GA

	X1	X2	X3	X4	X5	X6	X7	X8
Key	34	0	22	60	8	12	32	18
Genome	100010	000000	010110	111100	001000	001100	100000	010010

After the encoding process, we need to complete initialization. For the population size n, our choice is 60. The parameter depends on the problem. Then an initial population of 60 individual solutions is randomly generated.

Table 3. Example result of initialization.

	X1	X2	X3	X4	X5	X6	X7	X8
1	010011	101010	111011	011001	101101	101011	011100	100111
2	101101	011101	110110	101111	010111	111000	010101	111011
…				……				
n	101001	110101	011110	101011	010101	110010	100100	010111

In one generation, we have to select the random solutions which are close to the right solution. We use fitness function to evaluate the fitness of each individual. Fitness function on this question is defined as the correlation coefficient between the actual power consumption data and leakage model of 8 S-boxes.

$$Fitness = corr\left[Trace, \sum_{i=1}^{8} \mathrm{HD}(Sbox\ i)\right]$$

Where $\mathrm{HD}(Sbox\ i) \in \{0,1,2,3,4\}$.

The higher the correlation is, the better solution will be. Through summing up the Hamming distance of 8 S-boxes, the new leakage power model is generated.

Fitness function measures the quality of the represented solution. 60 individuals of each generation can get its own fitness. So we can sort them according to the ranking of fitness. Pick up the superior solutions to do the crossover and mutation operation. Table 4 shows the example of crossover operation and Table 5 shows the example of mutation operation. The positions changed by crossover and mutation operation use the font format in bold and italic.

Table 4. Example of crossover operation

010011	101010	111011	011001	101101	101011	011100	100111
101101	011101	110110	101111	010111	111000	010101	111011
Crossover Operation							
010011	101010	***110110***	011001	101101	***111000***	011100	100111
101101	011101	***111011***	101111	010111	***101011***	010101	111011

Table 5. Example of mutation operation

010011	101010	111011	011001	101101	101011	011100	100111
Mutation Operation							
0100***0***1	101***1***10	111011	011001	101101	101***1***11	011100	10***1***11***0***

It is worth tuning parameters such as the crossover probability and mutation probability to find reasonable settings for the problem which is working on. In [12], a crossover rate that is too high may lead to premature convergence of the genetic algorithm. A mutation rate that is too high may lead to loss of good solutions unless there is elitist selection. A very small mutation rate may lead to genetic drift. Crossover probability p_c is 0.7 and mutation probability p_m is 0.01.

After the selection, crossover and mutation operation, some new individuals is generated. Calculate fitness value of new individuals and pick up some superior to replace the inferior in last generation. Now one iteration is finished. The whole iteration is over when the best solution is found or reaches the maximum of generation.

During the evolution process, GA will guarantee the guessing key of each S-box can evolve to the right key of that S-box. Once a right S-box key is searched, the current solution is more superior to the ones which not reach the right S-box key. So the individuals that contain the right segment which maps the corresponding right S-box key will be kept generation by generation in that population. The survival of the fittest can be kept on and on. Because of the evolution, code of chromosome which represents the key can be optimized towards the global optimum solution. The global optimum solution picked up from the 2^{48} key space is the right 48-bit DES round key.

4 Experimental Validation of the Attack on DES

4.1 Details of Attack

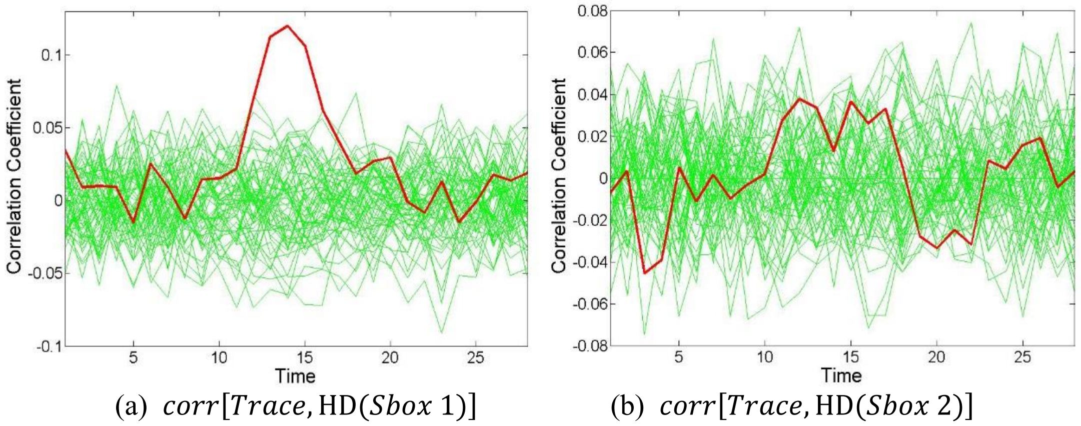
In order to verify the new leakage model is better than the classical leakage model, we acquire 9000 actual power consumption traces of DES algorithm from an IC card. The IC card which is used in experiment is the training card 8 provided by Riscure Company. The acquisition equipment is the Power Tracer equipment developed by Riscure Company.

After getting the power data, we choose the S-box as the attack target. We first analyze the simple case which compares the leakage model based on 2 S-boxes with leakage model based on 1 S-box. And we also illustrate the process which shows the distribution of solution and the convergence of the correlation coefficient.

In the following, we use GA to achieve the construction of new leakage model. The powerful ability of searching huge key space is unfolded. We confirm the result that the new leakage model based on 8 S-box is more efficient than classical leakage model based on one single S-box, whether it is in time domain or frequency domain.

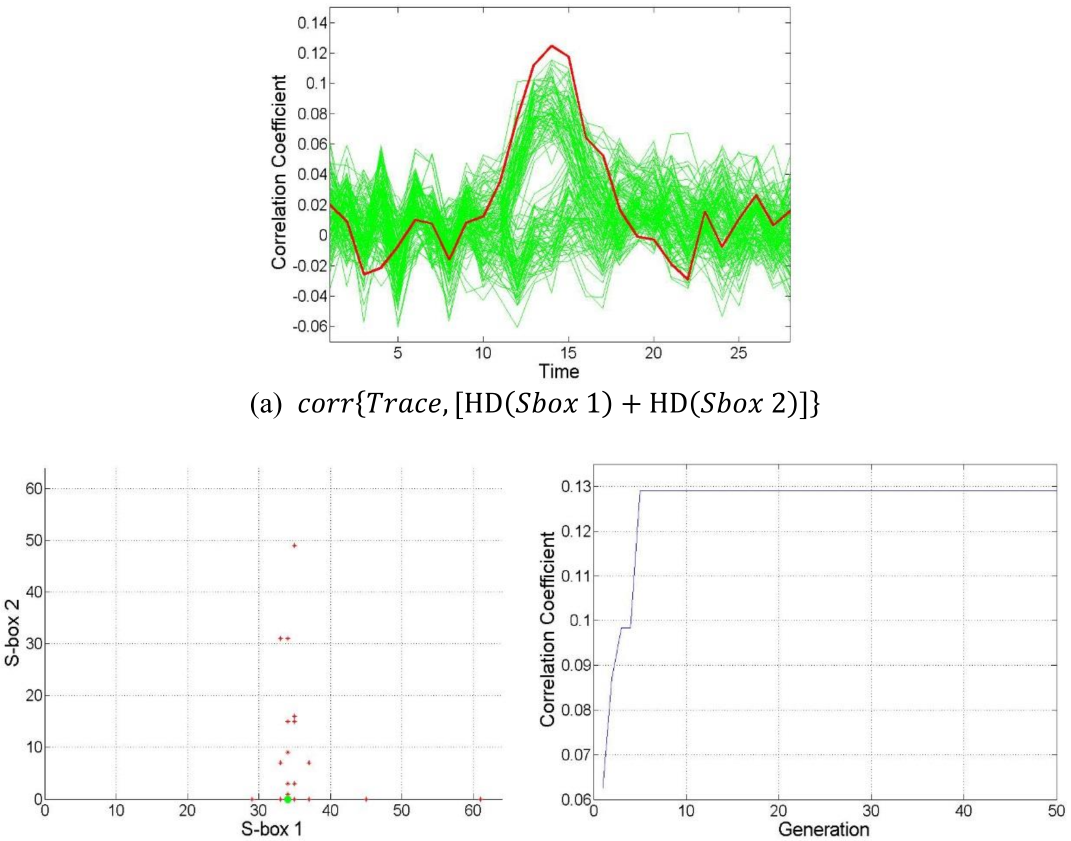
4.2 Experimental Results

As we suppose, leakage model of multi S-boxes is closer to actual power consumption. So, leakage model of 2 S-boxes is better than that of single S-box. We acquire 2000 traces power data from the IC card. Correlation calculated in Figure 2. (a) and (b) is based on the classical leakage model which aims at attacking S-box separately. In Figure 2. (a) . The right key of S-box 1 is separated from the other 63 candidate keys. That means we can get the key of S-box 1. While the right key of S-box 2 is buried in the other 63 candidate keys. This makes we can not get any valuable information about the key of S-box 2.



(a) $corr[Trace, \mathrm{HD}(Sbox\ 1)]$ (b) $corr[Trace, \mathrm{HD}(Sbox\ 2)]$

Fig.2. Attacking result of CPA based on leakage model of 1 S-box



(a) $corr\{Trace, [\mathrm{HD}(Sbox\ 1) + \mathrm{HD}(Sbox\ 2)]\}$

(b) Evolution of solution distributed in GA (c) Correlation varying curve in GA

Fig.3. Attacking result of CPA combined GA. It is based on leakage model of 2 S-boxes

We use the same 2000 traces acquired from the IC card to do the CPA based on leakage model based on 2 S-boxes. Figure 3. (a) shows that the key of S-box 1 and S-box 2 can be separated by the other candidate keys which means the right key of S-box 1 and S-box 2 are obtained at the same time. Figure 3. (b) displays the distribution of individuals in last generation. As we can see, the individuals is gathering around the right key which is the green spot of (34,0). According to the fitness function, the individuals can move to the region where the optimization solution is. Figure 3. (c) shows the correlation varying trend curve. Correlation is becoming higher and getting convergence rapidly. It is turn out that leakage model of 2 S-boxes is efficient than leakage model of a single S-box.

The same method is used for CPA based on leakage model of 8 S-boxes. GA begins with a population of individual solutions. And the algorithm searches the optimum matching path in parallel smartly. Finally, the optimization solution can be picked up from the key space which is as large as 2^{48}. Figure 4. shows the iteration of searching maximum correlation based on leakage model of 8 S-boxes by GA.

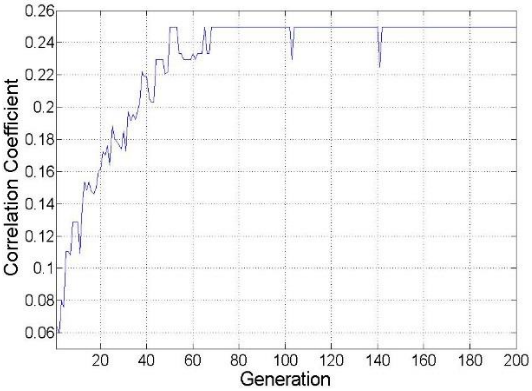
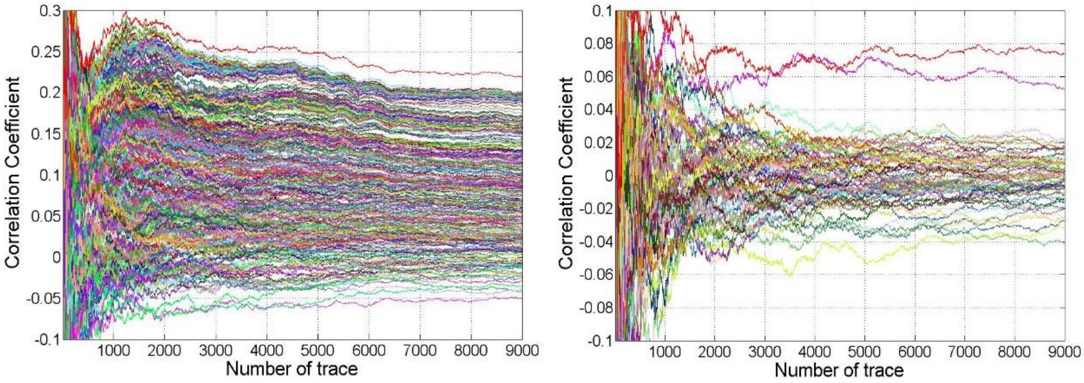


Fig.4. Iteration of searching maximum correlation coefficient based on leakage model of 8 S-boxes by GA

From the Figure 4, the generation of convergence can be less than 200. Usually the generation of convergence can be about 150 in average. That means, for each generation consisted of 60 individuals, the correlation in GA should be calculated for about 12,000 times which costs 9 seconds in average. The time can be ignored obviously. While the correlation of classical CPA attacking a single S-box each time should be calculated for 512 times which costs 0.4 seconds. Though the calculation of correlation in GA is about 20 times of that in CPA attacking single S-box each time, the searching key space of GA is 2^{48}, which is much larger than searching key space of 2^{6} in a brute force way.

In a word, CPA based on the leakage model of 8 S-boxes can get the right 48-bit DES round key in 5 minutes. The number of individual in population is 60 and the generation of convergence is about 150 in average.



(a) Leakage model of 8 S-boxes (b) Leakage model of 1 S-box

Fig.5. Relation between number of traces and result of CPA in time domain

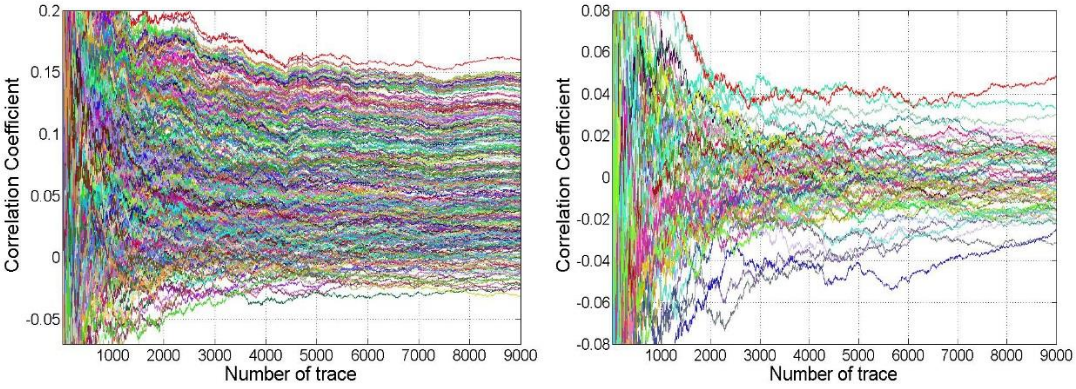
From the Figure 5 (a) and (b), we can see the correlation coefficient of the right key is separated from the others with the increasing number of traces. In Figure 5 (a), we find that it just needs about 2000 traces to get all 48-bit DES key of the first round, if we choose the leakage model of 8 S-boxes to do the CPA in time domain. The reason why correlation coefficient can range widely is that the searching solution maybe ranges from guessing 1 S-box right key to all the 8 S-box right key.

While Figure 5 (b) shows the number of trace is about 5000, if we choose the leakage model of a single S-box. The new leakage model can decrease about 60% number of traces in time domain.

So it is proved that the efficiency of leakage model based on multi S-box is enhanced and GA don't bring unacceptable calculation when searching huge key space of 2^{48}.

The leakage model based on 8 S-boxes is also suit for power attack in frequency domain. Power attack used in frequency domain is supported by the Parseval Theorem. Parseval Theorem reveals that the total energy contained in a waveform summed across all of time is equal to the total energy of the waveform's fourier transform summed across all of its frequency components.

As we know, clock randomized is the common and effective countermeasure to resist the CPA in time domain. In [13,14], CPA in frequency domain is better than CPA in time domain. While the problem is that interest power of first DES round is mixed with other unknown power at the same clock frequency. As we can predict that the leakage model based on 8 S-boxes is superior to that based on one single S-box.



(a) Leakage model of 8 S-boxes (b) Leakage model of 1 S-box

Fig.6. Relation between number of traces and result of CPA in frequency domain

From the Figure 6 (a) and (b), we can see the correlation coefficient of the right key is separated from the others with the increasing number of traces. In the Figure 6 (a), we find that it just need about 5000 traces to get all 48-bit DES key of the first round, if we choose the leakage model of 8 S-boxes to do the CPA in frequency domain. While Figure 6 (b) shows the number of trace is about 7500, if we choose the leakage model of a single S-box. The new leakage model can decrease about 33% number of traces in frequency domain.

We make a comparison between new leakage model and classical leakage model as the Table 6 shows. In a word, no matter what the case is in time domain or frequency domain, leakage model of multi S-box is much more powerful than the classical leakage model of a single S-box.

Table 6. Compare the number of power trace needed and time cost between the new leakage model and the classical leakage model.

Leakage model		8 S-boxes	1 S-box
Number of traces	Time domain	2,000	5,000
	Frequency domain	5,000	7,500
Running time	Time domain	9.1s	0.4s
	Frequency domain	12.7s	0.5s

5 Conclusion

In this paper, we put forward a new accurate leakage model which is based on the power consumption of multi S-box. The new leakage model has high SNR than the classical leakage model. We use genetic algorithm to achieve the construction of multi S-box leakage model in which searching space can be as large as 2^{48}. The improved leakage model reflects the actual power consumption of cryptographic devices accurately than that of classical leakage model based on a single S-box. We also experimentally validate the fact that leakage model of multi S-box is more efficient than the classical leakage model. While the quantity of calculation generated by genetic algorithm is acceptable.

This paper combines the genetic algorithm with power analysis attack. It has a profound effect on raising efficiency of power analysis. And we turn the construction of leakage model into the problem which searches the optimization solution in a huge space. We exploits the intelligent searching ability of genetic algorithm to improve the leakage model used for power attack.

Our work is of high scalability. New leakage model based on genetic algorithm has the quality of universality. The attack target of new leakage model is not only used for multi S-boxes which is working in parallel, but also used for some predicting power accumulation because of working state in parallel. Establishment of new leakage model can adjust to actual power consumption of cryptographic devices handily. And also the ability of intelligent searching space increases with the development of intelligent algorithm.

As an avenue for further research, this work connects the artificial intelligence algorithm with power analysis. Genetic algorithm is just one kind of artificial intelligence algorithm which can generate useful solutions to optimization and search problems. There is room for improving the speed of convergence and definition of optimization solution.

References

1. Brier, E., Clavier, C., Olivier, F.: Correlation power analysis with a leakage model. Cryptographic Hardware and Embedded Systems-CHES 2004, pp. 16-29. Springer Heidelberg (2004)
2. Coron, J.-S., Kocher, P., Naccache, D.: Statistics and secret leakage. In: Financial Cryptography, pp. 157-173. Springer, Heidelberg (2001)
3. Forrest, S., Mitchell, M.: What makes a problem hard for a genetic algorithm? Some anomalous results and their explanation. Machine Learning 13, 285-319 (1993)
4. Goldberg, D.E.: Genetic Algorithms and Walsh Functions: A Gentle Introduction. Clearinghouse for Genetic Algorithms, Department of Mechanical Engineering, University of Alabama (1988)
5. Goldberg, D.E.: Genetic algorithms in search, optimization, and machine learning. Addison-wesley Reading Menlo Park (1989)
6. Holland, J.H.: Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence. U Michigan Press (1975)
7. Kocher, P., Jaffe, J., Jun, B.: Differential power analysis. In: Advances in Cryptology—CRYPTO'99, pp. 388-397. Springer, Heidelberg (1999)
8. Kocher, P., Jaffe, J., Jun, B., Rohatgi, P.: Introduction to differential power analysis. Journal of Cryptographic Engineering 1, 5-27 (2011)
9. Mayer-Sommer, R.: Smartly analyzing the simplicity and the power of simple power analysis on smartcards. In: Cryptographic Hardware and Embedded Systems—CHES 2000, pp. 78-92. Springer, Heidelberg (2000)
10. Messerges, T.S., Dabbish, E.A., Sloan, R.H.: Investigations of power analysis attacks on smartcards. In: USENIX workshop on Smartcard Technology. (1999)
11. Oswald, E.: On side-channel attacks and the application of algorithmic countermeasures. A PhD Thesis in Graz University of Technology, IAIK . (2003)
12. Schlierkamp-Voosen, D.: Optimal interaction of mutation and crossover in the breeder genetic algorithm. In: Proc. of Fifth Int. Conf. on Genetic Algorithms (ICGA-93), Morgan Kaufmann, pp. 648. Citeseer, (1993)
13. Sugawara, T., Hayashi, Y.-i., Homma, N., Mizuki, T., Aoki, T., Sone, H., Satoh, A.: Spectrum analysis on cryptographic modules to counteract side-channel attacks. In: EMC, pp. 21-24. (2009)
14. Tiu, C.C.: A new frequency-based side channel attack for embedded systems. University of Waterloo (2005)
15. Whitley, L.D.: The GENITOR Algorithm and Selection Pressure: Why Rank-Based Allocation of Reproductive Trials is Best. In: ICGA, pp. 116-123. (1989)