Research on Side-Channel Attack Based on the Synergy between SNR and Convolutional Neural Networks

Zixin Liu¹ and Ting Zhu²

¹ Software College, East China University of Technology, Nanchang 330000, China.
² Foreign Languages Institute, Nanchang Institute of Technology, Nanchang 330000, China.
Email: yuhaitan@126.com

Abstract. The data encryption process in the encryption chip will be leaked by means of timing operation time, probing data collecting operation power and electromagnetic signal interception, which makes side channel attack possible. Nowadays, the research on template creation of template attacks has shifted from Gaussian distribution to the use of machine learning algorithms to create templates. With many parameters, it is difficult to find a suitable network structure. Based on many experimental studies, the experience of a convolutional neural network structure suitable for side channel analysis is summarized and proposed.

1. Introduction
With the popularization of computer science security, side-channel analysis has been a cue to attack the key by using the sensitive information leaked during the operation of the encryption device such as time consumption and power consumption. It has multiple implementations [1-3]. Generally, SPA simple power analysis [4] and DPA differential power analysis [5] can be classified as non-learning attacks. Template attack (TA template attack) [6] can be classified as a learned attack. Among them, the template attack method is mainly divided into two parts: analysis traces to create a template. In terms of attack efficiency, the template attack establishes a power consumption model for leaked information, which requires less attack to succeed. In the template attack, the template is mainly established by the multivariate Gaussian distribution, and the correct key is identified by matching the energy trace with the template of the multivariate Gaussian distribution. In addition to template attacks in this classic way, template attacks also use machine learning algorithms to create templates, such as Bayesian classification algorithms [7], support vector machines [8], and neural networks [9]. The use of neural networks for side-channel analysis in machine learning algorithms mainly includes multi-layer perceptions (MLP, Multi-Layer Perceptions) [10-12] and convolutional neural networks (CNN, Convolutional Neural Networks) [13-14].

Eleonora [13] proposed the effectiveness of CNNs against encryption algorithm attacks with jitter defense. This study shows that jitter defense fails under a convolutional neural network model. This article details the process of designing the "best" convolutional neural network structure and the best network structure parameters. The research method adopted is to adjust only one parameter at a time and fix other parameters. Through multiple training and attacks, find the optimal value of the parameter, then fix the hyperparameter, and then test the next hyperparameter. However, the influence of hyperparameters on the training effect is not independent of each other. An optimal hyperparameter obtained by testing with the remaining parameters given, but after the remaining parameters change, the parameter will no longer be optimal.

DPA CONTEST V2 data will be used to find out the key information leakage and training.
2. Convolutional Neural Networks
In recent years, CNN (Convolutional neural network) research has made great progress. CNN is a multi-layer neural network that generally consists of a convolutional layer, a pooling layer, and a fully connected layer. The convolution layer uses convolution processing to extract local features of the data and extracts more abstract features of the data through multiple convolution layers. The pooling layer is used for data reduction processing. The fully connected layer is a fusion of the features obtained by the convolution and a judgment on the probability of the features.

The structure of the convolutional network is shown in Figure 1. The convolutional layer generally includes multiple feature planes. A window ‘$M$’ on the input layer is used to obtain a feature on the feature layer through a convolution operation. The distance the window ‘$M$’ moves is called the step size. The window is convolved by moving steps to obtain other features on the feature layer. However, not every feature on the feature layer is effective, and a more effective feature can be obtained by pooling the feature ‘$X$’ on the feature layer (as shown in the Pooling Layer in Figure 1). This process also reduces the complexity of the model. The data features extracted by the convolution and pooling layers are finally combined non-linearly in the fully connected layer, and the probability distribution of the data for each category is output.

![Figure 1. CNN briefing](image)

3. Implementation Indexes
The reason why neural network hyperparameters are difficult to find is because the optimal value of hyperparameters will be affected by the settings of other hyperparameters. It needs to traverse all combinations of hyperparameters. Since there are many hyperparameters, more combinations will be. Such traversal experiments are basically not feasible in terms of cost. Two indexes were used in the experiment to measure the effect of model training and the effect of using the model to implement attacks. These two indicators are the verification accuracy of model training and the entropy of guessing during attack. The validation accuracy of model training is a commonly used indicator in machine learning, which expresses the model's ability to classify data. This indicator in the traditional machine learning field (such as image recognition) often directly reflects the model's ability to achieve its application goals (such as the ability to identify the category of the image).

But in side-channel attacks, data classification is not the goal. The goal is the ability to attack keys using a model. This ability is expressed using guessing entropy. CNN’s each layer uses a relu activation function and an average pooling function. The structure contains many parameters. In a convolutional neural network, the more parameters the network has, the easier it is for the parameters to overfit. The so-called overfitting means that the network model is very effective in identifying the training data, but the generalization ability is poor, and the ability to identify and verify the data is very poor. It is shown that with the progress of training, the accuracy of the training set continues to increase, and the loss continues to decrease, while the accuracy and loss of the validation set show the opposite trend. The training hyperparameters are affected by this, using a very small learning rate (1e-3) and a small training batch (500 traces per batch). This is because the small learning rate makes the occurrence of overfitting slower and the model degradation slows down. Using small batch training is an effective experience in deep learning to prevent overfitting Figure 2, Figure 3. However, this experience is not applicable in the training of energy trace data with very low SNR(Signal-to-Noise Ratio).
Ratio). Small training batches are more difficult to find the correct gradient optimization direction because the amount of data in each training is too small. Due to these factors, the training accuracy of CNN-best is very poor.

![Figure 2. CNN accuracy](image1)

![Figure 3. CNN cross entropy](image2)

4. Side-Channel Attack Based on CNN

4.1. Experimental Data

In the DPA Contest V2 data set, each power consumption curve is 3253 sample points, the bandwidth is 5 GHz, and the sampling rate is 5G sample/s. The FPGA device runs the AES-128 encryption algorithm in parallel at 24 MHz. The Public data set is a training data set, including 32 random keys, and each round of keys is 20000 random plaintexts. The Template data set is attack data including 32 fixed keys.

4.2. Signal-to-Noise Ratio and Guessing entropy

Signal-to-Noise Ratio (SNR), \( SNR = \frac{Var(E(signal))}{E(Var(signal))} \). The lower the signal-to-noise ratio is, the less features in the energy consumption are included, which denotes the leaking of the power curve.[11] Among them, MSNR represents the sampling point with the largest signal to noise ratio. At the same time, GE (Guessing entropy), Where \( p(K) \) is the score for guessing the key and \( p(K^*) \) is the score for the correct key. \( GE = |K \in \kappa | p(K) > p(K^*)| \). Figure 4 depicts the SNR.
4.3. Attack Process
The trace was pre-processed using regularization to increase the signal-to-noise ratio. In order to reduce the computational complexity, 2500-3000 sampling points (initial data is 0-3253 sampling points) are selected for training. In the process of neural network training, after calculating the intermediate value of the AES-128 algorithm, the validity of the side channel attack is checked by the guessing entropy of the CNN.

5. Result
The trace is filtered by normalization, the pre-processed SNR (Table 1) and attack guessing entropy is compared (Table 2).

- Attack object: Template data set (attack data containing 32 fixed keys).

| Key  | Key0  | Key4  | Key8  | Key12 | Key1 |
|------|-------|-------|-------|-------|------|
|      | 27.43%| 33.22%| 88.83%| 35.22%|
| Key1 | 11.05%| 42.78%| 40.99%| 34.01%|
| Key2 | 37.19%| 34.44%| 36.72%| 12.07%|
| Key3 | 47.38%| 24.63%| 70.46%| 62.26%|

| Pre-processing | 1rd | 2nd | 3rd | 4rd | 5rd | 6rd | 7rd | 8rd | 9rd | 10th |
|----------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| None           | 119.8 | 99.5 | 84.6 | 72.8 | 62.9 | 55.3 | 55.3 | 52.5 | 52.1 | 44.9 |
| MLP            | 79   | 53.5 | 31.5 | 29.7 | 28.3 | 13.9 | 12.5 | 14.9 | 14.5 | 13.7 |
| CNN            | 93.2 | 85.4 | 59.2 | 48.6 | 43.8 | 38.4 | 37.6 | 37.5 | 28.3 | 8.6  |

6. Conclusion
The attack parameters in CNN optimizing cannot be optimized fully, each dataset has unique parameter setting, we will continue finding the best parameter in side-channel attack. Different pre-processing methods have different effects on the data set, because of different noise distributions. In the following time, the parameter optimization in the side channel analysis of the hardware device will also be continued.

7. Acknowledgments
This work is supported in part by Optimization of GIS network architecture based on cloud computing 2.0, JETRCNGDSS201904.
8. References

[1] DU Z B, WU Z, WANG M, et al. Improved chosen-plaintext power analysis attack against SM4 at the round-output [J]. Journal on Communications, 2015.

[2] WU Z, WANG M, RAO J T, et al. Mutual information power analysis attack of HMAC based on SM3[J]. Journal on Communications, 2016.

[3] DU Z B, WU Z, WANG M, et al. Power analysis attack of dynamic password token based on SM3. Journal on Communications, 2017.

[4] KOCHER, P. C., Timing attacks on implementations of Diffie-Hellman, RSA, DSS, and other systems [C]. Annual International Cryptology Conference, 1996.

[5] KOCHER, P. C., Introduction to differential power analysis and related attacks [J]. http://www.cryptography.com/dpa/technical/index.html, 1998.

[6] CHARI, S., RAO, J. R., and ROHATGI, P., Template attacks [M]. Cryptographic Hardware and Embedded Systems-CHES 2002. ed: Springer, 2003.

[7] PICEK, S., HEUSER, A., and GUILLEY, S. Template attack versus Bayes classifier [J]. Journal of Cryptographic Engineering, 2017.

[8] BARTKEWITZ, T. and LEMKE-RUST, K., Efficient template attacks based on probabilistic multi-class support vector machines: Springer, 2013.

[9] MARTINASEK, Z., HAJNY, J., and MALINA, L., Optimization of power analysis using neural network [C]. International Conference on Smart Card Research and Advanced Applications, 2013.

[10] Zdenek Martinasek, Petr Dzurenda, and Lukas Malina. Profiling power analysis attack based on MLP in DPA contest V4.2. In 39th International Conference on Telecommunications and Signal Processing, TSP 2016.

[11] Zdenek Martinasek, Jan Hajny, and Lukas Malina. Optimization of power analysis using neural network. In Francillon and Rohatgi.

[12] Zdenek Martinasek, Lukas Malina, and K. Trasy. Profiling Power Analysis Attack Based on Multi-layer Perceptron Network. Computational Problems in Science and Engineering, 2015.

[13] Eleonora Cagli, Cecile Dumas, and Emmanuel Prouff. Convolutional neural networks with data augmentation against jitter-based countermeasures - profiling attacks without pre-processing. In Wieland Fischer and Naofumi Homma, editors, Cryptographic Hardware and Embedded Systems - CHES 2017 - 19th International Conference, Taipei, Taiwan, September 25-28, 2017.

[14] Benadjila R, Prouff E, Strullu R, et al. Study of Deep Learning Techniques for Side-Channel Analysis and Introduction to ASCAD Database [J]. ANSSI, France & CEA, LETI, MINATEC Campus, France. 2018.