120.147 Efficient Electromagnetic Side Channel Analysis by Probe Positioning using Multi-Layer Perceptron*

Anupam Golder[1], Baogeng Ma[1], Debayan Das[2], Josef Danial[2], Shreyas Sen[2], Arijit Raychowdhury[1]
[1]School of Electrical and Computer Engineering, Georgia Institute of Technology, USA
[2]School of Electrical and Computer Engineering, Purdue University, USA
anupamgolder@gatech.edu

ABSTRACT

In this work, we investigate a practical consideration for Electromagnetic (EM) side-channel analysis, namely, positioning EM probe at the best location for an efficient attack, requiring fewer traces to reveal the secret key of cryptographic engines. We present Multi-Layer Perceptron (MLP) based probe positioning and EM analysis method, defining it as a classification problem by dividing the chip surface scanned by the EM probe into virtual grids, and identifying each grid location by a class label. The MLP, trained to identify the location given a single EM trace, achieves 99.55% accuracy on average for traces captured during different acquisition campaigns.

CCS Concepts

• **Security and privacy → Embedded systems security**; **Side-channel analysis and countermeasures**;

KEYWORDS

EM Probe Positioning, Side-Channel Analysis, Multi-Layer Perceptron, Correlation Analysis

1 Introduction

When an Integrated Circuit (IC) is powered on, current flows between control, memory, I/O, and other data processing units during operations, causing a variation of EM field surrounding the chip [2], which can be picked up by inductive probes. Since the demonstration of first successful attack on Data Encryption Standard (DES) and Rivest-Shamir-Adleman (RSA) algorithm using localized EM radiations in [8], there has been an ever increasing interest on exploiting EM side-channels to reveal secret keys of cryptographic engines. Starting from Correlation EM Analysis in [8], much more powerful attack methods have been proposed, including practical template attack [7], and other profiling attacks based on deep learning techniques [15], [5], [3], making the attacks even more powerful.

Compared to Power Analysis, EM analysis is contactless, non-intrusive, and permits targetting specific locations [11]. For power analysis attacks, the most common approach is to insert a small-valued resistor in the power line, requiring modification of the Power Delivery Network. On the other hand, EM attack does not require any such modification. The downside is that EM measurements are much noisier compared to power measurements [8]. Also,

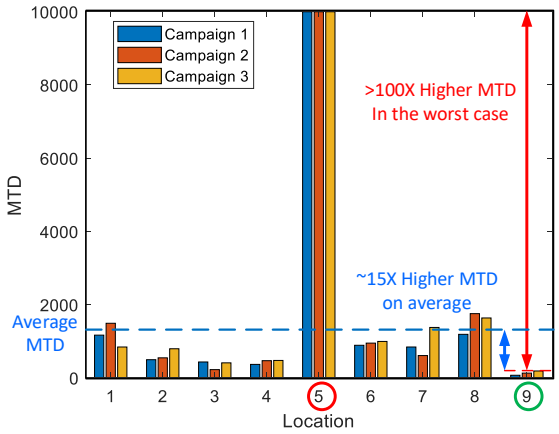

Figure 1: Location Dependency of MTD: MTD tends to be different at different locations, and almost the same at the same location, as observed for EM traces captured from 9 different locations around the targeted chip in 3 different acquisition campaigns, where location No. 5 (marked in red) always had the worst MTD, and location No. 9 (marked in green) had the best MTD. The average MTD across all locations is 15×, and in the worst case, >100× higher than the lowest MTD.

it adds a spatial degree of freedom to the EM wave capturing experiment, because EM radiations are highly location dependent, as illustrated in Fig. 1, which can both be advantageous (if the best location, such as, #9 in Fig. 1 is chosen), or disadvantageous (if the worst location, such as, #5 in Fig. 1 is chosen) to launch an attack.

But positioning the probe identically at the same location can be quite challenging [14], as small misalignment can result in significant differences in captured traces. In the scenarios, where security evaluation labs want to know the leakage location for their chips, or to evaluate the software/hardware countermeasures designed to thwart attacks, a probe positioning method based on already known location-dependent traces can be quite useful. Also, for an adversary, the best location to capture EM traces is of paramount importance, as at some locations, the required number of traces to reveal the secret key, usually reported as Minimum Traces to Disclosure (MTD) can be quite high (Fig. 1).

*This work was supported in part by the National Science Foundation (NSF) under Grant CNS 17-19235, and in part by Intel Corporation. Anupam Golder is also supported by the National Science Foundation (NSF) under Grant CNS 16-24731 - Center for Advanced Electronics through Machine Learning (CAEML) and its industry members.

Probe positioning using Neural Networks has recently been reported in [14] where the problem has been formulated as a regression problem. Compared to [14], the key contributions of this work are:

- In this work, the problem of location detection for probe positioning is formulated as a Classification problem instead of a Regression problem, from an adversary's perspective, where the ultimate goal is to obtain high Signal-to-Noise Ratio (SNR) signals, which reduces the attack duration approximately by 1.3-100 times.
- A Multi-layer Perceptron (MLP) has been trained to identify the probe position from just a single EM trace with >97.95% accuracy without any pre-processing, which has been verified for traces captured during 3 different acquisition campaigns on a 6×6 virtual grid covering the whole chip area.
- A multiple-trace majority voting strategy has been outlined to account for occasional misclassifications, which, with just 5 traces, can achieve 100% accuracy with the help of the trained MLP, in a simulated setting.
- High accuracy of MLP in location detection has been explained by high SNR of the EM traces for probe positioning problem, and rationale behind higher/lower MTDs observed at different locations has been related to separation between mean values for different Hamming Weight (HW) classes.

2 Background and Related Works

Most widely adopted method for probe positioning is visual positioning using microscope, and camera arrangement or a camera/laser combination [12], which takes a long time for accurate positioning. Manual probe positioning suffers from precision issue. Another way of probe positioning is to capture traces from each location and correlating them to a profiled trace captured previously to find the same location again, but this requires an exhaustive search over all possible locations, as pointed out in [14], thus making it a slow process. A better way to further improve this profiling is to profile the traces from the locations using machine learning techniques. In [14], a Convolutional Neural Network (CNN)-based regression problem approach has been adopted to scan and measure localized EM traces close to the surface of a decapsulated chip using micro-EM probe.

Although recapsulation of the chip is possible for industrially-equipped adversaries, as suggested in [8], we point out that an adversary may not have a very wide window of access to a target chip, or may not want to leave a mark behind, which might make decapsulating a chip prohibitive. Moreover, EM attack sensors [10] can be developed for secure IC, which can detect an approaching micro-EM probe in close proximity, but does not work if the probes are more than 200 μm away from the chip. This calls for evaluation of effectiveness of probe positioning method for probes positioned above the surface of the package, which may also have very large diameters compared to micro-EM probes. Larger diameter probes are frequently used in security evaluation of state-of-the-art secure ICs, for example, a probe with 0.4 inch (10-mm) loop diameter has been used in [16].

Machine Learning algorithms, specifically, Neural Networks has gained attraction from researchers coming from a wide variety of backgrounds, due to its success in computer vision and pattern recognition problems. In the Side-Channel analysis community, Neural Networks have been shown to be quite successful, exceeding the performance of Template Attacks [7], and working even in cases of severe misalignment, jitter [4], and masking countermeasures [9]. Two popular forms of Neural Networks, widely used in classification and regression problems are MLP, and CNN. In MLP, there can be one or more hidden layers between input and output layer, which are usually fully connected. Apart from the input, output, and hidden layer, several other layers can be used, such as, Batch normalization Layer, to re-normalize the data from the preceding layer to the next layer, Dropout Layer to randomly drop out outputs of several neurons from the preceding layer during training, thus aiding in generalization. Also, L2 regularization can be used to ensure that the weights do not grow exceptionally large. Non-linear activation functions, such as Rectified Linear Unit (ReLU) and Sigmoid, enable non-linear mapping between input and output. For classification problems, most widely used loss function is categorical cross-entropy, in situations where each sample can be labeled to one class only. During training, back propagation of errors through the layers adjusts weights to minimize loss with the help of optimizers such as Stochastic Gradient Descent with Momentum (SGDM), Adam, or RMSProp. Hyperparamaters, such as number of hidden layers, number of neurons, activation function, regularization methods, choice of optimizer, number of epochs, learning rate etc. have to be fixed prior to training, and optimized during training phase over several iterations..

Correlation EM Analysis proceeds by computing the correlation coefficients of individual time samples to hypothetical power consumption calculated by assuming a leakage model and using a known value (e.g., plaintext), and a guessed value (e.g., keybyte). Most common leakage model for microcontroller-based implementations is the HW model. We can evaluate if an implementation is vulnerable to Correlation Analysis by collecting a lot of traces, and computing the MTD.

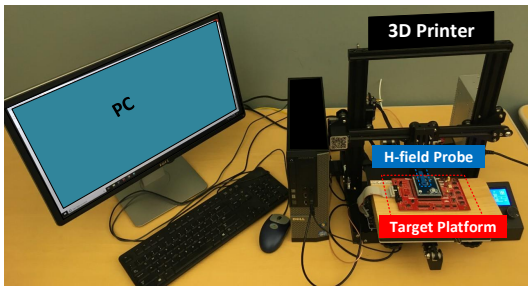

Figure 2: Experimental Setup shows the components: 3D printer for XY-scanning of H-field around the chip, an H-field probe, a Target board with trace capture hardware, and a PC to communicate with the capture hardware and 3D printer.

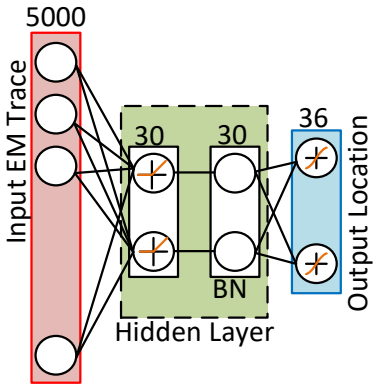

Figure 3: MLP Architecture: Input Layer takes in the raw EM traces without any pre-processing, and the Output Layer (activated by softmax) predicts the location. A single hidden layer is formed by 30 Neurons (with ReLU activation), followed by Batch Normalization Layer, and Dropout Layer with 50% dropout (not shown).

3 Probe Positioning with MLP

In this section, we elaborate the probe positioning method using MLP, its training method, accuracy, and the rationale behind such high accuracy.

3.1 Experimental Setup

To run experiments and validate the idea, we chose CW308T-XMega, an 8-bit Atmel AVR XMega128 microcontroller based Side-Channel Analysis (SCA) target from ChipWhisperer [13], running a software implementation of unprotected Advanced Encryption Standard (AES)-128 encryption algorithm at 7.37 MHz clock frequency. Unlike [14], we do not expose the die with nitric acid, as the goal is to exploit the EM radiation without any modification to the chip surface. We take our measurements from about 1 mm above the chip surface. The capture hardware accompanying the setup samples the EM waves at 4 times the clock frequency of the target, i.e., 29.48 MHz. EM waves were measured by a magnetic field (H) probe with loop diameter of 10 mm from TBPS01 EMC Near-Field Probe Set which is followed by a wideband low-noise amplifier with 40 dB gain in the passband. For XY scanning, we use Comgrow Creality Ender3, an affordable 3D printer costing 220 USD, as a low-cost alternative to probe station. The precision of the scanner is 100 μm. Fig. 2 illustrates the complete experimental setup.

3.2 Acquisition Campaigns and Splitting of Training and Test Sets

Fig. 9(a) shows the virtual 6×6 grid which we have used to capture EM traces around the chip covering an area of 12mm×12mm. We have captured EM traces from these 36 locations in 3 separate acquisition campaigns. The reasoning behind running 3 separate campaigns is to prove time-invariance of the learned model. In each campaign, the probe traversed through each of the 36 locations, and captured 10,000 EM traces with 5000 samples each from each location for random plaintexts (thus enabling subsequent MTD analysis) and fixed key. The reason behind collecting a huge number of traces from each location is to test the robustness of the idea. Then we have split the total number of traces in two parts:

(1) 3 separate training and validation sets for 3 acquisition campaigns, with 36,000 traces in each (10% of total number of traces), 1/9th of which has been used for validation.

(2) 3 separate test sets for 3 acquisition campaigns, with 324,000 traces in each set (90% of total number of traces).

3.3 MLP Architecture

We do not use Batch Normalization or averaging as pre-processing steps on the raw input traces, which has been done in [14]. In [14], 20 traces have been averaged to obtain one trace for both training and test sets, which should increase SNR. As the operating frequency is relatively low in our setting (hence it does not get affected by jitter), and the traces are already perfectly aligned, we adopt Multi-Layer Perceptron as our choice of machine learning algorithm.

MLP architecture used in this work is shown in Fig. 3. The Input Layer consists of 5000 neurons, which is equal to the number of samples in the raw EM traces. The Output Layer has 36 neurons, corresponding to the number of locations around the chip, where the traces have been captured from. The activation function chosen for this layer is softmax. A single Hidden Layer between Input and Output Layer has been used with ReLU activation, followed by Batch Normalization Layer and a Dropout Layer with percentage dropout of 50%.

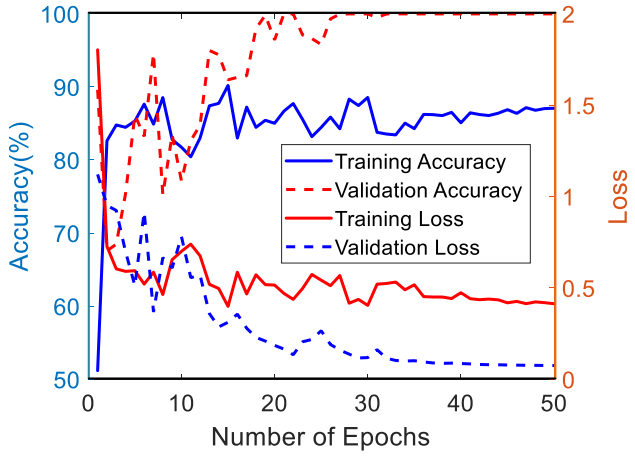

Figure 4: Training History of MLP: Accuracy and Loss; Higher accuracy and Lower loss for validation set compared to training set can be attributed to Dropout Layer, which is only activated during training.

3.4 Training of MLP

We have used the training and validation sets (comprising a total of 36,000 traces) to train the MLP. Adam optimizer with a learning rate of 0.005 and a batch size of 256 has been used to train the

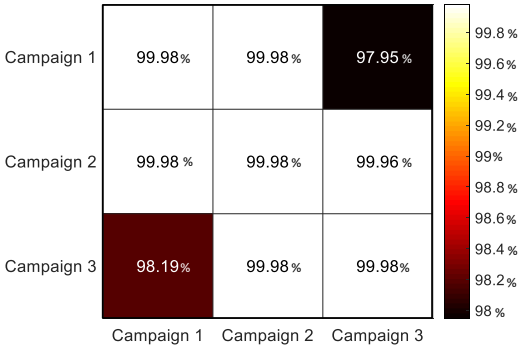

Figure 5: Accuracy Heatmap: obtained by training the MLP with traces from one campaign and testing on a separate test batch, including the test batches from the other two. For the same campaign, the test set does not include any trace from the training set.

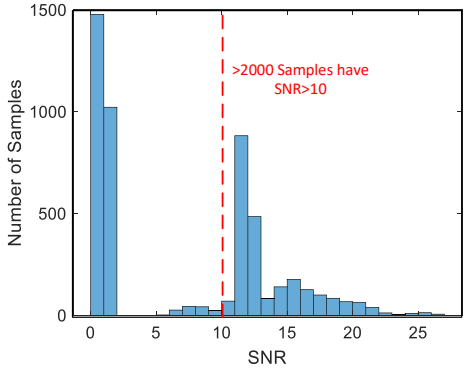

Figure 6: Histogram of SNR: SNR of the traces for location detection scenario has been computed using Eqn.(1). We can see that >2000 samples (more than 40%) have SNR>10. Also, it is apparent that SNR degradation is different for different time samples.

network for 50 epochs. Number of neurons in the hidden layer has been optimized by searching over a range from from 10 to 50. It has been observed that for more than 30 neurons, the test accuracy for the same campaign saturates to 99.98%. As such, for all results mentioned in this paper, 30 neurons have been chosen for the hidden layer. The models have been developed in Python using keras [6] with tensorflow [1] backend. Fig. 4 shows the training progress of the MLP for 50 epochs. The validation accuracy is higher and validation loss is lower than respective metrics for the training set due to the dropout layer, which is activated while calculating training accuracy and loss, and turned off during testing.

3.5 Performance of MLP

We have tested performance of the MLP for the traces captured in different acquisition campaigns. We have used the already separated test set, while evaluating test accuracy for the same campaign. The resulting accuracy metrics are shown in Fig. 5, from which we can see that the lowest accuracy is 97.95% and occurs when the model is trained with acquisition campaign 1 and tested on acquisition campaign 3. The maximum accuracy is 99.98%, and the average accuracy across all training scenarios and campaigns is 99.55%.

3.6 Explanation and Implication of High Accuracy in Location Detection

Such high accuracy mentioned in the previous section for different acquisition campaigns, on one hand, proves the success of the devised method, but a plausible explanation is necessary. To investigate into that, we have computed the SNR, SNR_i for each sample, $traces_{i,i=1,2,\ldots,5000}$ of the 5000-samples long trace from a training set of 36,000 traces for 36 different locations using the equation:

$$SNR_i = \frac{Var[E[trace_i|location]]}{E[Var[trace_i|location]]} \ldots\ldots\ldots(1)$$

where E[.] denotes expected value and Var[.] denotes the variance. The resulting SNRs have been summarized in the histogram plot in Fig. 6. We can see that, out of 5000 samples, more than 2000 samples have SNR>10 for this classification problem. This explains the high accuracy of the trained model in identifying the location for unseen traces. This also means that the traces at different locations across the chip are very much different, suggesting that a profiling attack will be very hard to implement if the profiling location and the attack location are not the same.

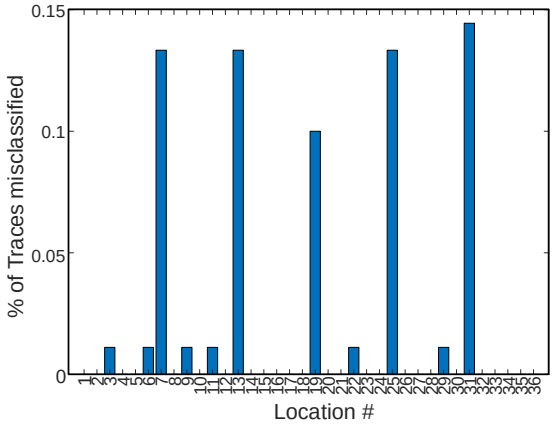

Figure 7: Misclassified labels: For test traces of Campaign 3, the highest percentage of misclassified trace for any location stays below 0.15%.

3.7 Improving Probe Positioning Accuracy using Multiple Traces

In the previous sections, we have focused on location detection using only a single EM trace. In this section, we further investigate into the misclassified labels, to see if any location has a dominance

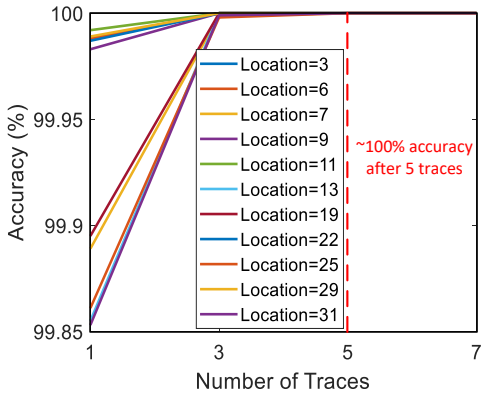

Figure 8: Multi-Trace Accuracy: Uniform random sampling from 9000 test traces per location for locations which have lower than 100% accuracy in Fig. 7, and computing the majority voting for multiple traces reveals that 5 traces are enough to reach an accuracy level of 100% **(For test traces of Campaign 3, experimented in a simulated setting).**

over the errors. Fig. 7 shows that even in the worst case, a single location has only about 0.15% error rate for test traces in campaign 3, when the model is trained on traces from campaign 3, which suggests that if we capture multiple traces, the accuracy is likely to improve. To test this strategy, we create a simulated environment, where we randomly sample 1,3,5,or 7 traces from the test traces uniformly, for the locations which have <100% test accuracy, and run this test 100,000 times. For each run, we compute the majority label, and compare it to the true label, and average the accuracy over the 100,000 runs. Fig. 8 shows that such multi-trace majority voting strategy quickly converges to 100% accuracy, requiring at best 5 traces. This multi-trace strategy does not require prohibitively large number of traces, thus remains feasible. This can always compensate for the errors produced by the trained models, as long as, the required number of traces does not invalidate the benefit of such a probe positioning method.

4 Attack Efficiency Enhancement with Probe Positioning

In this section, we show that this probe positioning can effectively increase the efficiency of an attack, and provide a reasoning behind such improvement by choosing the best location.

4.1 Attack Model

For this experiment, we assume that the adversary is a weak adversary, thus does not have access to industry-level equipment. Also, we show a non-profiled attack, based on Correlation analysis, instead of a profiled attack. The assumption is that the adversary can profile the location-dependent EM traces on a profiling device, and can also perform a Correlation EM Analysis based on HW leakage model of the SBox output, to produce a map of MTDs, as shown in Fig. 9. But such a method would not work if the MTDs significantly vary for each location over time. We investigated if they remain more or less consistent across different acquisition campaigns, and Fig. 9(c)-(d) show that such time-invariance is present, as seemingly, there is not much difference in MTD maps for 3 different campaigns, which ensures that this will be a feasible approach. It is evident from the figure that the lowest MTDs are at the bottom right locations, and higher MTDs are along the regions from bottom left corner to the top right corner. The highest MTD is for location (X,Y) = (3,4) in the virtual grid. At that location, we could not reveal the secret key even after capturing 10,000 traces.

In a sense, this attack model uses profiling, but not to reveal the secret key, but to identify the current location, and to go to the best location to capture the trace. In the attack phase, the positioning system will capture a few trace from a random location around the chip, identify the location, find its position in the learned co-ordinate system of the profiling device, and move to the best location to launch an attack.

4.2 Attack Efficiency

The average MTD across all locations for campaign1 is 1595, whereas the MTD at the best location is 103 on average, and at the worst location is >10000. At the second best location, the average MTD is 131. This translates to 1.3-100× improvement in efficiency of the attack.

In our experimental setting, it takes about 25 milliseconds to capture a single trace comprising of 5000 samples. So, at the best location, the attack would require approximately 2.5 seconds, to capture about 100 traces. Thus lower number of traces to be captured directly translates to less time required for an adversary to perform an attack. The motivation behind a perfect positioning instead of a coarse visual positioning is that, as can be seen in Fig. 9, a slight movement of the probe can result in significant increase in MTD, due to the fact that, the best location and the worst location to launch an attack are not further apart.

4.3 Explanation of variation of MTD with location

To further analyze variation of MTD with location, we chose to analyze two of the extremes, namely, the best location, and the worst location. In this analysis, we have tried to find the samples which provide the highest separation between the 9 Hamming Weight class means in each of the aforementioned cases. The reasoning behind this analysis is that, the sample with the highest separation between the class means is the sample which most frequently shows up as the highest correlating sample in the traces, while calculating MTD. We have observed that such a sample is different for the two different locations (Sample #2056 and Sample #284 for the best and the worst location, respectively).

In Fig. 10(a)-(b), we see the fitted Gaussian distributions for the aforementioned samples, when they have been categorized into their respective Hamming Weight classes. Such distributions conform to our intuition that, the hamming weight values of 4 and 5 occur most frequently, and 0 and 8 the least frequently, for uniformly sampled random plaintext values and a fixed key. From Fig. 10(a), we can see that Hamming Weight class means are well-separated at the best location, which suggests that as we increase the number of traces, the sample means will converge to the true

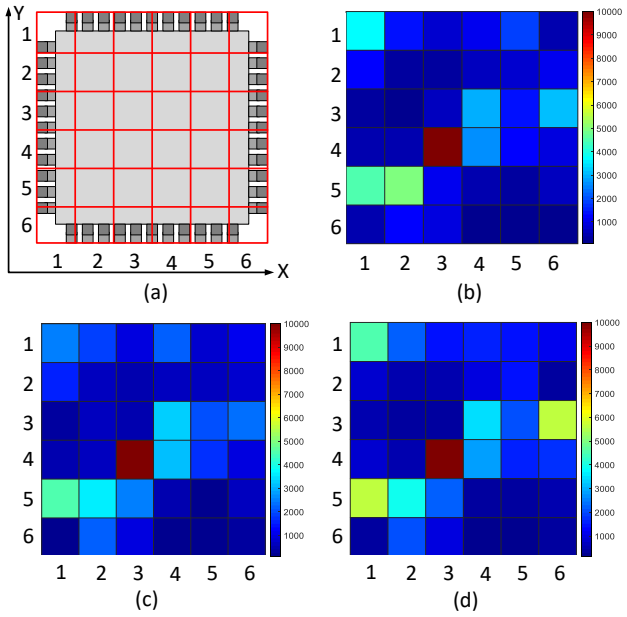

Figure 9: (a) shows Virtual Grid around Chip. (b)-(d) show the MTD map for 3 different acquisition campaigns. This illustrates that the MTD map remains almost the same with respect to time.

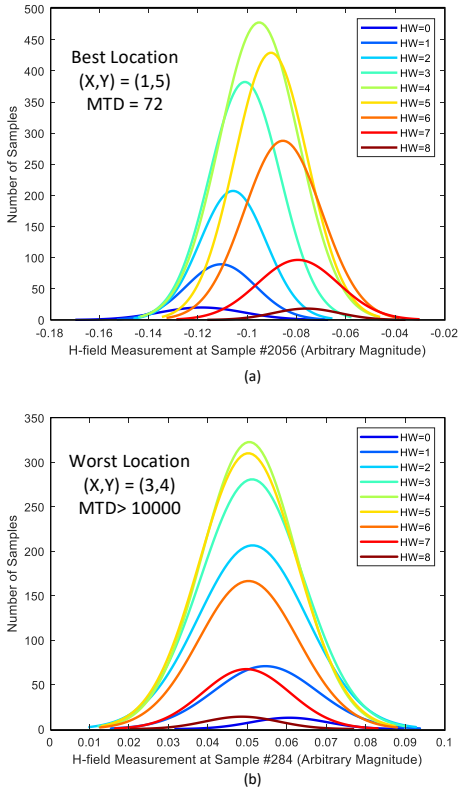

Figure 10: Fitted Gaussian Distributions of samples with different Hamming Weights from measured data for the best class separating sample at the (a)Best location (b)Worst location. This reveals that MTD increases at the worst location due to severe overlapping between distributions of different Hamming Weights.

mean, and the correct key will emerge. Also note that, the means decrease linearly from Hamming Weights 0 to 8. This suggests that a Hamming Weight leakage model for the target platform was justified. On the other hand, from Fig. 10(b), we can see that, at the worst location, the distributions completely overlap, and this is why, even after 10,000 traces, the correct key was not revealed.

5 Conclusion And Future Work

In this work, we have thoroughly investigated the location dependency of EM traces, and by leveraging that fact, showed how an MLP-based probe positioning method may aid an adversary to launch an efficient attack, requiring fewer traces than the choice of a random location, and hence, shorter time. However, we admit that, even this scenario is idealistic, because the traces will be very different if the probe height from the chip is changed, or the probe positioning does not start from one of the locations of the virtual grid.

Although there is a considerable amount of interest in profiling attacks due to their more powerful nature, we note that, they will also be susceptible to such location dependency of EM traces. One way to fix this issue is to position the probe at the best location each time. Another way can be to make a location-invariant profiling model, which will be a very interesting direction for future work, but based on our analysis, is suspected to be a very hard problem.

References

[1] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, et al. 2016. Tensorflow: a system for large-scale machine learning. In *OSDI*, Vol. 16. 265–283.

[2] D. Agrawal, B. Archambeault, J. Rao, and P. Rohatgi. 2002. The EM side—channel (s). In *International Workshop on Cryptographic Hardware and Embedded Systems*. Springer, 29–45.

[3] R. Benadjila, E. Prouff, R. Strullu, E. Cagli, and C. Dumas. 2018. Study of Deep Learning Techniques for Side-Channel Analysis and Introduction to ASCAD Database. Cryptology ePrint Archive, Report 2018/053.

[4] E. Cagli, C. Dumas, and E. Prouff. 2017. Convolutional neural networks with data augmentation against jitter-based countermeasures. In *International Conference on Cryptographic Hardware and Embedded Systems*. Springer, 45–68.

[5] M. Carbone, V. Conin, M. Cornélie, F. Dassance, G. Dufresne, C. Dumas, E. Prouff, and A. Venelli. 2019. Deep learning to evaluate secure RSA implementations. *IACR Transactions on Cryptographic Hardware and Embedded Systems* (2019), 132–161.

[6] F. Chollet et al. 2018. Keras: The python deep learning library. *Astrophysics Source Code Library* (2018).

[7] P.A. Fouque, G. Leurent, D. Réal, and F. Valette. 2009. Practical electromagnetic template attack on HMAC. In *International Workshop on Cryptographic Hardware and Embedded Systems*. Springer, 66–80.

[8] K. Gandolfi, C. Mourtel, and F. Olivier. 2001. Electromagnetic analysis: Concrete results. In *International workshop on cryptographic hardware and embedded systems*. Springer, 251–261.

[9] R. Gilmore, N. Hanley, and M. O'Neill. 2015. Neural network based attack on a masked implementation of AES. In *2015 IEEE International Symposium on Hardware Oriented Security and Trust (HOST)*. IEEE, 106–111.

[10] N. Homma, Y. Hayashi, N. Miura, D. Fujimoto, D. Tanaka, M. Nagata, and T. Aoki. 2014. Em attack is non-invasive?-design methodology and validity verification

of em attack sensor. In *International Workshop on Cryptographic Hardware and Embedded Systems*. Springer, 1–16.

[11] J. Longo, E. De Mulder, D. Page, and M. Tunstall. 2015. SoC it to EM: electromagnetic side-channel attacks on a complex system-on-chip. In *International Workshop on Cryptographic Hardware and Embedded Systems*. Springer, 620–640.

[12] N. Mai-Khanh, T. Iizuka, S. Nakajima, and K. Asada. 2017. High-sensitivity micro-magnetic probe for the applications of safety and security. In *2017 7th International Conference on Integrated Circuits, Design, and Verification (ICDV)*. IEEE, 10–15.

[13] C. O'Flynn and Z.D. Chen. 2014. Chipwhisperer: An open-source platform for hardware embedded security research. In *International Workshop on Constructive Side-Channel Analysis and Secure Design*. Springer, 243–260.

[14] B. Richter, A. Wild, and A. Moradi. 2019. Automated Probe Repositioning for On-Die EM Measurements. Cryptology ePrint Archive, Report 2019/923. https://eprint.iacr.org/2019/923.

[15] P. Robyns, P. Quax, and W. Lamotte. 2018. Improving CEMA using Correlation Optimization. *IACR Transactions on Cryptographic Hardware and Embedded Systems* 2019 (Nov. 2018), 1–24.

[16] A. Singh, M. Kar, V. C. K. Chekuri, S. K. Mathew, A. Rajan, V. De, and S. Mukhopadhyay. 2019. Enhanced Power and Electromagnetic SCA Resistance of Encryption Engines via a Security-Aware Integrated All-Digital LDO. *IEEE Journal of Solid-State Circuits* (2019).