Paradigm Shift of Machine Learning to Deep Learning in Side Channel Attacks -A Survey

Conference Paper · March 2022

DOI: 10.1109/IMTIC53841.2021.9719689

3 authors, including:

Mehwish Shaikh
Mehran University of Engineering and Technology
11 PUBLICATIONS **4** CITATIONS
SEE PROFILE

Salahuddin Saddar
Mehran University of Engineering and Technology
19 PUBLICATIONS **23** CITATIONS
SEE PROFILE

Paradigm Shift of Machine Learning to Deep Learning in Side Channel Attacks – A Survey

Mehwish Shaikh
Department of Software Engineering
Mehran University of Engineering and Technology, Jamshoro.
mehwishshaikh173@gmail.com

Qasim Ali Arain
Department of Software Engineering
Mehran University of Engineering and Technology, Jamshoro.
qasim.arain@faculty.muet.edu.pk

Salahuddin Saddar
Department of Software Engineering
Mehran University of Engineering and Technology, Jamshoro.
salahuddin.saddar@faculty.muet.edu.pk

***Abstract*— A side-channel attack is a sort of computer security attack depending on data obtain from program implementation rather than flaws in the program itself. Additional data sources that might be exploited include timing data, power consumption, electromagnetic leakage, and even sound. Since the 1990's, when side channel attacks were first introduced, a lot of effort is done in improving their effectiveness and efficiency. Many machine learning methods were designed for side-channel attacks. The researchers have recently noticed a growing trend in the SCA community to use deep learning techniques, which has resulted in more precise side-channel studies, even when countermeasures are in effect. Machine learning algorithms, on the other hand, have the drawback of requiring human engineering to function and some performance fluctuations in some cases. Recent research has focused on using deep learning techniques to extract characteristics from data automatically. Concepts of side channel attacks, machine learning, deep learning, and current breakthroughs in deep learning-based side-channel attacks are described in this paper. This discussion provides an overview of contemporary machine and deep learning research in the context of SCA. Machine Learning is not completely negated, or Deep Learning is not completely supported but their competitiveness and trend is discussed as a baseline to new challenges in context of side channel attack. Thus, a paradigm shift of ML to DL in SCAs is the focus of this paper.**

Keywords—side channel attack, machine learning, deep learning

I. INTRODUCTION

A side channel attack is an implementation-specific attack that uses an indirect measurement to get knowledge about the secret value utilized during calculation by exploiting the fact that different inputs cause the algorithm implementation to behave differently. Any attack that retrieves a secret value via indirect measurements of a calculation such an auxiliary method is termed as side-channel attack. The power attack used to gather information about the executed instructions during the computation of an embedded CPU, the amount of time it took to perform the calculations (timing attacks), sound, written content is revealed by acoustic attacks on keyboards, and the electromagnetic emissions created are all examples of popular indirect measures.

Figure 1 demonstrates how side channel or information leakage happen. The input/output and a cryptographic operation is performed on the main channel of commodity hardware and with this normal flow, the side channel information is also related e.g., data instruction, timing, power, sound, radiations etc. So having the side channel information in the first place is not a problem but using this side channel information to extract the main channel information is the real problem and this is how the information is leak.

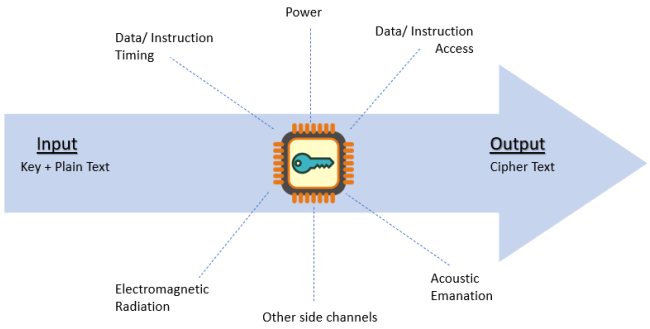

Figure 1 Side Channel Information

Attack channels have received attention since Paul Kocher released the first side-channel attack on multiple public key cryptosystems. SCAs are concerned with how the secret key is recovered in software or hardware.
Countermeasures for side-channel assaults are divided into two categories.

- Reduce or eliminate information leakage.
- Remove link between the confidential information and the leaked information.

Two groups of Side-Channel Attacks [1]:

- Profiling SCAs are effective because they anticipate the ultimate target to fine-tune all of the attack's settings ahead of time. There are two steps to a profiling SCA. The attacker initially obtains a duplicate device and examines the physical release. Second, it attacks the target device with a key-recovery attack. Template attacks and stochastic models are examples of this type of attack (Linear Regression Analysis).
- The non-profiling SCA is a passive opponent and has only access to the physical leaks captured on the target device. In order to recover the secret key used, it performs statistical research to detect the association between the leaking data and this sensitive variable. Differential power analysis, correlation power analysis, and mutual information analysis are non-profiling attacks.

ML is widely used in variant fields like image recognition, robotics, natural language processing and is expected to become even more important in the future for autonomous systems. In addition, ML approaches for SCAs have been

978-1-6654-8294-3/21/$31.00 ©2021 IEEE

used in a great number of articles presented in recent years. In a typical classification task, training examples consisting of input data vectors (features) and related output measurements are presented to the system (labels). During training, the algorithm produces predictions based on the input data and corrects itself if the predictions do not correspond to the predicted labels. The goal is to create a prediction model that generalizes well to new inputs, producing the correct result even when the training data is missing. Unsupervised learning, on the other hand, is used to solve problems for which no outcome labels are available. The learning technique seeks to extract relevant aspects of the data set's underlying structure by clustering the input data set into different groups. Machine learning and deep learning are divided into three categories: supervised learning, which uses data with labels, unsupervised learning, which uses data without labels, and semi-supervised learning, which uses data with and without labels.

Deep learning algorithms have been demonstrated to be particularly useful for evaluating the security of embedded systems in recent study. Many researchers were interested in using deep learning approaches to undertake side-channel analysis because they had strong results with them. In the profiled SDAs, (CNNs) are a deep learning technology that has proven to be a potential paradigm. One of the key advantages of using deep learning for SCA is that it is easy to perform a probabilistic attack that scales with the number of traces used.

II. Survey Overview

The key machine learning and deep learning techniques are briefly discussed to provide background information to the reader that are suggested for side channel attack environment in Section 3. We review the literature on machine learning and deep learning surveys and show an analogy that how trend of machine learning is moving towards deep learning in context of side channel attack in Section 4. The discussion on this shift in Section 5. In Section 6, we also address some key findings and research directions, as well as draw some conclusions.

III. BACKGROUND

In this part, some most widely used machine learning and deep learning algorithms when dealing with side channel attacks are elaborated.

A. MACHINE LEARNING

It is a kind of data analysis in which artificial intelligence develops predictive algorithms. It's a subset of artificial intelligence based on the premise that computer systems can automatically learn, spot patterns, and draw conclusions without human intervention [2].

A.01 Steps of Machine Learning

Following are the steps of machine learning in Figure 2.

Figure 2 Steps of Machine Learning

- Data collection and preparation [3] - From deciding where to collect data to putting it up and preparing it for feature selection and engineering, everything is covered.
- Feature engineering and feature selection - This includes all data alterations from the time it is prepared until it is consumed by the model.
- Selecting a ML method and training it with the purpose of achieving a "better than baseline" outcome that can be improved.
- Evaluating our model entails both metric selection and the actual evaluation.
- Model tweaking, regularisation, and hyperparameter tuning - In an iterative process, we progress from a "just okay" model to our stellar performance in an iterative process.

Machine learning is classified into two categories due to its complexity: supervised and unsupervised learning. Each machine learning algorithm serves a certain purpose and performs a specific task, providing specific outputs and utilizing diverse types of data.

Supervised Learning

Learning is supervised since the data is known. The data is fed into a machine learning algorithm, which trains the model. You can add unknown data to the model after it has been trained with known data to receive a new response. Classification and regression are the two types of supervised learning.

- Classification: Organizing the data into the categories defined on the data set tailored to specific characteristics.
- Regression: Predicting or concluding the data's other features based on the data's available features.

a. Support Vector Machines (SVM)

SVM is used for classifying and predicting data. It uses the kernel trick to transform your data, then uses these modifications to build an ideal boundary between the various outputs. The kernel function used by an SVM is critical to its performance because it determines the updated feature space in which classification will take place. In SCA circumstances, the radial basis and the linear kernel function were both heavily used. were both extensively utilized. Before choosing how to partition your data using the labels or outputs you provide, it performs some extremely complex data transformations.

b. Decision Tree

Internal nodes store dataset information, whereas, branched nodes show decision rules, and each leaf node delivers the result in a decision tree. The dataset's attributes inform the judgments or testing. It's a visual representation of all possible solutions to a problem or decision based on a set of criteria. The C4.5 method is a well-known decision tree building algorithm that was also employed in SCAs. It was created to address several issues that might arise while training a tree, such as generalization and dealing with input that contains missing attribute values. Based on the gain ratio metric, C4.5 selects decision features to divide the tree into multiple branches.

c. Random Forest

To solve a complicated problem and enhance the performance of the model, it employs ensemble learning. Random Forest is a classifier that improves the predicted accuracy of a dataset by averaging the results of numerous decision trees on distinct subgroups of the dataset. The random forest estimates the end-result according to majority votes by integrating the predictions of each tree, rather than relying on a single decision tree. As the number of trees in the forest grows, the model gets more accurate, and overfitting is avoided.

d. K-Nearest Neighbors (KNN)

A non-parametric lazy learning algorithm is the KNN algorithm. It is non-parametric since no assumptions are made. Rather than assuming that the structure of the model is common, it is entirely based on the data presented.

A "lazy learning" algorithm is one that does not make any generalizations. As a result, this method requires very little training to implement. As a result, all of the training data is used in testing when using KNN.

Unsupervised Learning

The training data in unsupervised learning is unknown and unlabeled, implying that no one has ever looked at it before. The input cannot be led to the algorithm without the aspect of known data, which is where the word "unsupervised" comes from. The model is trained using this data, which is input into the Machine Learning algorithm. The trained model tries to find a pattern and respond accordingly. Clustering and association are two further types of unsupervised learning.

- Clustering: Finding data groups that are comparable to one another when the data's intrinsic groupings are unknown.
- Association: Identifying the relationships and connections between data in a single data set.

e. K-Means

It's an iterative clustering method for finding the greatest value for each iteration. First, the number of clusters to be employed must be established. Using this clustering procedure, the data points must be divided into k groups. A higher k, on the other hand, denotes smaller groupings with more granularity. The smaller the k value, the larger the group and the less granular it is.

The algorithm establishes a set of "labels" because of its prior work. Each data point is given to one of the k groups. Each group is determined by identifying a center for each group in k-means clustering. The cluster's heart, the centroids, captures and incorporates the points nearest to them.

f. Hierarchical Clustering

An algorithm that creates a hierarchy of clusters is known as hierarchical clustering. Each piece of data is first assigned to its own cluster. Two clusters that are near together will be in the same cluster in this situation. The procedure comes to an end when there is just one cluster left.

B. DEEP LEARNING

Artificial neural networks are the subject of deep learning, a form of machine learning (ANNs). ANNs are algorithms that are inspired by the brain's structure and function. Processing capacity have often limited the complexity of neural networks [4].

There are three primary benefits to using deep learning for SCA:

1. Working directly on raw data: Instead of depending on human-engineered characteristics and assumptions, models learn directly from raw power consumption or electromagnetic traces. This makes assaults easier to develop, decreases the need for domain-specific expertise, and, like computer vision applications, will likely lead to more efficient attacks in the long run. You don't need to resynchronize the traces or do feature selection, unlike typical attack techniques.
2. Direct Attack point targeting: Without employing approximate models, models can learn to predict specific intermediate values directly. As with the first point, this simplifies the assault design, making it more adaptable and efficient in the end.
3. Natural Probabilistic Attack: Because the model output scores on numerous power traces can be immediately integrated to rank every conceivable byte value from the most likely correct to the least likely, the models' predictions can be utilized to mount an efficient probabilistic assault.

In contrast to machine learning, which requires humans to pick features, deep learning architectures can automatically extract features, leading to more extensive machine learning models.

B.01 Steps of Deep Learning

Following are the steps of deep learning in Figure 3 [5].

Figure 3 Steps of Deep Learning

i. Dataset preprocessing

Dimension reduction occurs during the data preprocessing step, which involves the machine learning models. Because the feature engineering of the data is so complex, the preprocess is essential. When implementing deep learning to profile SCA, the dimension reduction stage can be omitted (except in the case where the dimension is too large) and the training dataset can be directly preprocessed. The training set is used to train the model, whereas the verification data set is used to evaluate the signal.

ii. Model Selection and Training

Various deep neural networks are used to train the models. e.g. MLP, CNN, and RNN

iii. Verification and Assessment

Several assessment frameworks are often used to examine the effectiveness of the operation or to establish the appropriate parameters for the model family. The goal of these

approaches is to provide a precision estimate for a metric, independent of the training and test sets Dtrain and Dtest, but reliant on the scale.

B.02 Models

g. Multi-Layer Perceptron

A feedforward neural network that creates a set of results from a set of input data is known as a multilayer perceptron (MLP). A directed graph is built between the input and output layers of an MLP by numerous layers of input nodes. Back propagation is a technique used by MLPs to train their networks.

h. Denoising Auto-Encoder

An auto-encoder is a neural network that uses backpropagation to closely match input values. The input layer x, hidden layer y, and output layer z make up the three layers of an AE. It reduces the input data x into the hidden layer y to provide a more realistic depiction. After compression, the input and output layers should contain fewer neurons than the hidden layer. The more representation there is, the more important it is, and the network may learn about the data's hidden features as a result. The higher representation is then used by the AE to rebuild the input feature x into the output data z. The entire process of learning is unsupervised.

i. Convolution Neural Network

CNNs are a type of neural network inspired by biological processes in animals' visual brains. Originally, they were intended for two-dimensional convolutions. They're mostly employed for image classification, but they've recently been shown to be effective in classifying time series data like music and speech. Some believe they might be beneficial in side-channel analysis.

j. Recurrent Neural Network

It is an artificial neural network that works with time series or sequenced input. Ordinary feed forward neural networks are designed to deal with data that is unrelated to one another. If we have data in a series where one data point is dependent on the preceding data point, we must modify the neural network to account for these dependencies. RNNs have memory, which allows them to preserve the states or information of previous inputs in order to construct the next output in the sequence.

k. Long Short-Term Memory Network

LSTMs, a type of Recurrent Neural Network, can be used to learn and memorize long-term dependencies (RNN). The human brain has a natural ability to recall information from the past over a long period of time. LSTMs are data-tracking systems that keep track of information throughout time. They're useful for time-series prediction since they remember previous inputs. The four interconnected layers of LSTMs form a chain-like structure that communicates in a unique way.

IV. LITERATURE REVIEW

Countering the side channel attacks analysis also requires the use of machine learning and deep learning techniques. A lot of study has been done in this field, and new approaches have been offered. Machine learning techniques are deployed from 2011 [6,7] and gradually deep learning techniques were adopted in 2016 but the DL is still limited in scope of SCAs. In this section, the literature of recent studies is discussed.

[8] provided an overview of power based SCAs, as well as a comprehensive study of the various machine learning (ML) approaches utilized in SCAs vs cryptographic implementations, also a broad comparison of different ML techniques. In conclusion, machine learning approaches represent a powerful alternative to traditional side-channel evaluation methods. SCAs are complex attacks that rely on information gathered from cryptographic device implementation. Profiling side-channel attacks has greatly increased in recent years [9] because it outlines the worst-case security assumptions. Two notable ways for profiled side channel attacks are machine and template based. Template attacks TA is mainly focused on high-dimensional statistical modelling under Gaussian noise assumptions, whereas ML includes tools for solving problems in more flexible side-channel attack scenarios [10]. Many researchers were interested in using deep learning approaches to undertake side-channel analysis since they were successful. Regardless of the nature of the countermeasures, from a worst-case scenario standpoint, the deep learning model is suited for testing implementations [11]. A concept of DeepSCA is introduced by Shourong Hou [12] that greatly reduces the amount of traces on the side channels necessary to undertake successful attacks on datasets that are significantly desynchronized, outperforming the published optimized CNNs model. Convolutional Neural Networks (CNNs), a deep learning approach has shown to be a credible paradigm in the profiling SCAs. The growth of convolutional neural networks was highlighted by Stjepan Picek et al. [13] They compared the performance of numerous machine learning methods and evaluated scenarios for profiled attacks. Convolutional neural networks showed to be an excellent solution for profiled SCA because of the good results they produced. Only in circumstances where the noise degree is low, measurements and features are large, the CNN architectures outperform ML approaches. The findings of this research imply that machine learning can perform on equivalent with (or even better than) CNNs (at a substantially lower computing cost). The results prefer approaches like Random Forest and XGBoost when using the guessing entropy metric, indicating that more tests need to correctly measure the strengths of convolutional networks. [14] CNN models should be selected in the context of SCA because they are nearly as efficient as MLP networks in the case of perfectly synced observations but outperform them in the presence of desynchronization. Several parametrization possibilities were addressed, as well as a wide range of benchmarks that were utilized to either experimentally validate choices or make the best decision. Batch normalization, dropout, and weight decay were recommended as three strategies to use [15]. It has been demonstrated that by properly combining these strategies, It is feasible to increase attack performance by over 55%, respect to the number of signatures required to retrieve the secret. Deep learning approaches are based on some high stability that make it possible to recover keys

successfully. Different approaches for extracting features (CNN and AE) and exploiting sample time dependency (RNN, LSTM) were used. [16] The DL-based attacks proposed in this study were compared to TA and ML. The evaluation of the number of traces necessary during the attack phase to obtain unity guessing entropy with a fixed amount of profiling data-set was undertaken on three different data-sets. The solution has a clear advantage in breaking both unprotected and protected AES implementations, according to practical results. [17] provided an end-to-end profiling attack technique by using CNN-based attacks against various sorts of misaligned data. In addition, misaligned side channel traces were fitted to two Machine Learning (ML) Data Augmentation techniques. Wang, H says that it is critical to train and test neural network models on traces taken from several boards, as well as employing various implementations of the cryptographic technique under attack. Otherwise, it's all too easy to inflate the trained network's classification accuracy [18].

Table 1 shows paradigm shift of machine learning to deep learning in recent studies as most of the researchers focuses on deep learning and trying to explore it in the context of side channels to prevent attacks from powerful cryptographic applications.

Paper	Determination	ML	DL
[16] (2016)	• Ability of deep learning to recover key using side channels information.		✓
[17] (2017)	• Profiling attack strategy based of CNN • CNNs efficiently manage high-dimensional data		✓
[9] (2018)	• Study of ML methods that performs efficiently in SCA. • Machine and Template based attacks are prevalent types of profiled side channel attacks.	✓	
[14] (2018)	• CNN performance similar to MLP.	✓	✓
[13] (2018)	• On a profiled SCA, machine learning algorithms are compared. • The CNN architecture offers a clear advantage over ML techniques only when the low degree noise, large measurements and features. • Simple ML approaches perform similar at a lower computational cost.		✓
[8] (2019)	• SCAs employ machine learning approaches to counter cryptographic systems.	✓	
[12] (2019)	• Proposed a novel CNNs architecture called DeepSCA		✓
[18] (2019)	• Importance of training and testing traces in order to preserve classification accuracy.		✓
[10] (2020)	• The Negative Log's Minimization when training deep neural networks, the likelihood loss function is used.		✓
[11] (2021)	• The implications when using deep learning techniques on the mitigation of SCA. • Batch normalization, dropout, and weight decay are three techniques presented that have not previously been employed in a side-channel situation.		✓

Table 1 Survey of recent studies of side channel attacks supporting machine learning or deep learning techniques

V. DISCUSSION

This survey aims to discuss machine learning and deep learning competitiveness in terms of side channel attacks. Deep learning is a subclass of machine learning that employs a set of approaches to model architectures with multiple processing layers to represent high-level abstractions in data. Deep learning techniques, according to current study, have outperformed other existing machine learning algorithms in a range of domains, including side channels. This research aims to depict drift towards DL from ML in detection and mitigation frameworks of side channel attacks. Regardless of advantages, DL have some limitations.

- There are certain deep learning models to choose from. Only the algorithms AE, MLP, CNN, and RNN are employed today.
- There isn't enough study done on encryption methods with profiled attacks protection. The structural benefit of the deep learning model is being able to deal with protected encryption algorithms, which can be designed in this regard.
- Only a few studies have been done on non-profiled side channel attacks.

VI. CONCLUSION AND FUTURE DIRECTIONS

Side channel attack is an attack that retrieve secret data using indirect measures. Such measures are countermeasures using machine and deep learning technologies. This research

discusses the background of machine learning and deep learning techniques on side channel attacks, along with the recent studies on ML DL based SCAs which shows paradigm shift towards deep learning due to more accurate and precise results and also tightened in security algorithms for countermeasures of side channel attacks. However, machine learning is still performing better in some cases and deep learning still have limitations in some cases. We can not say ML is superior to DL or vice versa but the studies show growing trends towards deep learning-based side channel attacks.

A future study may involve searching more patterns about deep learning because recent research shows limitations to deep learning-based side channel attacks. Furthermore, no research has been done to yet on using ML DL approaches to overcome devices that have greater side-channel countermeasures. For side-channel analysis, it is suggested that customized learning algorithms be developed. This paper lays the groundwork for latest research into new deep learning techniques in order to improve their challenge adaptation.

REFERENCES

[1] O. Bronchain, F. Durvaux, L. Masure and F. -X. Standaert, "Efficient Profiled Side-Channel Analysis of Masked Implementations, Extended," in IEEE Transactions on Information Forensics and Security, vol. 17, pp. 574-584, 2022, doi: 10.1109/TIFS.2022.3144871.

[2] Alam, M., Bhattacharya, S., & Mukhopadhyay, D. (2021). Victims Can Be Saviors. ACM Journal On Emerging Technologies In Computing Systems, 17(2), 1-31. doi: 10.1145/3439189

[3] M. Shaikh, P. H. Ali Qureshi, M. Shaikh, Q. A. Arain, A. Zubedi and P. Shaikh, "Security Paradigms In SDLC Requirement Phase — A Comparative Analysis Approach," 2021 International Conference on Engineering and Emerging Technologies (ICEET), 2021, pp. 1-6, doi: 10.1109/ICEET53442.2021.9659614.

[4] D. Kwon, H. Kim and S. Hong, "Non-Profiled Deep Learning-Based Side-Channel Preprocessing with Autoencoders," in IEEE Access, vol. 9, pp. 57692-57703, 2021, doi: 10.1109/ACCESS.2021.3072653.

[5] Song, S., Chen, K., & Zhang, Y. (2019). Overview of side channel cipher analysis based on Deep Learning. Journal of Physics: Conference Series, 1213, 022013. doi:10.1088/1742-6596/1213/2/022013

[6] Hou, S., Zhou, Y., Liu, H., Zhu, N.: Wavelet support vector machine algorithm in power analysis attacks. Radioengineering **26**(3), 890–902 (2017)

[7] Lerman, L., Bontempi, G., Markowitch, O.: The bias-variance decomposition in profiled attacks. J. Cryptogr. Eng. **5**(4), 255–267 (2015). https://doi.org/10.1007/s13389-015-0106-1

[8] Hettwer, B., Gehrer, S., & Güneysu, T. (2019). Applications of machine learning techniques in side-channel attacks: a survey. Journal Of Cryptographic Engineering, 10(2), 135-162. doi: 10.1007/s13389-019-00212-8

[9] Batina L., Djukanovic M., Heuser A., Picek S. (2021) It Started with Templates: The Future of Profiling in Side-Channel Analysis. In: Avoine G., Hernandez-Castro J. (eds) Security of Ubiquitous Computing Systems. Springer, Cham. https://doi.org/10.1007/978-3-030-10591-4_8

[10] You, H., & Kocayusufoglu, F. (2018). Side Channel Attack with Machine Learning. koclab. Retrieved from http://koclab.cs.ucsb.edu/teaching/cren/project/2018/You+Kocayusufoglu.pdf

[11] Loïc Masure, Cécile Dumas, Emmanuel Prouff. A Comprehensive Study of Deep Learning for SideChannel Analysis. IACR Transactions on Cryptographic Hardware and Embedded Systems, IACR, 2019, 2020 (1), pp.348-375. ff10.13154/tches.v2020.i1.348-375ff.

[12] Hou, S., Zhou, Y., & Liu, H. (2019). Convolutional Neural Networks for Profiled Side-channel Analysis. Radioengineering, 27(3), 651-658. doi: 10.13164/re.2019.0651

[13] Picek, S. (2018). On the Performance of Convolutional Neural Networks for Side-channel Analysis. Retrieved 12 July 2021, from https://eprint.iacr.org/2018/004.pdf

[14] Benadjila, R. (2018). Study of Deep Learning Techniques for Side-Channel Analysis and Introduction to ASCAD Database. Retrieved 12 July 2021, from https://eprint.iacr.org/2018/053.pdf

[15] Robissout, D., Bossuet, L., Habrard, A., & Grosso, V. (2021). Improving Deep Learning Networks for Profiled Side-channel Analysis Using Performance Improvement Techniques. ACM Journal On Emerging Technologies In Computing Systems, 17(3), 1-30. doi: 10.1145/3453162

[16] Maghrebi, H., Portigliatti, T., Prouff, E.: Breaking cryptographic implementations using deep learning techniques. In: Security, Privacy, and Applied Cryptography Engineering - 6th International Conference, SPACE 2016, Hyderabad, India, December 14-18, 2016, Proceedings. (2016) 3–26

[17] Cagli, E., Dumas, C., Prouff, E.: Convolutional neural networks with data augmentation against jitter-based countermeasures - profiling attacks without preprocessing. In: Cryptographic Hardware and Embedded Systems - CHES 2017 - 19th International Conference, Taipei, Taiwan, September 25-28, 2017, Proceedings. (2017) 45–68

[18] Wang, H., Brisfors, M., Forsmark, S., & Dubrova, E. (2019). How Diversity Affects Deep-Learning Side-Channel Attacks. Retrieved from https://eprint.iacr.org/2019/664.pdf