Feature selection methods for non-profiled side-channel attacks on ECC

Bhasin, Shivam; Ravi, Prasanna; Jungk, Bernhard; Jap, Dirmanto; Najm, Zakaria

2018

Ravi, P., Jungk, B., Jap, D., Najm, Z., & Bhasin, S. (2018). Feature selection methods for non-profiled side-channel attacks on ECC. 2018 IEEE 23rd International Conference on Digital Signal Processing (DSP). doi:10.1109/ICDSP.2018.8631824

<https://hdl.handle.net/10356/103427>

<https://doi.org/10.1109/ICDSP.2018.8631824>

© 2018 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. The published version is available at:
<https://doi.org/10.1109/ICDSP.2018.8631824>

Downloaded on 26 Apr 2023 10:30:00 SGT

Feature Selection Methods for Non-Profiled Side-Channel Attacks on ECC

Prasanna Ravi*, Bernhard Jungk†, Dirmanto Jap*, Zakaria Najm*‡ and Shivam Bhasin*

*Temasek Laboratories@NTU, Singapore

Email: {prasanna.ravi, djap, zakaria.najm, sbhasin}@ntu.edu.sg

†Independent Researcher

Email: bernhard@projectstarfire.de

‡Delft University of Technology (TU Delft)

Abstract—Elliptic curve cryptography (ECC) is a public key cryptosystem which is widely used for different real world applications. With the introduction of side-channel attacks, there is a growing concern regarding the security of such implementations. Indeed, side-channel attacks have been reported to break even the theoretically secure ciphers due to the exploit in the physical leakage. The non-profiled side-channel attacks, especially are considered more serious than the profiled counterpart, as the former can work in almost black box setting. Several attacks have been proposed, however, one of the main issue normally encountered is regarding the selection of relevant features from the side-channel signal. For ECC implementation, normally the side-channel measurements will contain lots of irrelevant points which could hinder the effectiveness of the attack. For profiling scenario, these features can be determined, since the attacker has full knowledge, however, for black box non-profiled setting, this might pose an issue. In this work, we investigate different feature selection approaches to improve the accuracy for non-profiled attacks on ECC. We demonstrate the effectiveness of proposed methods on real measurements from FPGA and microcontroller targets, achieving accuracy comparable to profiled case (88.6% and 98.4% respectively).

Index Terms—Cryptography; Side-Channel Attacks; Feature Selection; Clustering

I. INTRODUCTION

Elliptic Curve Cryptography is based on the difficulty of computing the discrete logarithm problem over elliptic curves (ECDLP). The classical cryptanalysis mainly focuses on efficiency improvements to compute the ECDLP, but so far, there is no efficient classical algorithm to compute the ECDLP for large enough key sizes. On the other hand, in practice, other attacks are more relevant, for instances side-channel analysis (SCA) [1].

SCA normally exploits the physical properties of the implementation. When conducting SCA, the attacker observes one or few specific physical traits of the target implementation such as the time, power consumption or electromagnetic emanations (EM) and derives the key from these observations using statistical methods. For ECC, there are many proposed attacks, which exploit different properties of the algorithms [2].

For most of the attacks, the attacker needs to know the (public) input data, or has to perform a profiling step with known secret inputs to successfully derive a secret key. A simple approach to protect against many of such attacks is randomization based countermeasures. In contrast, unsupervised attacks

have the potential to render most of these countermeasures ineffective, because they do not require any knowledge about the input data, neither the public or the private part. The attacks are applied on randomized scalars with minimum assumptions. The original scalar can be deterministically computed once randomized scalar is recovered.

As the measurement signal be performed in several milliseconds and contain millions of sample points, direct application of the attack technique is not optimal. The main challenge is to identify few relevant sample points or features in the measured signal which would lead to a successful attack. In the present paper, we explore different approaches for feature selection and compare these to a profiled attack. While success rate of profiled attack represents the best case, bit-wise processing of scalar brings random guess success to 50% (worst case). In this paper, we evaluate a pool of feature selection methods commonly used in side-channel testing, in context of unsupervised attacks.

The rest of the paper can then be organized as follow: In Section II, we provide a brief description of the related work. Section III, we provide the description of the proposed approach. Practical experiments are discussed in Section IV and finally, in Section V, we conclude the paper.

II. RELATED WORK

The application of unsupervised clustering or feature selection algorithms is a relatively new topic. One of the first publications used the k -means clustering algorithm [3]. Instead of using a two step approach, feature selection first and then clustering the traces with a reduced set of features, they directly apply the k -means clustering on the set of traces. To improve the success rate, the authors use several low-noise localized EM measurements. An improved version of this attack applied PCA for dimensional reduction to improve the success rate further [4]. The main drawback of the general approach is, that it is not applicable in a high-noise environment.

The more recent work [5] proposes a two step approach. They basically follow the framework from [6] and adapt it to ECC. In principle, the procedure is iterative in nature. A preliminary leakage assessment step is used to identify a first approximate of the leaking samples in the trace. In a second step, the identified preliminary features are used to cluster

the traces. Then, based on this clustering a refinement step is performed, which leads to better set of features. The final attack on the actual trace to attack is carried out using the refined set of features. This attack framework works very well on both ECC and RSA. The two papers report very high success rates in many cases for both ECC and RSA. The main target in [5] is only on key dependent processes. However, in general, there might be some noise attributed to other data dependent leakage which might not be interesting to the attacker. Hence, in this work, we consider an alternative approach which might help to mitigate this problem.

III. METHODOLOGY

In this section, we will briefly describe different statistics we propose to use for feature selection, as well as the classification method. Based on these methods, we then choose sufficient number of relevant features and conduct the attack.

A. Trace Characterization

Computations in an Elliptic Curve Scalar Multiplication (ECSM) operation are segmented and performed in a loop, where each segment is corresponding to one bit of the secret scalar. Since the implementations of the ECSM operation are recommended to be fairly regular across all key bits to avoid timing side channels and SPA attacks, it allows us to analyze trace segments corresponding to only one bit of the scalar (which we henceforth refer to as *traces* for brevity) across multiple scalars for feature selection. The resulting selected features thus apply to the traces corresponding to other key bits as well. We acquire a set of traces $\mathcal{T}$ with n traces t^i for $i \in [0, n-1]$ and we denote a sample of length L with index j as t_j^i for $j \in \{0, L-1\}$. Each of the traces correspond to one of the two labels (0 or 1) equal to the corresponding secret scalar bit processed.

B. Feature selection

In a supervised attack scenario where the attacker knows the labels corresponding to the different traces in the trace set $\mathcal{T}$, the attacker can partition the traces into two sets $\mathcal{T}_0$ and $\mathcal{T}_1$ and apply the commonly used univariate Welch's t -test (also known as TVLA [7]) and select those samples as features that correspond to a high TVLA value. But, in an unsupervised scenario, where the attacker does not know the labels, the attacker can follow two approaches. The first approach is a label-dependent approach, wherein first he uses certain clustering algorithms such as the univariate k -means [8] over all the samples of the trace. Further, a cluster of two classes $c_{0,j}$ and $c_{1,j}$ for each sample j is obtained. The clusters are then evaluated using the following approaches.

- Difference-of-Mean (DoM) [9]: The difference of the means of the two clusters are calculated as $\text{DoM}_j = m_{0,j} - m_{1,j}$ and samples with high DoM are selected as features.
- Welch's t -test [10]: The technique is the same as that of the profiled case, except that the t -test is applied over

the clusters with predicted labels ($P_{0,j}$ & $P_{1,j}$) instead of actual labels.

- Normalized Inter-Class Variance (NICV) [11]: NICV is another metric used for leakage detection similar to the TVLA (t-test) metric and a sample with high value of NICV can be classified as a feature.

While these methods require preliminary estimation of the labels for each sample, the attacker can also attempt to select features solely based on the distribution of the data. For this, we consider the following approaches:

- Variance: One can compute the variance of the samples for a fixed index j as Var_j and repeat the same for all samples and select those samples with high variance as features. This is based on the argument made by Clavier et al. [12] that samples corresponding to high activity due to manipulated data have a large variance compared to samples with low or constant activity independent of the algorithmic input.
- Range: Computing the range for each sample j as $\text{Range}_j = \max_{1 \leq i \leq n} t_j^i - \min_{1 \leq i \leq n} t_j^i$ across traces follows a similar rationale as variance, but has the possibility of giving rise to many outliers.

The features that we are interested in are those that help distinguishing traces based on the value of the corresponding processed key bit, which can be alternatively termed as *key-dependent* features. But, it is important to know that there are also samples that depend on the value of the intermediate data processed, which can be termed as *data-dependent* features. While combining key-dependent features will increase the success probability of key-recovery, data-dependent features are simply random and add noise due to randomness of the intermediate data, which is independent of the processed key bit. Since all the above mentioned approaches look at each sample individually in a univariate manner, it is definitely possible that some of the irrelevant and random data-dependent samples are also selected as features.

On the assumption that, there are multiple key dependent leakage points in a single trace, we claim that all the key-dependent features observable across all traces in the trace-set are very highly correlated. For example, consider a set of P key-dependent features of a trace t processing bit k as $(f_1, f_2, \dots, f_{P-1})$, the time samples of trace t corresponding to these features $(t_{f_1}, t_{f_2}, \dots, t_{f_{P-1}})$ all will belong to the same class k of their corresponding univariate clusters $(c_{k,f_1}, c_{k,f_2}, \dots, c_{k,f_{P-1}})$. But, the same cannot be said for a data dependent feature as it is very unlikely to find two highly correlated data-dependent leakage points on the same trace due to the randomness of the intermediate data. Thus, we propose to adopt a multi-variate approach wherein we use the covariance metric to identify key-dependent features at two different indices (positions) on the trace.

For all pairs of possible indices u, v with $u, v \in \{0, L-1\}$, we calculate the covariance of vectors t_u^i and t_v^i with i running from $\{0, n-1\}$, which we denote as $\text{Cov}_{u,v}$. We thus build a corresponding covariance matrix for the trace set $\mathcal{T}$ of

size $(n \times L)$. We identify those entries in the matrix which have a very high magnitude for covariance and consider the corresponding pairs to be highly correlated. A high positive value at entry (u, v) indicates that the position of the class k is the same in the corresponding univariate clusters c_u and c_v and thus $\text{Sign}(\text{DoM}_u) = \text{Sign}(\text{DoM}_v)$. A high negative value otherwise indicates that the position of class k is switched in the respective univariate clusters leading to opposing signs for their DoMs. But, it is a well known fact in side channel analysis that any point in the trace is always very highly correlated with itself and its neighboring points. Thus, barring the neighboring pairs of points, which are the entries near the diagonal of the covariance matrix, an entry with a high magnitude of covariance conveniently far from the diagonal can be considered to be key-dependent features with very high confidence. Unlike the previous works on unsupervised attacks on ECC [3], [4], [5], we are able to distinguish key dependent features from data dependent features which will subsequently increase the attack success rates.

C. Classification Phase

Once the key-dependent features have been identified, most of the previous works [3], [4], [5] propose to utilize the multi-dimensional k -means classification algorithm over the identified features for classification. It might also be possible that the number of selected features might be too high for the k -means algorithm, thus bumping into a clear case of the *curse of dimensionality*. We here propose a much simpler technique for classification which only requires calculation of the mean of a subset of the selected features. From the selected set of $\tilde{P}$ relevant features $\mathcal{F} = (f_0, f_1, f_2, f_{\tilde{P}-1})$ from the feature selection phase, we select only that subset of features which pairwise correlate to a very high positive value, through which we get a clear partition of the feature set resulting in disjoint subsets $\mathcal{F}_0$ and $\mathcal{F}_1$. The attacker can choose either of the partitions and calculate the mean of the samples at the respective features of the selected partition ($\mathcal{F}_0$ or $\mathcal{F}_1$), thereby assigning a score of s_i for the trace t_i for all $i \in \{0, 1\}$. Assuming we have approximately equal number of traces corresponding to both the key bits (which is easily maintained by the randomness of the key), we calculate the mean of all the assigned scores s_i to be the global threshold for classification. This leaves us with only two possible values of the key depending on position of the class in the resulting cluster.

IV. EXPERIMENTS

In this section, we describe the implementation details of the targetted ECSM operation, the devices used, the evaluation setup and the corresponding experimental results. We target two different ECSM implementations, one on a hardware and the other on a software target to show the applicability of our attack.

A. Hardware Implementation and Evaluation Setup

It is a custom straightforward hardware design of an ECSM operation on a twisted Edwards curve Ed25519 [13]

implemented using the Joye's double and add ladder [14] and extended projective coordinates [15]. The design is implemented on Virtex-5 FPGA on the standard side-channel evaluation board SASEBO-GII. The access pattern of the registers directly depends on the key bit used and thus, the difference in address of the registers used manifests as key-dependent leakage. The implementation is constant-time and devoid of conditional branches to protect against timing side-channel attacks. The implementation run at 24 MHz and measurements were performed using a Lecroy 610Zi oscilloscope synchronised with a suitable trigger from the design. A set of 4000 traces (2000 for each bit) with a sampling rate of 500 MSam/s were collected using high-sensitivity EM probe.

B. Software Implementation and Evaluation Setup

The protected version of the popular μNacl library, which implements the ECSM operation on Curve25519 using the Montgomery ladder and randomized projective coordinates, is tested on ATmega2560 on an ARDUINO MEGA board running at 16 MHz. The design uses *cswap* (conditional swap) operation leading to key bit leakage (refer to [16]). The measurement setup is the same as before.

C. Experimental Results

We performed the feature selection methods as described in Section III-B, and we compared the results with the profiled scenario. *Accuracy* is used as comparison metric to evaluate the performance of different feature selection methods. We obtained 300 features based on each method. For profiled case, the features are obtained using t-test method (TVLA) on the labeled dataset. We also include the covariance based method to consider the bivariate feature selection scenario. The experiments are then conducted on 2 sets of data, collected from FPGA and ATmega respectively. The measurements from ATmega were misaligned when acquired leading to failure of attack. Thus, traces were realigned for further analysis. For profiling based attack, from the set of traces, the traces are divided for training and testing using 80:20 ratio (3200 traces for training and 800 for testing), whereas for non-profiled scenario, all the traces are used for the attack.

The accuracy results are then presented in Figure 1. As observed in the figure, the features obtained from the profiled TVLA approach yields the highest accuracy (96% for FPGA implementation and 99.9% for realigned microcontroller implementation). Most of the applied techniques surpass random guess. Among different methods, covariance based approach performs consistently better than the other approach. For the FPGA implementation, the achieved accuracy is 88.6%, whereas for aligned microcontroller implementation, the achieved accuracy is 98.43%. For other statistics, however, the results are less consistent. For example, in FPGA case, interval range gives the second best accuracy (84.75%), followed by variance (81%) and DoM (80%). However, for microcontroller case, their accuracies are in the range of 55 - 65%, whereas, taking the t-test gives accuracy of 93.6%

Fig. 1: Accuracy of different feature selection methods, on different target implementations

Hence, in these examples, we have shown that the proposed covariance approach can result in better accuracy, and has consistent performance across different target implementations.

D. Discussions

One of the issue that might arise for unsupervised feature selection is to determine the number of features used for the attack. Unlike profiling attack, in unsupervised scenario, there is no label available to evaluate the accuracy directly. One idea for evaluating the features selected is to use the internal metric evaluation commonly used in clustering. We then choose Silhouette index [17] and DB index [18]. Normally, these metrics are used to evaluate number of clusters, however, in [5], they apply these metric to evaluate the quality of the cluster (since the initialization of the cluster is randomized) to determine the best resulting cluster. Similarly, we use these metrics to determine the optimal number of features. We conduct preliminary experiments on FPGA traces for the same and the optimal number of features returned by both the metrics (maximum value for Silhouette and minimum value for DB index) agrees with the best accuracy results. The results are presented in Figure 2. However, for the same experiments on the AVR microcontroller, the optimal number of features returned did not exactly correlate with the best accuracy. This could be due to misalignment in the traces. A good method for selecting number of features is still an open question.

Fig. 2: Comparison between accuracy and returned estimation from Silhouette and DB index

We could alternatively use deep learning to deal with misalignment. As shown in [19], in the presence of jitter countermeasure, Convolutional Neural Network (CNN) [20] can help to identify the targeted leakage parts in the traces. In our case, we targeted the misaligned traces (similar effect to jitter countermeasure). In profiling scenario, we could achieve the accuracy of 97 - 100% (as compared to 51.9% for profiling case with misaligned traces) without the need for realignment. Hence, a potential future work is to optimize the CNN algorithm for the misalignment case and extend it to the case of unsupervised scenario.

V. CONCLUSION

In this work, we have investigated different approaches for improved feature selection from ECC traces. It has been shown that choosing relevant features could help to improve the performance of the classification process. It has been shown that considering individual feature during the unsupervised phase might lead to the selection of non-optimal features unrelated to the targeted secret. Here, we proposed using covariance method (considering two points or bivariate) and showed that the proposed approach could minimize the selection of irrelevant feature and could help improving the accuracy of the classification. We also discussed some potential approaches which could further improve feature selection which we consider for future work.

REFERENCES

- [1] P. C. Kocher, "Timing attacks on implementations of diffie-hellman, rsa, dss, and other systems," in *Advances in Cryptology - CRYPTO '96, 16th Annual International Cryptology Conference, Santa Barbara, California, USA, August 18-22, 1996, Proceedings*, ser. Lecture Notes in Computer Science, N. Kobitz, Ed., vol. 1109. Springer, 1996, pp. 104–113. [Online]. Available: https://doi.org/10.1007/3-540-68697-5_9
- [2] J. Danger, S. Guilley, P. Hoogvorst, C. Murdica, and D. Naccache, "A synthesis of side-channel attacks on elliptic curve cryptography in smart-cards," *J. Cryptographic Engineering*, vol. 3, no. 4, pp. 241–265, 2013. [Online]. Available: <https://doi.org/10.1007/s13389-013-0062-6>
- [3] J. Heyszl, A. Ibing, S. Mangard, F. D. Santis, and G. Sigl, "Clustering algorithms for non-profiled single-execution attacks on exponentiations," in *Smart Card Research and Advanced Applications - 12th International Conference, CARDIS 2013, Berlin, Germany, November 27-29, 2013. Revised Selected Papers*, ser. Lecture Notes in Computer Science, A. Francillon and P. Rohatgi, Eds., vol. 8419. Springer, 2013, pp. 79–93. [Online]. Available: https://doi.org/10.1007/978-3-319-08302-5_6
- [4] R. Specht, J. Heyszl, M. Kleinstueber, and G. Sigl, "Improving non-profiled attacks on exponentiations based on clustering and extracting leakage from multi-channel high-resolution EM measurements," in *Constructive Side-Channel Analysis and Secure Design - 6th International Workshop, COSADE 2015, Berlin, Germany, April 13-14, 2015. Revised Selected Papers*, ser. Lecture Notes in Computer Science, S. Mangard and A. Y. Poschmann, Eds., vol. 9064. Springer, 2015, pp. 3–19. [Online]. Available: https://doi.org/10.1007/978-3-319-21476-4_1
- [5] E. Nascimento and L. Chmielewski, "Applying horizontal clustering side-channel attacks on embedded ECC implementations," in *Smart Card Research and Advanced Applications - 16th International Conference, CARDIS 2017, Lugano, Switzerland, November 13-15, 2017, Revised Selected Papers*, ser. Lecture Notes in Computer Science, T. Eisenbarth and Y. Tegliah, Eds., vol. 10728. Springer, 2017, pp. 213–231. [Online]. Available: https://doi.org/10.1007/978-3-319-75208-2_13
- [6] G. Perin and L. Chmielewski, "A semi-parametric approach for side-channel attacks on protected RSA implementations," in *Smart Card Research and Advanced Applications - 14th International Conference, CARDIS 2015, Bochum, Germany, November 4-6, 2015. Revised Selected Papers*, ser. Lecture Notes in Computer Science, N. Homma and M. Medwed, Eds., vol. 9514. Springer, 2015, pp. 34–53. [Online]. Available: https://doi.org/10.1007/978-3-319-31271-2_3
- [7] G. Goodwill, B. Jun, J. Jaffe, and P. Rohatgi, "A testing methodology for side-channel resistance validation," 2011.
- [8] S. P. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Information Theory*, vol. 28, no. 2, pp. 129–136, 1982. [Online]. Available: <https://doi.org/10.1109/TIT.1982.1056489>
- [9] P. C. Kocher, J. Jaffe, and B. Jun, "Differential power analysis," in *Advances in Cryptology - CRYPTO '99, 19th Annual International Cryptology Conference, Santa Barbara, California, USA, August 15-19, 1999, Proceedings*, ser. Lecture Notes in Computer Science, M. J. Wiener, Ed., vol. 1666. Springer, 1999, pp. 388–397. [Online]. Available: https://doi.org/10.1007/3-540-48405-1_25
- [10] B. L. Welch, "The Generalization of 'Student's' Problem when Several Different Population Variances are Involved," *Biometrika*, vol. 34, no. 1/2, pp. 28–35, 1947. [Online]. Available: <http://dx.doi.org/10.2307/2332510>
- [11] S. Bhasin, J. Danger, S. Guilley, and Z. Najm, "NICV: normalized inter-class variance for detection of side-channel leakage," *IACR Cryptology ePrint Archive*, vol. 2013, p. 717, 2013. [Online]. Available: <http://eprint.iacr.org/2013/717>
- [12] C. Clavier and L. Reynaud, "Improved blind side-channel analysis by exploitation of joint distributions of leakages," in *Cryptographic Hardware and Embedded Systems - CHES 2017 - 19th International Conference, Taipei, Taiwan, September 25-28, 2017, Proceedings*, ser. Lecture Notes in Computer Science, W. Fischer and N. Homma, Eds., vol. 10529. Springer, 2017, pp. 24–44. [Online]. Available: https://doi.org/10.1007/978-3-319-66787-4_2
- [13] D. J. Bernstein, N. Duif, T. Lange, P. Schwabe, and B. Yang, "High-speed high-security signatures," *J. Cryptographic Engineering*, vol. 2, no. 2, pp. 77–89, 2012. [Online]. Available: <https://doi.org/10.1007/s13389-012-0027-1>
- [14] M. Joye, "Highly regular right-to-left algorithms for scalar multiplication," in *Cryptographic Hardware and Embedded Systems - CHES 2007, 9th International Workshop, Vienna, Austria, September 10-13, 2007, Proceedings*, ser. Lecture Notes in Computer Science, P. Paillier and I. Verbauwhede, Eds., vol. 4727. Springer, 2007, pp. 135–147. [Online]. Available: https://doi.org/10.1007/978-3-540-74735-2_10
- [15] H. Hisil, K. K. Wong, G. Carter, and E. Dawson, "Twisted edwards curves revisited," in *Advances in Cryptology - ASIACRYPT 2008, 14th International Conference on the Theory and Application of Cryptology and Information Security, Melbourne, Australia, December 7-11, 2008. Proceedings*, ser. Lecture Notes in Computer Science, J. Pieprzyk, Ed., vol. 5350. Springer, 2008, pp. 326–343. [Online]. Available: https://doi.org/10.1007/978-3-540-89255-7_20
- [16] M. Hutter and P. Schwabe, "Nacl on 8-bit AVR microcontrollers," in *Progress in Cryptology - AFRICACRYPT 2013, 6th International Conference on Cryptology in Africa, Cairo, Egypt, June 22-24, 2013. Proceedings*, ser. Lecture Notes in Computer Science, A. Youssef, A. Nitaj, and A. E. Hassanien, Eds., vol. 7918. Springer, 2013, pp. 156–172. [Online]. Available: https://doi.org/10.1007/978-3-642-38553-7_9
- [17] P. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, no. 1, pp. 53–65, Nov. 1987. [Online]. Available: [http://dx.doi.org/10.1016/0377-0427\(87\)90125-7](http://dx.doi.org/10.1016/0377-0427(87)90125-7)
- [18] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 1, no. 2, pp. 224–227, 1979. [Online]. Available: <https://doi.org/10.1109/TPAMI.1979.4766909>
- [19] E. Cagli, C. Dumas, and E. Prouff, "Convolutional neural networks with data augmentation against jitter-based countermeasures - profiling attacks without pre-processing," in *Cryptographic Hardware and Embedded Systems - CHES 2017 - 19th International Conference, Taipei, Taiwan, September 25-28, 2017, Proceedings*, ser. Lecture Notes in Computer Science, W. Fischer and N. Homma, Eds., vol. 10529. Springer, 2017, pp. 45–68. [Online]. Available: https://doi.org/10.1007/978-3-319-66787-4_3
- [20] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, 2015. [Online]. Available: <https://doi.org/10.1016/j.neunet.2014.09.003>
- [21] W. Fischer and N. Homma, Eds., *Cryptographic Hardware and Embedded Systems - CHES 2017 - 19th International Conference, Taipei, Taiwan, September 25-28, 2017, Proceedings*, ser. Lecture Notes in Computer Science, vol. 10529. Springer, 2017. [Online]. Available: <https://doi.org/10.1007/978-3-319-66787-4>