# Machine Learning and Implementation Attacks

Stjepan Picek
stjepan@computer.org

CROSSING Summer School on Sustainable Security & Privacy,
Darmstadt, September 12, 2019

# Outline

# Outline

# Where to use Machine Learning in Cryptology

- Machine learning is data driven approach.
- It seems more difficult to use such techniques for design.
- Additional benefit from using them in attacks: it is easy to validate the solution.

# Where to use Machine Learning - Classical Applications

- **Side-channel attacks**.
- **Fault injection**.
- Modeling attacks on PUFs.
- Detecting Hardware Trojans.
- Machine learning over encrypted data.

# Where to use Machine Learning - Exotic Applications

- Factoring numbers.
- Design of ciphers.

# Outline

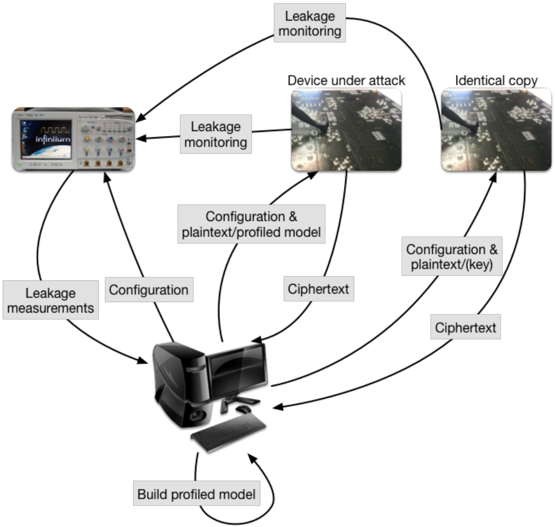
# Implementation Attacks and SCA

### Implementation attacks

Implementation attacks do not aim at the weaknesses of the algorithm, but on its implementation.

- **Side-channel attacks** (SCAs) – passive, non-invasive attacks.
- SCAs – one of the most powerful category of attacks on crypto devices.
- Profiled attacks – the most powerful among SCAs.
- Within profiling phase the adversary estimates leakage models for targeted intermediate computations, which are exploited to extract secret information in the actual attack phase.

# Profiled Attacks



# SCA and Profiling Attacks

Table: Overview of profiling side-channel attacks used in literature (up to March 2019 and limited to symmetric key crypto).

| Algorithm | Reference |
| --- | --- |
| Naive Bayes and its variants | [1, 2, 3, 4, 5, 6] |
| Random Forest | [2, 3, 4, 7, 8, 6, 9, 10, 11, 12, 13, 14] |
| Rotation Forest | [15, 4, 5, 16] |
| XGB | [5] |
| MultiBoost | [15] |
| Self-organizing maps | [9] |
| Support Vector Machines | [15, 4, 7, 8, 6, 17, 18, 9, 10, 11, 12, 19, 13, 20, 16] |
| Multivariate regression analysis | [21, 11, 12] |
| Multilayer Perceptron | [2, 3, 5, 7, 8, 6, 22, 23, 24, 25, 26, 27, 28] |
| Convolutional Neural Networks | [8, 5, 7, 29, 30, 22, 28] |
| Autoencoders | [8] |
| Recurrent Neural Networks | [8] |
| Template Attack and its variants | [1, 15, 4, 7, 8, 29, 30, 6, 17, 9, 10, 11, 12, 19, 13, 28, 16] |
| Stochastic attack | [11, 12, 7] |

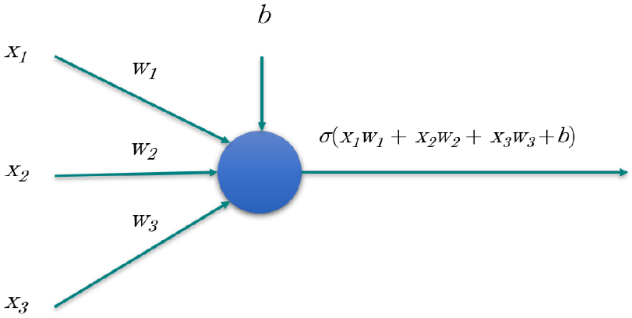
# Profiled Attacks

- Template Attack is the most powerful attack from the information theoretic point of view.
- Some machine learning techniques (supervised learning) also belong to the profiled attacks.
- Deep learning has been shown to be able to reach top performance even if the device is protected with countermeasures.

# Outline

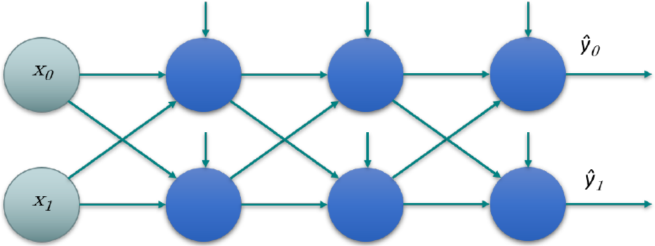
# Deep Learning

- Let us build a neural network.

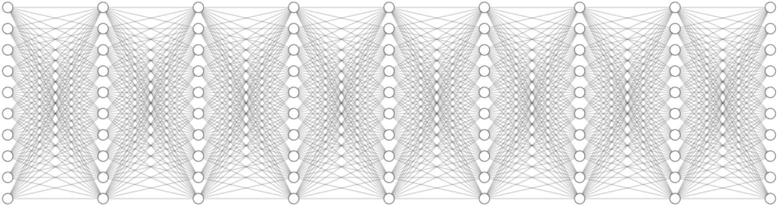


# Deep Learning

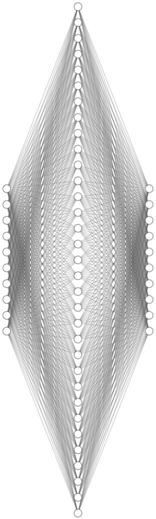
- Let us continue adding neurons.



# Multilayer Perceptron - "Many" Hidden Layers



# Multilayer Perceptron - One Hidden Layer



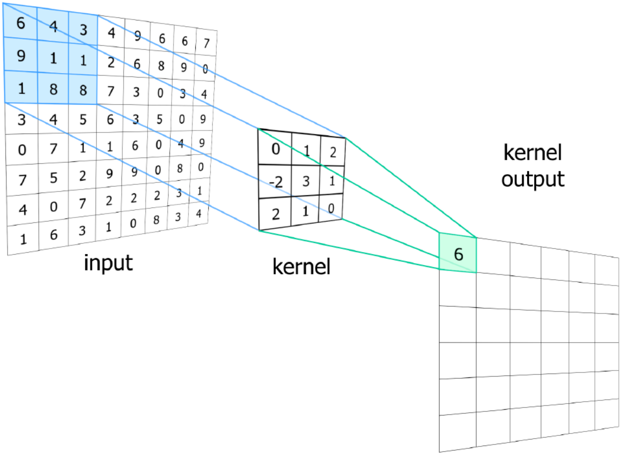
# Universal Approximation Theorem

- A feed-forward network with a single hidden layer containing a finite number of neurons can approximate continuous functions on compact subsets of $\mathbb{R}^n$.
- Given enough hidden units and enough data, multilayer perceptrons can approximate virtually any function to any desired accuracy.
- Valid results if and only if there is a sufficiently large number of training data in the series.

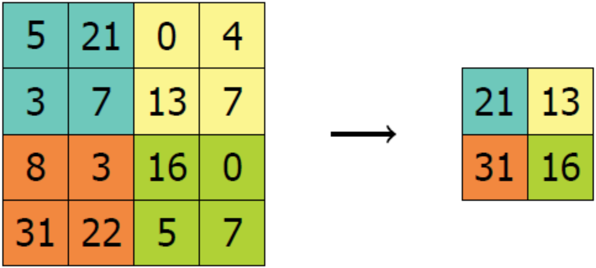
# Convolutional Neural Networks

- CNNs represent a type of neural networks which were first designed for 2-dimensional convolutions.
- They are primarily used for image classification but lately, they have proven to be powerful classifiers in other domains.
- From the operational perspective, CNNs are similar to ordinary neural networks: they consist of a number of layers where each layer is made up of neurons.
- CNNs use three main types of layers: convolutional layers, pooling layers, and fully-connected layers.

# Convolutional Neural Networks - Convolution Layer



# Convolutional Neural Networks - Pooling



# Design Principle - VGG Like CNN

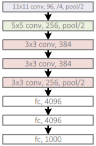
$$\mathtt{net} = \mathtt{fc}_{\theta,\mathtt{softmax}} \circ \prod_{p=1}^{P} \mathtt{fc}_{\theta^p,\mathtt{ReLU}} \circ \prod_{q=1}^{Q} \left(\mathtt{pool}_{\mathtt{Max}} \circ \prod_{r=1}^{R_q} \mathtt{conv}_{\phi^r,\mathtt{ReLU}}\right), \tag{1}$$

$$\mathtt{conv}_{\phi,\sigma}(X) = \sigma(\phi * X), \tag{2}$$

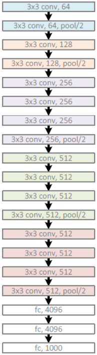
$$\mathtt{fc}_{\theta,\sigma}(x) = \sigma(\theta^\intercal x). \tag{3}$$

# Common Architectures

AlexNet, 8 layers
(ILSVRC 2012)




VGG, 19 layers
(ILSVRC 2014)




GoogleNet, 22 layers
(ILSVRC 2014)

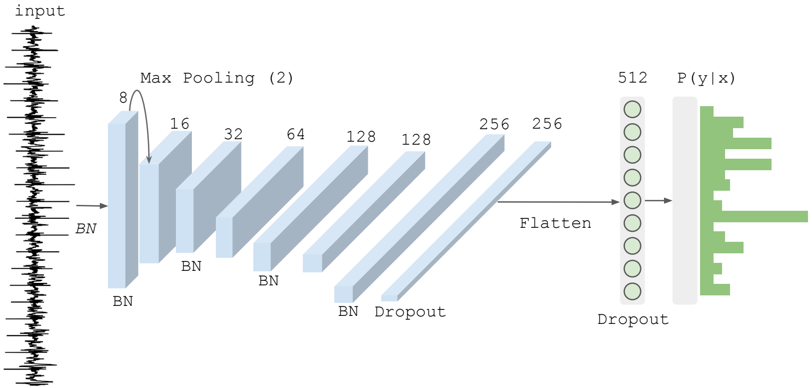
# More Complex Architectures

AlexNet, 8 layers
(ILSVRC 2012)

VGG, 19 layers
(ILSVRC 2014)

ResNet, 152 layers
(ILSVRC 2015)

# Convolutional Neural Network in SCA



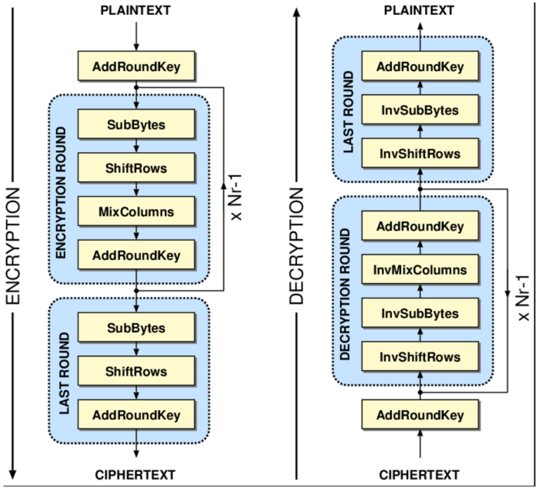
# Making the Architectures Even More Powerful

- To reduce the overfitting of the model, we introduce noise to the training phase.
- Since in our case, the input normalization is also learned during the training process via the BN layer, we added the noise tensor after the first BN.

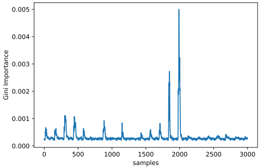
$$X^* = BN_0(X) + \Psi, \quad \Psi \sim \mathcal{N}(0, \alpha). \tag{4}$$

- The noise tensor follows the normal distribution.

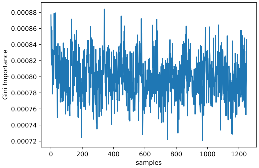
# AES


PLAINTEXT
AddRoundKey
ENCRYPTION ROUND
SubBytes
ShiftRows
MixColumns
AddRoundKey
x Nr-1
LAST ROUND
SubBytes
ShiftRows
AddRoundKey
CIPHERTEXT
ENCRYPTION
DECRYPTION
PLAINTEXT
LAST ROUND
AddRoundKey
InvSubBytes
InvShiftRows
DECRYPTION ROUND
AddRoundKey
InvMixColumns
InvSubBytes
InvShiftRows
x Nr-1
AddRoundKey
CIPHERTEXT

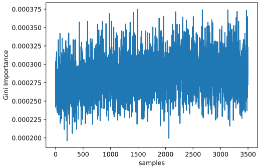
# Datasets



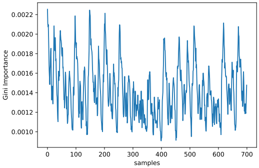

(a) DPAcontest v4 dataset




(b) AES_HD dataset

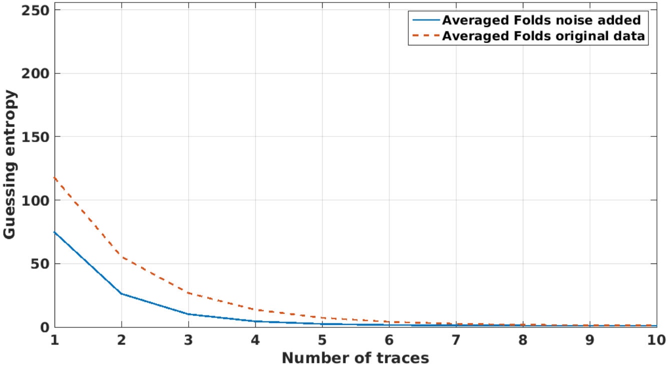



(c) Random Delay dataset



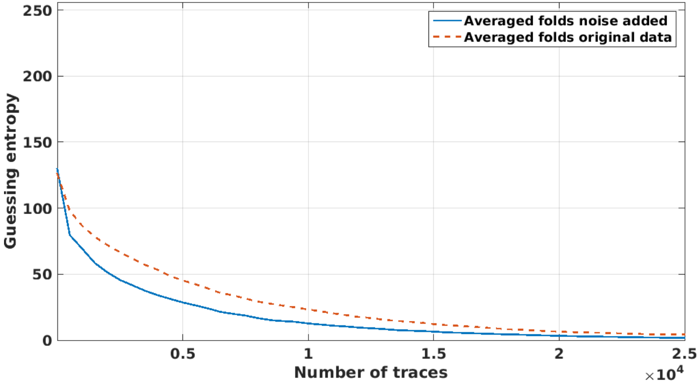

(d) ASCAD dataset

# Results DPAv4



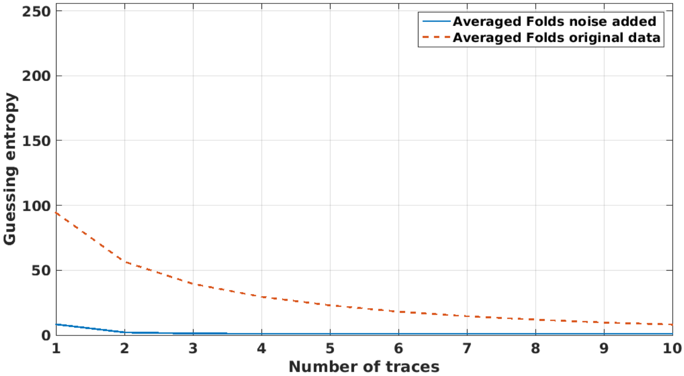

(e) RD network averaged

# Results AES_HD



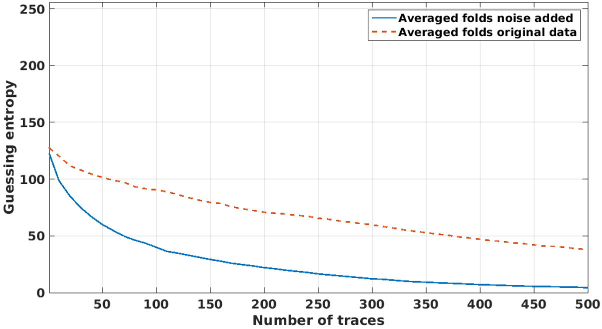

(f) ASCAD network averaged

# Results AES_RD




(g) RD network averaged

# Results ASCAD

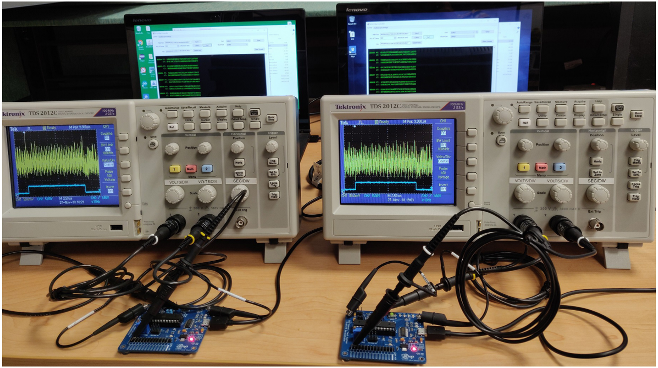



(h) ASCAD network averaged

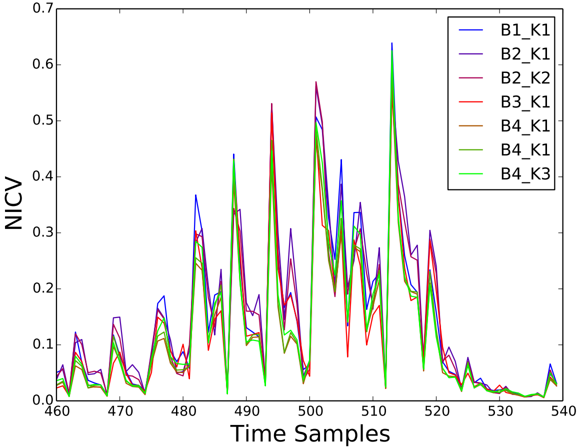
# Profiling Attacks and Portability

- There are two devices: one for training and the second one for attack.
- Two devices, different keys.
- Usually, we make our lives simpler and assume only one device and the same key.
- It is the same?

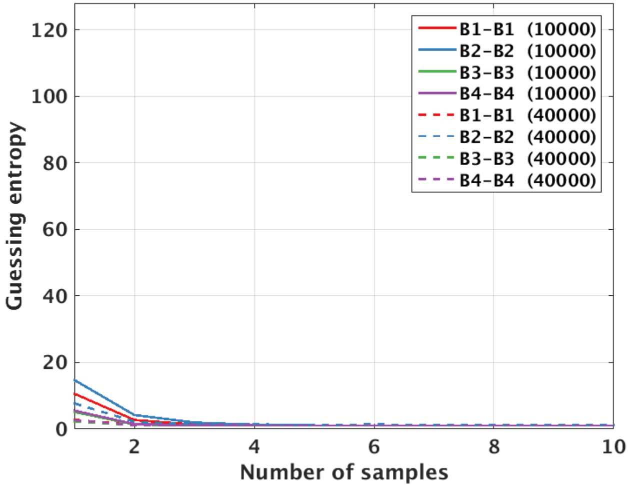
# Setup



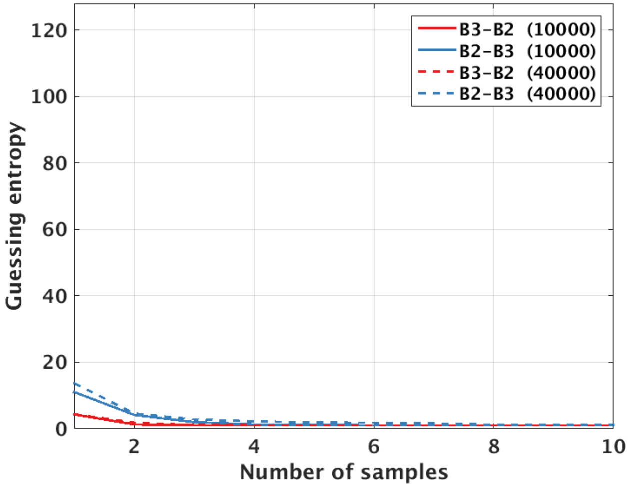
# NICV


NICV
Time Samples
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
460
470
480
490
500
510
520
530
540
B1_K1
B2_K1
B2_K2
B3_K1
B4_K1
B4_K1
B4_K3

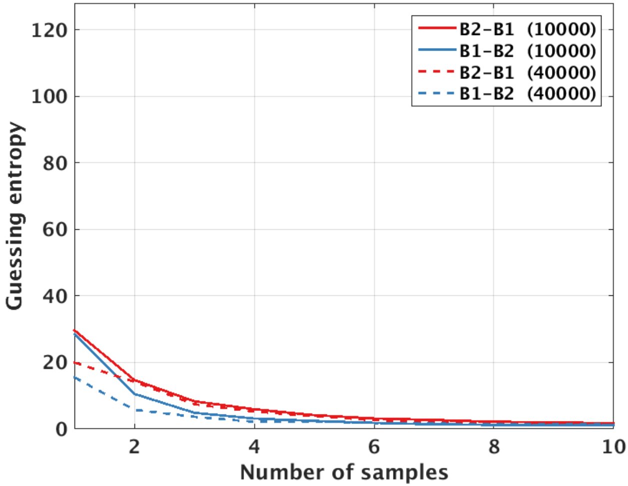
# Same Key and Device



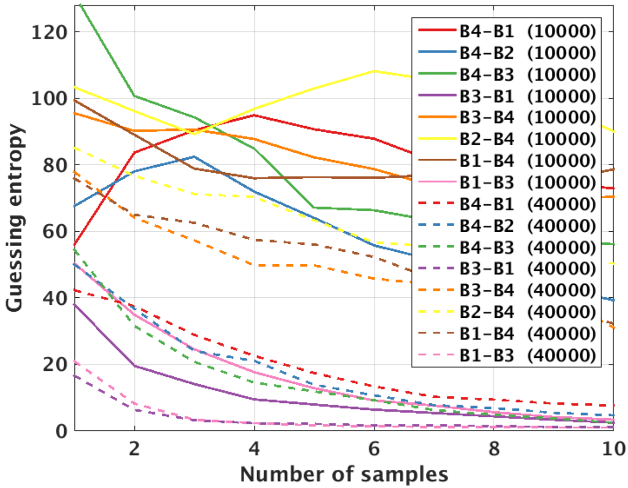
# Different key and Same Device



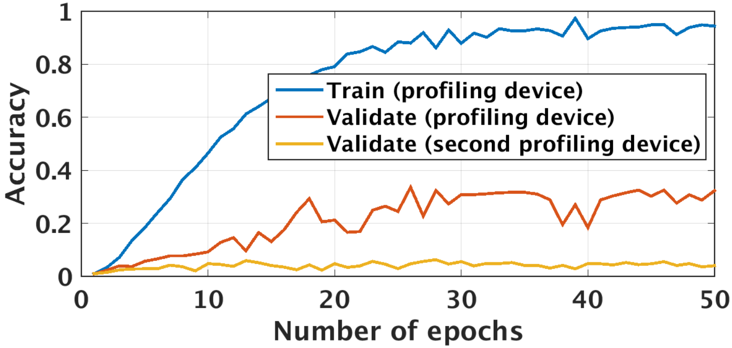
# Same Key and Different Device



# Different key and Device



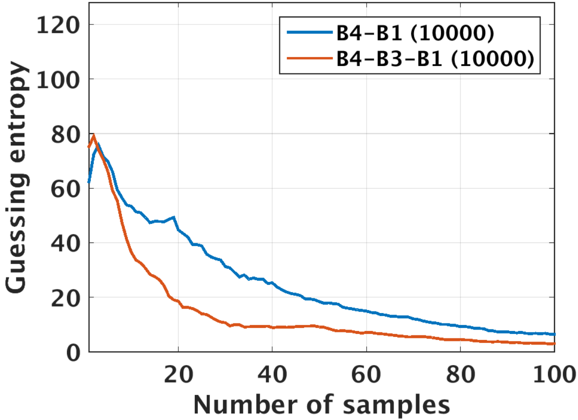
# Validation


Accuracy
1
0.8
0.6
0.4
0.2
0
Train (profiling device)
Validate (profiling device)
Validate (second profiling device)
0
10
20
30
40
50
Number of epochs

# Multiple Device Model

- Instead of validating on the same device as training, we need one more device!
- Separate devices for train, validation, attack.
- If we do not have a third device, we can use artificial noise.

# Multiple Device Model



# Problems and "Problems"

- Selection of machine learning techniques and hyper-parameter tuning.
- Portability.
- Lack of datasets.
- Reproducibility and explainability.
- Still no clear connection between machine learning and side-channel analysis metrics.
- Countermeasures.
- Academia vs. industry perspective.
- ...

# Introduction

- A fault injection (FI) attack is successful if after exposing the device to a specially crafted external interference, it shows an unexpected behavior exploitable by the attacker.
- Insertion of signals has to be precisely tuned for the fault injection to succeed.
- Finding the correct parameters for a successful FI can be considered as a search problem.
- The search space is typically too large to perform an exhaustive search.

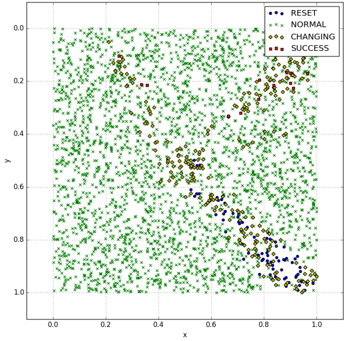
# Verdict classes

- FI testing equipment can output only verdict classes that correspond to successful measurements.
- Several possible classes for classifying a single measurement:
  1. NORMAL: smart card behaves as expected and the glitch is ignored
  2. RESET: smart card resets as a result of the glitch
  3. MUTE: smart card stops all communication as a result of the glitch
  4. INCONCLUSIVE: smart card responds in a way that cannot be classified in any other class
  5. SUCCESS: smart card response is a specific, predetermined value that does not happen under normal operation

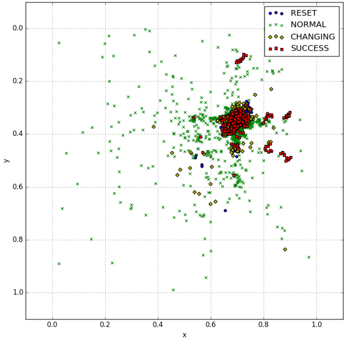
# Approaches

- Random search and exhaustive search.
- For voltage glitching and EMFI, we can use various heuristics, like genetic algorithms.
- Approaches as exhaustive search cannot work: would last 29 000 years.
- For laser FI, the situation is more complex as laser can easily break the target so we use deep learning.

# EMFI and Keccak

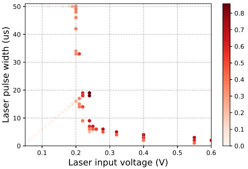



(i) Random search

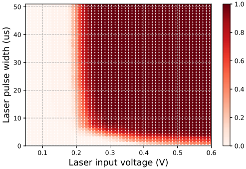



(j) GA and local search

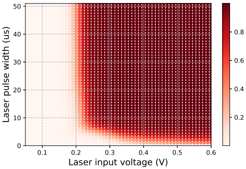
# LFI and DES




(k) Characterization

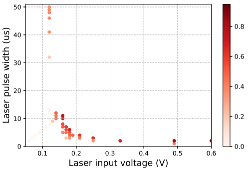



(l) Exhaustive search

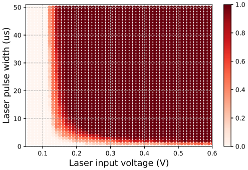



(m) Prediction with neural network

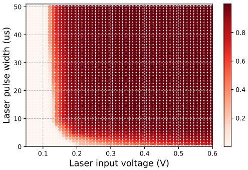
# LFI and AES




(n) Characterization




(o) Exhaustive search




(p) Prediction with neural network

# Outline

# Conclusions

- Machine learning (and even wider, artificial intelligence) play important role in cryptography.
- Currently, attacks perspective seem to be more developed.
- In implementation attacks, machine learning represents even the most powerful option.
- Still, our state-of-the-art techniques are usually much simpler than in other domains.
- There are some specific parts one does not encounter in other domains, but much of the knowledge is transferable.
- What do new attacks teach us about improving the countermeasures?

# Questions?

Thanks for your attention! Q?

Picek, S., Heuser, A., Guilley, S.:
Template attack versus Bayes classifier.
Journal of Cryptographic Engineering **7**(4) (Nov 2017) 343–351

Heuser, A., Picek, S., Guilley, S., Mentens, N.:
Side-channel analysis of lightweight ciphers: Does lightweight equal easy?
In: Radio Frequency Identification and IoT Security - 12th International Workshop, RFIDSec 2016, Hong Kong, China, November 30 - December 2, 2016, Revised Selected Papers. (2016) 91–104

Heuser, A., Picek, S., Guilley, S., Mentens, N.:
Lightweight ciphers and their side-channel resilience.
IEEE Transactions on Computers **PP**(99) (2017) 1–1

Picek, S., Heuser, A., Jovic, A., Legay, A.:
Climbing down the hierarchy: Hierarchical classification for machine learning side-channel attacks.
In Joye, M., Nitaj, A., eds.: Progress in Cryptology - AFRICACRYPT 2017: 9th International Conference on Cryptology in Africa, Dakar, Senegal, May 24-26, 2017, Proceedings, Cham, Springer International Publishing (2017) 61–78

Picek, S., Samiotis, I.P., Kim, J., Heuser, A., Bhasin, S., Legay, A.:
On the performance of convolutional neural networks for side-channel analysis.
In Chattopadhyay, A., Rebeiro, C., Yarom, Y., eds.: Security, Privacy, and Applied Cryptography Engineering, Cham, Springer International Publishing (2018) 157–176

Picek, S., Heuser, A., Alippi, C., Regazzoni, F.:
When theory meets practice: A framework for robust profiled side-channel analysis.
Cryptology ePrint Archive, Report 2018/1123 (2018) `https://eprint.iacr.org/2018/1123`.

Picek, S., Heuser, A., Jovic, A., Bhasin, S., Regazzoni, F.:
The curse of class imbalance and conflicting metrics with machine learning for side-channel evaluations.
IACR Trans. Cryptogr. Hardw. Embed. Syst. **2019**(1) (2019) 209–237

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

Lerman, L., Bontempi, G., Markowitch, O.:
Power analysis attack: An approach based on machine learning.
Int. J. Appl. Cryptol. **3**(2) (June 2014) 97–115

Lerman, L., Poussier, R., Bontempi, G., Markowitch, O., Standaert, F.:
Template Attacks vs. Machine Learning Revisited (and the Curse of Dimensionality in Side-Channel Analysis).
In: COSADE 2015, Berlin, Germany, 2015. Revised Selected Papers. (2015) 20–33

Lerman, L., Medeiros, S.F., Bontempi, G., Markowitch, O.:
A Machine Learning Approach Against a Masked AES.
In: CARDIS. Lecture Notes in Computer Science, Springer (November 2013) Berlin, Germany.

Lerman, L., Bontempi, G., Markowitch, O.:
A machine learning approach against a masked AES - Reaching the limit of side-channel attacks with a learning model.
J. Cryptographic Engineering **5**(2) (2015) 123–139

Lerman, L., Bontempi, G., Ben Taieb, S., Markowitch, O.:
A time series approach for profiling attack.
In Gierlichs, B., Guilley, S., Mukhopadhyay, D., eds.: Security, Privacy, and Applied Cryptography Engineering, Berlin, Heidelberg, Springer Berlin Heidelberg (2013) 75–94

Najm, Z., Jap, D., Jungk, B., Picek, S., Bhasin, S.:
On comparing side-channel properties of AES and chacha20 on microcontrollers.

In: 2018 IEEE Asia Pacific Conference on Circuits and Systems, APCCAS 2018, Chengdu, China, October 26-30, 2018, IEEE (2018) 552–555

Picek, S., Heuser, A., Jovic, A., Ludwig, S.A., Guilley, S., Jakobovic, D., Mentens, N.:
Side-channel analysis and machine learning: A practical perspective.
In: 2017 International Joint Conference on Neural Networks, IJCNN 2017, Anchorage, AK, USA, May 14-19, 2017. (2017) 4095–4102

Lerman, L., Medeiros, S.F., Veshchikov, N., Meuter, C., Bontempi, G., Markowitch, O.:
Semi-supervised template attack.
In Prouff, E., ed.: Constructive Side-Channel Analysis and Secure Design, Berlin, Heidelberg, Springer Berlin Heidelberg (2013) 184–199

Heuser, A., Zohner, M.:
Intelligent Machine Homicide - Breaking Cryptographic Devices Using Support Vector Machines.
In Schindler, W., Huss, S.A., eds.: COSADE. Volume 7275 of LNCS., Springer (2012) 249–264

Hospodar, G., Gierlichs, B., De Mulder, E., Verbauwhede, I., Vandewalle, J.:
Machine learning in side-channel analysis: a first study.
Journal of Cryptographic Engineering **1** (2011) 293–302 10.1007/s13389-011-0023-x.

Bartkewitz, T., Lemke-Rust, K.:
Efficient template attacks based on probabilistic multi-class support vector machines.
In Mangard, S., ed.: Smart Card Research and Advanced Applications, Berlin, Heidelberg, Springer Berlin Heidelberg (2013) 263–276

Hospodar, G., De Mulder, E., Gierlichs, B.:
Least squares support vector machines for side-channel analysis.
Center for Advanced Security Research Darmstadt (01 2011) 99–104

Sugawara, T., Homma, N., Aoki, T., Satoh, A.:

Profiling attack using multivariate regression analysis.
IEICE Electron. Express **7**(15) (2010) 1139–1144

Timon, B.:
Non-profiled deep learning-based side-channel attacks.
Cryptology ePrint Archive, Report 2018/196 (2018) https://eprint.iacr.org/2018/196.

Pfeifer, C., Haddad, P.:
Spread: a new layer for profiled deep-learning side-channel attacks.
Cryptology ePrint Archive, Report 2018/880 (2018) https://eprint.iacr.org/2018/880.

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

Martinasek, Z., Hajny, J., Malina, L.:
Optimization of power analysis using neural network.
In Francillon, A., Rohatgi, P., eds.: Smart Card Research and Advanced Applications, Cham, Springer International Publishing (2014) 94–107

Yang, S., Zhou, Y., Liu, J., Chen, D.:
Back propagation neural network based leakage characterization for practical security analysis of cryptographic implementations.
In Kim, H., ed.: Information Security and Cryptology - ICISC 2011, Berlin, Heidelberg, Springer Berlin Heidelberg (2012) 169–185

Martinasek, Z., Zeman, V.:
Innovative method of the power analysis.
Radioengineering **22**(2) (2013)

Hettwer, B., Gehrer, S., Güneysu, T.:

Profiled power analysis attacks using convolutional neural networks with domain knowledge.
In Cid, C., Jr., M.J.J., eds.: Selected Areas in Cryptography - SAC 2018 - 25th International Conference, Calgary, AB, Canada, August 15-17, 2018, Revised Selected Papers. Volume 11349 of Lecture Notes in Computer Science., Springer (2018) 479–498

Kim, J., Picek, S., Heuser, A., Bhasin, S., Hanjalic, A.:
Make some noise: Unleashing the power of convolutional neural networks for profiled side-channel analysis.
Cryptology ePrint Archive, Report 2018/1023 (2018) https://eprint.iacr.org/2018/1023.

Cagli, E., Dumas, C., Prouff, E.:
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. (2017) 45–68