Side channel attack
an approach based on machine learning

L. LERMAN & G. BONTEMPI & O. MARKOWITCH

Université Libre de Bruxelles
Faculty of Sciences
Department of Computer Sciences
Machine Learning Group & Cryptography and Security Service

February 24, 2011

Context

Cryptography is used since a long time for confidentiality purposes

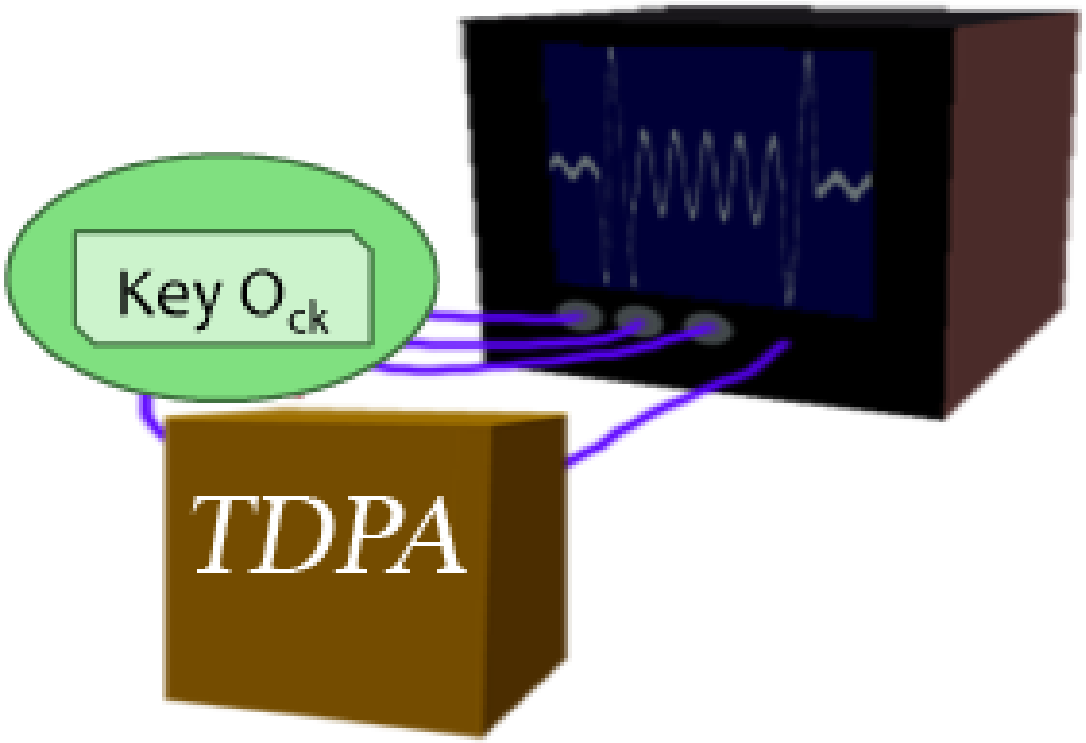
- Mobile phones
- Banks
- Cars
- Government

Side channel attacks

Reduction in cryptography security in real situation
Possibility to find the secret key when we focalize on a side channel

- Timing attack (Kocher - 1996)
- Electromagnetic attack (Gandolfi, Mourtel & Olivier - 2001)
- Power monitoring attack (Kocher, Jaffe & Jun - 1999)

Power monitoring attack

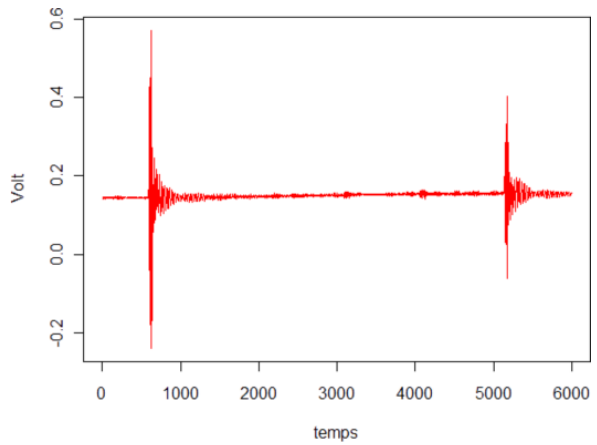


Notations

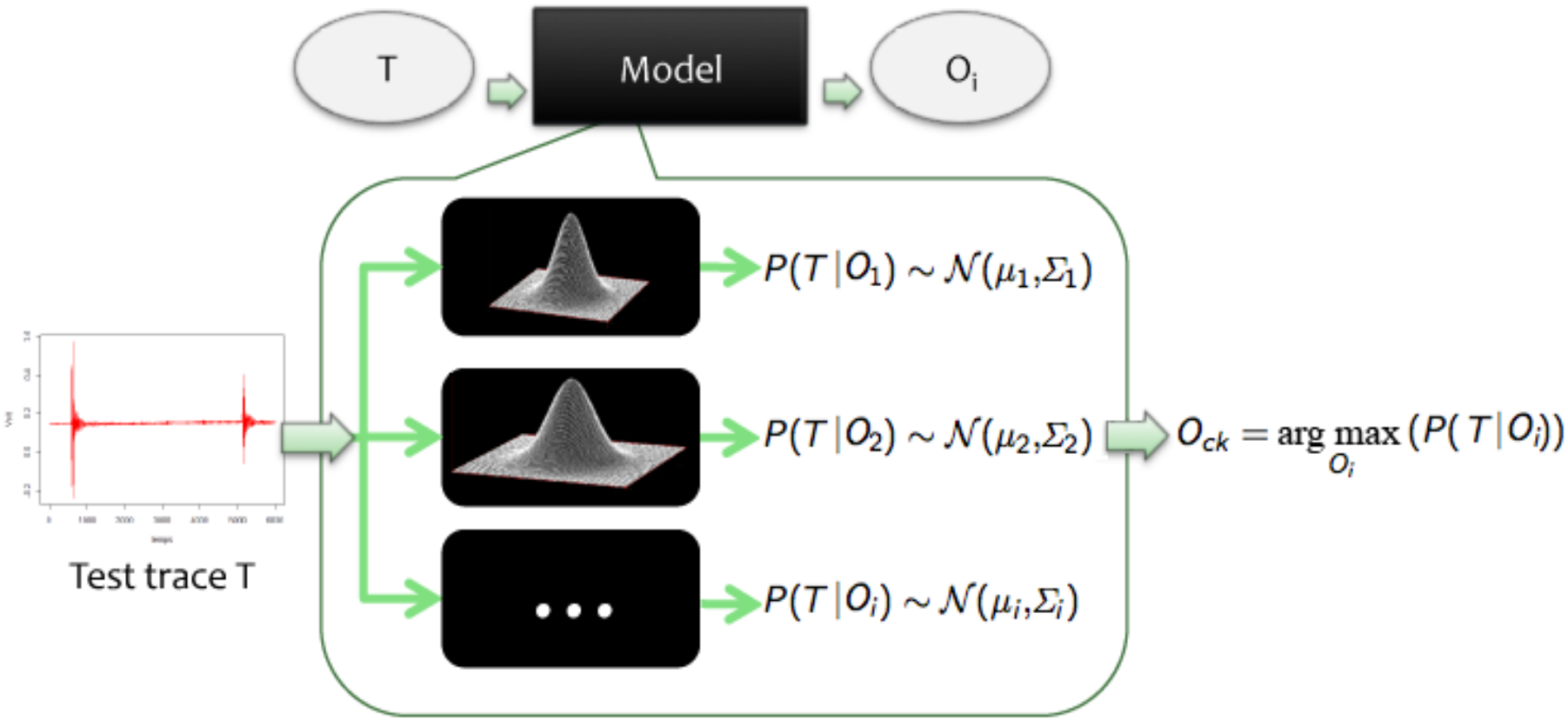
Trace (T) = power consumption, in volts, during an encryption (multidimensional problem)

$O_i = i^{th}$ key

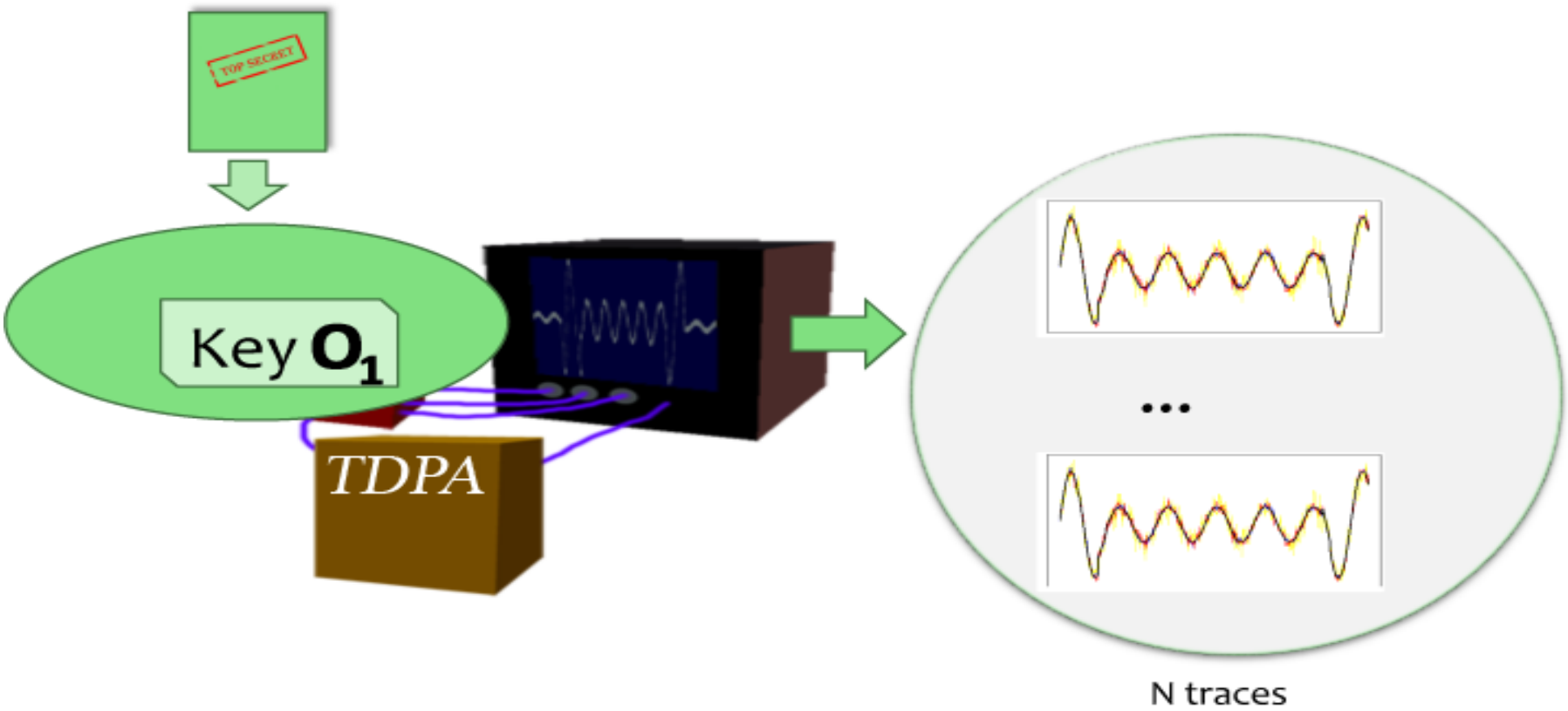
O_{ck} = correct key



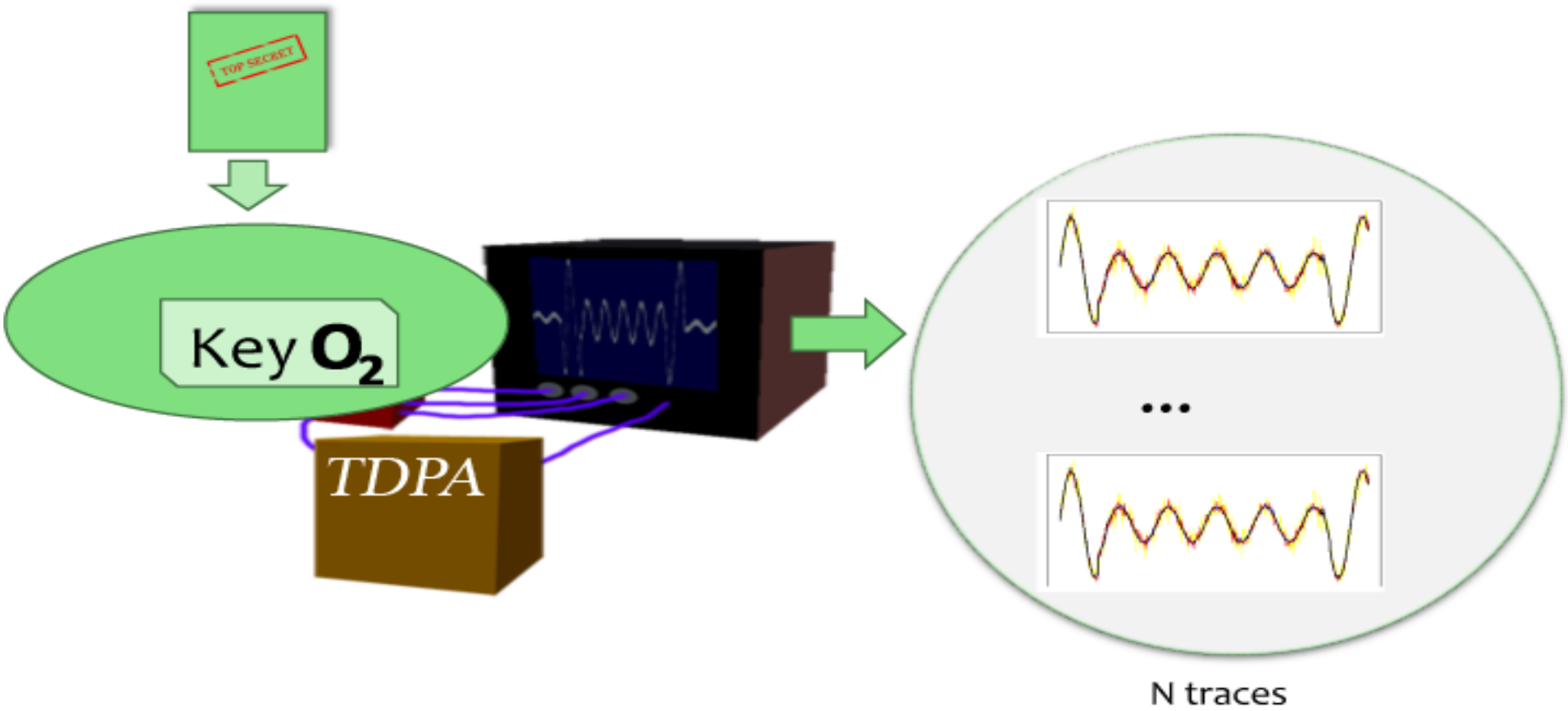
Template Based DPA



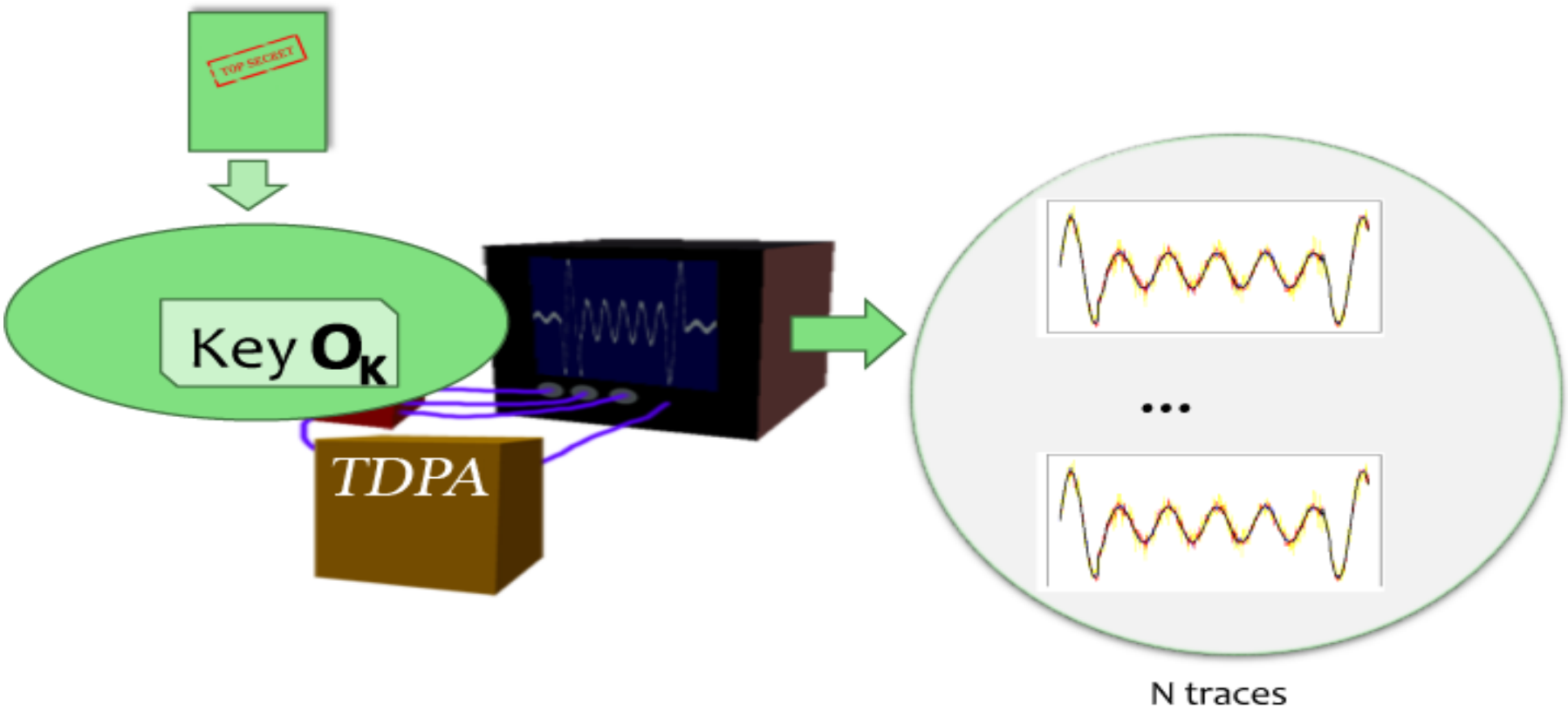
Sub models



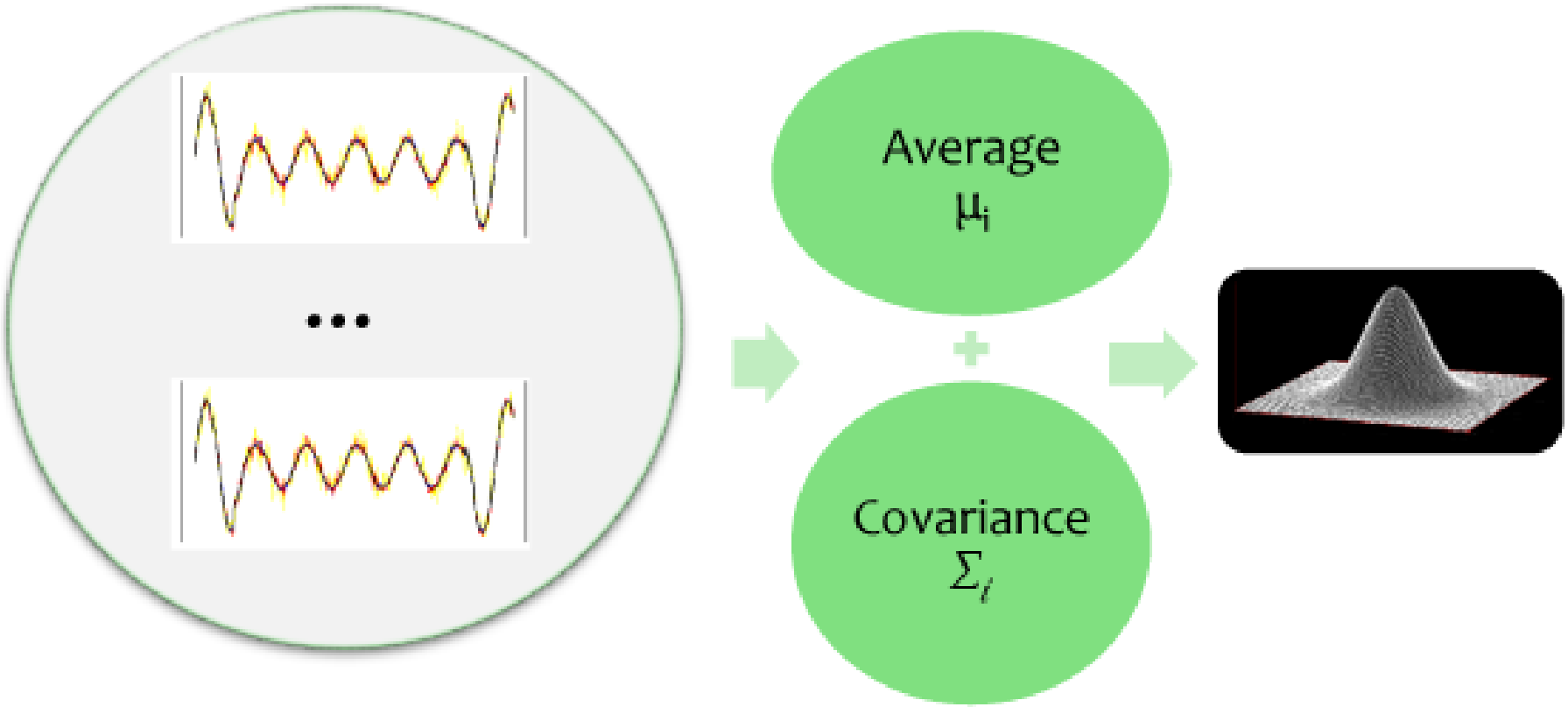
Sub models



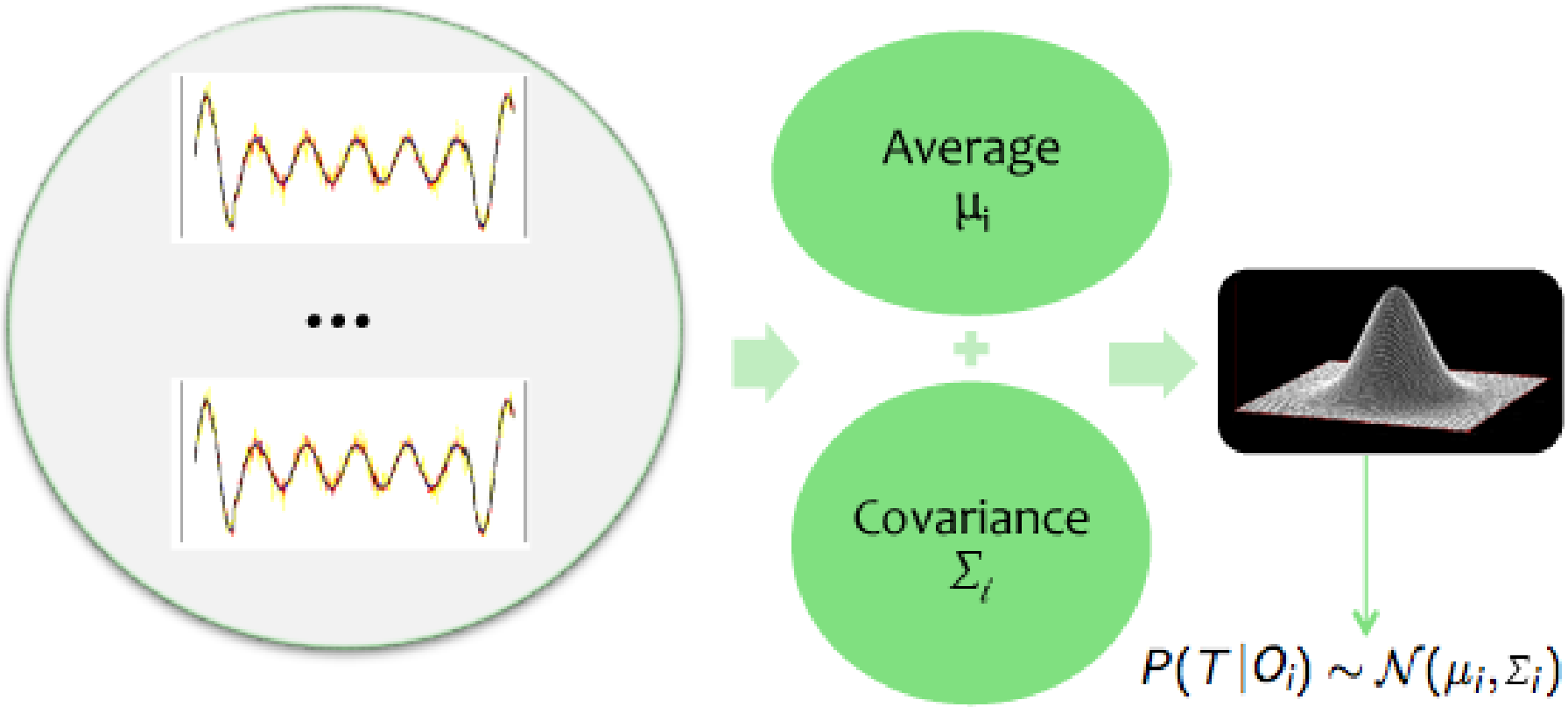
Sub models



Sub models



Sub models


Average
μ_i
Covariance
Σ_i
$P(T|O_i) \sim \mathcal{N}(\mu_i, \Sigma_i)$

Problem

if we have 2^{56} possible values for a key (encoded on 56 bits), do we have 2^{56} sub-models ?

No! We can create two sub-models per bit
Each one linked to a value of a bit

Pros & cons

Pros

- Efficiency in the attack
- Requires only one trace to predict a key
- No knowledge about the plaintext
- No knowledge about the cryptographic schema (e.g. 3DES)
- Theoretically it takes all the information in a trace

Pros & cons

Cons

- Access is needed to the cryptographic device
- Attack specific to a cryptographic device
- Assumption about the Gaussianity of data

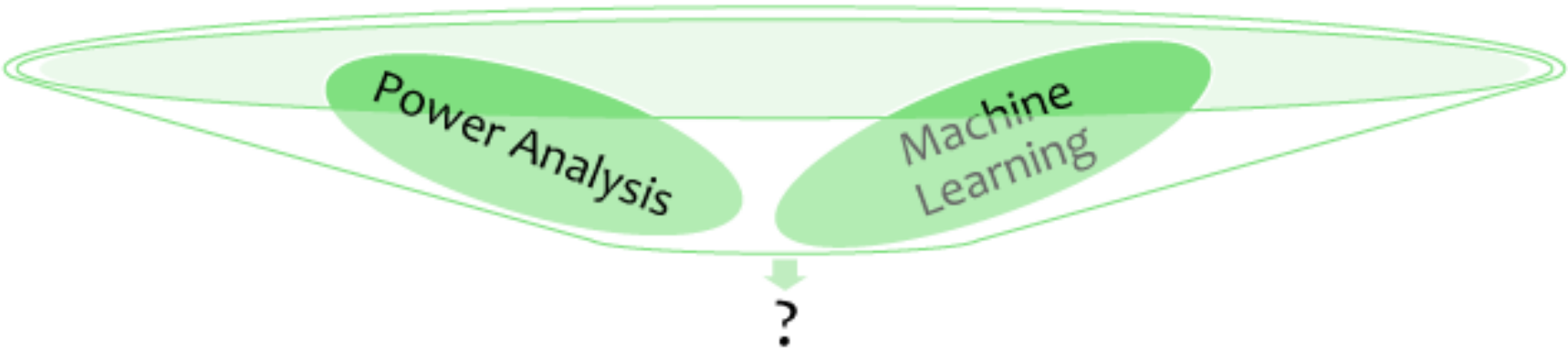
Question

When considering power consumption analysis, can we design an attack which is more efficient than the template based DPA?

Possible answer

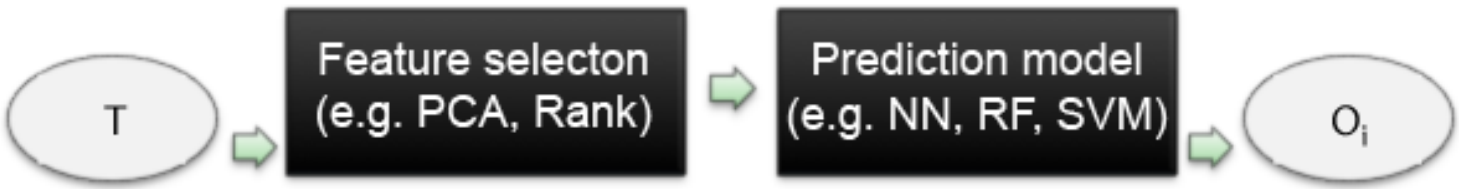
Interdisciplinary combination

Power analysis attack associated with machine learning

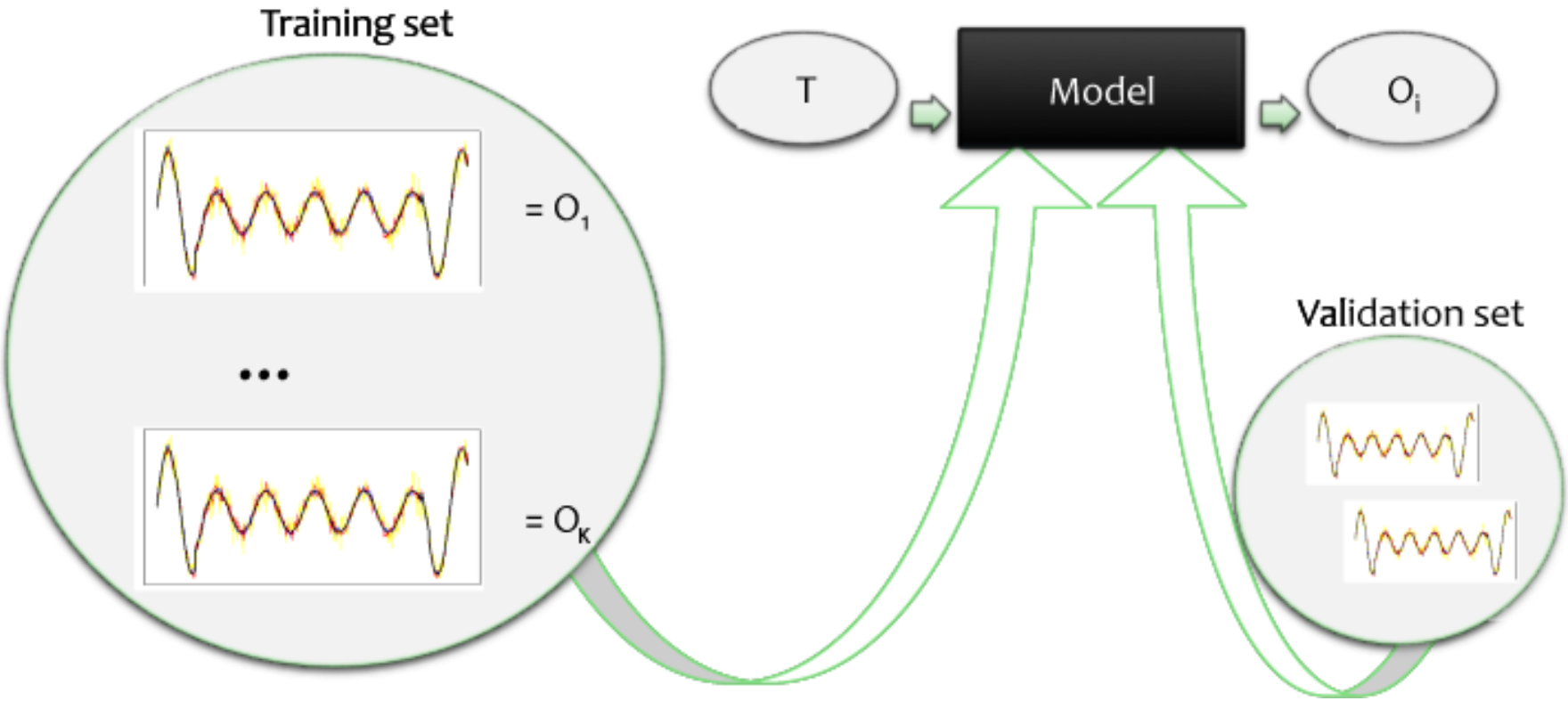


Machine learning

Machine learning



Machine learning



Motivations & disadvantages

Motivations

- Few applications of machine learning to cryptanalysis
- Feasible in practice
- Possibility of avoiding the assumption on the Gaussianity of data
- Reduction of the dimensionality
- Availability of several off-the-shelf algorithms of machine learning

Motivations & disadvantages

Disadvantages

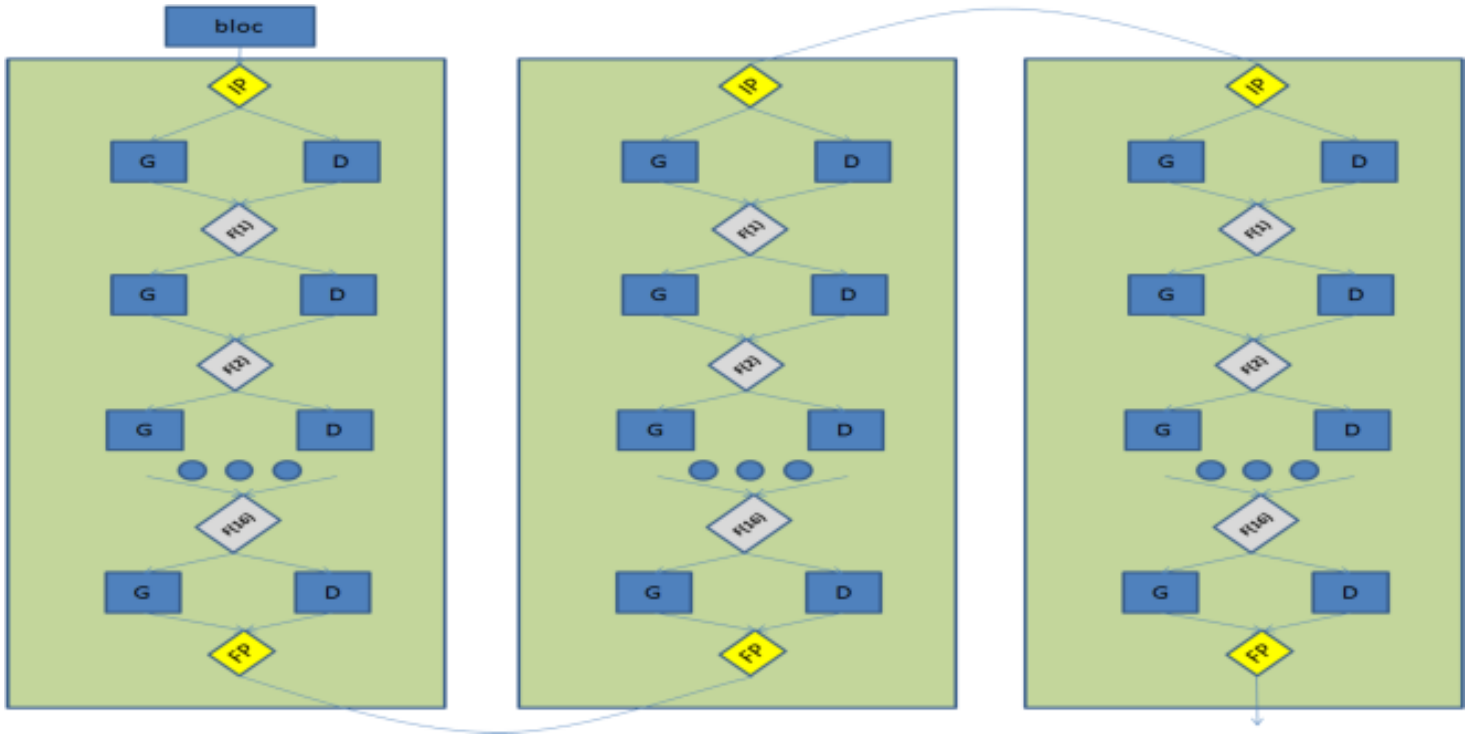
- Worse than template Based DPA if traces follow a parametric Gaussian distribution
- Often seen as a black box

Implementation of the attack

Implementation done thanks to the help of Atos Worldline (Service Data Encryption Peripheral (DEP) , Belgium)

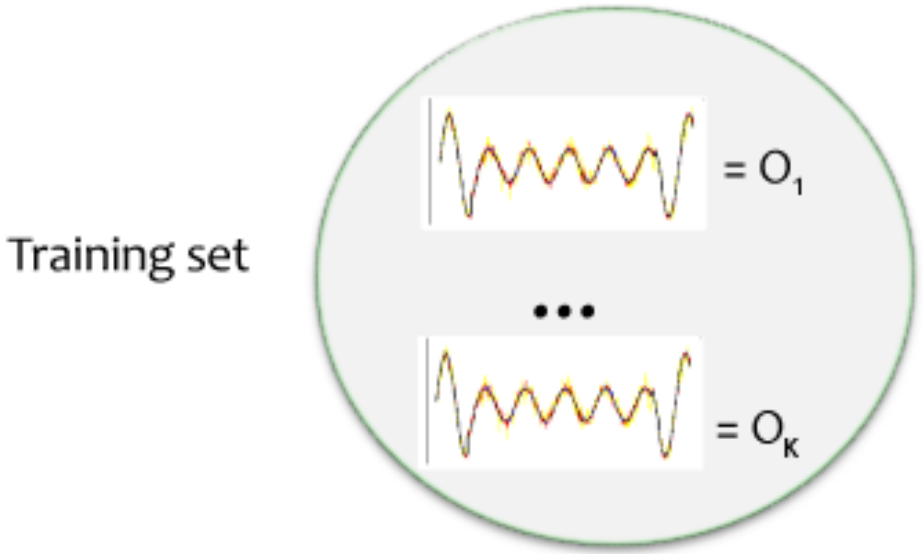
Cryptographic device

Cryptographic device: 3DES on a Xilinx SPARTAN XC3s5000
24 bytes key (three distinct 56 bit keys)



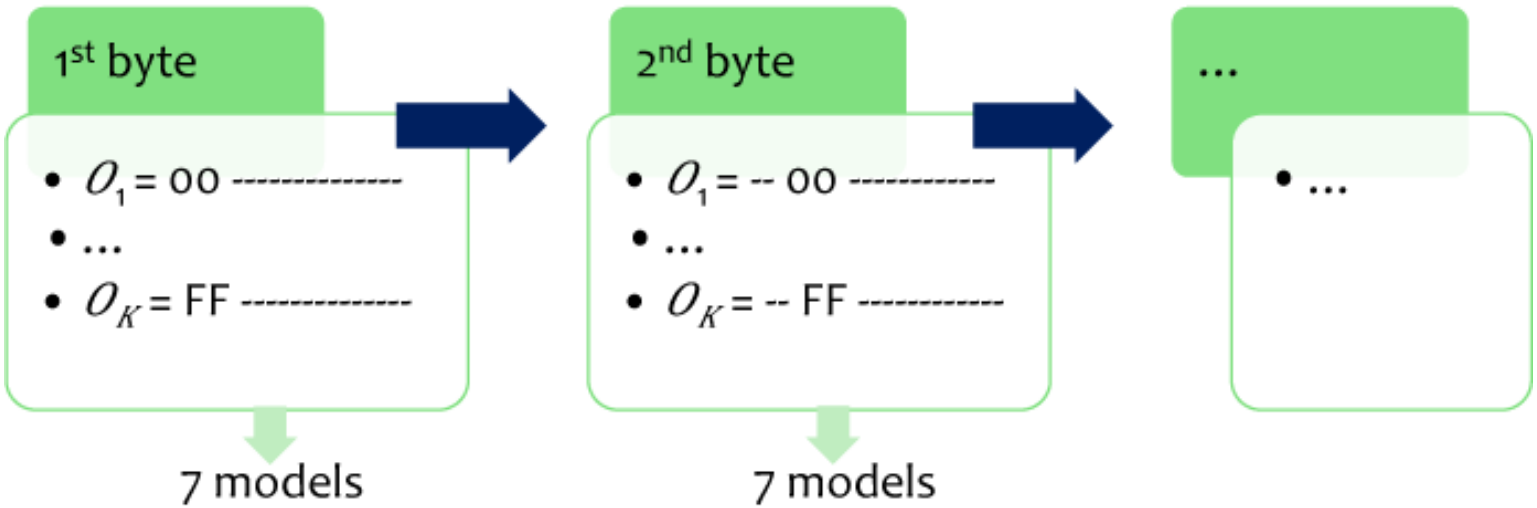
Collection of data

- 400 traces T_i per each key O_j attacked (N = 400)
- Encrypted message is constant and random.
- Noise reduction: $T_j = \frac{1}{N} \sum_{i=1}^{N} T_i$
- One T_j per each key O_j

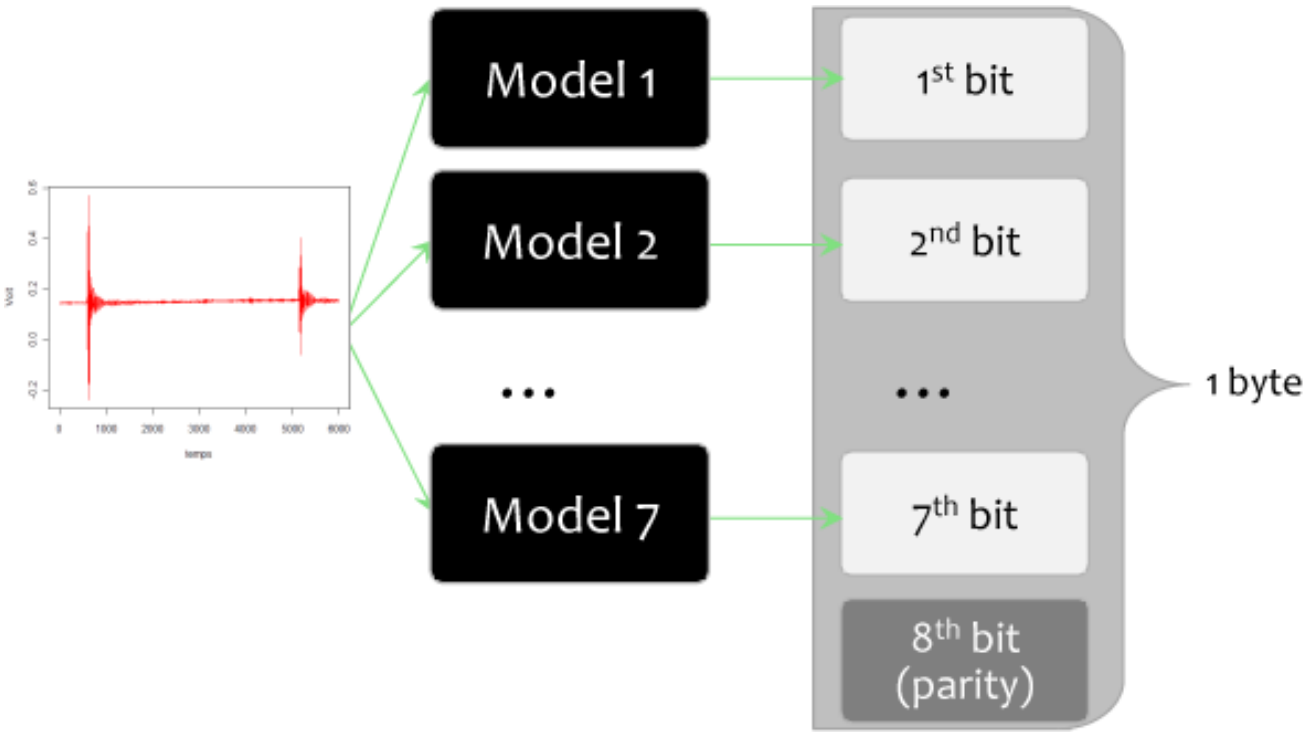


Collection of data

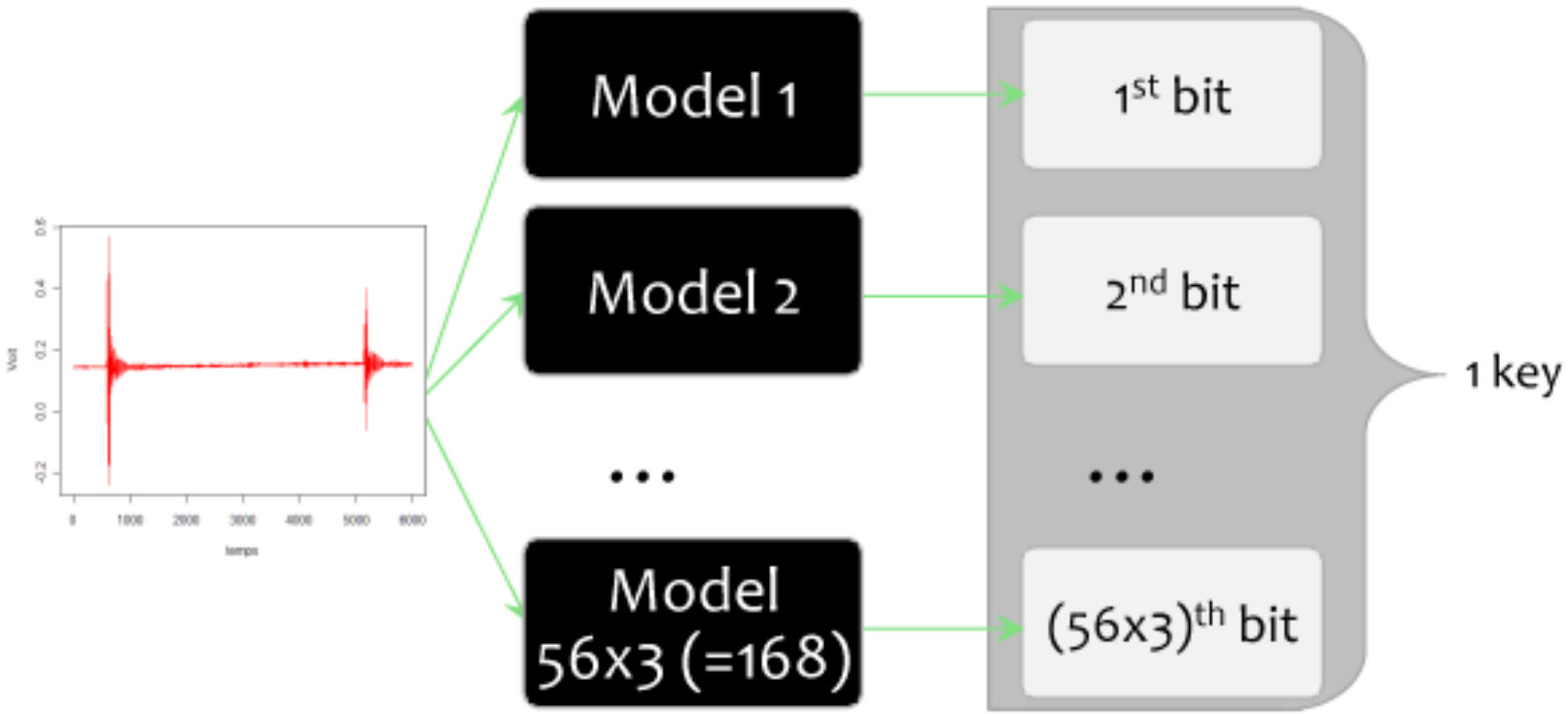
1 byte attacked at a time



Collection of data



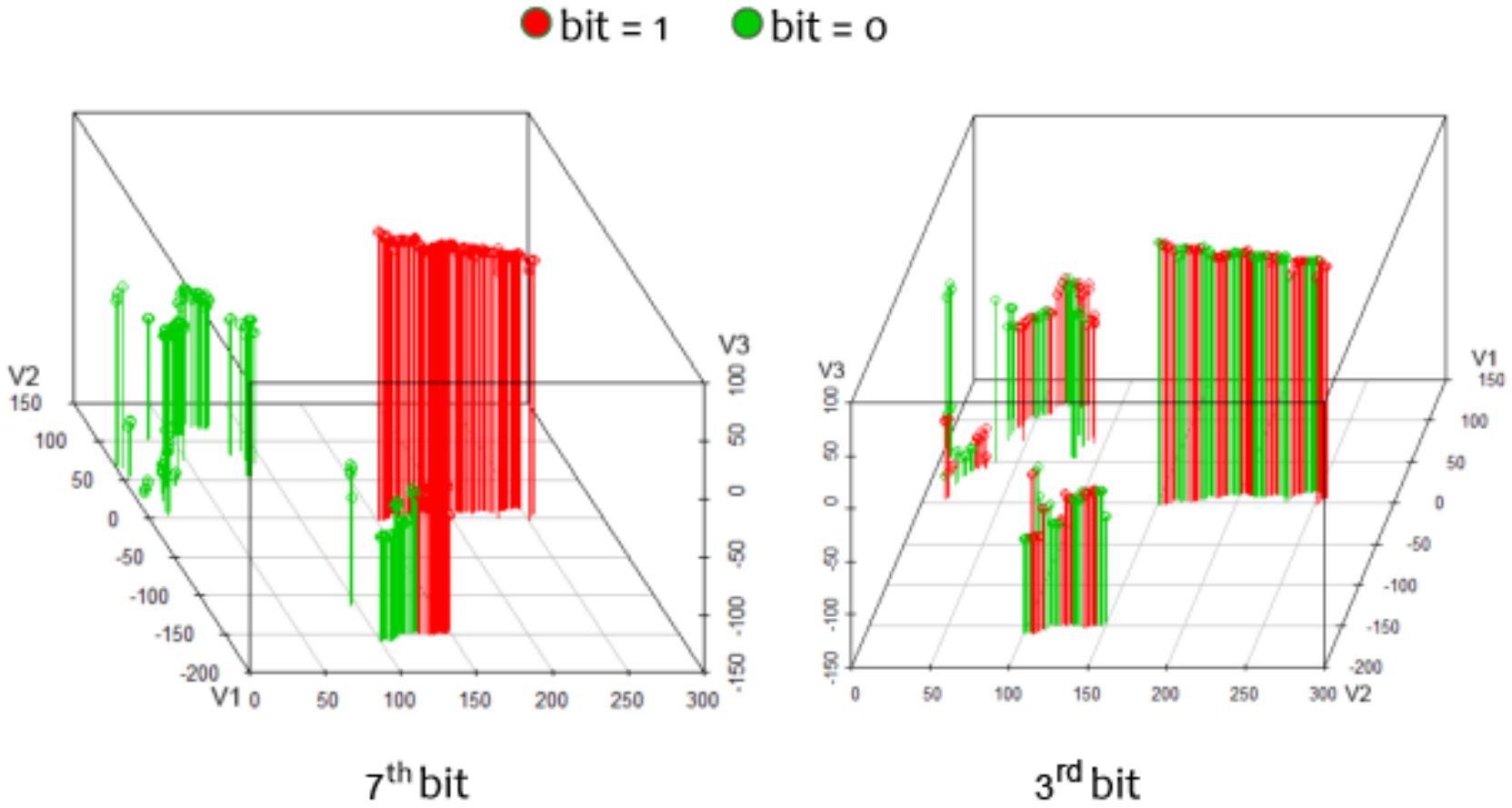
Collection of data



Strategy

- 3D visualization of traces (thanks to 3 components of the PCA)
- Selection of a model
- Attack on all bytes of the key

3D visualization



Selection of a model

Prediction models algorithm

- SOM (Kohonen - 2001)
- SVM (Cortes & Vapnik - 1995)
- RF (Breiman - 2001)

Feature selections algorithm

- Nosel
- Rank
- SOM (Kohonen - 2001)
- PCA (Pearson - 1901)

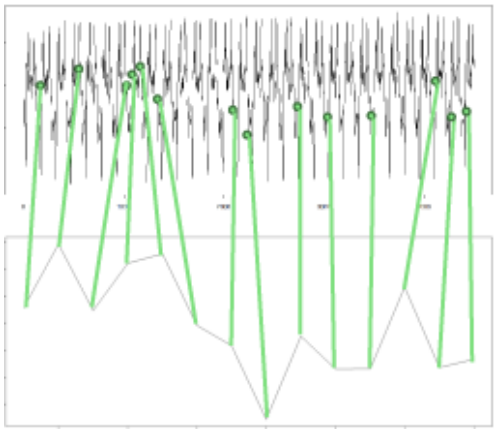
Selection of a model

Models

- SOM
- SVM
- **RF**

Feature selections

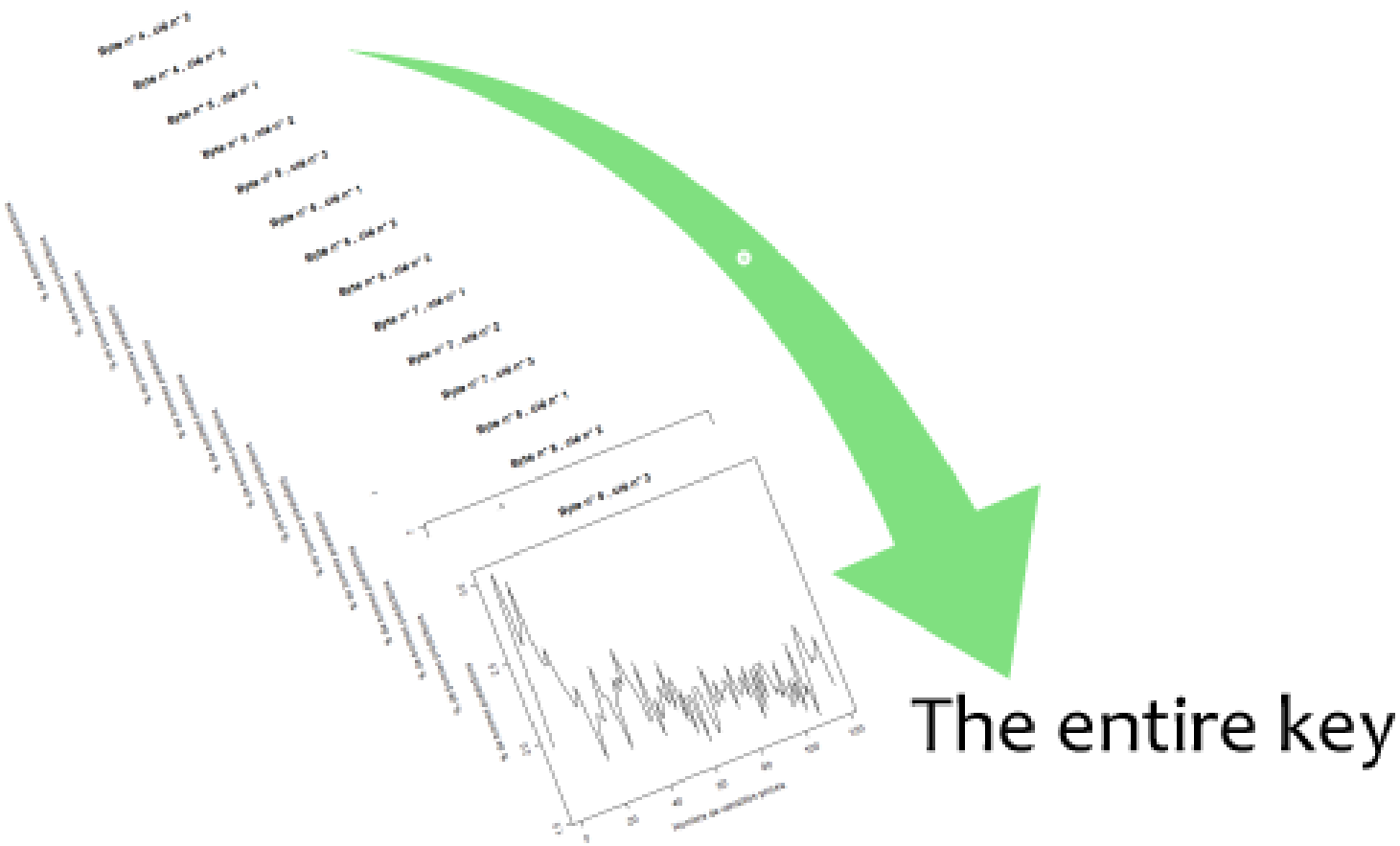
- Nosel
- Rank
- SOM
- **PCA**



	7^{th}bit	6^{th}bit	5^{th}bit	4^{th}bit	3^{rd}bit	2^{nd}bit	1^{st}bit	1 byte
%	96.09	92.58	90.63	85.55	75.39	58.98	50	15.33

(estimated by the leave-one-out)

Attack on all bytes of the key



Attack on all bytes of the key

	7^{th}bit	6^{th}bit	5^{th}bit	4^{th}bit	3^{rd}bit	2^{nd}bit	1^{st}bit
1st byte	78.13	65.63	77.34	60.16	60.16	53.13	50.00
2nd byte	85.16	75.00	67.97	50.00	57.03	50.00	50.00
...	...	...	...	...	...	...	...
μ	86.66	76.47	66.89	59.54	56.90	51.79	51.40
σ	8.44	7.39	6.09	5.71	5.71	3.45	2.88

Template Based DPA vs RF/PCA

Template Based DPA/mRMR

- 5.80% of good answers
- 35 dimensions
- $>$ 59 points: technique is not reliable
- Shrinkage estimation (Schäfer & Strimmer - 2005) makes possible of $>$ 59 but this has no remarkable effects in terms of accuracy

Template Based DPA vs RF/PCA

Number of guesses to do on average before finding the right key

- Template Based DPA/mRMR (on 1 byte): 21 keys
- RF/PCA (on 1 byte): 11 keys

Contributions

- Proposition of a new attack based on machine learning
- Implementation of the new attack on a real setting
- Better techniques of attack compared to template Based DPA

Future works

- Larger portions of the key
- Assessing the impact of the coded message on the prediction accuracy
- Varying the cryptographic device
- Varying the number of measurements during learning and validation process
- Adoption of specific learning techniques for the classification of time series
- Fusion of different measurements

Thank you

Thank you

Liran LERMAN (llerman@ulb.ac.be)

Gianluca Bontempi (gbonte@ulb.ac.be)

Olivier Markowitch (olivier.markowitch@ulb.ac.be)