Side-Channel Attacks Against the Human Brain: the PIN Code Case Study

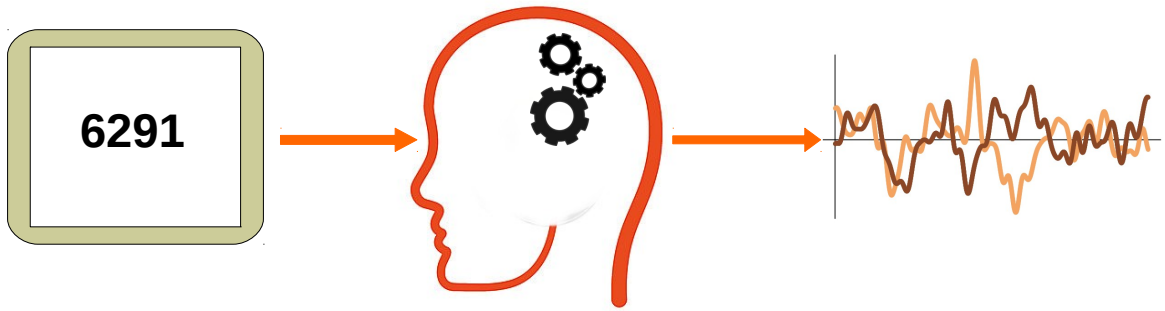
J. Lange, **C. Massart**, A. Mouraux and F.-X. Standaert

UCL Crypto Group, Belgium.

Cosade 2017, Paris, France

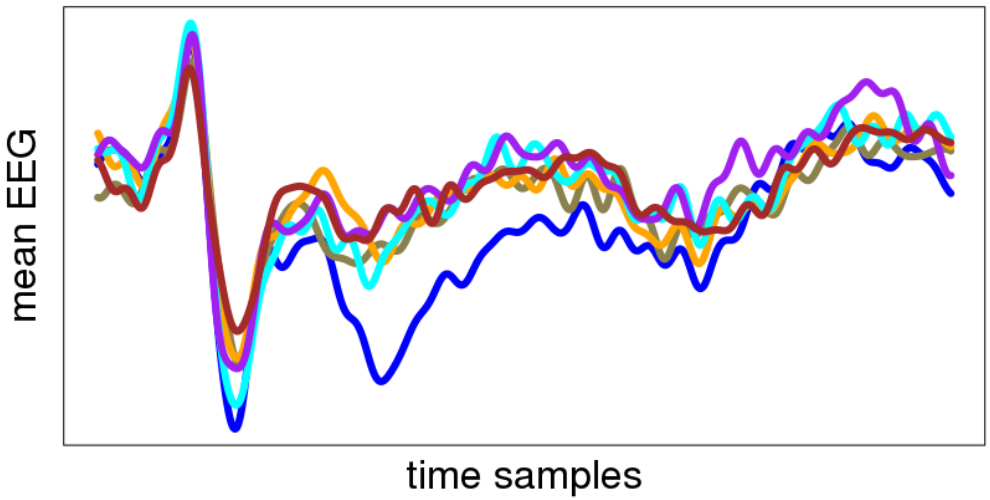
Introduction

- Side-channel attacks allow PIN code recoveries
  - e.g., Le Bouder et al., *A Template Attack against Verify PIN Algorithms*, SECRYPT 2016
- Can we apply them to BCIs & EEG signals?



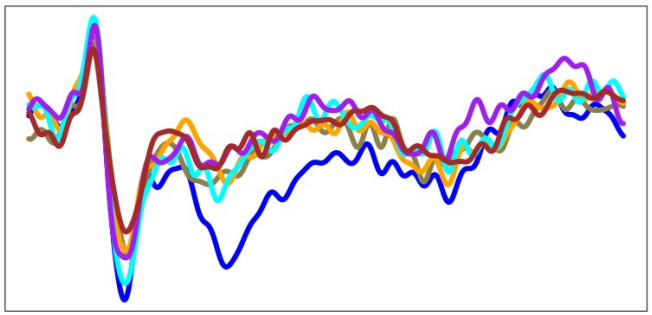
Motivation

- Feasibility shown by Martinovic et al. (USENIX 2012)
  - i.e., there is exploitable information in EEG signals



- BCIs more and more commercialized (e.g., for gaming)

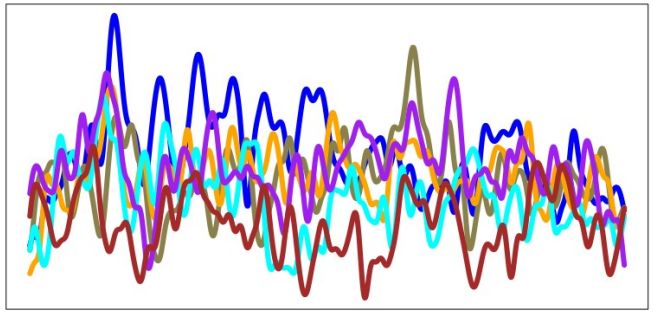
Challenge: low (& irregular) SNR



Range of the
Standard Deviation

≈

Differences in the
mean traces

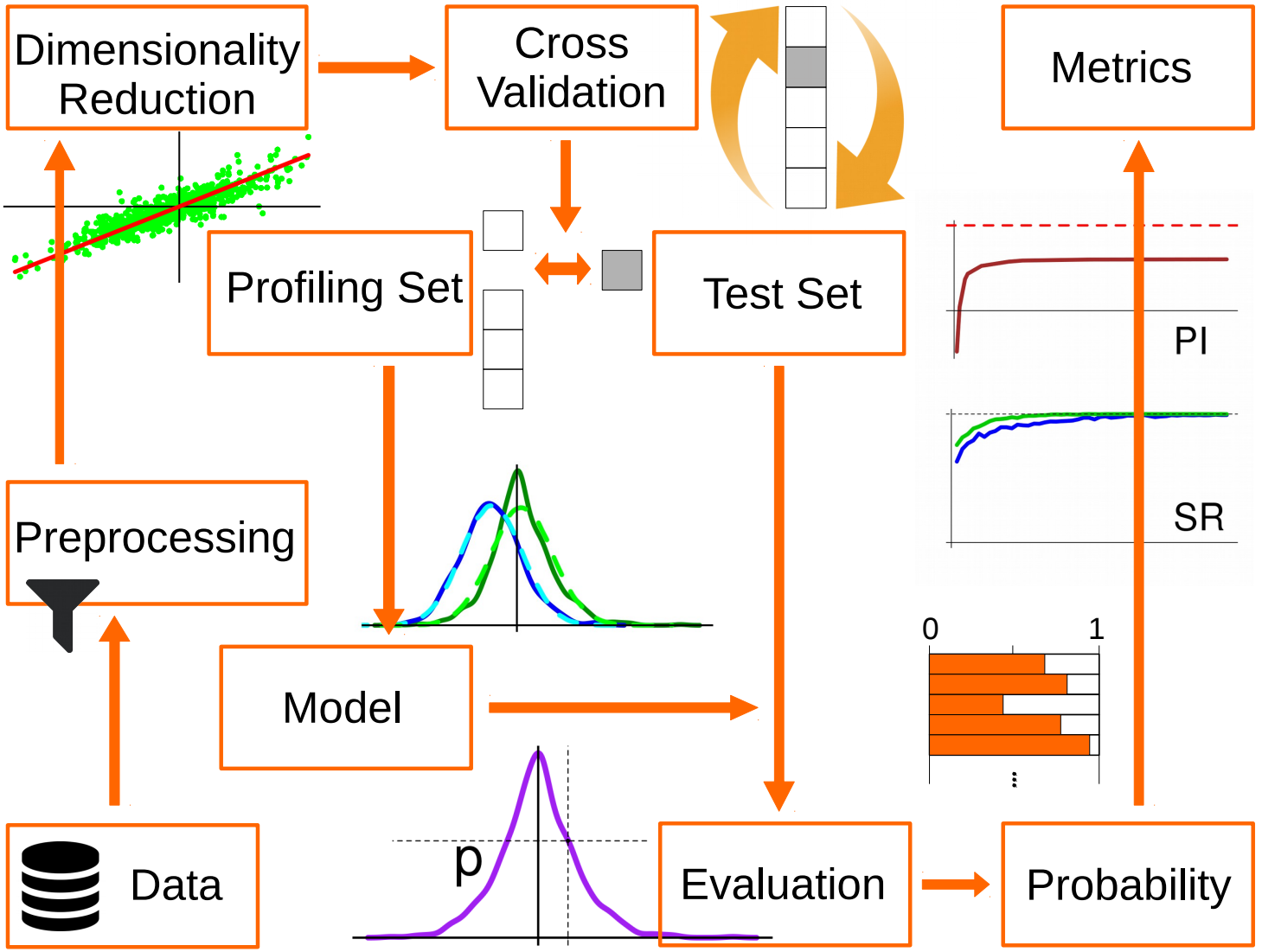


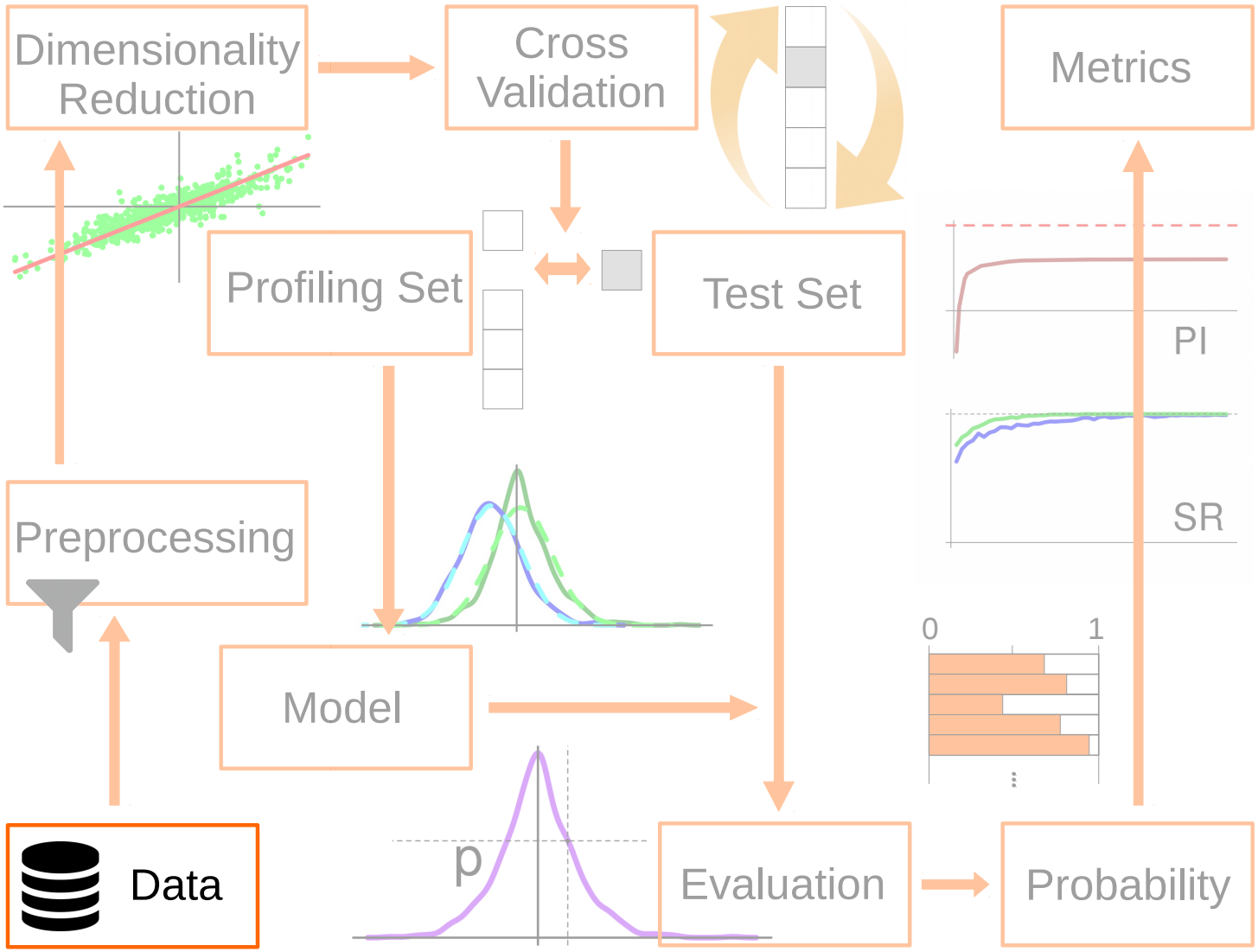
Main questions

- Can we extract PINs exactly (or only partially)?
- Can we extract them with sufficient confidence?
- How do supervised (aka profiled) and unsupervised (aka non-profiled) attacks compare?
- *How similar/different are different subjects?*
- *What are the consequences for security & privacy?*

Note: results can be viewed as positive or negative!

Related works: semantic associations and incongruities


Dimensionality Reduction
Cross Validation
Metrics
Profiling Set
Test Set
PI
SR
Preprocessing
Model
0
1
p
Data
Evaluation
Probability

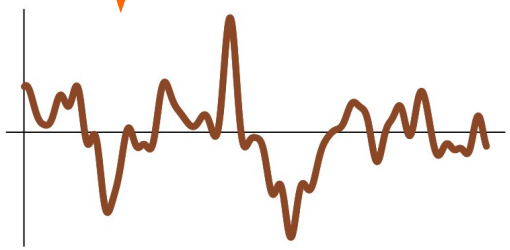

Dimensionality Reduction
Cross Validation
Metrics
Profiling Set
Test Set
PI
SR
Preprocessing
0
1
Model
p
Data
Evaluation
Probability

The data

8 subjects

pins

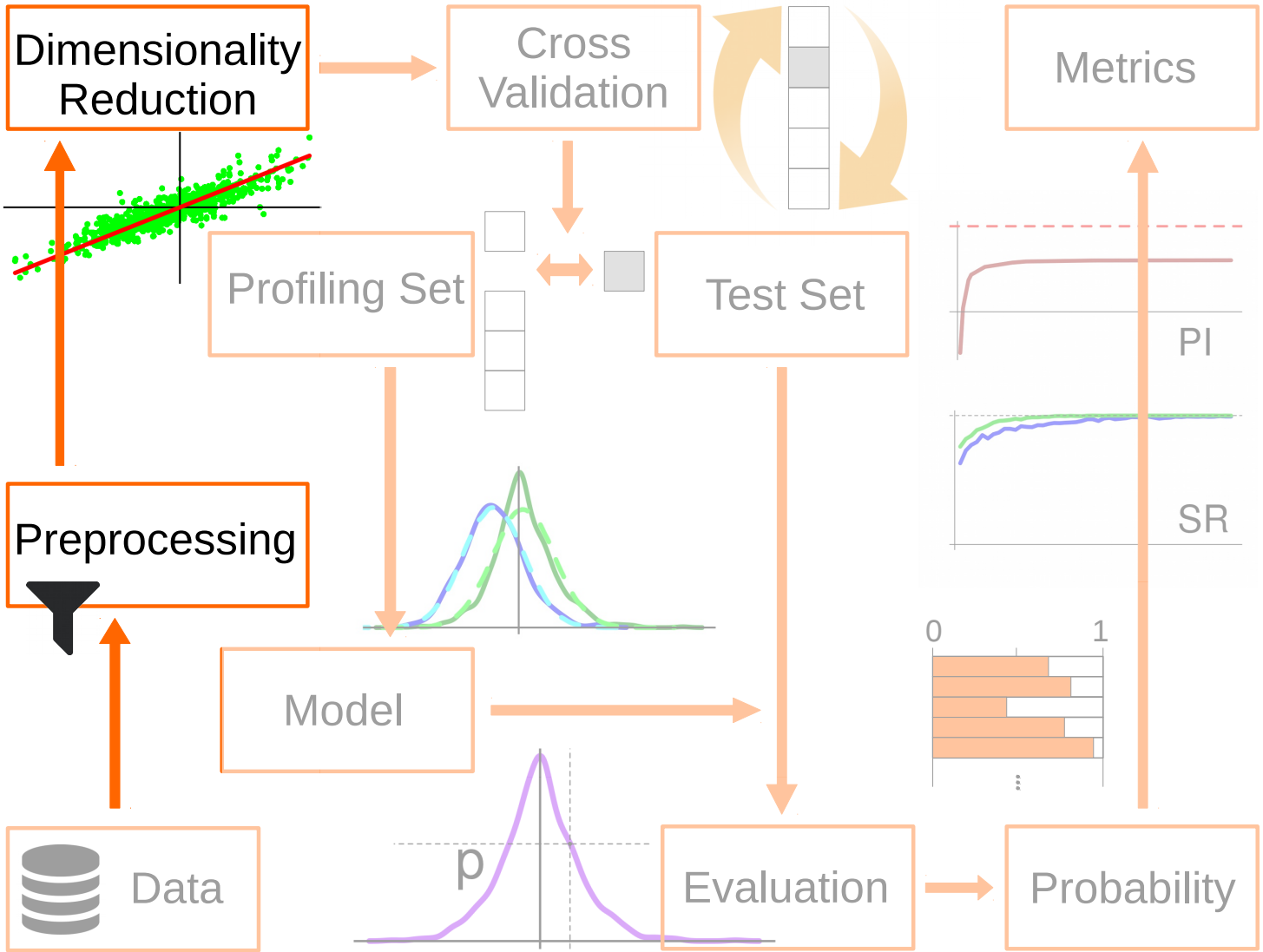
Headset :
32 electrodes



1000 x 1 ms

150 traces

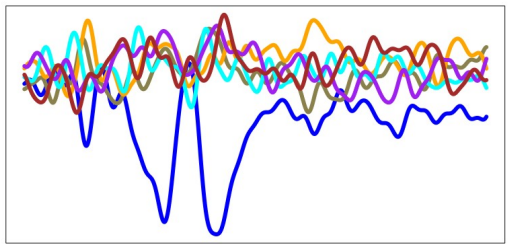
- per pin
- per subject
- per electrode


Dimensionality Reduction
Cross Validation
Metrics
Profiling Set
Test Set
PI
SR
Preprocessing
Model
0
1
p
Data
Evaluation
Probability

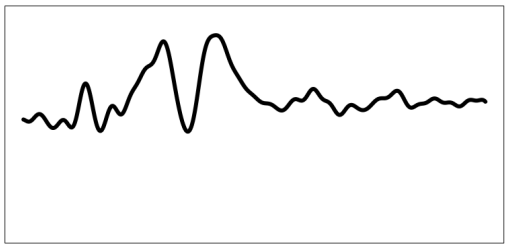
Dimensionality reduction

Principal Component Analysis (PCA)

Mean traces

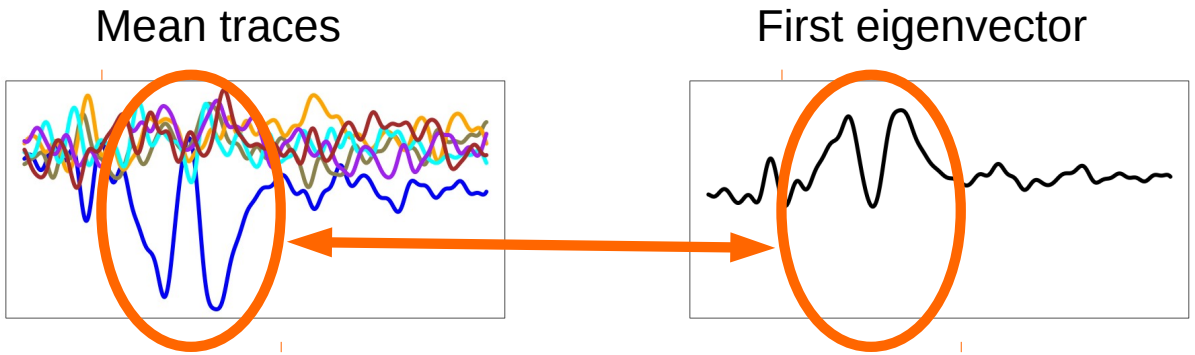


First eigenvector



Dimensionality reduction

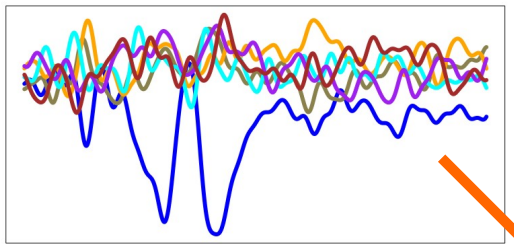
Principal Component Analysis (PCA)



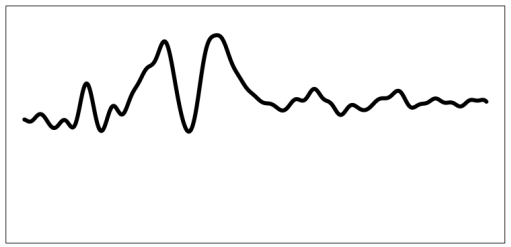
Dimensionality reduction

Principal Component Analysis (PCA)

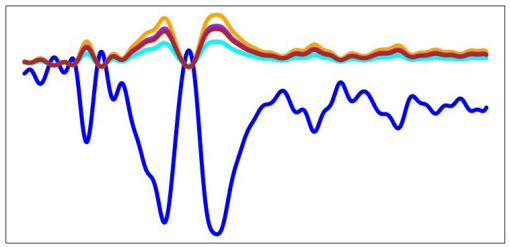
Mean traces



First eigenvector

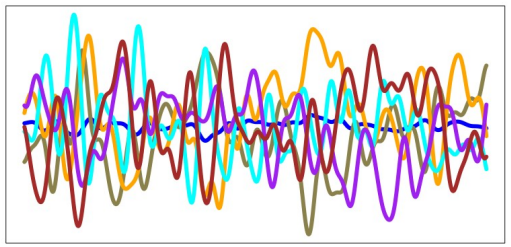


Reconstructed signal (1D)



+

Reconstructed signal (other Ds)

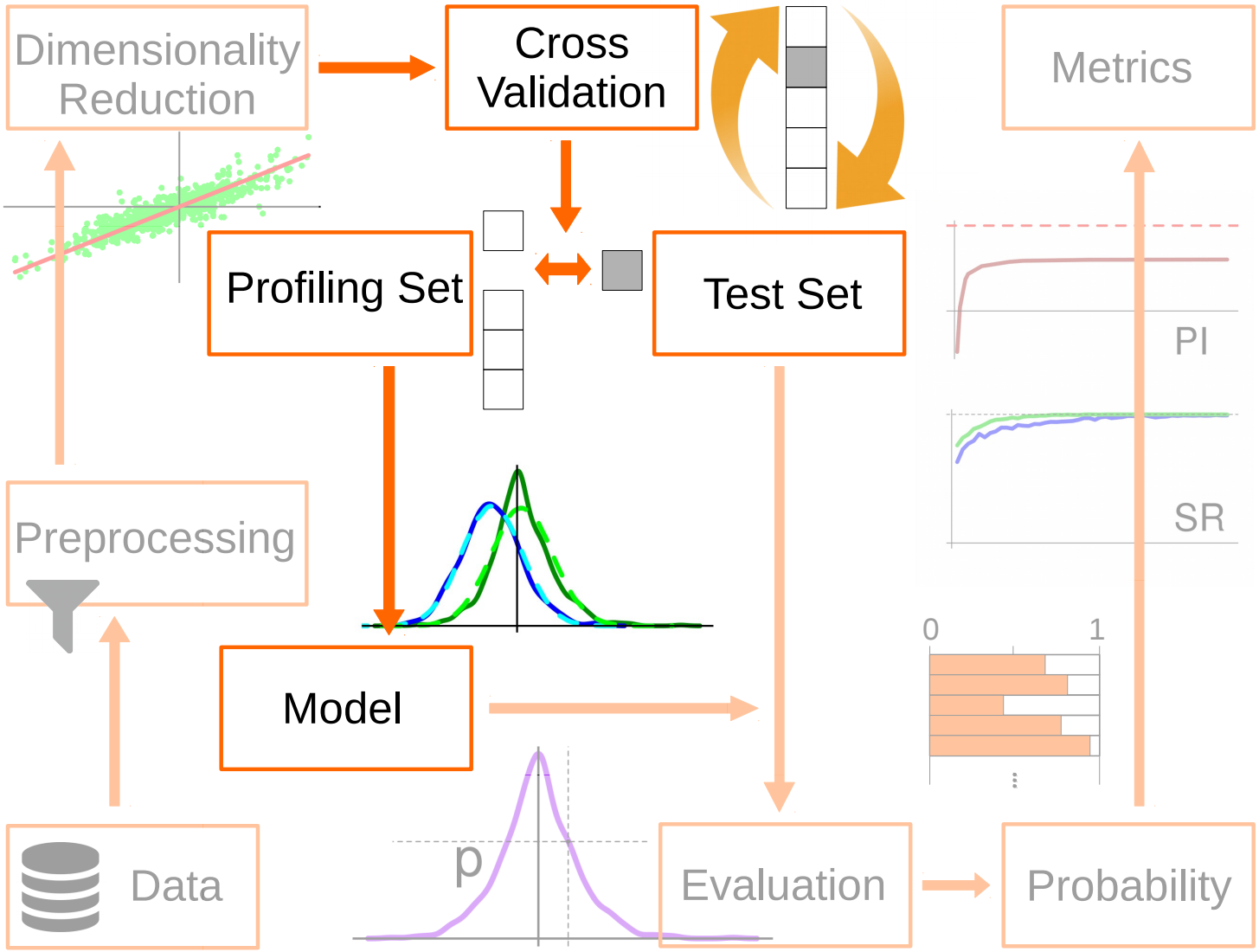


Principal Component Analysis

\+ Average PCA: one dimension is sufficient
\+ Projects 1000 dimensions to a single one

– Estimating means becomes expensive with many PINs

$\Rightarrow$ Raw PCA also studied in the paper
(Requires more dimensions and outliers management)


Dimensionality Reduction
Cross Validation
Metrics
Profiling Set
Test Set
PI
SR
Preprocessing
Model
0
1
Data
p
Evaluation
Probability

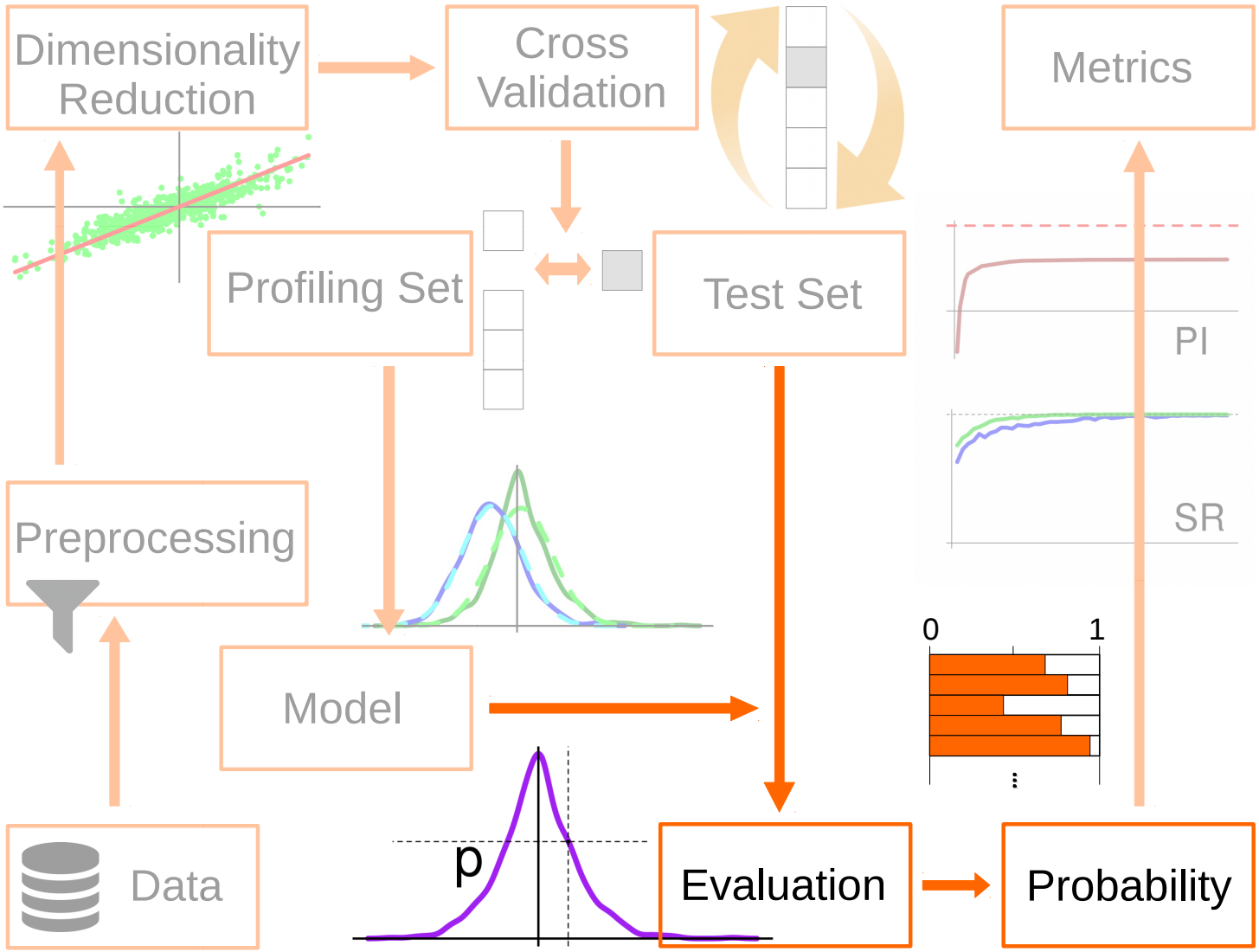
Profiling/modeling

- **Gaussian** estimation with mean $\hat{\mu}$ and variance $\hat{\sigma}$:

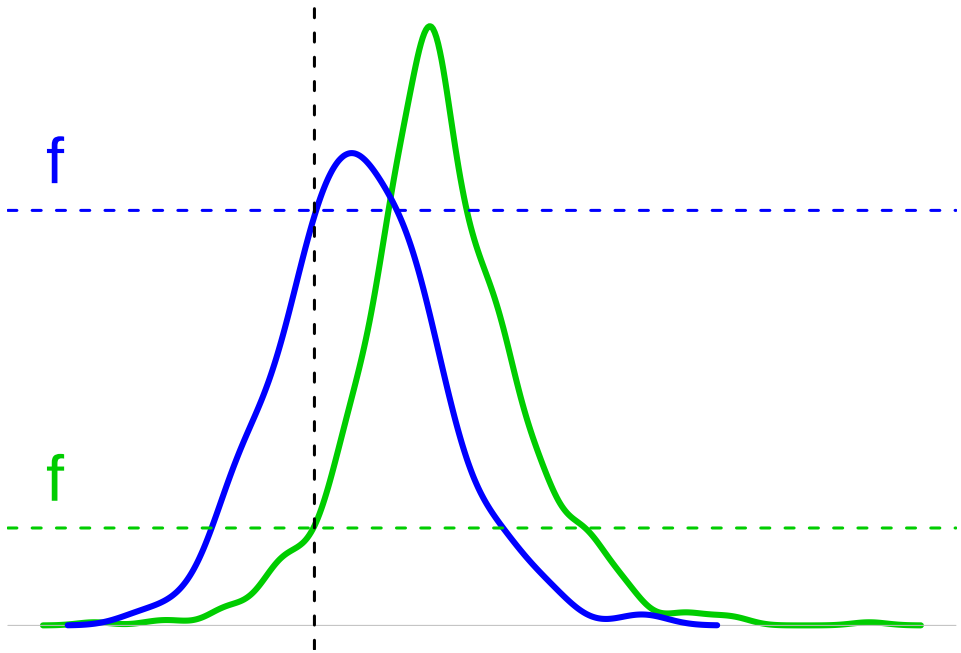
$$\hat{f}_{\mathrm{g}}(x) = \frac{1}{\sqrt{2\pi\hat{\sigma}}} \exp\left(-\frac{1}{2}\left(\frac{x-\hat{\mu}}{\hat{\sigma}}\right)^2\right)$$

- **Kernel** density estimation with bandwidth parameter h and samples $x_1, \ldots, x_n$:

$$\hat{f}_{\mathrm{k}}(x) = \frac{1}{nh}\sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-x_i}{h}\right)^2\right)$$


Dimensionality Reduction
Cross Validation
Metrics
Profiling Set
Test Set
PI
SR
Preprocessing
Model
0
1
Data
p
Evaluation
Probability

Model evaluation



Probability generation

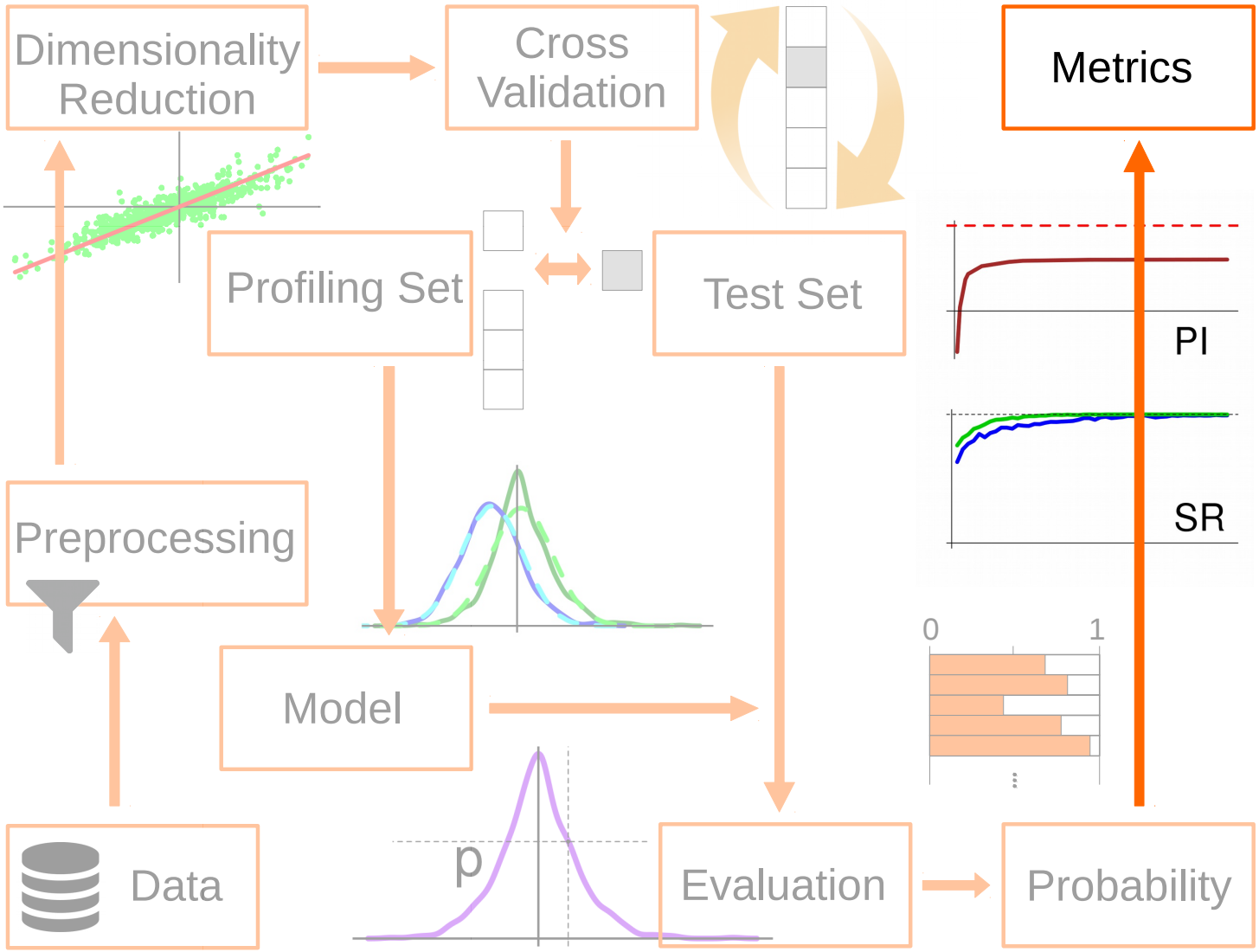
- From the estimated PDFs

$$\hat{\mathbf{f}}_{model}[\mathbf{o} \mid \mathbf{p} = \text{correct PIN}] = f,$$

$$\hat{\mathbf{f}}_{model}[\mathbf{o} \mid \mathbf{p} = \text{incorrect PIN}] = f.$$

- Produce probabilities thanks to Bayes

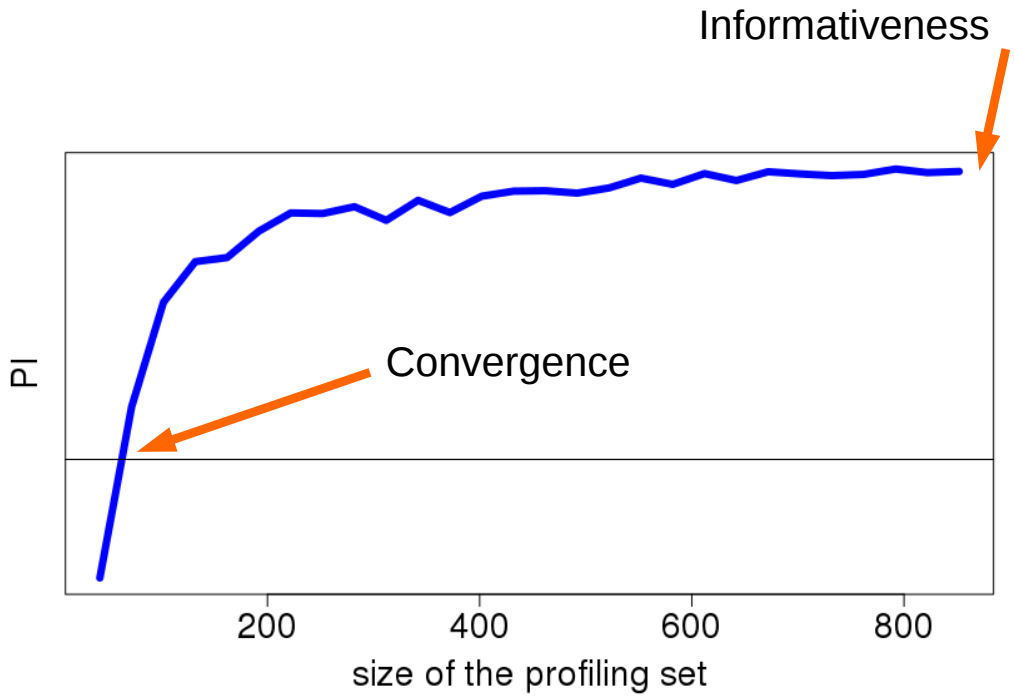
$$\hat{\mathbf{Pr}}_{model}[\mathbf{p} \mid \mathbf{o}] = \frac{\hat{\mathbf{f}}_{model}[\mathbf{o} \mid \mathbf{p}] \cdot \mathbf{Pr}[\mathbf{p}]}{\sum_{p^\star} \hat{\mathbf{f}}_{model}[\mathbf{o} \mid \mathbf{p}^\star] \cdot \mathbf{Pr}[\mathbf{p}^\star]}$$


Dimensionality Reduction
Cross Validation
Metrics
Profiling Set
Test Set
PI
SR
Preprocessing
Model
0
1
Data
p
Evaluation
Probability

Metrics

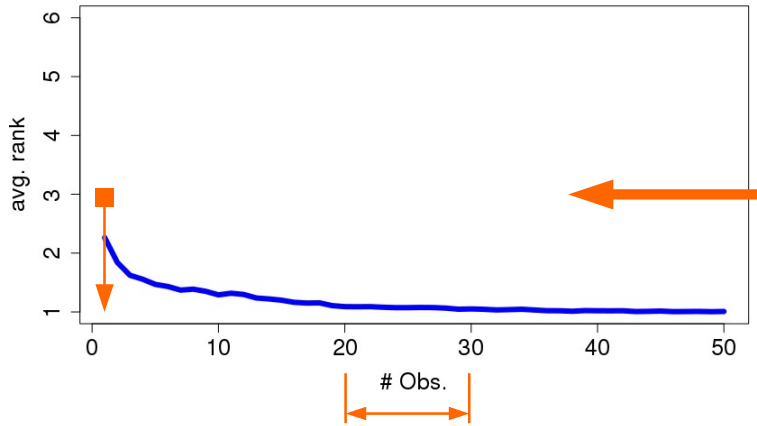
- **Perceived Information (PI)**: amount of information extracted from the observations (given a model)
- **Success Rate**: probability of correct classification (estimated for correct and incorrect PIN values)
- **Average rank**: average position of the correct PIN value in the sorted list of 6 possible ones

Profiled attacks: PI



- Convergence reached after $\approx$ 200 to 400 traces

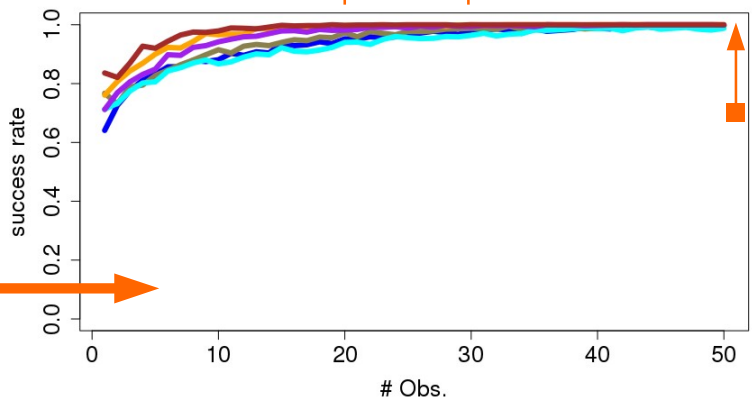
Profiled attacks: SR & avg. rank



Rank 3 to 1

Need about 20-30 traces

Success Rate from
65-85% to 100%



Profiled attacks: summary

- PIN recovery for most subjects (7 out of 8)
  - Failure due to another "distinguishable" event
  - Seems inherent to the investigated setup

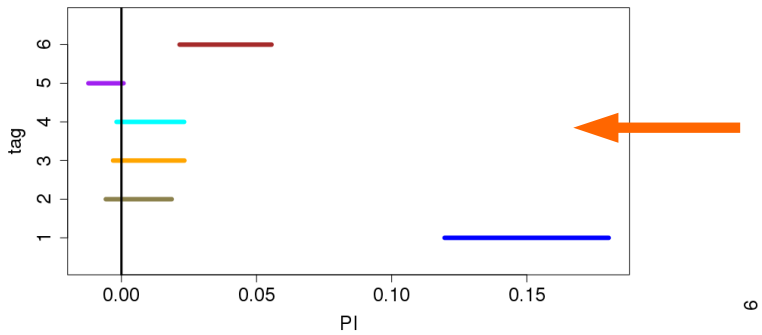
$\Rightarrow$ We minimized false negatives (to allow enumeration)

- Answers our first questions: PINs can be extracted from EEG signals partially and with good confidence
  - $\exists$ scenarii where this can be damaging
- Profiling more expensive than online attack
  - Given a good model is available

Unprofiled attacks

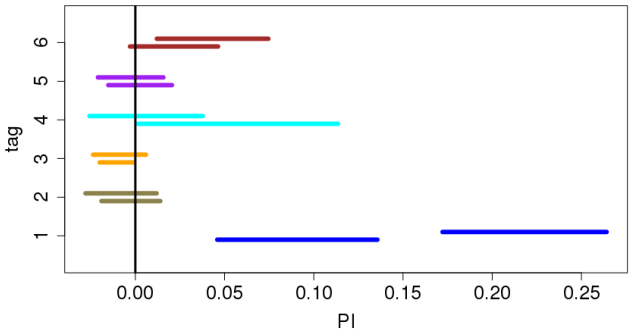
- Consider each PIN to be the correct one
- Estimate the PI on-the-fly for each case
  - And compute confidence intervals
- Correct PIN is expected to have the highest PI

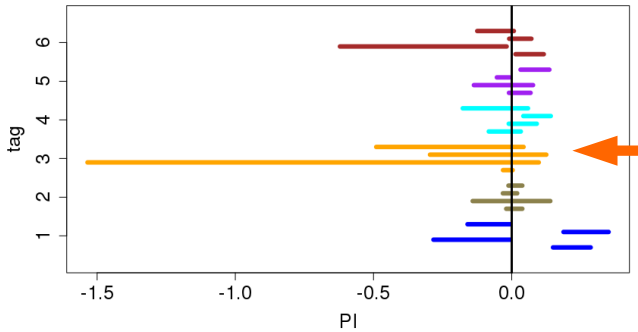
Unprofiled attacks: results



≈ 1 x 900

≈ 2 x 450





≈ 4 x 225

- Nicely correlated with the profiling cost (slide 21)

Other results

- **Portability**: attack one subject with a model built from others' data: less successful (5 out of 8)
- **Privacy**: target the subjects' identities instead of their PIN: positive results obtained for all users

Conclusions

- Information available and exploitable with confidence
- Yet not sufficient for full (4-digit) PIN recovery
- Mostly because of signal instability / subjects' focus
- Biggest risk here: reduction of the guessing entropy

Conclusions

- Information available and exploitable with confidence
- Yet not sufficient for full (4-digit) PIN recovery
- Mostly because of signal instability / subjects' focus
- Biggest risk here: reduction of the guessing entropy

More generally...

- Targets of smaller cardinality would be more worrying
- Privacy is also more worrying (unbounded data)
- Motivation for MPC, FHE, ...
- Much more research needed

Thanks !