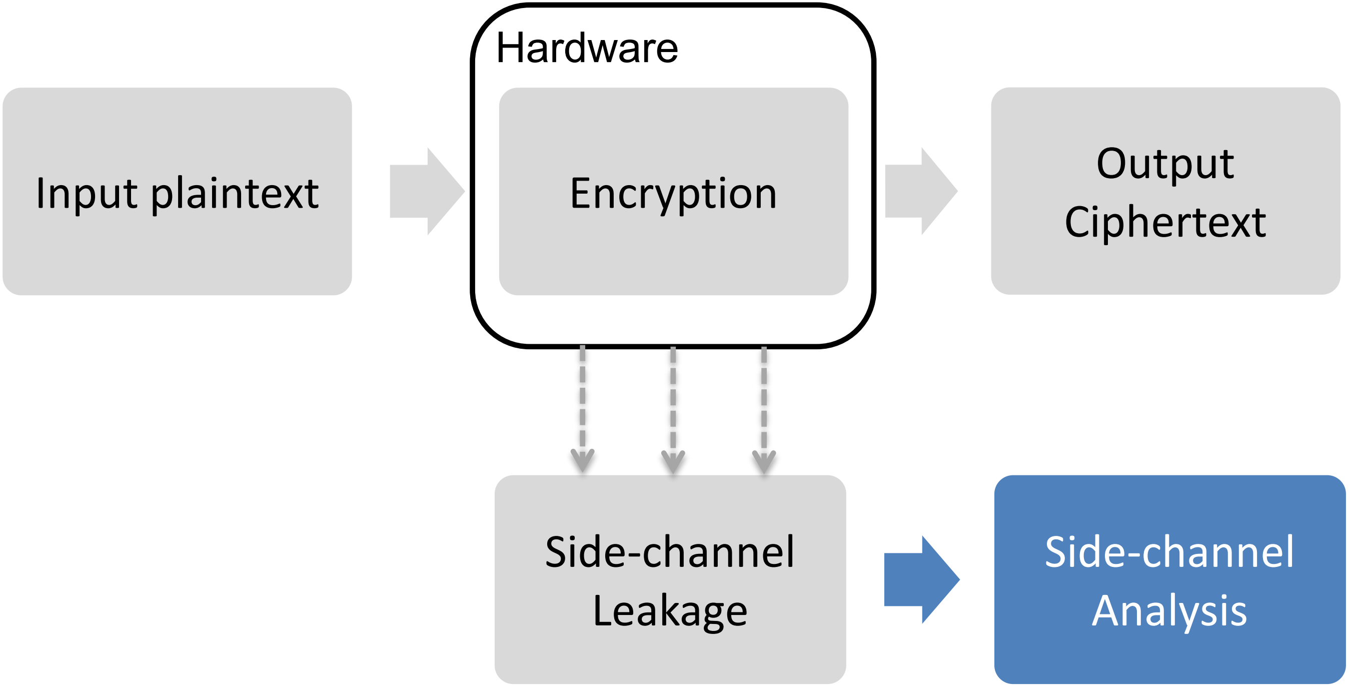
On the Evaluation of Deep Learning-based Side-channel Analysis

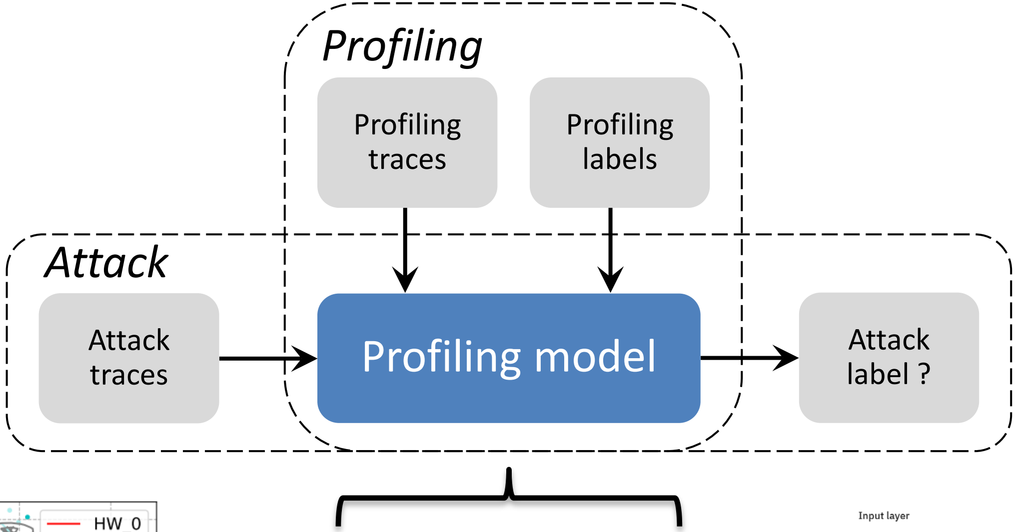
Lichao Wu[1], Guilherme Perin[1,2] and Stjepan Picek[2, 1]

[1]*Delft University of Technology, The Netherlands*
[2]*Radboud University, The Netherlands*

Side-channel Analysis (SCA)



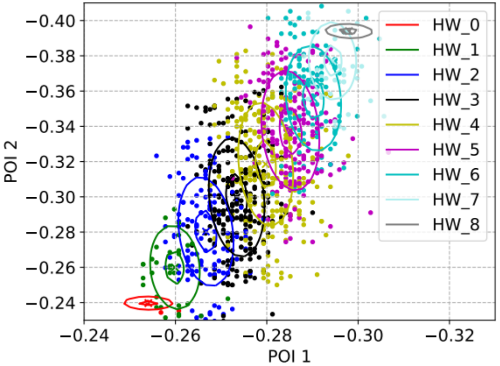
Profiling Side-channel Attack

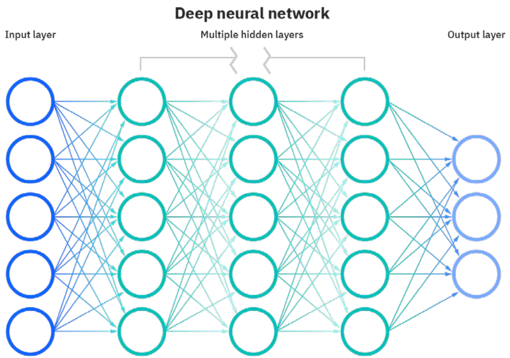


Template attack

Deep learning attack

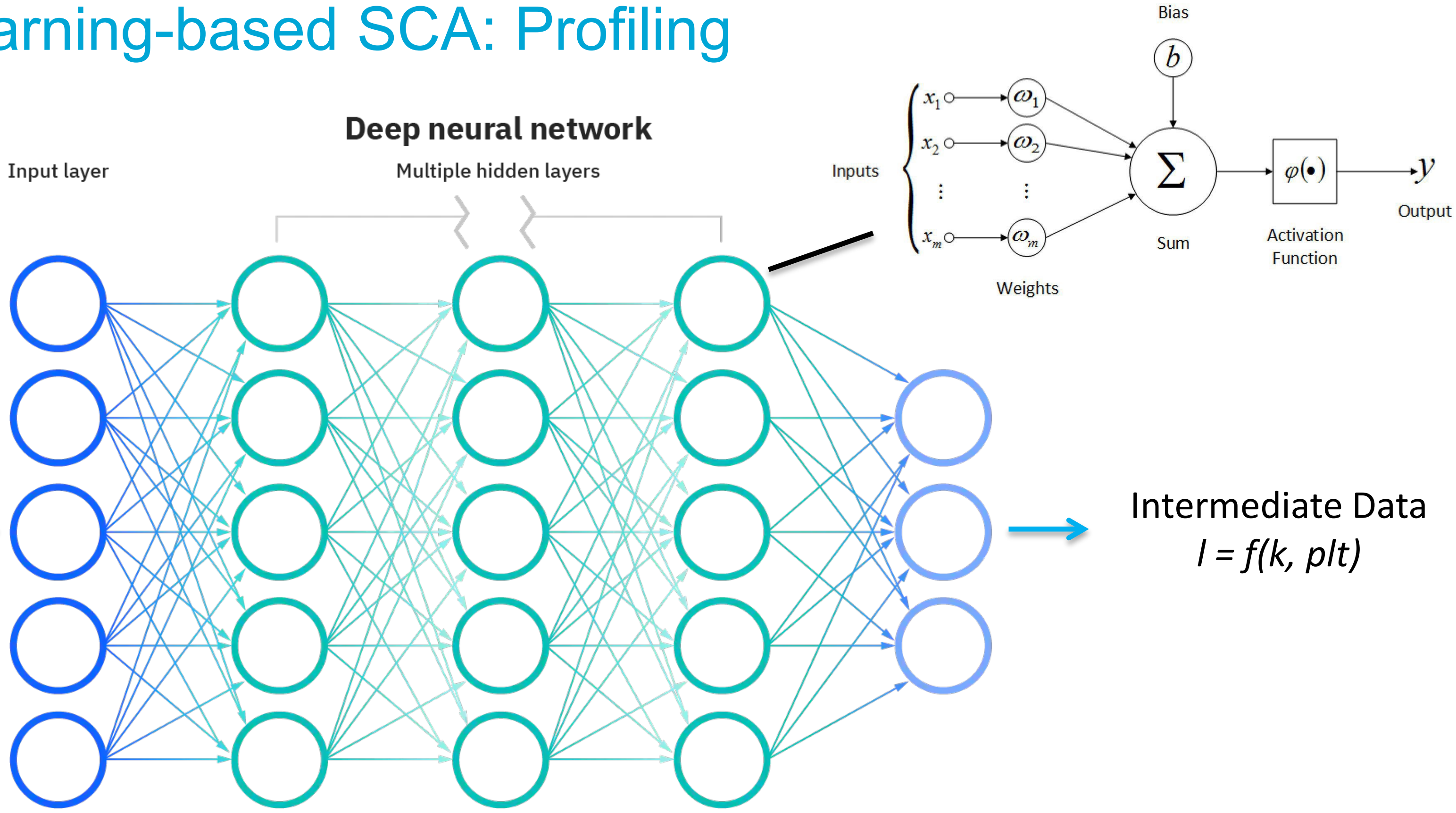
...





Deep Learning-based SCA: Profiling

Profiling Traces

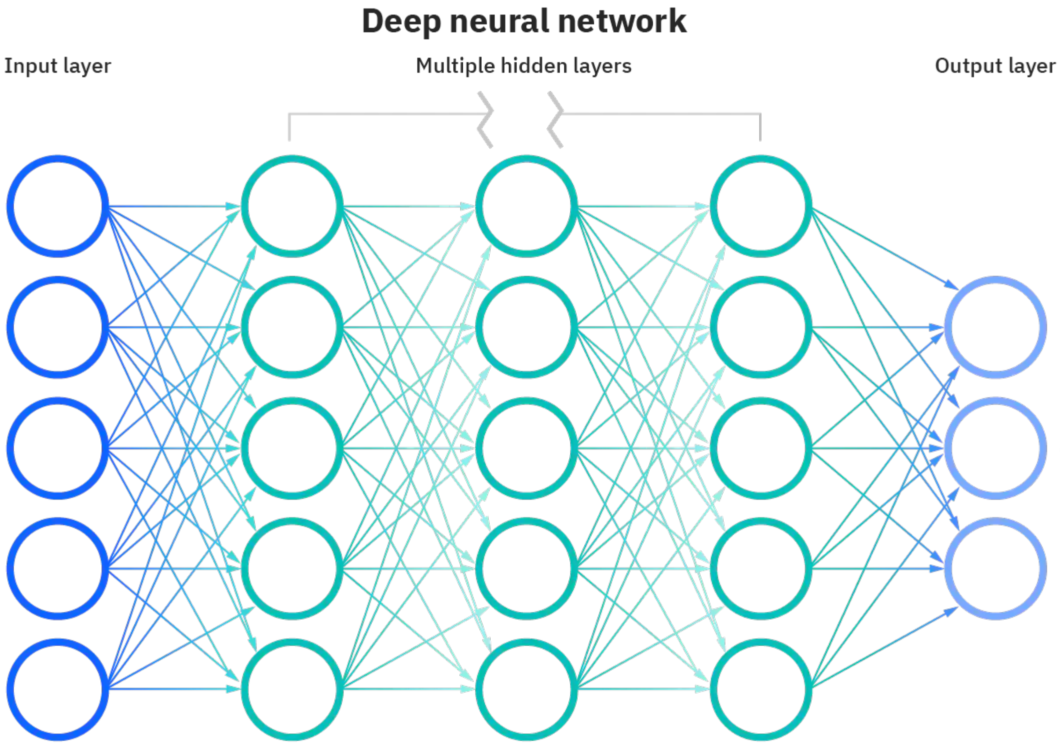


Intermediate Data
I = f(k, plt)

TUDelft

Deep Learning-based SCA: Attack

Attack Traces



Prob(l), l∈L

Deep Learning-based SCA: Post-processing

Prob(l), l∈L $\xrightarrow{L = f(k,\ plt)}$ Prob(k), k∈K $\xrightarrow{argmax}$ Secret key get!!!

Deep Learning-based SCA: Post-processing

$$
\begin{tikzcd}
\mathrm{Prob(l),\ l}\in\mathrm{L} \arrow[r, "L = f(k{,}\ plt)"] & Prob(k),\ k\in K \arrow[d, "+"] & \\
\mathrm{Prob(l),\ l}\in\mathrm{L} \arrow[r, "L = f(k{,}\ plt)"] & Prob(k),\ k\in K \arrow[d, "+"] & \\
\mathrm{Prob(l),\ l}\in\mathrm{L} \arrow[r, "L = f(k{,}\ plt)"] & Prob(k),\ k\in K \arrow[d, "+"] & \\
\mathrm{Prob(l),\ l}\in\mathrm{L} \arrow[r, "L = f(k{,}\ plt)"] & Prob(k),\ k\in K \arrow[r, "argmax"] & \text{Secret key get!!!}
\end{tikzcd}
$$

Deep Learning-based SCA: Post-processing

Prob(l), l∈L $\xrightarrow{L\,=\,f(k,\ plt)}$ *Prob(k), k∈K*

↓ +

Prob(l), l∈L $\xrightarrow{L\,=\,f(k,\ plt)}$ *Prob(k), k∈K*

↓ +

Prob(l), l∈L $\xrightarrow{L\,=\,f(k,\ plt)}$ *Prob(k), k∈K*

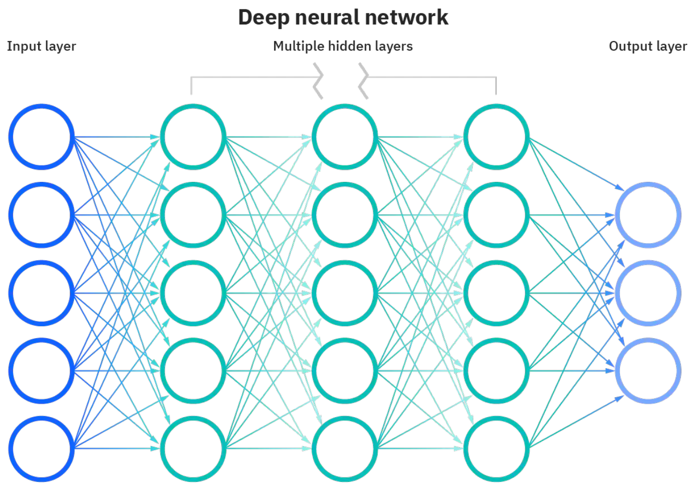
↓ +

Prob(l), l∈L $\xrightarrow{L\,=\,f(k,\ plt)}$ *Prob(k), k∈K* $\xrightarrow{argsort}$ Key Rank

Key Rank ↑↑↑ *average* → Guessing Entropy

Algorithmic Randomness of DL-based SCA

- Initialization
  - Random weight & bias initialization
- Regularization
  - Dropout layer
- Optimization procedure
  - Stochastic gradient descent (SGD)
  - Mini-batches
  - Limited-memory Broyden–Fletcher–Goldfarb–Shannoalgorithm (L-BFGS)
- Others
  - Random architecture of the model



TUDelft

The Goal of This Work

- Investigate the influence of algorithmic randomness on the attack performance
  - Mean
  - Standard deviation
- Investigate the most appropriate summary statistic for evaluating the attack performance
  - Arithmetic mean $\overline{x} = \frac{1}{z}\sum_{i=1}^{z} x_i$
  - Geometric mean $\check{x} = \left(\prod_{i=1}^{z} x_i\right)^{\frac{1}{z}}$
  - Medium mean $\tilde{x} = \frac{x_{\frac{z}{2}} + x_{\frac{z}{2}+1}}{2}$
- Investigate how a different number of independent experiments (key rank evaluations) in the attack phase influences attack performance

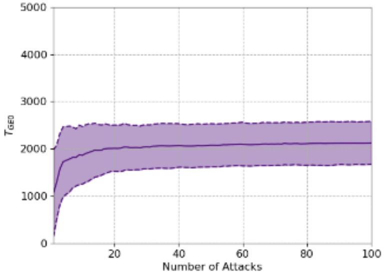
TU Delft

Experimental Results

- The Influence of Algorithmic Randomness
- The Best Summary Statistic for Evaluating the Attack Performance

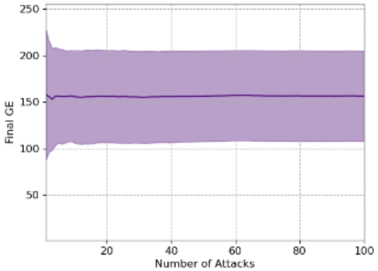
The Influence of Algorithmic Randomness

Initialization



(a) T_{GE0}: Random initialization of weights and biases of a well-performing model.

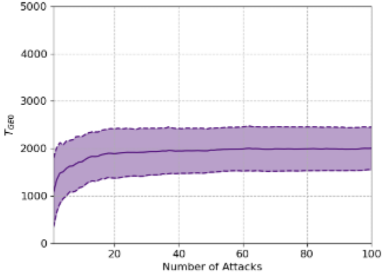
DL Architecture



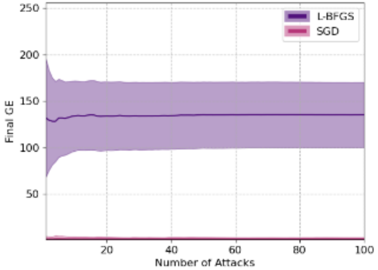
(b) GE: Random initialization of weights and biases of a bad-performing model.

Initialization

Regularization



(c) T_{GE0}: Regularization techniques (dropout on well-performing model.)

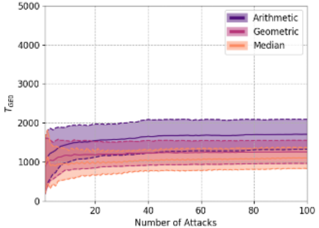


(d) GE: Optimization techniques (SGD, L-BFGS on well-performing model.)

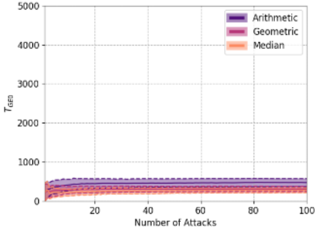
Optimization Procedure

TUDelft

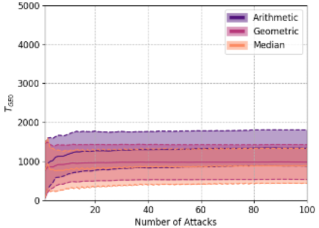
The Best Statistic for Evaluating the Attack Performance: on the ASCAD Fixed Key Dataset



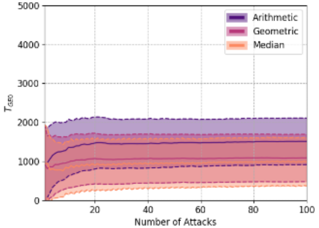
(a) Random MLP with the HW leakage model.



(b) Random MLP with the ID leakage model.

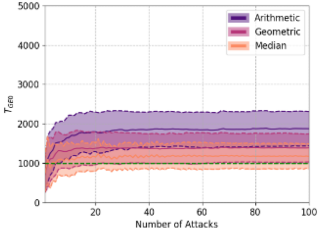


(c) Random CNN with the HW leakage model.

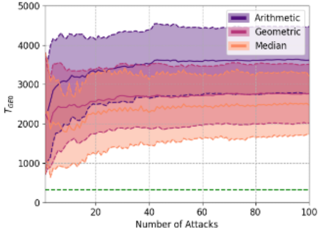


(d) Random CNN with the ID leakage model.

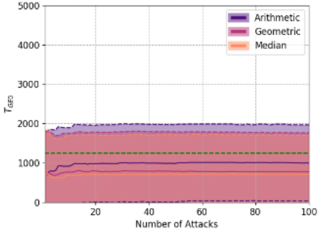
Random Architectures



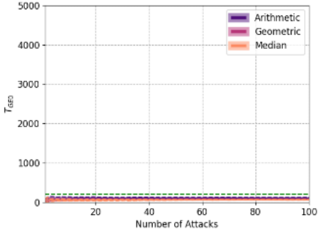
(a) State-of-the-art MLP with the HW leakage model.



(b) State-of-the-art MLP with the ID leakage model.



(c) State-of-the-art CNN with the HW leakage model.

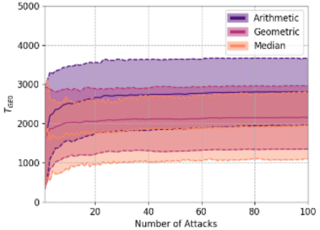


(d) State-of-the-art CNN with the ID leakage model.

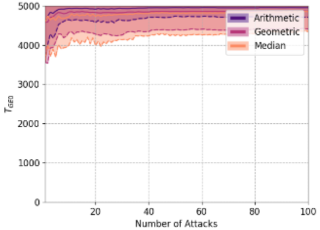
SotA Architectures

TU Delft

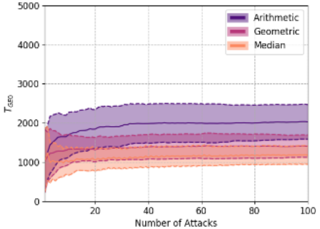
The Best Statistic for Evaluating the Attack Performance: on the ASCAD Random Keys Dataset



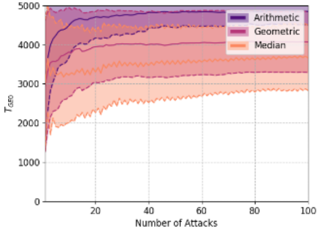
(a) Random MLP with the HW leakage model.



(b) Random MLP with the ID leakage model.

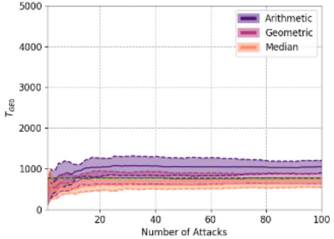


(c) Random CNN with the HW leakage model.

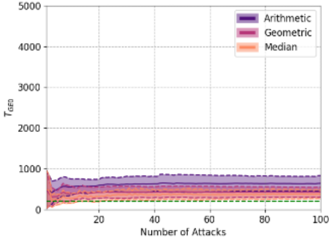


(d) Random CNN with the ID leakage model.

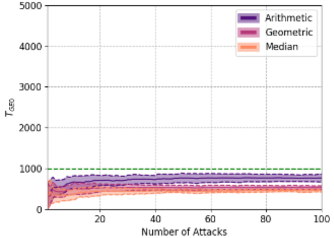
Random Architectures



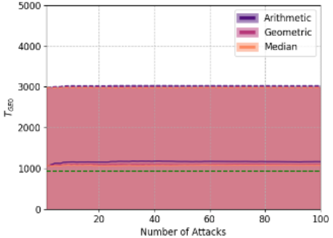
(a) State-of-the-art MLP with the HW leakage model.



(b) State-of-the-art MLP with the ID leakage model.



(c) State-of-the-art CNN with the HW leakage model.

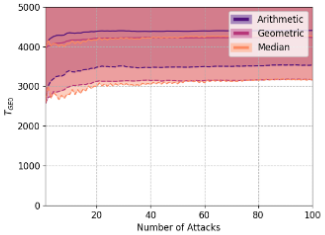


(d) State-of-the-art CNN with the ID leakage model.

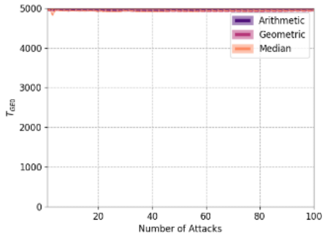
SotA Architectures

TUDelft

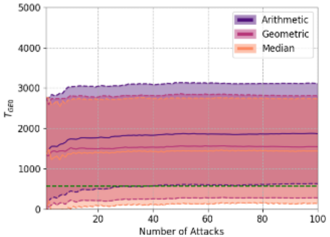
The Best Statistic for Evaluating the Attack Performance: on the CHES_CTF 2018 Dataset



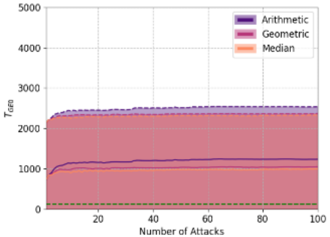
(a) Random MLP with the HW leakage model.



(b) Random CNN with the HW leakage model (most of the attacks failed to converge).



(a) State-of-the-art MLP with the HW leakage model.



(b) State-of-the-art CNN with the HW leakage model.

Random Architectures

SotA Architectures

TUDelft

Conclusions

- Deep learning-based SCA can show different attack results due to algorithmic randomness and skewed distribution of attack results
- The median mean, instead of arithmetic mean, is the best choice since it is not affected by outliers and thus represents a resistant measure of a center
- Large number of independent experiments to average the attack performance does not increase the stability of results
- For state-of-the-art models, a large standard deviation indicates the low stability. Thus, the performance of such models could be questionable when facing practical challenges such as devices' portability
- We emphasize the necessity of reporting the averaged performance over a number of profiling models with different weight initialization so that the actual attack performance can be reliably estimated

TU Delft

Future Works

- Consider dataset randomness and use more summary statistics
- It would be interesting to compare the results for line plots (as commonly used) and boxplots when depicting the GE results

TU Delft

Thanks for your attention!