The Need for MORE: Unsupervised Side-channel Analysis with Single Network Training and Multi-output Regression

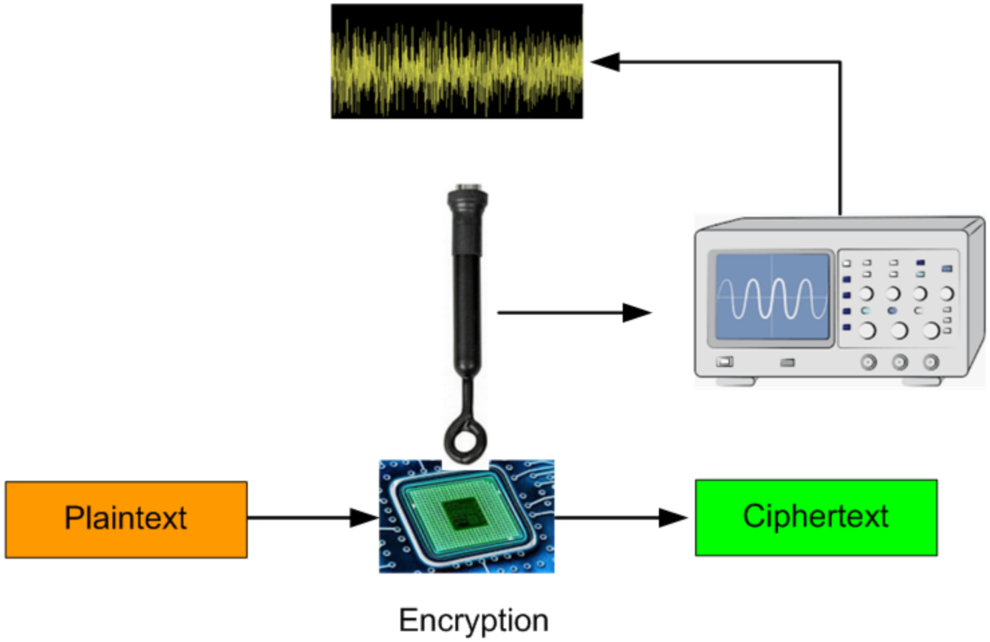
Ioana Savu[1], Marina Krček[2], Guilherme Perin[3], Lichao Wu[4], and Stjepan Picek[4]

[1]NXP Semiconductors, [2]Delft University of Technology, [3]Leiden University, [4]Radboud University, The Netherlands

Introducing MORE: Multi-Output Regression Enhanced

Enhanced attack performance over MOR in non-profiling SCA

Side-channel attacks (SCA)



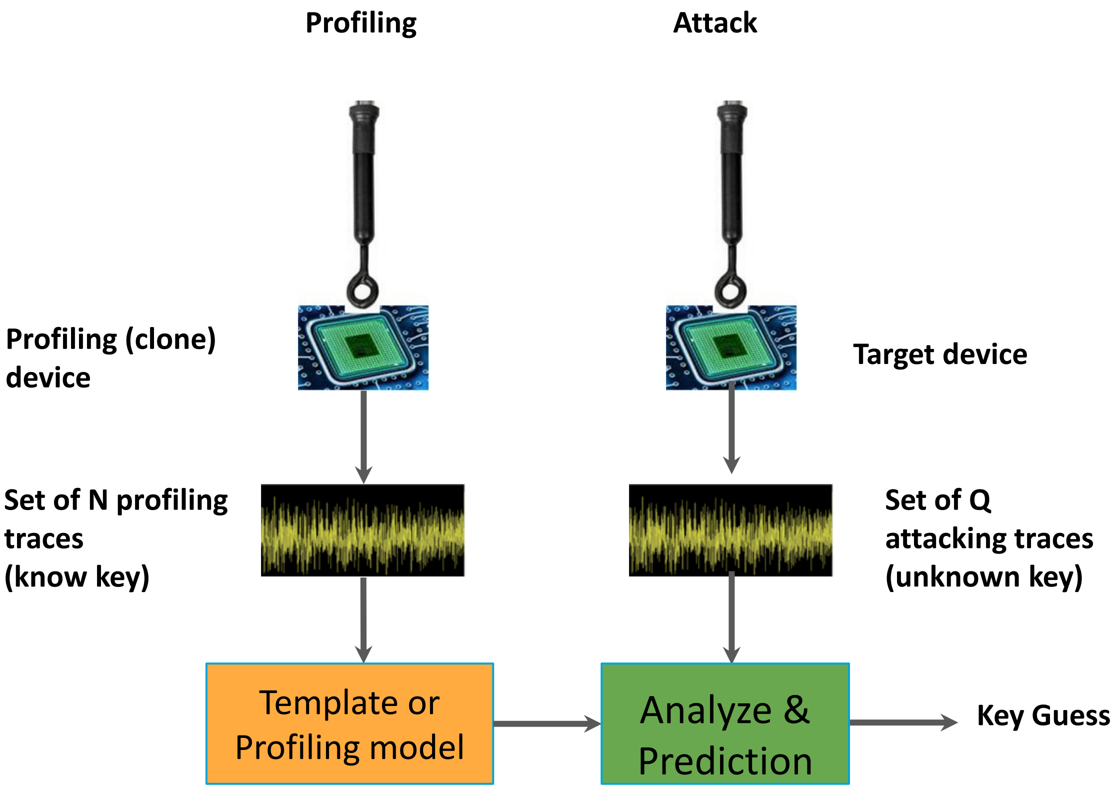
Power consumption

Electromagnetic leaks

Sound

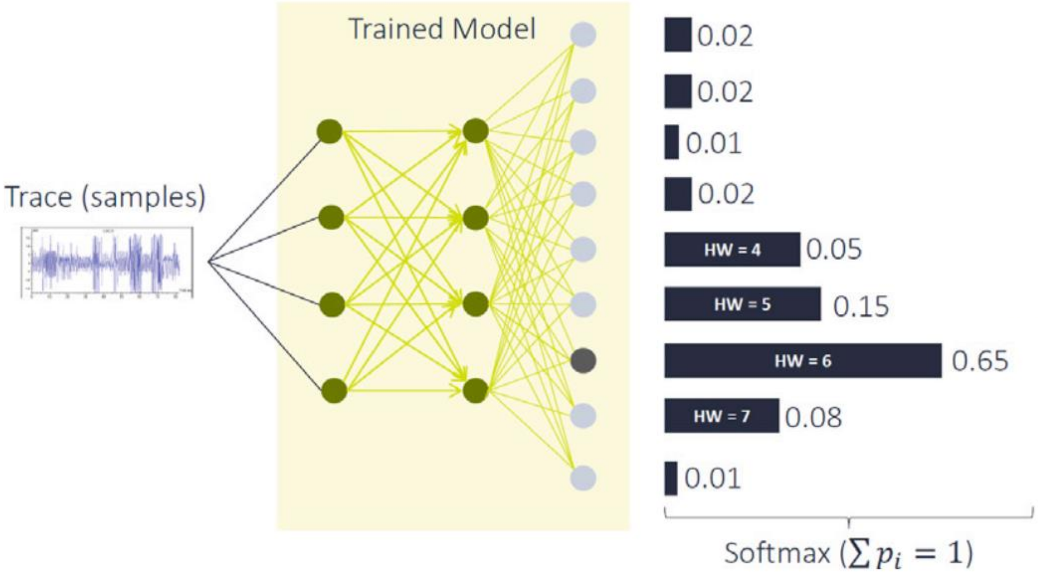
Timing

Profiling vs. non-profiling SCA



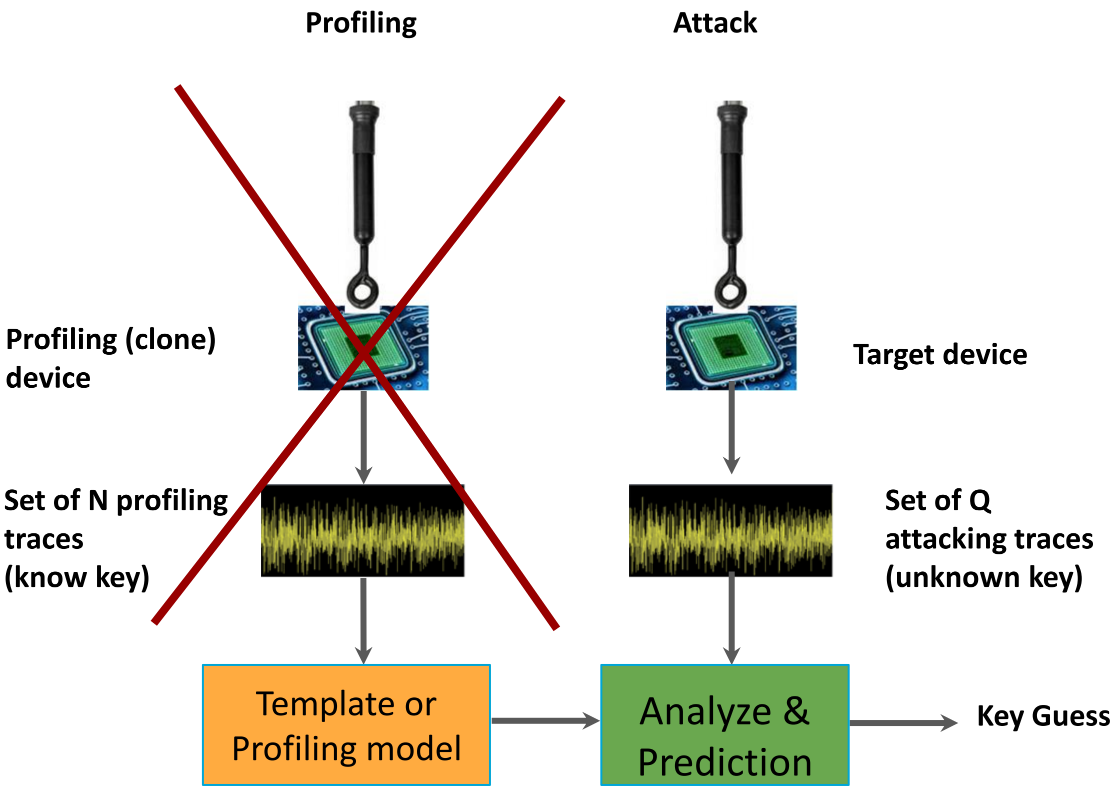
Deep Learning (DL) for profiling SCA

Supervised learning, classification



Kubota, Takaya, et al. "Deep learning side-channel attack against hardware implementations of AES." *Microprocessors and Microsystems* 87 (2021): 103383.

Profiling vs. non-profiling SCA



Deep Learning (DL) for non-profiling SCA

Differential Deep Learning Analysis (DDLA) by Timon (2019) [1]

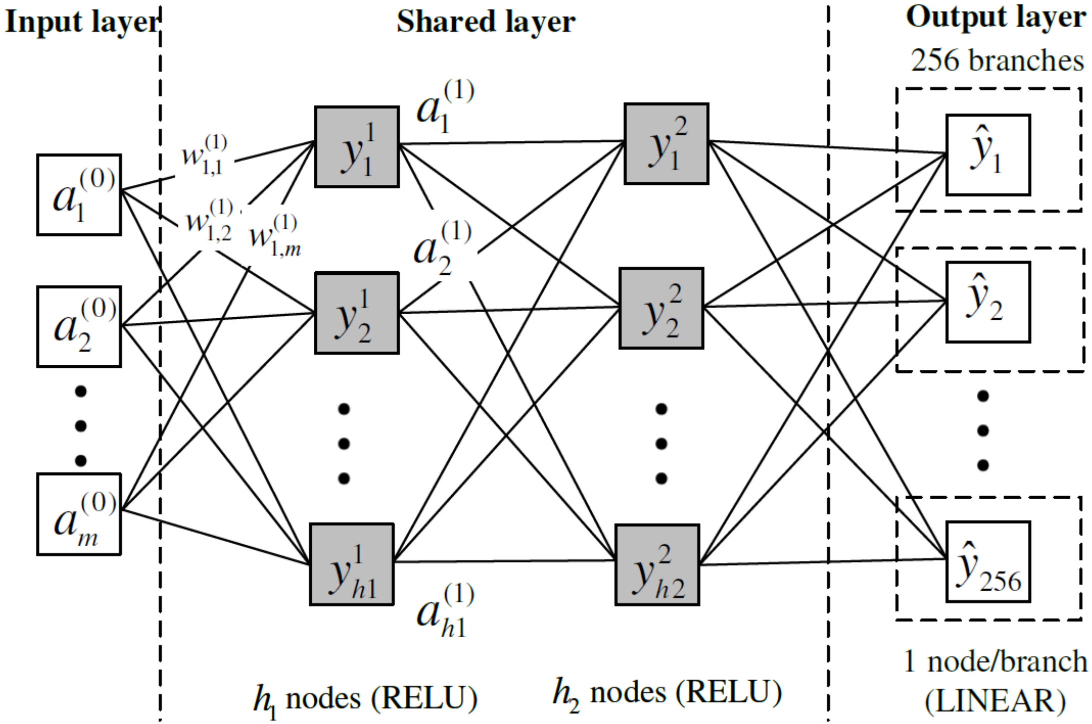
- Network trained for each key hypothesis (one key byte - training 256 times)
- Several solutions afterwards were proposed to decrease the time consumption [2]

Multi-output Learning (MOL) [3]

- Model trained to predict multiple outputs from a single input simultaneously
- Both classification (MOC) and regression (MOR)
- Lower execution time and better performance

[1] Timon, B.: Non-profiled deep learning-based side-channel attacks with sensitivity analysis. CHES 2019
[2] Kwon, D., Hong, S., Kim, H.: Optimizing implementations of non-profiled deep learning-based side-channel attacks. IEEE Access 2022
[3] Do, N.T., Le, P.C., Hoang, V.P., Doan, V.S., Nguyen, H.G., Pham, C.K.: MO-DLSCA: Deep Learning Based Non-profiled Side Channel Analysis Using Multi-output Neural Networks. ATC 2022

Multi-output Regression (MOR)



Do, N.T., Le, P.C., Hoang, V.P., Doan, V.S., Nguyen, H.G., Pham, C.K.: MO-DLSCA: Deep Learning Based Non-profiled Side Channel Analysis Using Multi-output Neural Networks. ATC 2022

Multi-output Regression (MOR)

Uses ID and HW leakage model

Non-profiling attacks hypothesize labels

- Supervised learning can be applied using the data (SC traces) and hypothesized labels

Loss function: mean squared error (MSE)

Key distinguisher: lowest MSE (loss)

Only one of the outputs is related to the correct key → find an outlier in the loss value

MOR Enhanced (MORE)

Loss functions - that could help emphasize the outliers

Key distinguisher - objective during hyperparameter tuning and attack

Validation set - mitigate overfitting

Loss functions

Common regression loss functions:

- MSE, MAE, Huber

Proposed:

- Pearson correlation

$$\mathcal{L}_{Pearson} = 1 - |\rho(\mathbf{y}(k), \hat{\mathbf{y}}(k))|, \ k \in \mathcal{K}$$

- Z-score normalization

$$\mathcal{L}_{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y_i})^2 \quad \rightarrow \quad \mathcal{L}_{Z-scoreMSE} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{y_i - \mu(\mathbf{y})}{\sigma(\mathbf{y})} - \frac{\hat{y_i} - \mu(\hat{\mathbf{y}})}{\sigma(\hat{\mathbf{y}})} \right)^2$$

Key distinguishers

Loss function distinguisher

$$k^* = \arg\min_k \mathcal{L}(k), \;\; k \in \mathcal{K}$$

Pearson correlation distinguisher

$$k^* = \arg\max_k \rho(\mathbf{y}(k), \hat{\mathbf{y}}(k)), \;\; k \in \mathcal{K}$$

Mitigate overfitting

Typically, as in MOR, all collected traces are used for training and attack

Proposed method:

- Two datasets: training and validation
- Attack on validation set

Loss function benchmark

MLP models only, on <u>ASCADf</u> and ASCADr, 1000 models

z-MSE at least 65% higher SR (excl. corr)

Validation set improved SR for the z-MSE for HW, decreases SR for others

Distinguisher	LM	Set	MSE	z-MSE	MAE	z-MAE	Huber	z-Huber	Corr	z-Corr
loss	HW	Validation	15.68%	**94.95%**	8.14%	3.75%	16.29%	18.46%	72.8%	77.1%
		Training	27.31%	**92.5%**	9.21%	6.25%	18.35%	8.54%	75.3%	77.5%
	ID	Validation	0%	**74.1%**	0.6%	0.3%	0%	0.8%	41.2%	44%
		Training	25.7%	**77.6%**	35.8%	14%	36.1%	10.4%	45.4%	50%
Pearson	HW	Validation	45.9%	**94.9%**	37.7%	6.7%	48.5%	29%	72.8%	77.1%
		Training	51%	**92.5%**	43.3%	10.1%	55.9%	35.7%	75.3%	77.5%
	ID	Validation	25.7%	**73.9%**	29.2%	22.8%	29%	13.7%	41.2%	44%
		Training	37.5%	**77.6%**	43.4%	26.1%	44.5%	16.1%	45.4%	50%

Loss function benchmark

ASCADr

z-MSE at least 74% higher SR (excl. corr)

Validation set improves SR for z-MSE in all cases, and for most loss functions when Pearson distinguisher is used

Distinguisher	LM	Set	MSE	z-MSE	MAE	z-MAE	Huber	z-Huber	Corr	z-Corr
loss	HW	Validation	3.9%	**91.2%**	1.5%	1%	2.2%	8.2%	44.7%	50%
		Training	30.9%	**77.2%**	15.9%	9%	20.8%	12.2%	42.9%	44.6%
	ID	Validation	1.2%	**22.2%**	1.3%	0%	0%	0.4%	14.7%	13.8%
		Training	4.5%	**9.4%**	2.6%	3.6%	0%	2.2%	6.9%	3.4%
Pearson	HW	Validation	36.6%	**90.2%**	28.4%	8.7%	39%	18.4%	44.7%	50%
		Training	31.5%	**77.2%**	27.5%	8.7%	33.4%	21.6%	42.9%	44.6%
	ID	Validation	10.8%	**20.5%**	8.8%	3.9%	10%	3.5%	14.7%	13.8%
		Training	4.3%	**9.4%**	3.5%	4.8%	2%	5.4%	6.9%	3.4%

Adaptability of loss functions

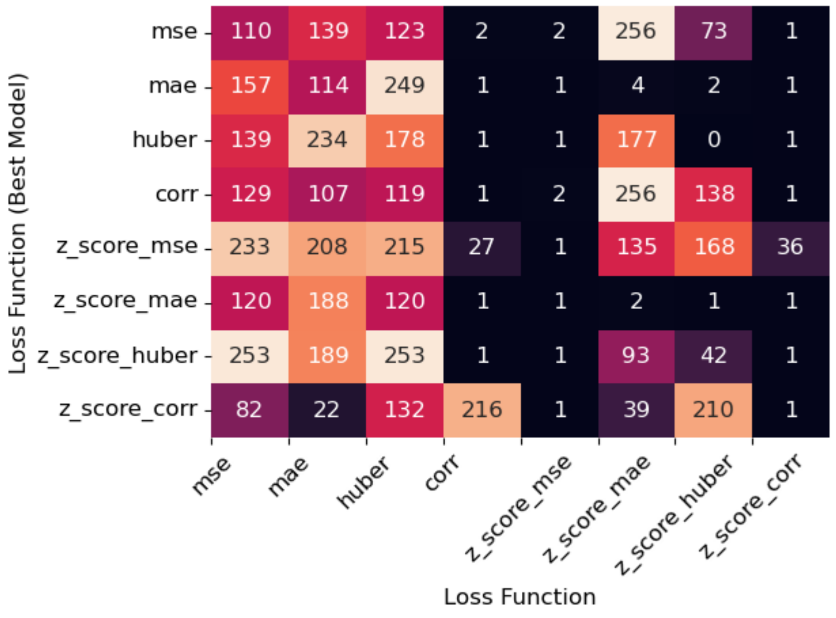
ID/HW, MLP/CNN, loss function - with each combination 100 models

Select the best model with each loss function

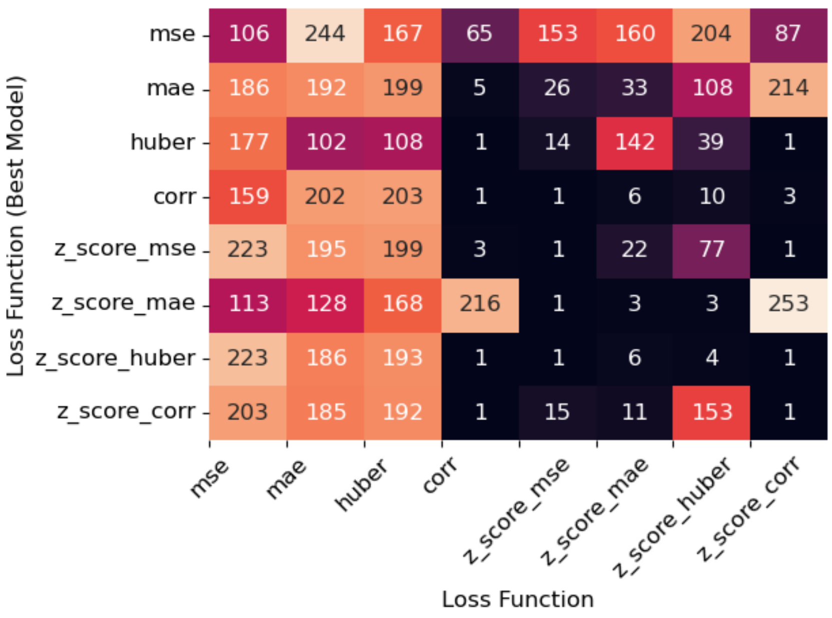
Retrain the best model with all other loss functions

- Which loss function performs better across different hyperparameter configurations

ASCADf, loss function key distinguisher

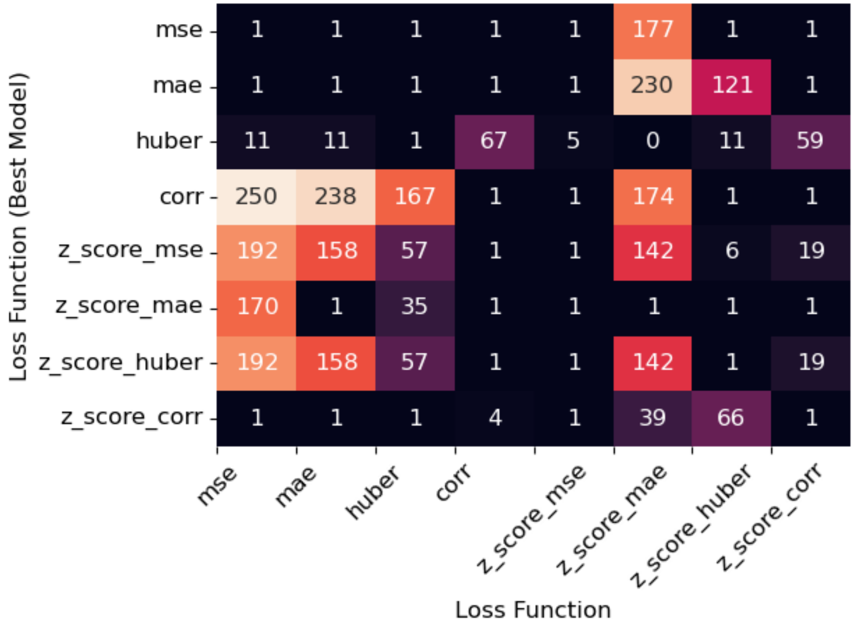


HW

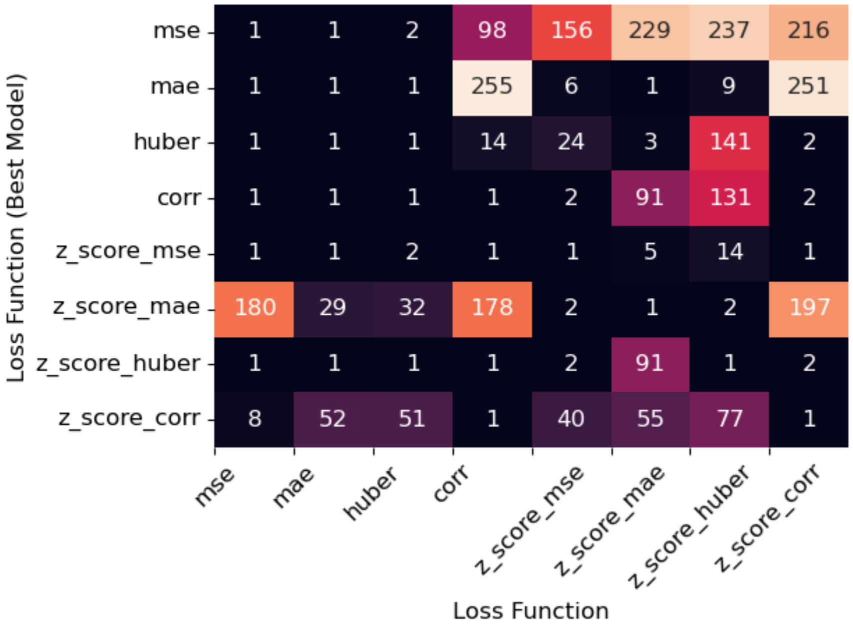


ID

ASCADr, Pearson correlation as key distinguisher



HW



ID

MORE

- Z-score normalized MSE as the loss function
- Pearson correlation for key distinguisher
- Validation set

On average, MORE provides a 3.9x higher SR

	HW/HD		ID	
Dataset	MOR	MORE	MOR	MORE
ASCADf	27.3%	92.5%	25.7%	74.1%
ASCADr	33.9%	77.2%	4.5%	22.2%
AES_HD	10.2%	77.3%	11.3%	58%
AES_RD	22.01%	75.8%	0.7%	1.1%

Size of MORE networks

In profiling SCA, smaller NNs can work well (esp. on public dataset)

MORE has a harder task of predicting 256 outputs at the same time

MORE might need a more complex NNs

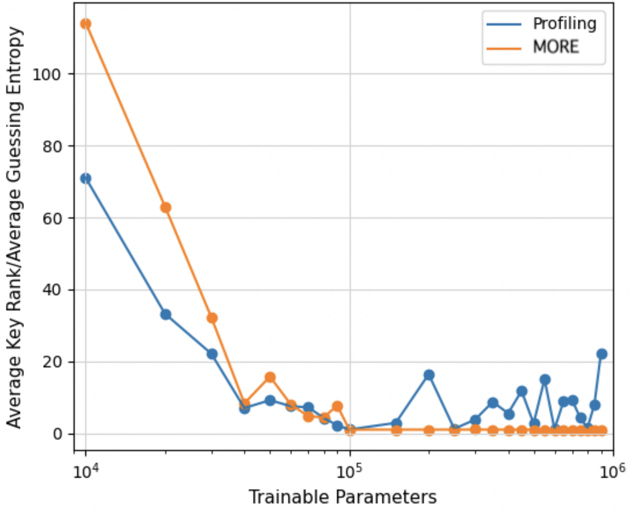
Hyperparameter search space needs to be adapted

Experiment:

3000 NNs, from 10k - 1M trainable parameters

Both CNNs and MLPs

Each network as MORE and as Profiling model



ASCADr, MLP, HW

MORE and Data augmentation

- Random time shift with uniform distribution
- Adding Gaussian noise (mean 0, std 1)

Augmented dataset:

- Adding 10k more traces, or
- Double the number of traces if less than 10k available

ASCADf, ASCADr, AES_HD and AES_RD

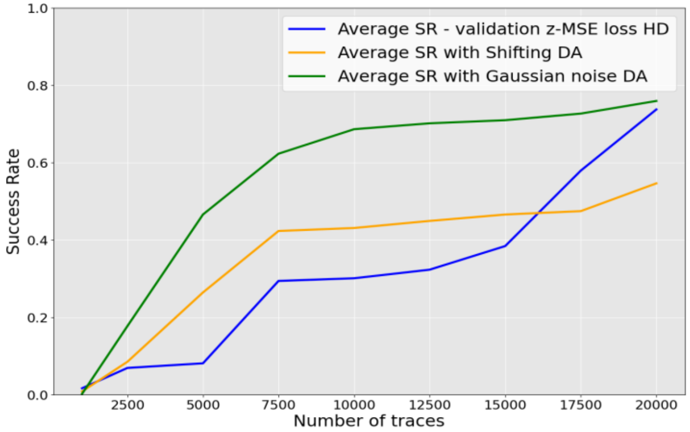
500 random NNs → average SR

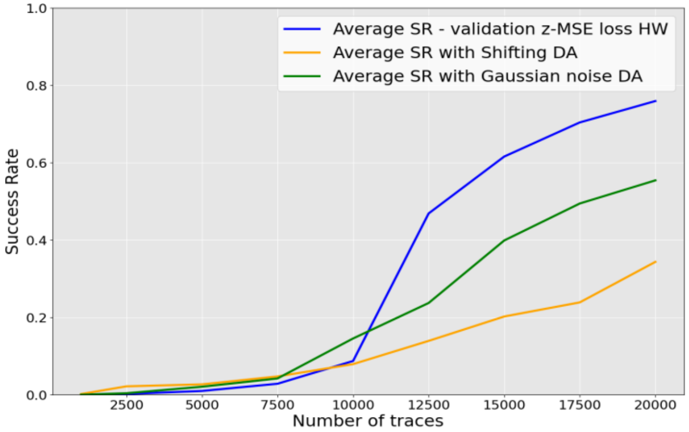
MORE and Data augmentation

Gaussian noise improve SR when less than 10k traces

Shifting DA commonly lower SR

AES_HD and AES_RD





Conclusions and future work

MORE is an improvement over MOR through changes in

- Loss function,
- Key distinguisher, and
- Usage of validation set

MORE achieves 3.9x higher SR than MOR

NNs for MORE should be larger (in trainable parameters) than for profiling SCA

DA and ensembles help when fewer traces are available

Future work:

- Other DA methods and ensembles
- Regularization techniques

Thank you! Q?