Technologies for Defense and Security

Enhancing Performance of Deep Learning Based Non-profiled SCA Using Multi-Output and Transfer Learning

Van-Phuc Hoang, Ngoc-Tuan Do, and Huu Minh Nguyen

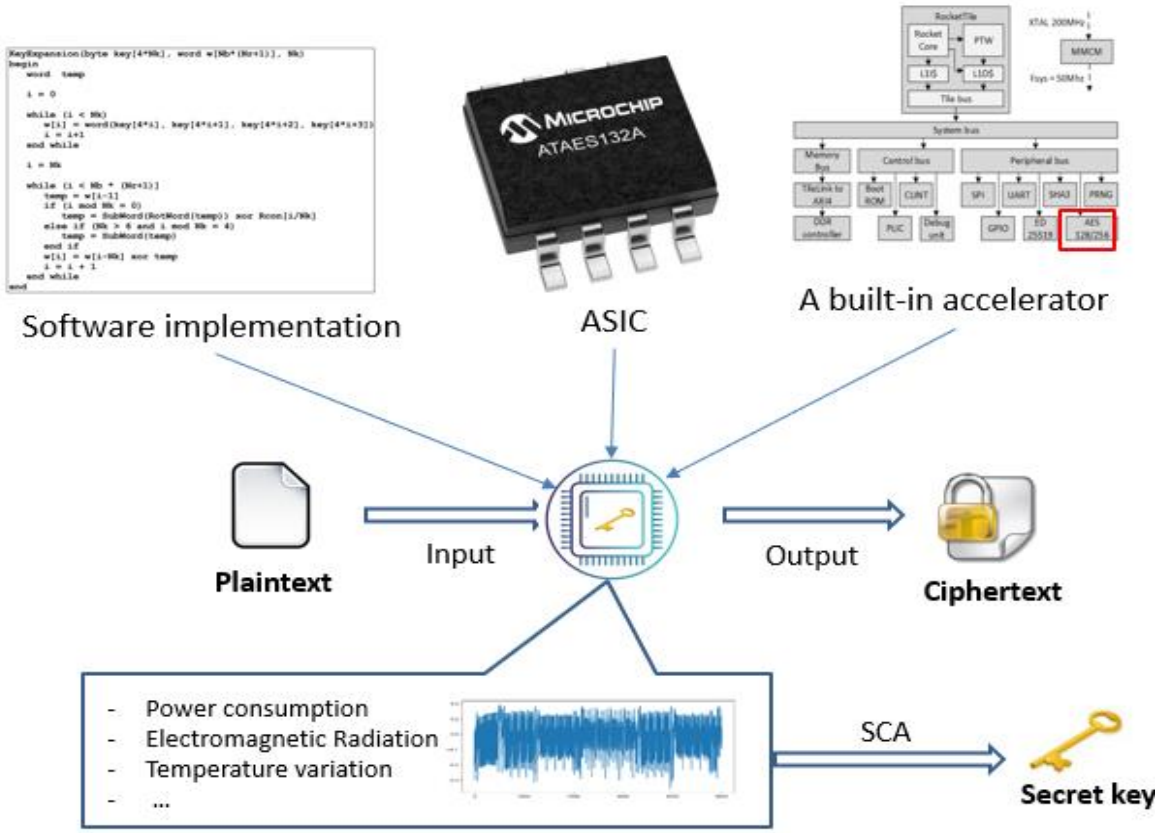
November, 2024

Outline

- Side-channel attack (SCA) on cross-targets
- Enhancing performance of DL non-profiled based SCA
- Experimental results
- Conclusions and future works

Side-channel attack (SCA) on cross-targets

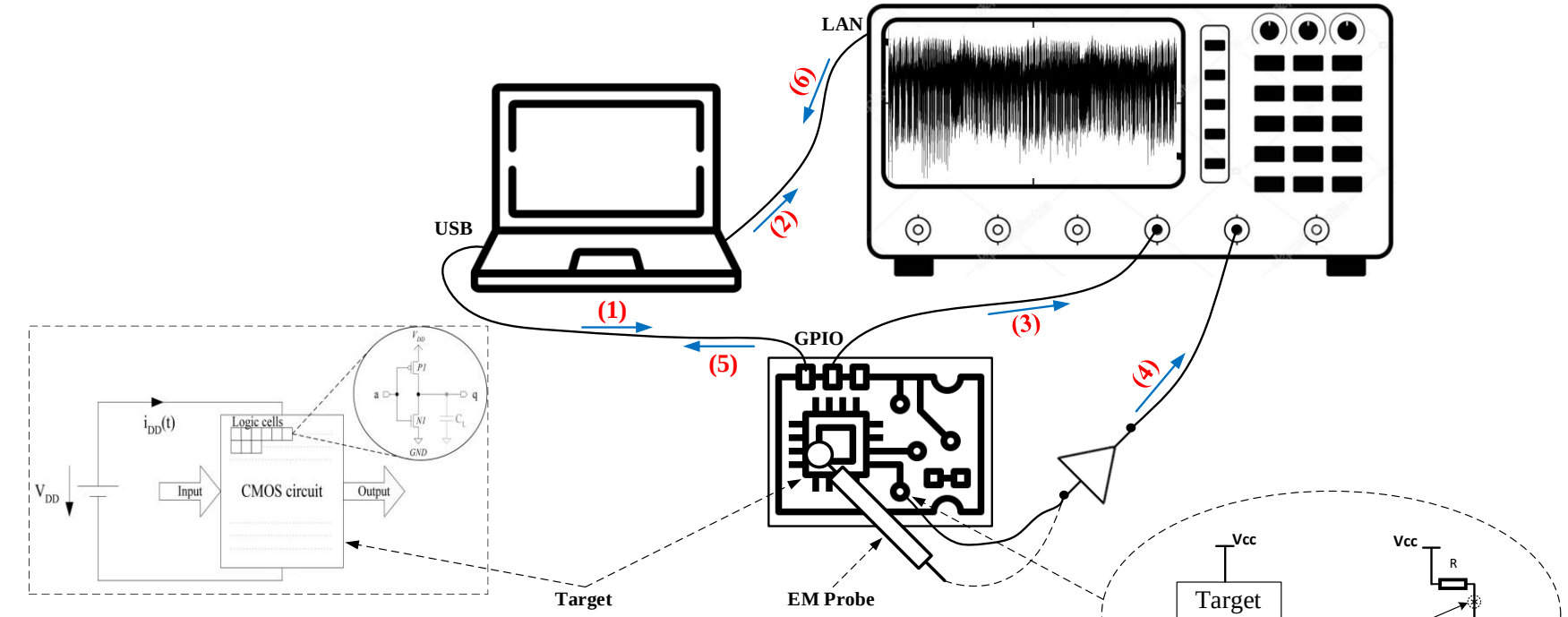
Side channel attack?



- Goal of SCA: to analyze the collected data to infer the secret key used during encryption.
- Advanced statistical and machine learning techniques are applied to derive patterns and recover secret key.
- SCA exploits the physical characteristics of devices → Effective even against theoretically secure algorithms like AES.
- Attackers do not need direct access to software or internal architecture of cryptographic device (monitoring emissions or power consumption can be sufficient).

Side-channel attack (SCA) on cross-targets

Side channel data and measurement setups



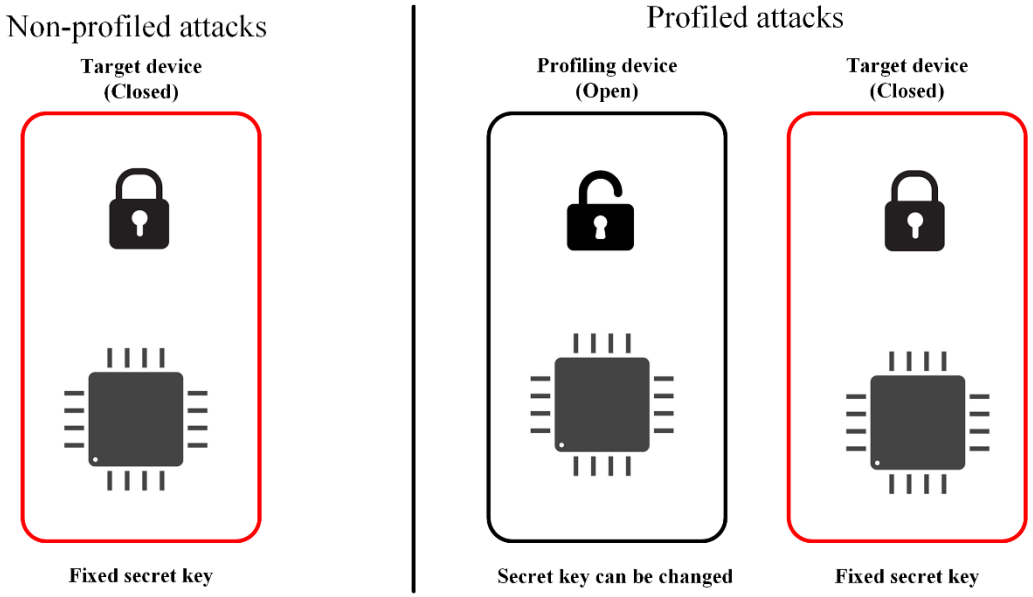
Side channel data of CMOS circuit

Measurement points

(1) The PC sends plaintexts to the target
(2) The PC sends commands to the oscilloscope
(3) The target send trigger signal to the oscilloscope
(4) The oscilloscope collects Power/EM data
(5) The target sends Ciphertexts back to the PC
(6) The oscilloscope sends ADC data to the PC

Side-channel attack (SCA) on cross-targets

Classification of Side channel attacks

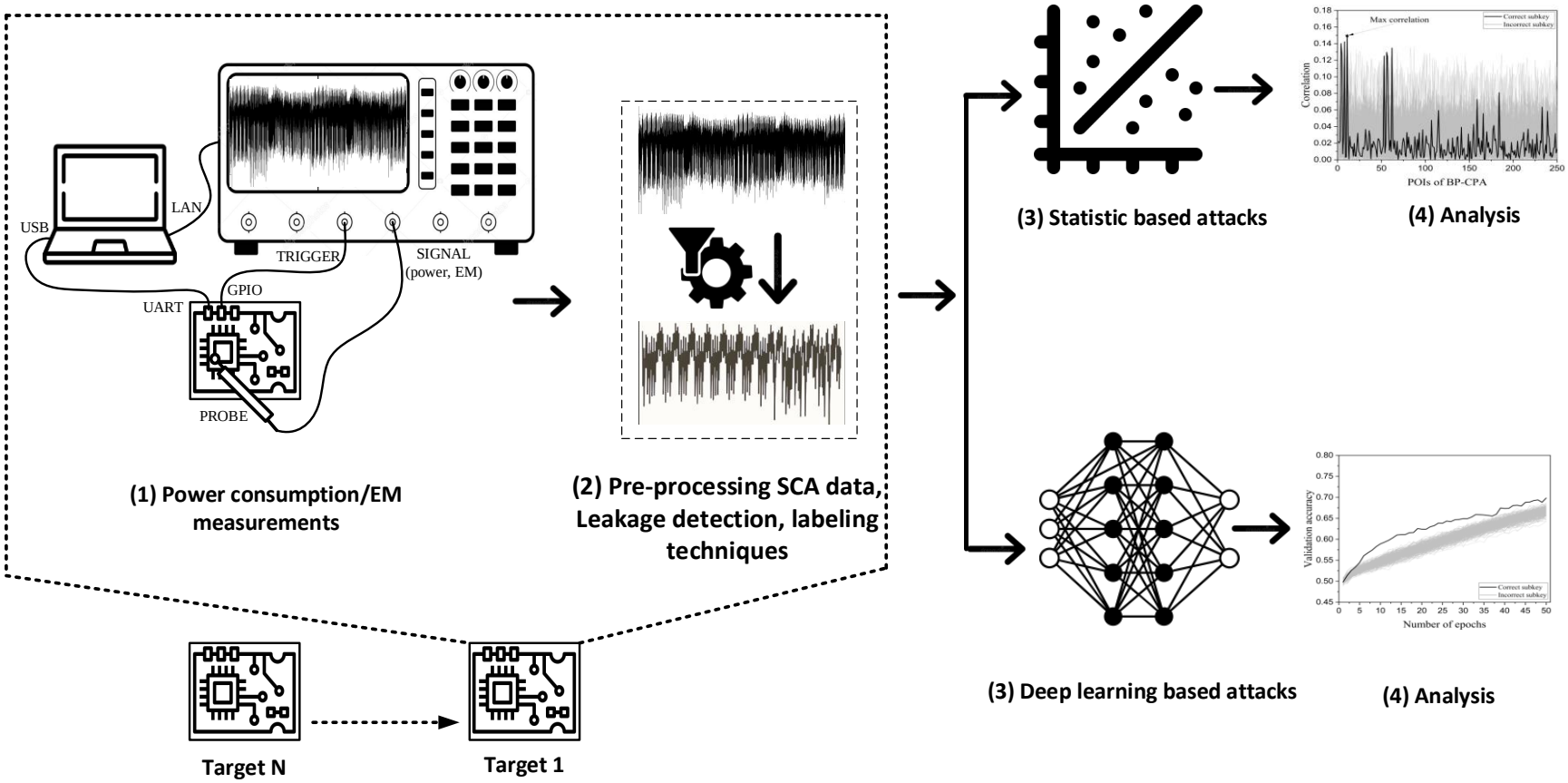


- ✓ Non-profiled attacks exploit the relationship between real power consumption and the power consumption model.
- ✓ Only target device.

- ✓ Profiled attacks take place in two stages: the profiling stage and key extraction stage.
- ✓ Target device and reference device.

Side-channel attack (SCA) on cross-targets

Attacking cross-targets

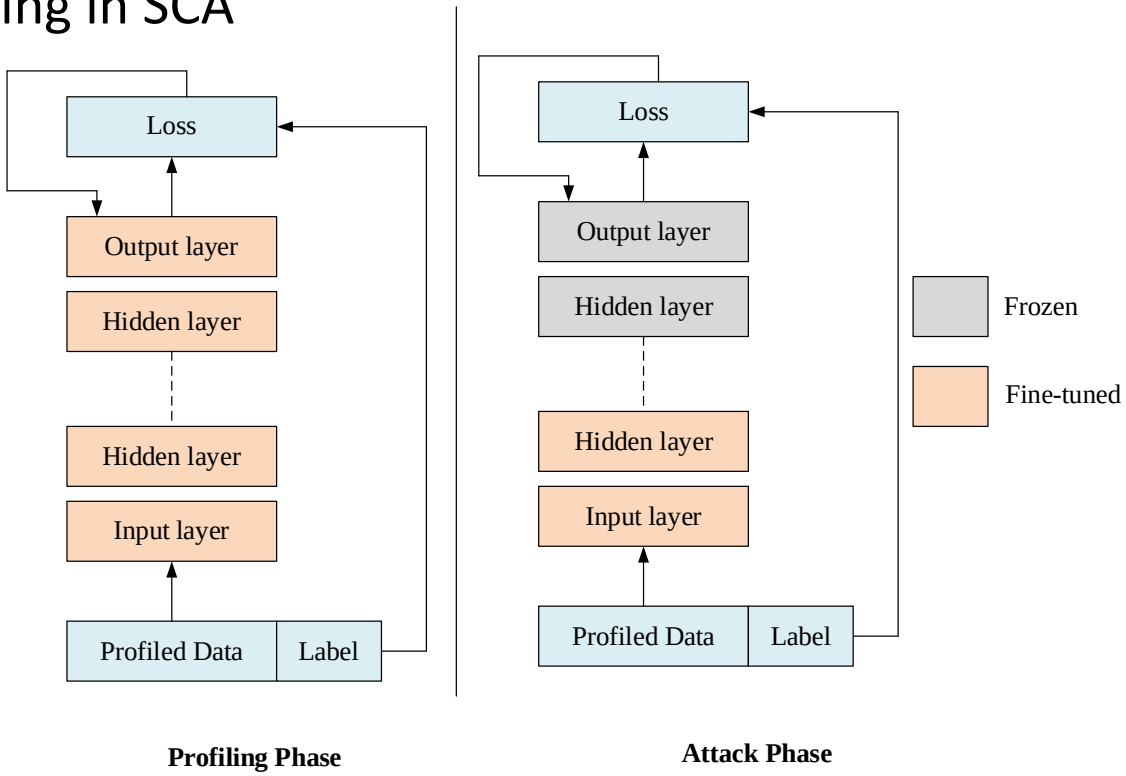


In non-profiled scenario, attack procedure must repeatedly perform all steps for each target, even for a family of targets

→ time-consuming and costly SCA evaluation process.

Enhancing performance of DL non-profiled based SCA

1. Transfer learning in SCA

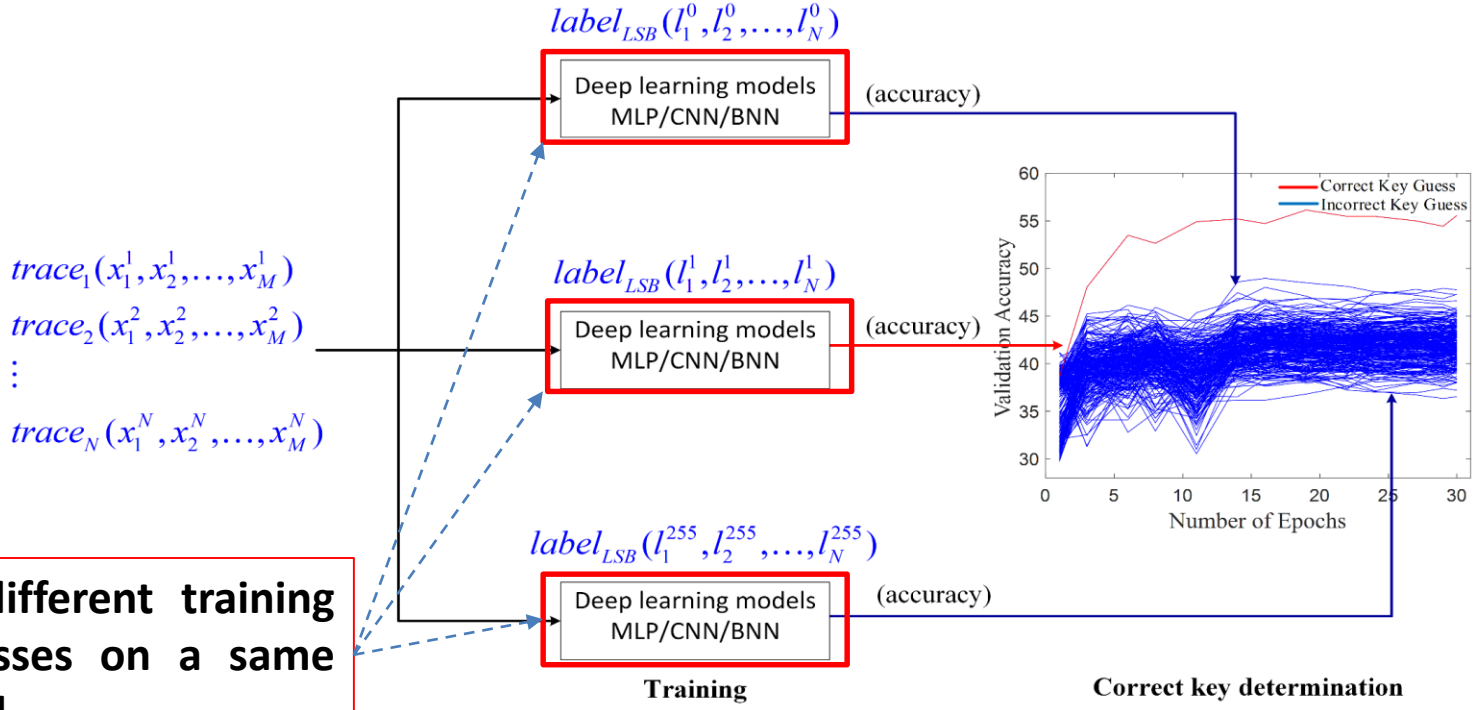


Transfer learning based profiled SCA

- Leveraging pre-trained models and knowledge gained from similar tasks, these techniques can significantly reduce amount of data and time required to execute an attack.
- There has never been a deep learning based non-profiled using transfer learning.

Enhancing performance of DL non-profiled based SCA

1. Differential deep learning analysis (DDLA)



256 different training processes on a same model.

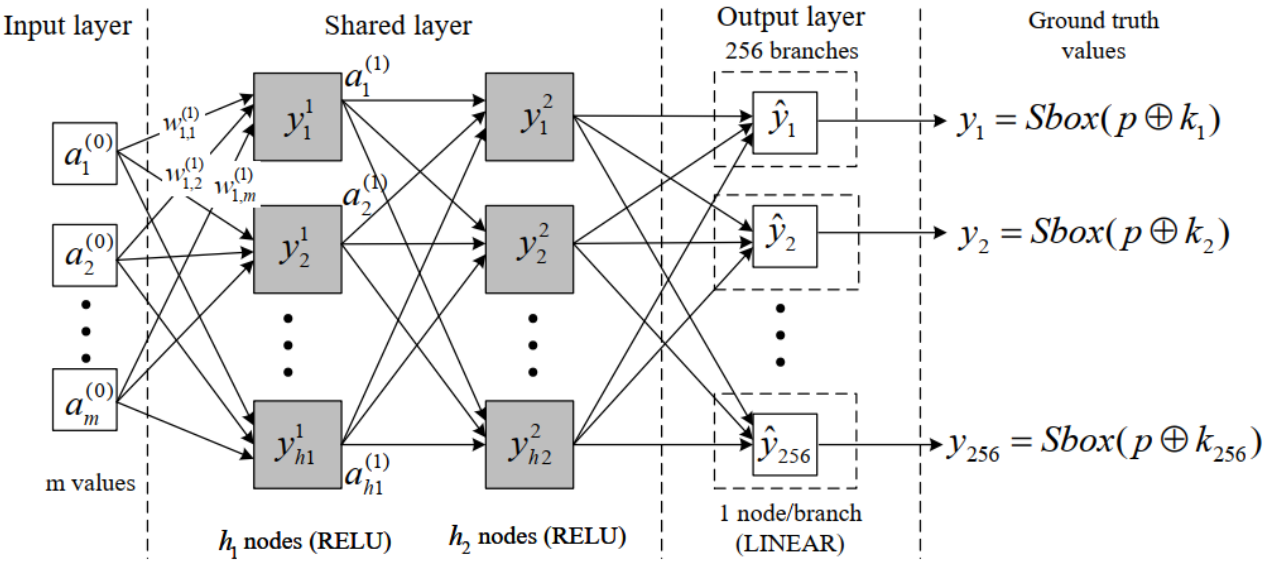
DDLA attack procedure [*]

+ DDLA requires the attacker repeatedly perform the training process to observe the training metrics, which are then used to determine the correct subkey byte.
+ It can not apply transfer learning due to the re-train 256 other models for a new target.

[*] Timon, B.: Non-profiled deep learning-based side-channel attacks with sensitivity analysis. IACR Trans. Cryptogr. Hardw. Embedd. Syst **2019**(2), 07–131 (2019)

Enhancing performance of DL non-profiled based SCA

2. Multi-output deep learning based non-profiled SCA



Multi-output regression based non-profiled SCA [*]

$$M_{MO} : T \rightarrow (y_1, y_2, \ldots, y_{256})$$

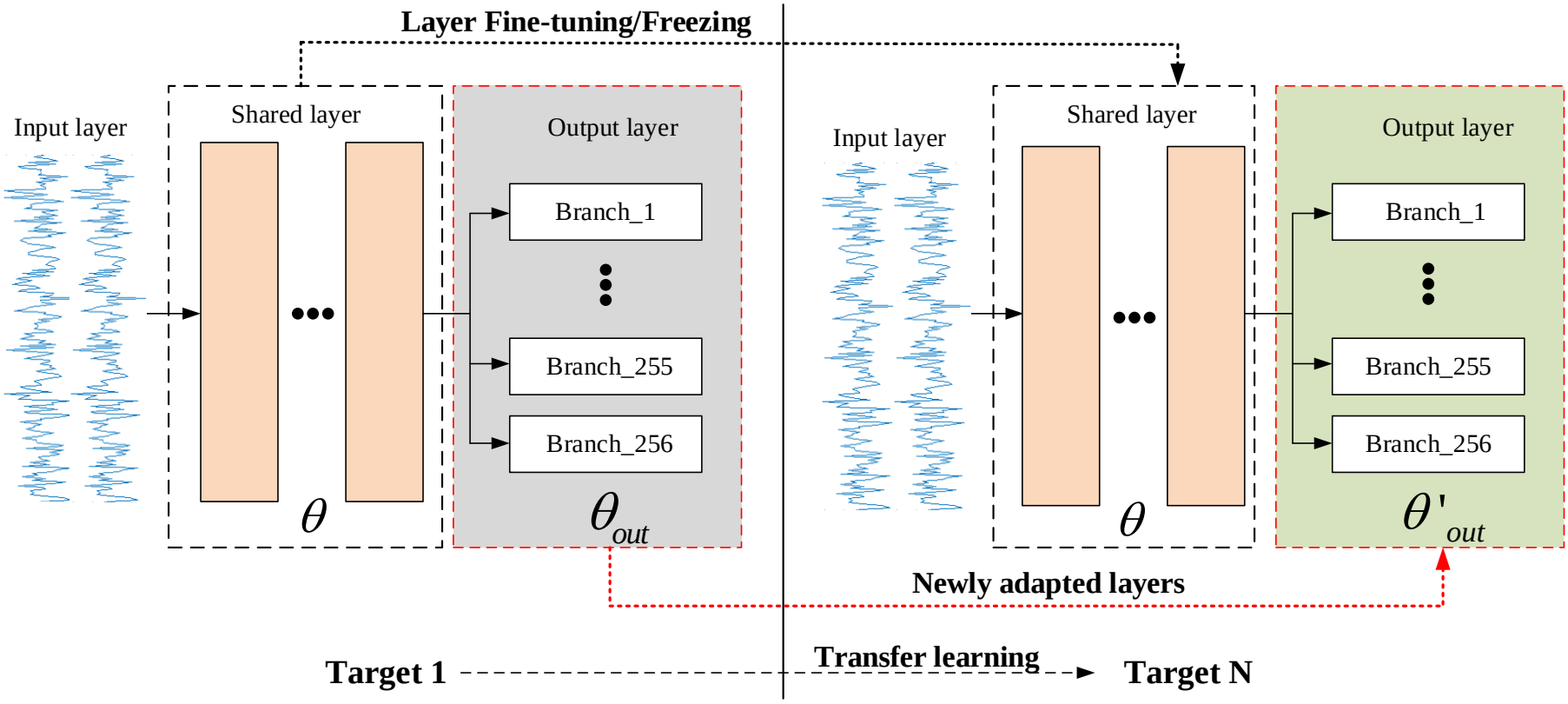
$$\hat{y}_{i,j} = f_{out}(M_{MO}(x_j; \theta); \theta_{out_i})$$

Multi-output regression neural network can simultaneously estimate all key hypotheses in a single training process

[*] Do, NT., Hoang, VP. & Doan, V.S. A novel non-profiled side channel attack based on multi-output regression neural network. *J Cryptogr Eng* **14**, 427–439 (2024). https://doi.org/10.1007/s13389-023-00314-4

Enhancing performance of DL non-profiled based SCA

2. Proposed transfer learning based non-profiled SCA



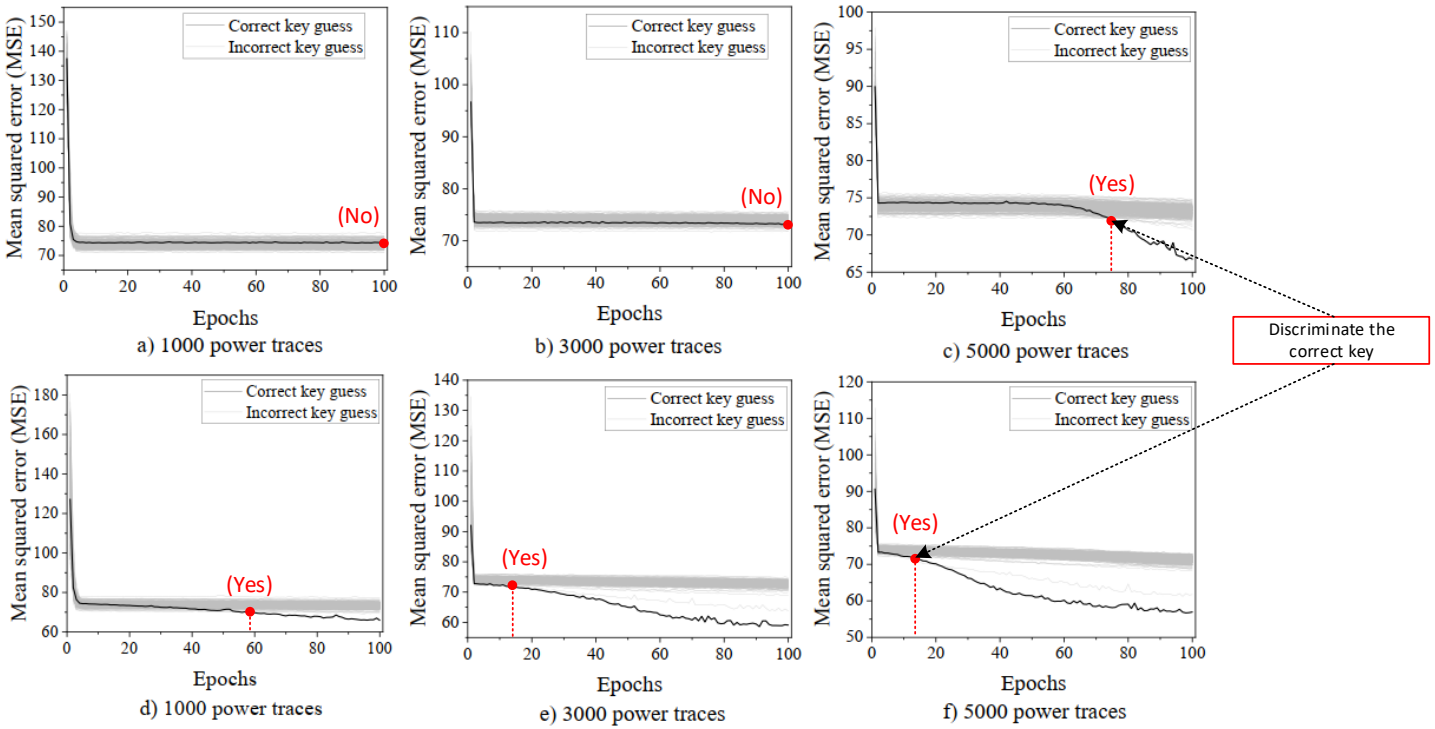
$$\hat{y}_{i,j} = f_{out}(M_{MO}(x_j;\theta);\theta_{out_i})$$

$$y'_{i,j} = f'_{out}(M'_{MO}(x'_j;\theta);\theta'_{out_i})$$

Pre-trained model's knowledge is transferred to a new task, enabling adapted model to leverage learned features while being fine-tuned for the new dataset, allows the model to retain useful representations and efficiently adapt to new side-channel traces or variations.

Experimental results

Case1: Data collected from same target

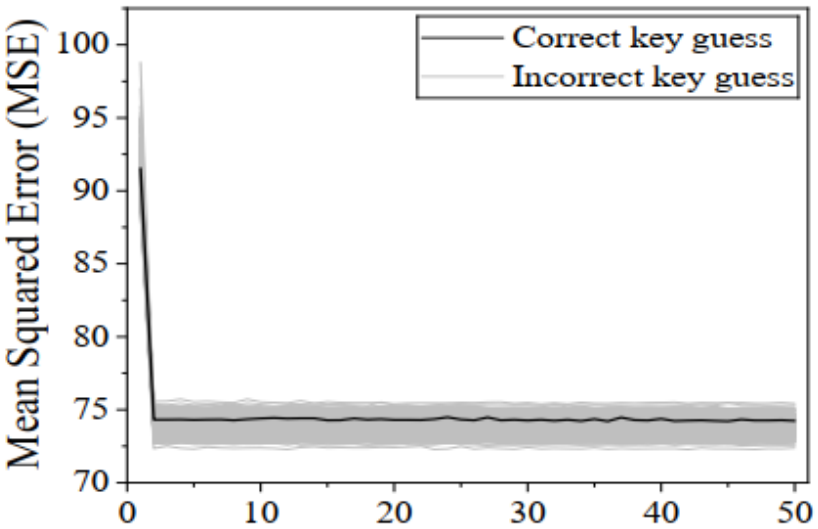


Attack results on ChipWhisperer board using MOR and proposed technique. a,b,c) MOR model; d,e,f) Transfer learning along with MOR

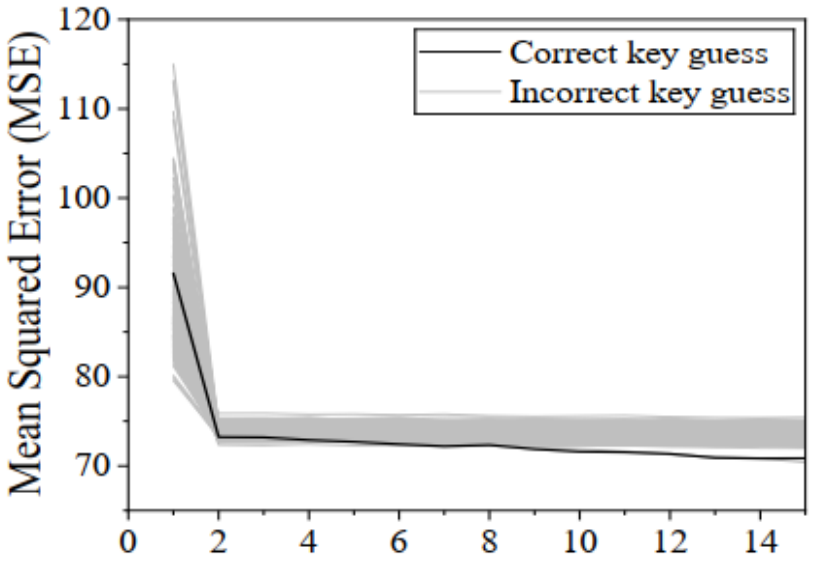
- Transfer learning based attacks achieve better discrimination of correct and incorrect key guess.
- Transfer learning could be used to mount attack on fewer power traces and number of training epochs.

Experimental results

Case1: Data collected from same target



a) 5000 power traces



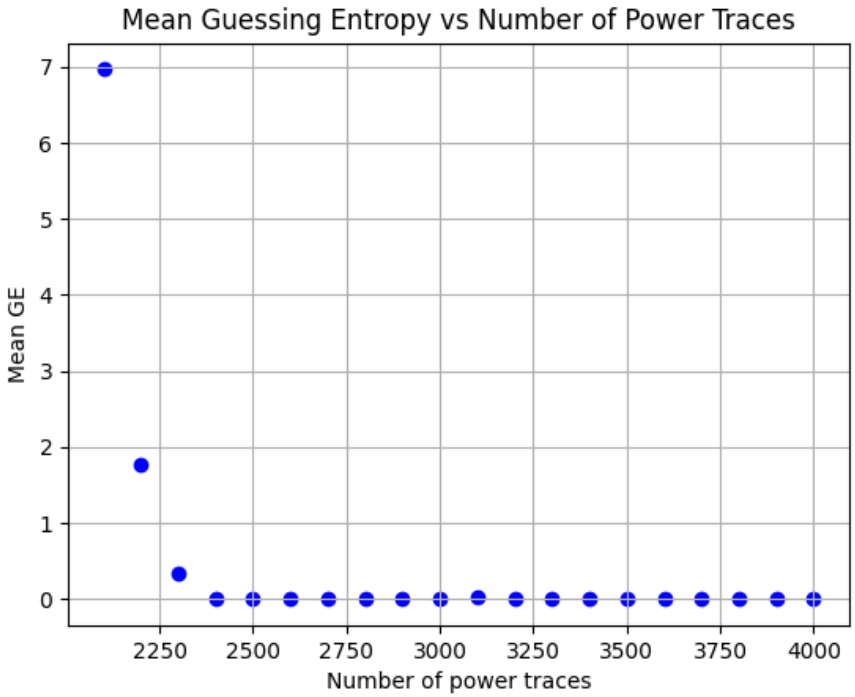
b) 500 power traces

Attack results of transfer learning based and normal MOR attacks on different number of power traces and epochs. a) MOR; b) Transfer learning based MOR

- Correct key can be clearly discriminated at epoch number of 15 with transfer learning based MOR.
- MSE metrics of all guess keys are almost unchanged in the case of MOR model.
- Proposed technique outperforms the original MOR model regarding both number of measurements and number of training epoch.

Experimental results

Case 2: Data collected from different target

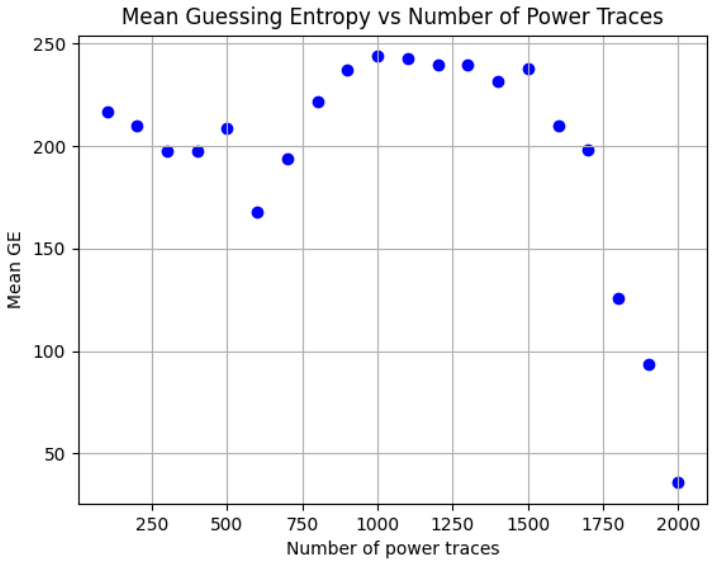


Attack results on CW board 1 using different number of power

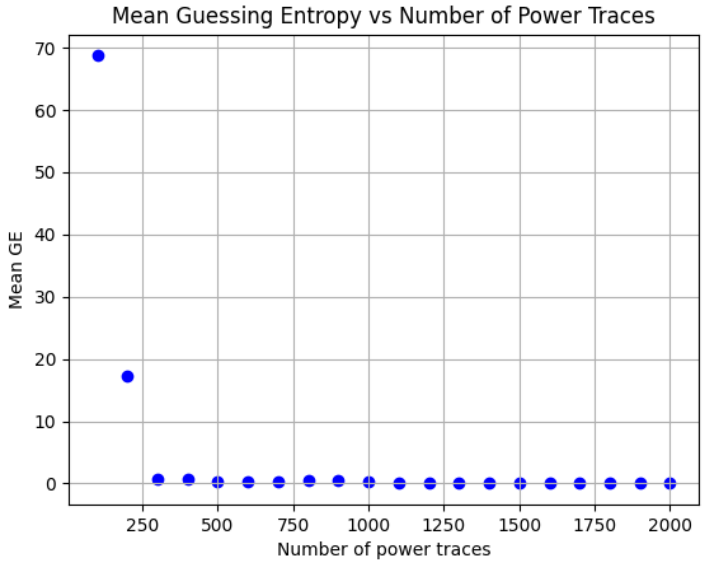
- Attack AES-128 on CW board 1 successfully with approximately 2500 power traces.
- The trained model is saved for applying transfer learning to CW board 2

Experimental results

Case 2: Data collected from different target



a) Without transfer learning

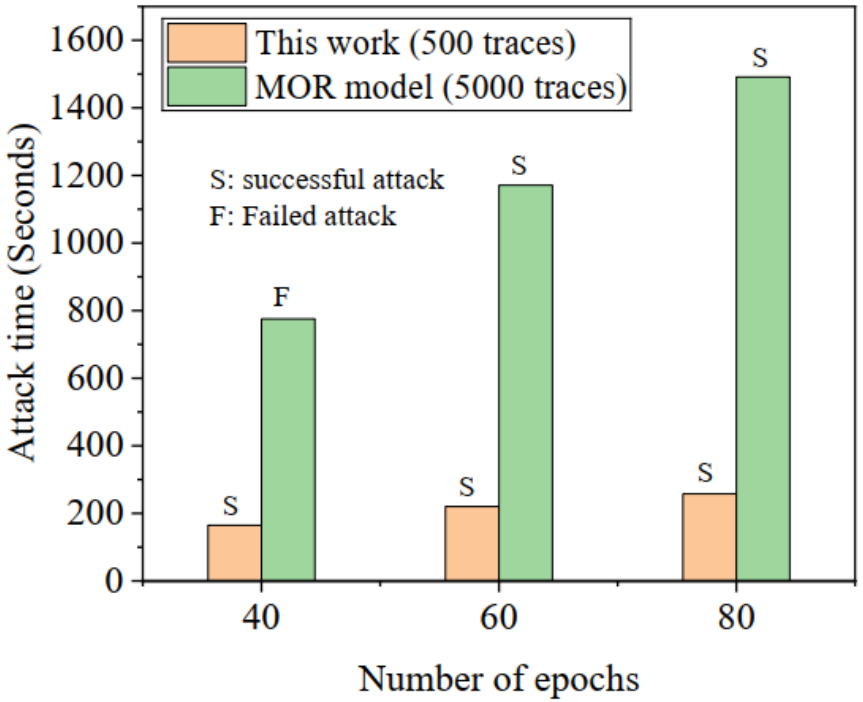


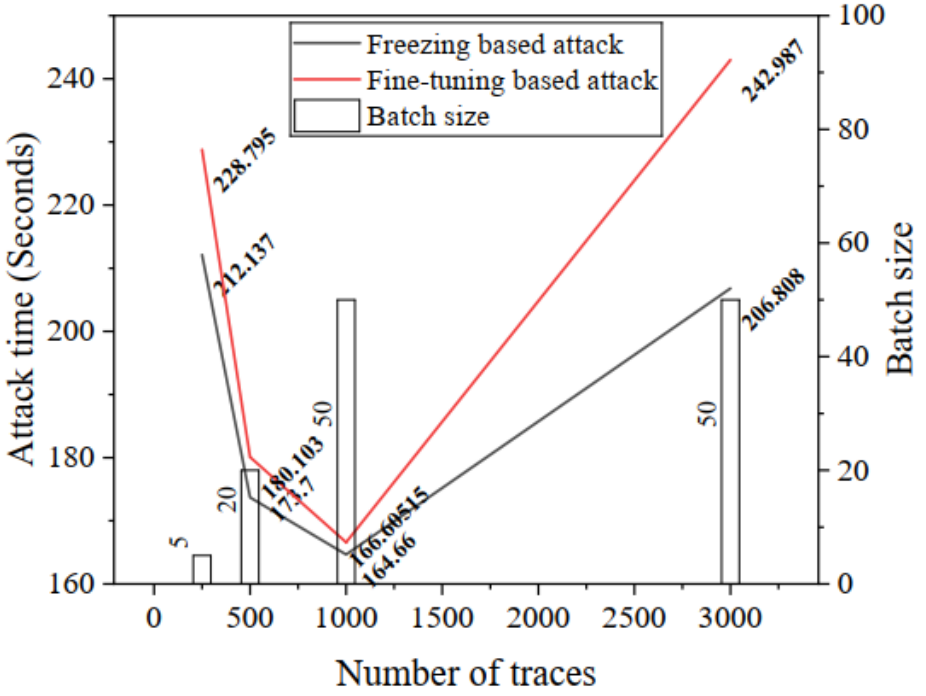
a) Based on transfer learning

Attack results of W/WT transfer learning based MOR attacks on CW Board 2 (different targets).

- Normally, the model without transfer learning can not reveal the secret key with less than 2000 power traces.
- Transfer learning based method requires only 300 traces for taking the correct key.

Experimental results





- Since the number of required power traces for transfer learning-based MOR is less than for the original model, the execution time of transfer learning-based attacks is significantly reduced.
- Experimental results show that the attack time of the freezing-based approach is slightly higher than that of the fine-tuning-based approach.
- Time varies depending on the batch size.

Conclusions and Future Works

❖ Conclusions

✓ Proposed multi-output deep learning-based side-channel analysis (MO-DLSCA) with transfer learning greatly enhances non-profiled SCA.

✓ It reveals the subbyte key in 2nd target using just 10% of power traces needed for 1st target, reducing attack time by 5.7 times.

→ Potential for optimizing non-profiled SCA on similar devices, offering a significant improvement over traditional MODLSCA methods.

❖ Future works

- Investigating other DL architectures in SCA domain.

- Developing online DL-based SCA methods for reducing attack time & determining minimum measurements for an attack.

- Developing efficient SCA countermeasures to account for both traditional and DL-based attacks.

THANK YOU FOR YOUR ATTENTION!

Acknowledgement

This publication is the output of the ASEAN IVO project, "Artificial Intelligence Powered Comprehensive Cyber-Security for Smart Healthcare Systems (AIPOSH)", and financially supported by NICT, Japan.