
CSIT
CENTRE
FOR SECURE
INFORMATION
TECHNOLOGIES
DEEPSECURITY: APPLYING
DEEP LEARNING TO
HARDWARE SECURITY
QUEEN'S
UNIVERSITY
BELFAST
ESTP1845
Máire O'Neill, Shichao Yu, Anh-Tuan Hoang
December 2022

DeepSecurity: Applying Deep Learning to Hardware Security

Overall Goal

To investigate the use of Deep Learning for security verification in EDA tools, specifically in relation to *Hardware Trojan detection* and *Side channel analysis* to allow non-security experts to receive feedback on how to improve the security of their designs prior to fabrication.

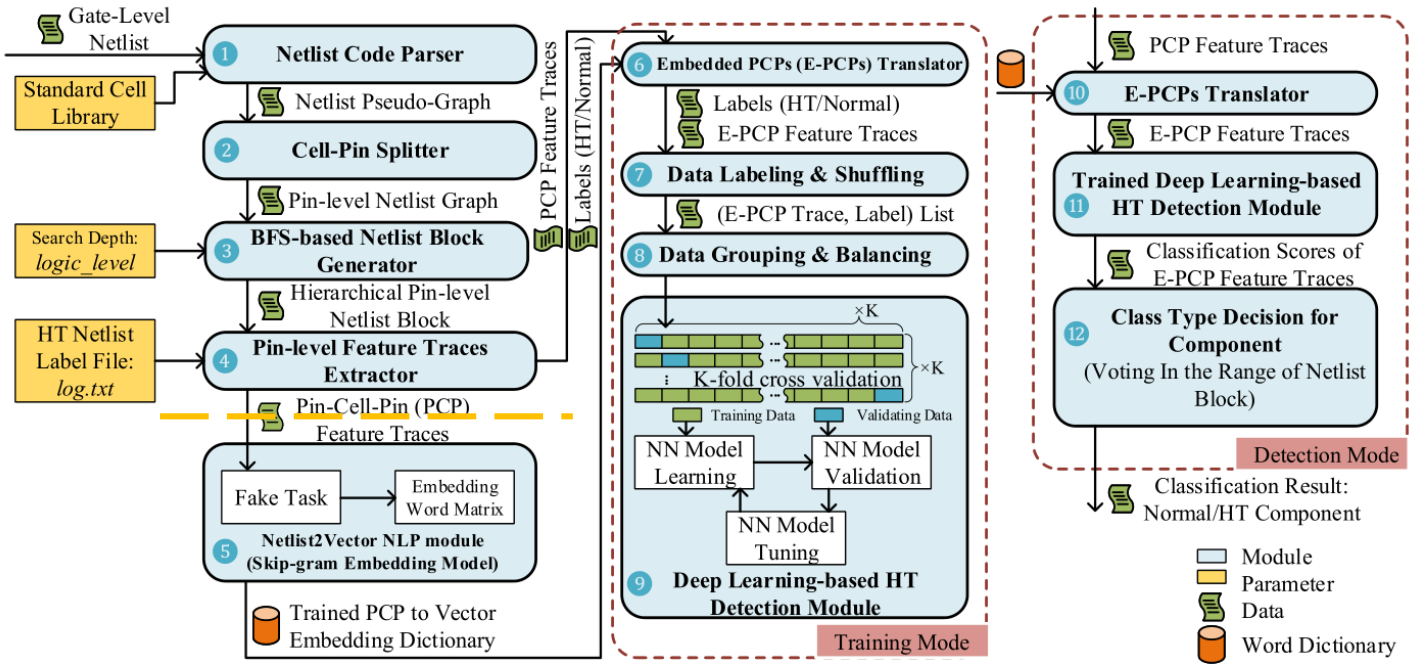
CSIT CENTRE FOR SECURE INFORMATION TECHNOLOGIES

QUEEN'S UNIVERSITY BELFAST

Hardware Trojan Detection

Implementation

The Architecture of the DL-based HT Detection System



- A data-driven HT detection that can effectively detect HTs without any prior knowledge of the circuits;
- To the best of the authors' knowledge, this is the first HT detection system based entirely on DL;
- This is also the first time NLP has been applied on raw gate-level netlist data for HT detection.

Hardware Trojan Detection

[1] S. Yu, C. Gu, W.Liu, M O'Neill, Deep Learning-based Hardware Trojan Detection with Block-based Netlist Information Extraction, **IEEE Transactions on Emerging Topics in Computing**, October 2021

[2] M. Xue, C. Gu, W Liu, S. Yu, M. O'Neill, Ten years of Hardware Trojans; A Survey from the Attacker's perspective, **IET computer and Digital Techniques**, pp 231-246, Oct 2020

[3] Y. Dou, S. Yu, C. Gu, M. O'Neill, C. Wang, W. Liu, Security Analysis of Hardware Trojans on Approximate Circuits, **ACM Great Lakes Symposium on VLSI** (GLSVLSI), Sept 2020.

[4] S. Yu, C. Gu, W. Liu and M. O'Neill, A Novel Feature Extraction Strategy for Hardware Trojan Detection, **IEEE Intl Symposium on Circuits and Systems (ISCAS),** May 2020.

CSIT CENTRE FOR SECURE INFORMATION TECHNOLOGIES

Next steps

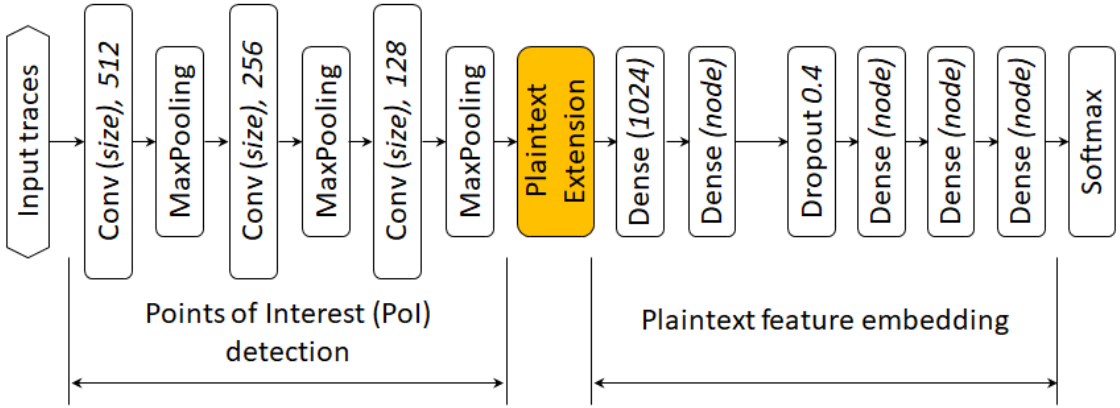
- Generic HW Trojan detection approaches for the design-stage
- ML-based approaches to detect Trojans at other stages of the IC manufacturing process
- Ensuring trustworthiness of DL-based HT Detection Systems against adversarial attacks
- Integration of Trustworthy DL-based HT Detection System into a security verification framework in EDA Tools

Applying Deep Learning to Side-channel Analysis

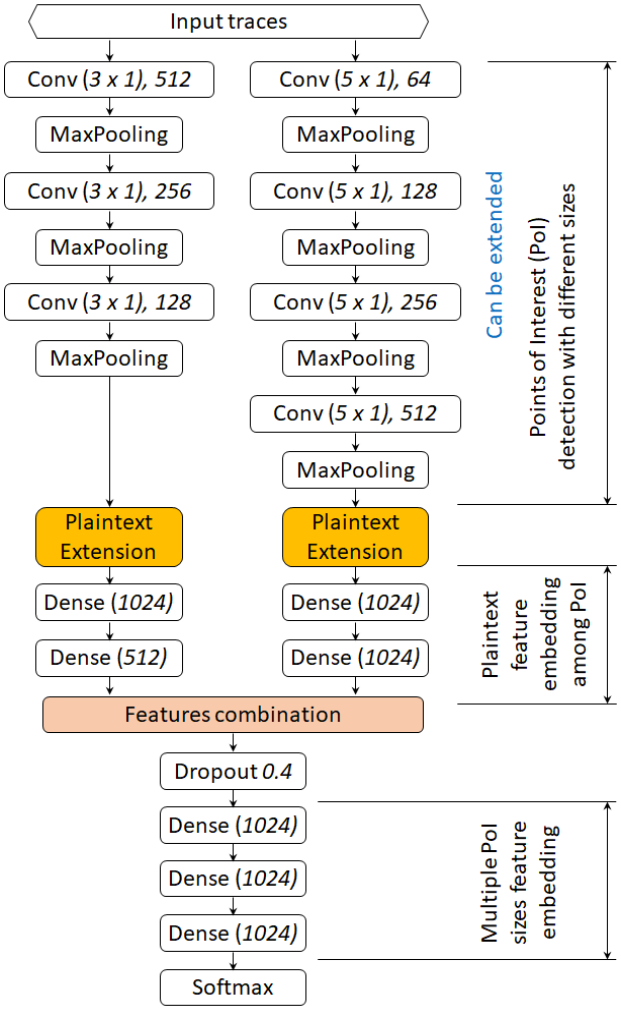
QUEEN'S UNIVERSITY BELFAST

CNNP models

One kernel size



Two kernel sizes

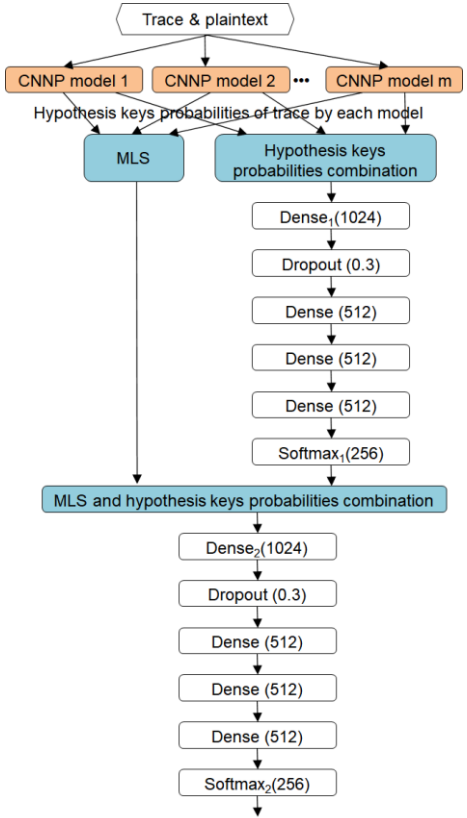


- Models: one or two kernel sizes
- Database: ASCAD (v1) with variable key dataset
- Classification: output of SBox
- Model inputs: traces and plaintext
- Model output: probability of each key value
- Attacking result: key rank 2 after 40 traces

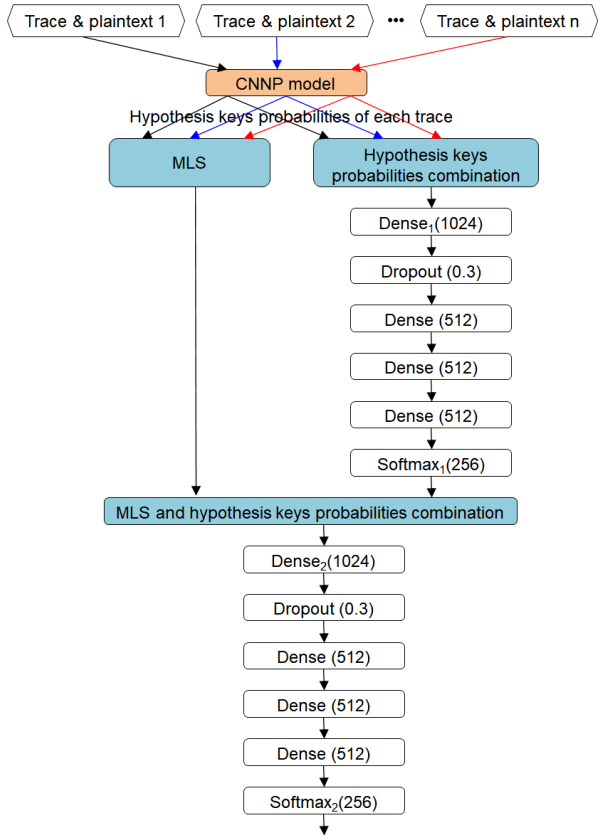
Reinforcement learning using Stacked Ensembles Model

- Utilize the previous research on CNNP model
- Two step training:
  - Train CNNP models by pairs of traces and plaintexts
  - Train SEM models (reinforcement) by the CNNP output probabilities and MLS of hypothesis keys
- CNNP sub-model:
  - One to three CNNP sub-models with single or two convolutional filter sizes
- Input:
  - Single or multiple traces of the same 8-bit plaintext
  - 8-bit plaintext
- Output:
  - Probability of each hypothesis key

Stacked Ensembles Models

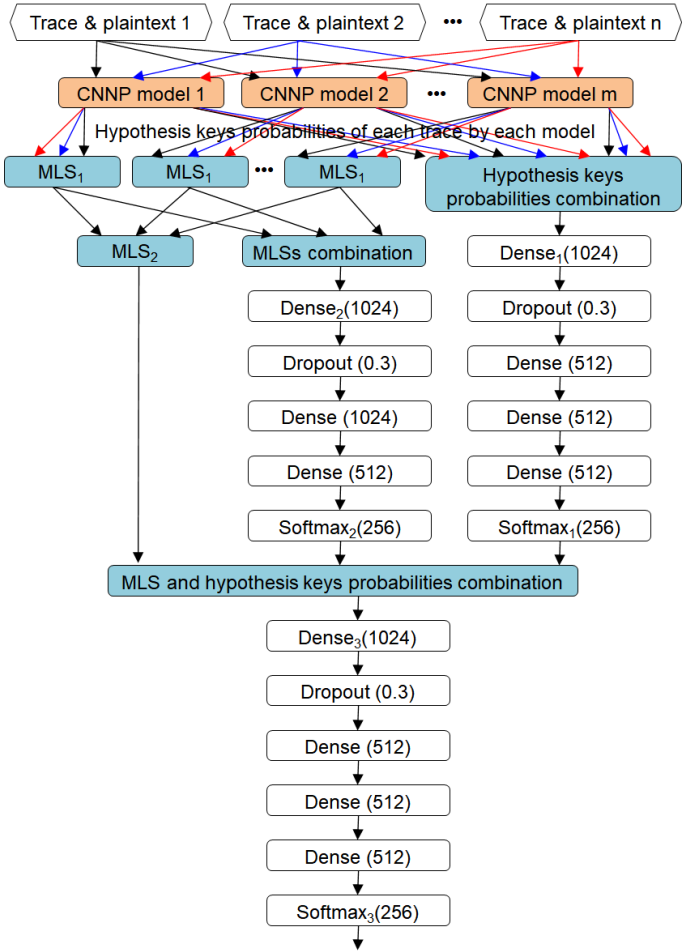


Single traces multiple CNNP models



Multiple traces single CNNP model

Stacked Ensembles Models

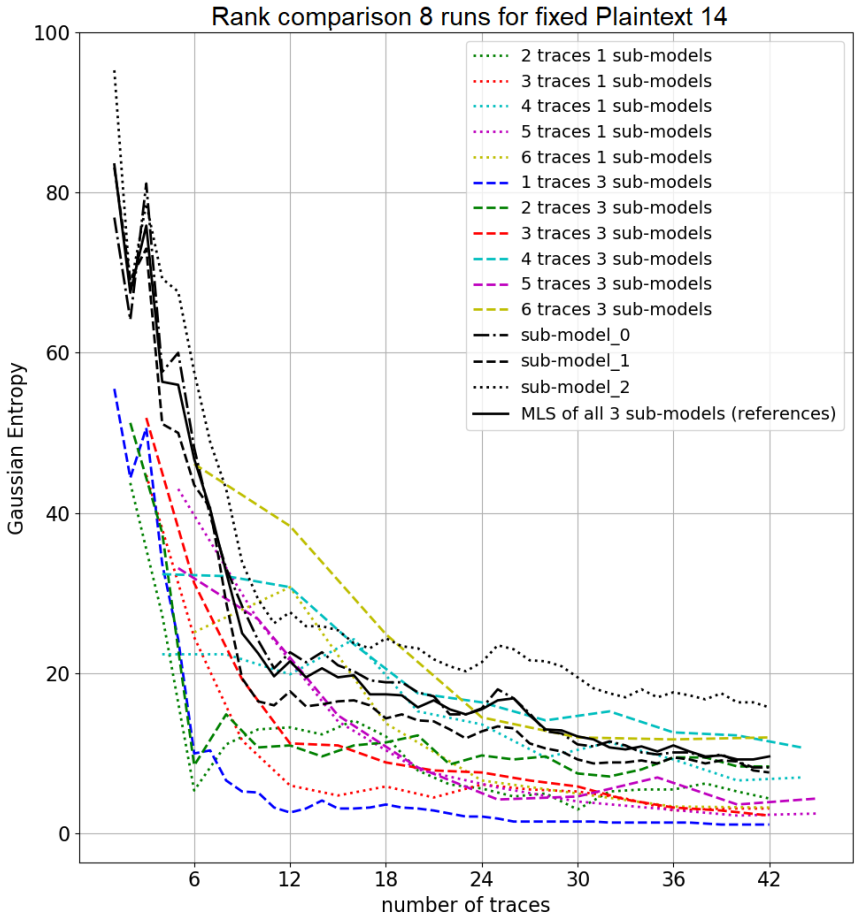


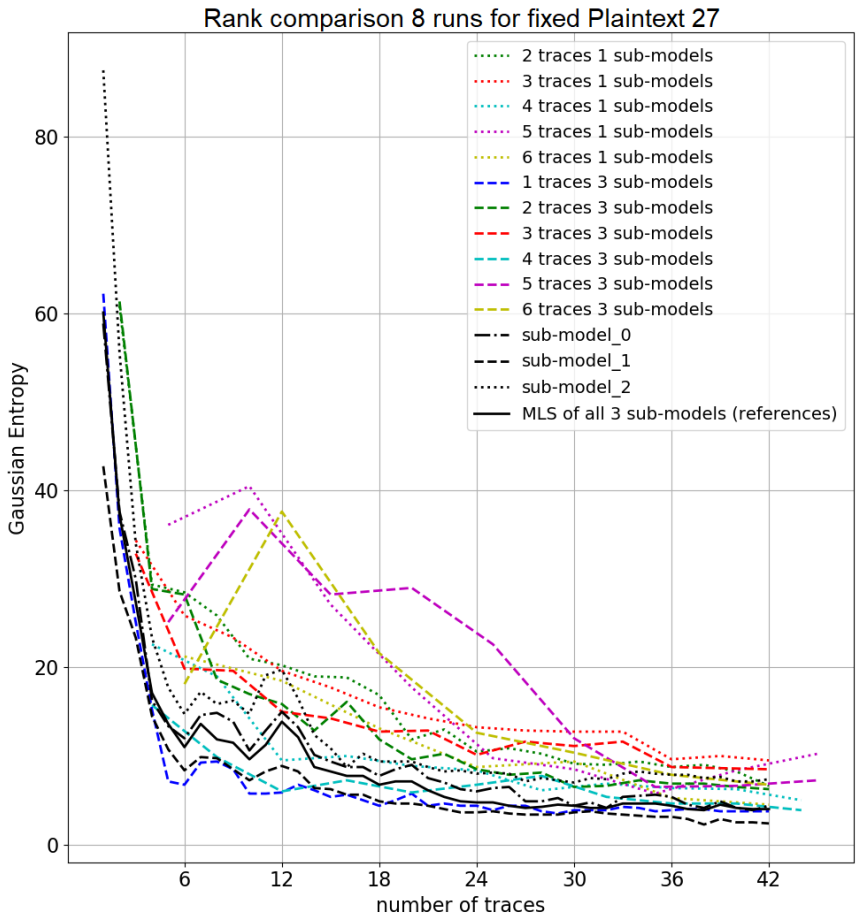
Multiple traces multiple CNNP model

Model testing

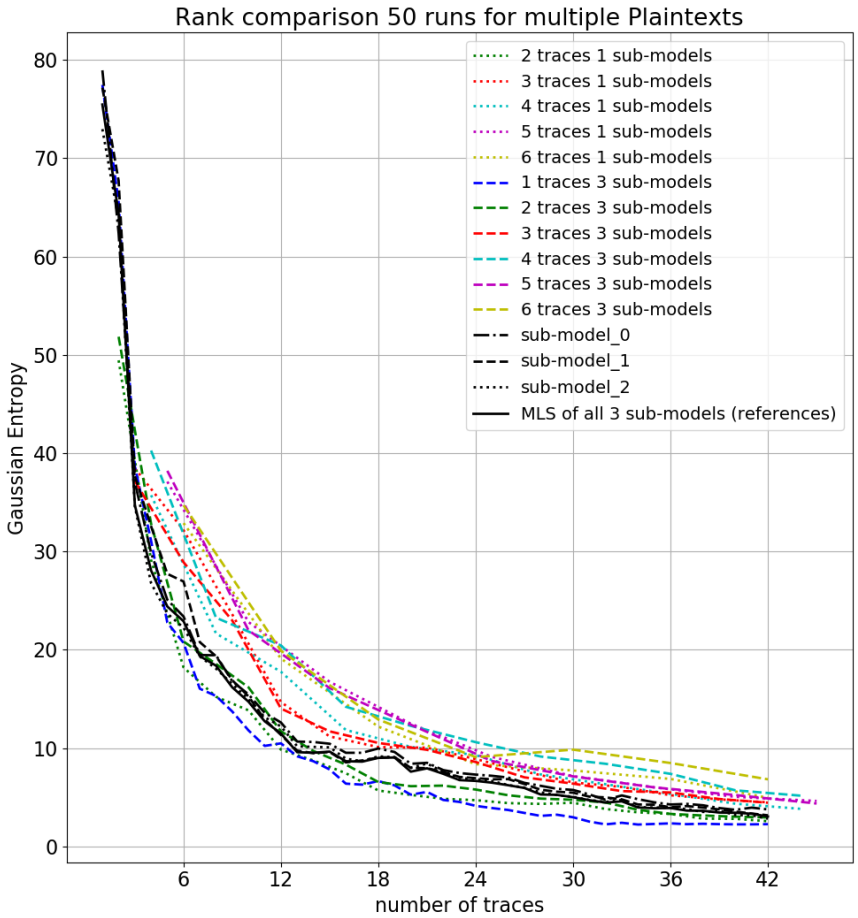
- CNNP based Stacked ensembles models:
  - Inputs: from 1 to 6 traces with the same plaintext
  - Sub models: 1 to 3 CNNP sub-models
- ASCAD variable key dataset (v1):
  - Training set:
    - 200,000 traces with 1,400 sample points
    - Grouped by the same number of input traces of the stacked ensembles structure for training
  - Testing set:
    - 100,000 traces with the same length
    - Grouped by the same method

Single kernel size SECNNP model evaluation with fixed Plaintext



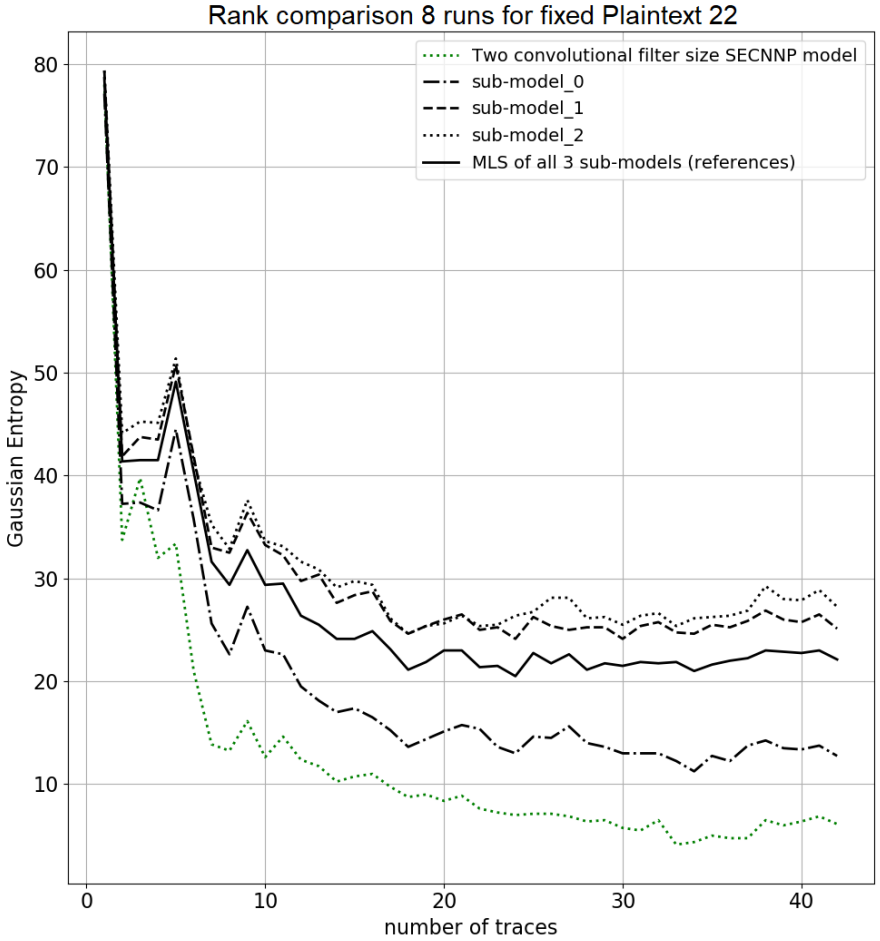


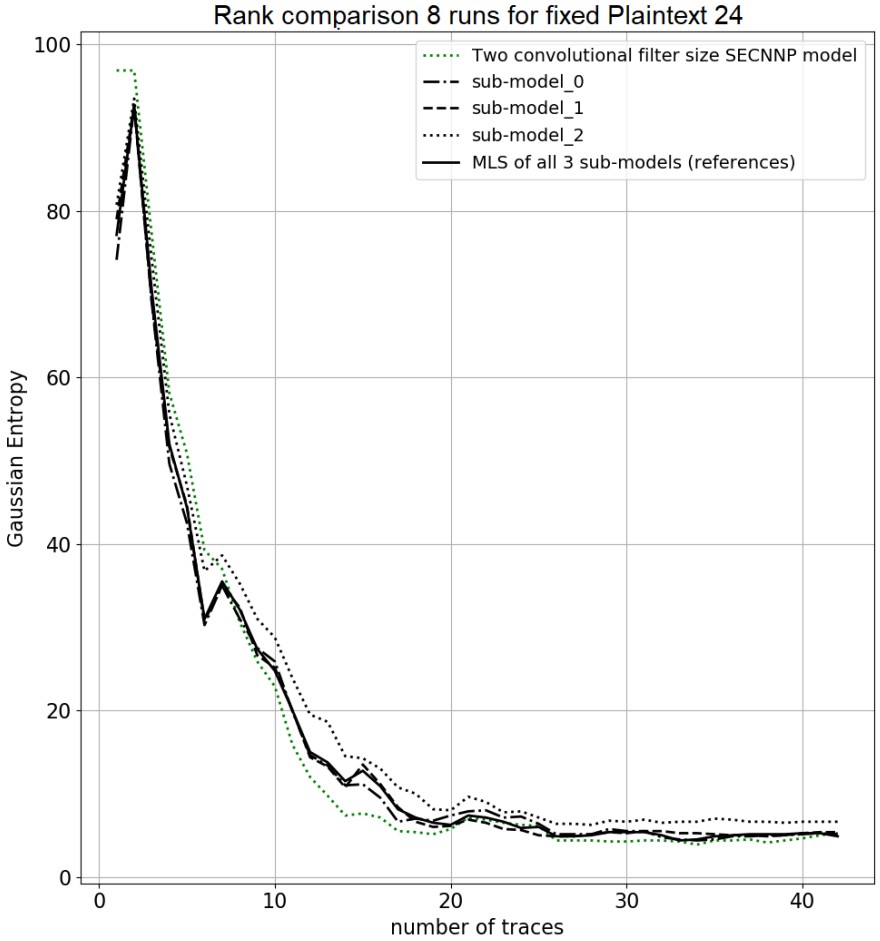
Single kernel size SECNNP model evaluation with multiple Plaintexts



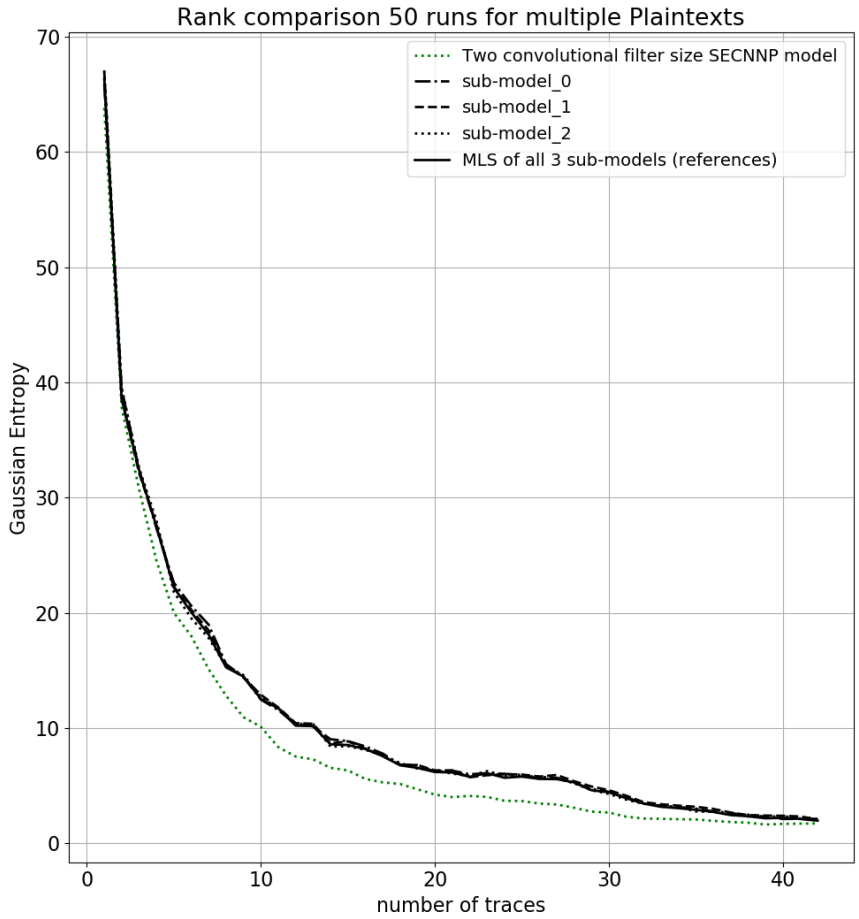
- Sub-models are the same single convolutional filter size model with different training epochs.
- SEM models are built from 1 ~ 6 traces and 1 or 3 sub models.
- One and two-trace input SEM with multiple sub-models are compatible and achieved better results than other SEM models and references.

Two kernel sizes SECNNP model evaluation with fixed Plaintext



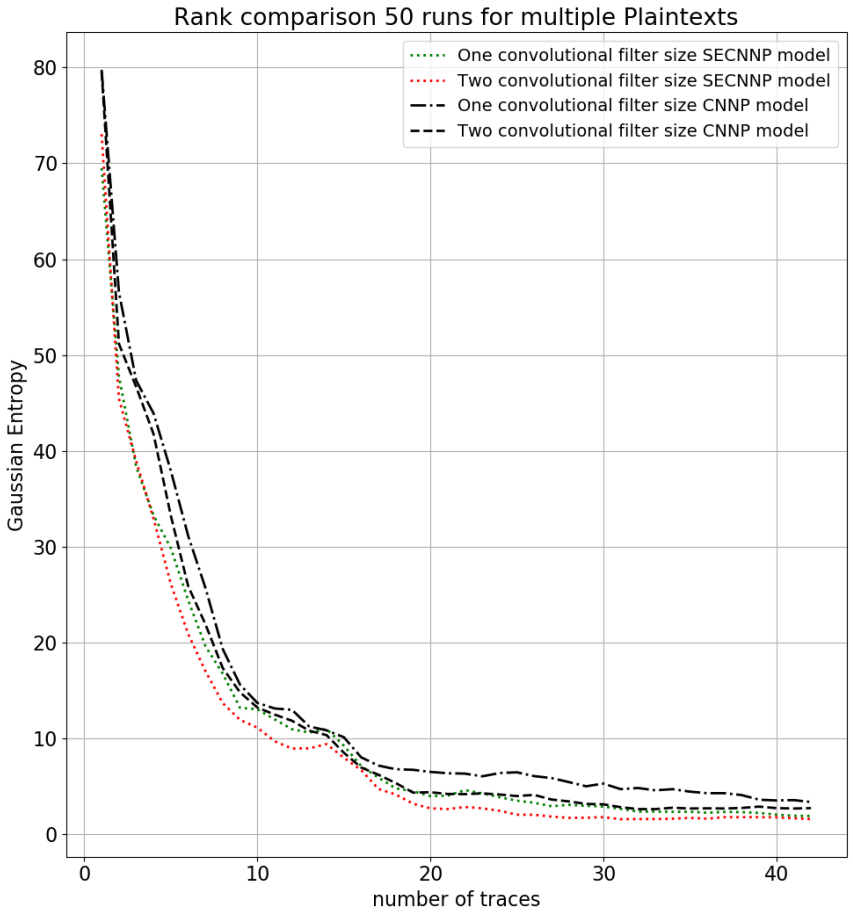


Two kernel sizes SECNNP model evaluation with multiple Plaintexts



- Sub-models are the same two convolutional filter sizes model but different training epochs.
- SEM models are built from single input trace and 3 CNNP sub-models.
- SEM achieved better results than the reference CNNP sub-models.

One vs Two kernel sizes SECNNP model evaluation with multiple Plaintexts



- SEM models contains 1 trace and 3 sub models.
- Three sub-models are the same model with one or two convolutional filter sizes with different training epochs.
- SEM model built from two convolutional filter sizes sub-model reduces the number of required trace to a half compared with the referred sub-model.

Side-channel Analysis (SCA)

A-T. Hoang, N. Hanley, M.O'Neill,
Plaintext: A Missing Feature for Enhancing the Power of Deep Learning in Side-Channel Analysis?
IACR Transactions on Cryptographic Hardware and Embedded Systems (TCHES), 2020(4), 49-85

A-T Hoang, N. Hanley, A. Khalid, D. S. Kundi, M. O'Neill
Stacked Ensemble Model for Enhancing the DL based SCA. **Proceedings of the 19th International Conference on Security and Cryptography, (SECRYPT),** 2022, 59-68

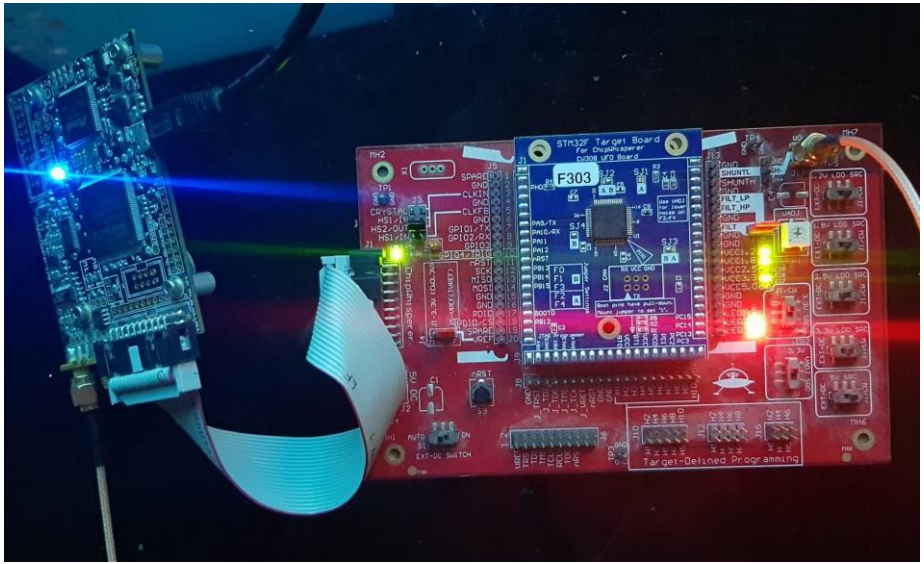
A-T Hoang, N. Hanley, A. Khalid, D. S. Kundi, M. O'Neill
Stacked Ensemble Model Evaluation on DL based SCA. **Springer – Communications in Computer and Information Science (CCIS) book series, 2023.**

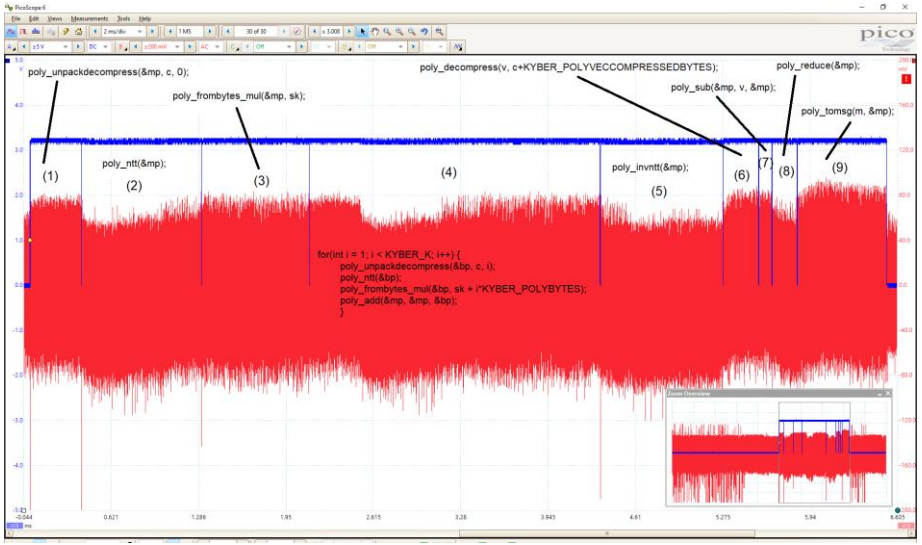
CSIT CENTRE FOR SECURE INFORMATION TECHNOLOGIES

Next steps

- Deep learning based SCA applied to PQC implementations

Profiling SCA platform using ChipWhisperer for Kyber





Thank you