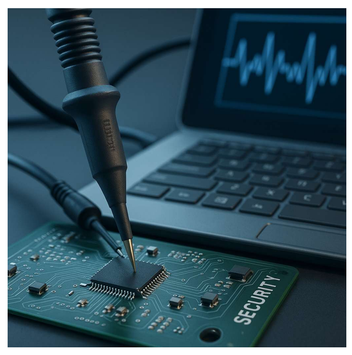
Exploring the Performance of BERT Transformers on ASCAD Variable-Key Dataset

Ali Alper Sakar[+*] & Elif Bilge Kavun[+~]
+ {ali.sakar, elif.kavun}@barkhauseninstitut.org
* sakar01@ads.uni-passau.de
~ elif_bilge.kavun@tu-dresden.de

Introduction & Motivation

- AES implementations leak information via power traces, leading to Side-Channel Attacks (SCAs)
- SCAs have become more threatening with use of Deep Learning (DL) models
- Traditional DL models (CNNs) are good but limited in modeling long-range dependencies

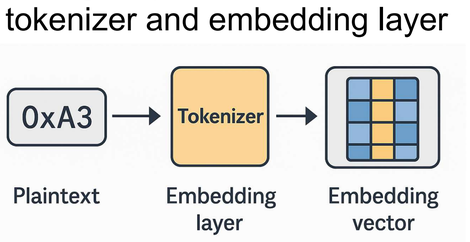


- **BERT Transformers** are a powerful sequence model not yet fully explored in SCA
- **variable-key scenario = real-world setting**.
- We demonstrate, *for the first time*, **full key recovery with BERT** on **ASCADv1 Variable-Key**

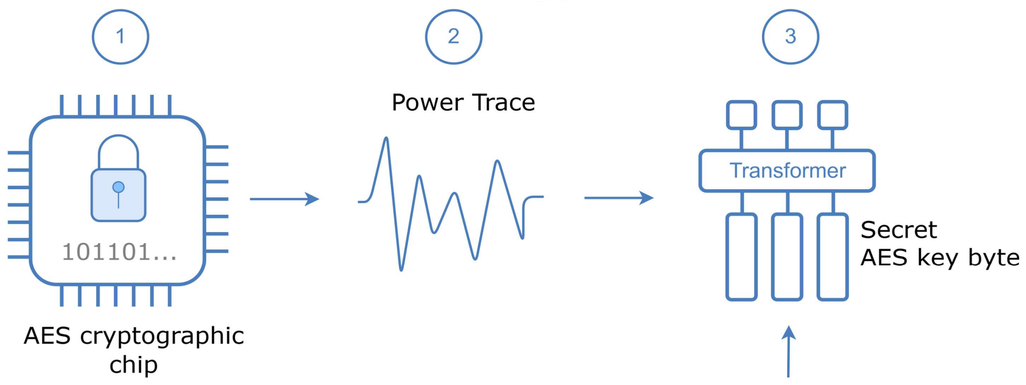
Methodology

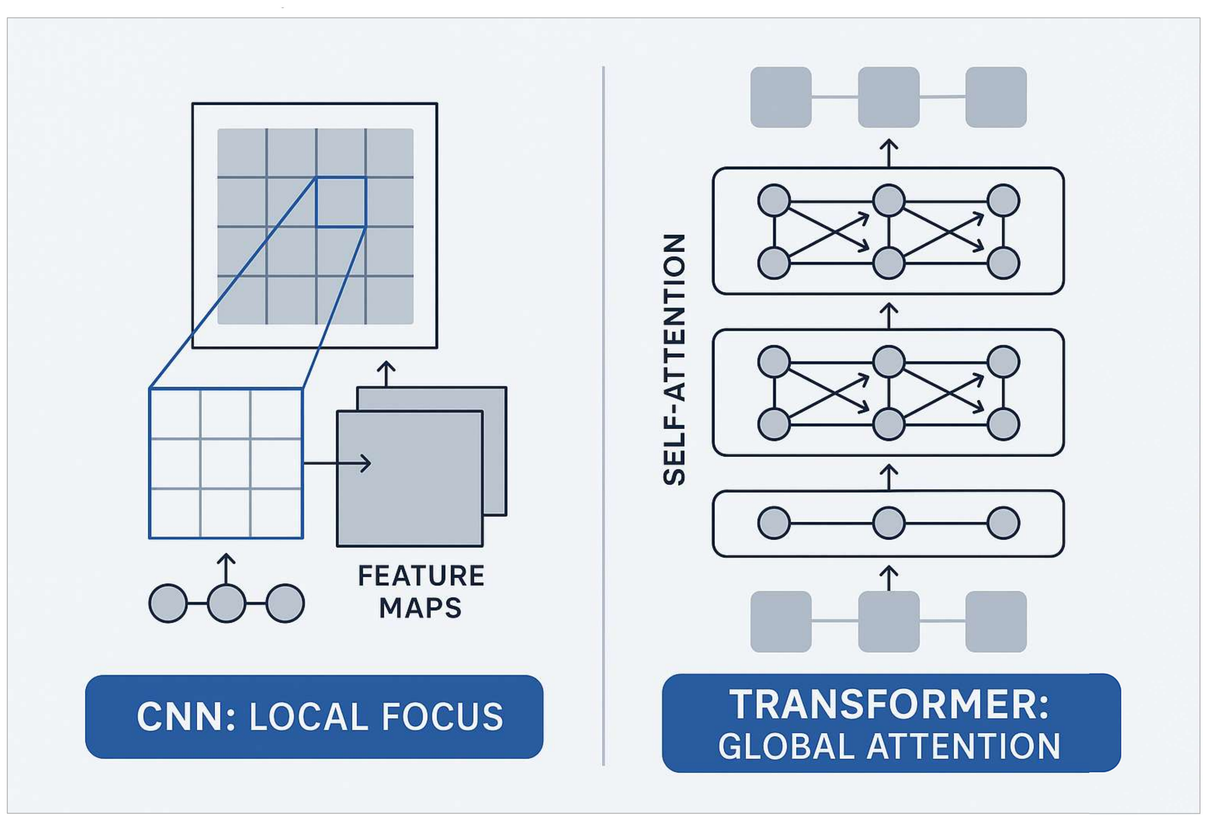
- We propose a Transformer-based attack pipeline that fuses power trace and plaintext embeddings to predict AES key bytes in a variable-key setting
- The attack model receives both power traces and corresponding plaintext bytes as input
- Our methodology includes:

➢ **Plaintext Embedding**
- Three strategies (HEX, BYTE, ASCII) are evaluated
- HEX-based tokenization yields the best key rank performance
- Each byte is encoded as a string (e.g., '0xA3') and passed through a pretrained BERT tokenizer and embedding layer



➢ **Trace Fusion Strategy**
- Traces are projected into the same embedding dimension via a learnable linear layer, enabling concatenation with BERT embeddings
- Trace and plaintext embeddings are concatenated and passed through fully connected layers for classification

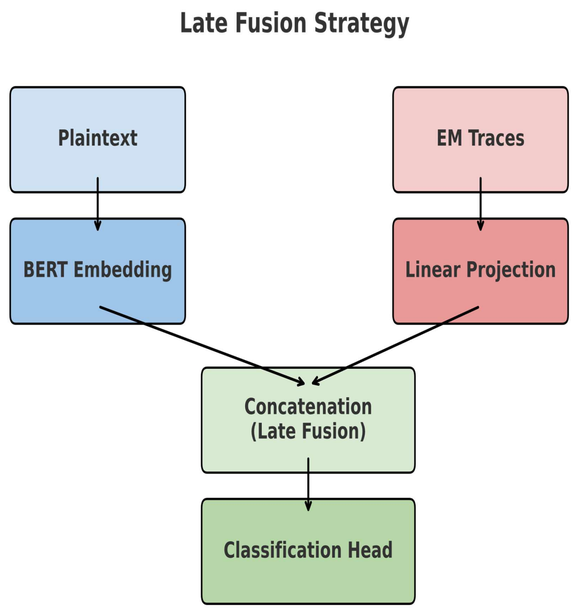




CNNs focus locally; Transformers attend globally

Architecture Highlights

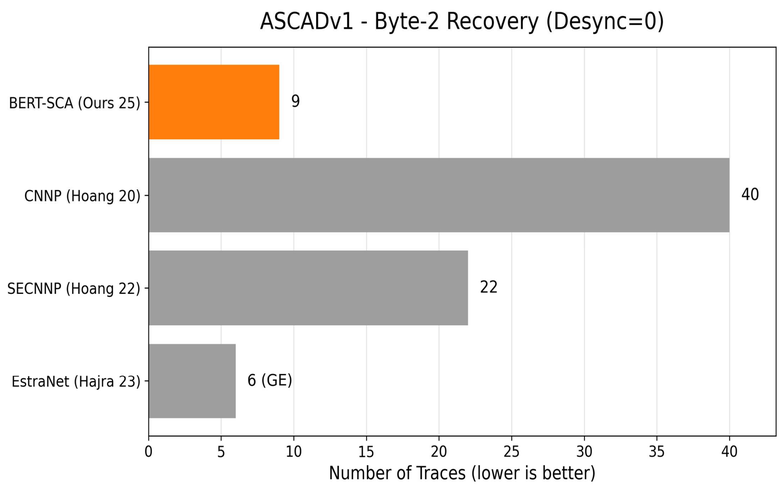
- BERT backbone (12-layer, 12-head, pre-trained) for plaintext embedding
- Trace encoder + feature fusion to combine EM traces with BERT embeddings
- Custom classification for robust key-byte prediction



Late fusion architecture combining plaintext embeddings (BERT) with EM trace features for key-byte classification

Results & Evaluation

- We benchmarked our BERT-SCA model against top DL approaches evaluated on the ASCAD variable-key dataset
- Our model achieves key-byte extraction with as few as 9 traces
  - Approaching recent Transformer models
  - Outperforming the best CNN results
- Uniquely, we demonstrate full 16-byte AES key recovery, even under strong desynchronization, a first for BERT-based models
- These results highlight our model's strength in both early-byte inference and full-key generalization



Our model achieves rank-0 recovery in 9 traces, outperforming CNNs; only EstraNet reports lower GE-based numbers (not strict rank-0).

- Desync Robustness (BERT-SCA)
  - 9 traces (Desync=0)
  - 305 traces (Desync=50)
  - 444 traces (Desync=100)

Novelty: First Transformer-based model combining early-byte efficiency and full-key generalization across all desync levels

First *BERT-SCA* with Full Key Recovery on ASCADv1 Variable-Key Dataset

Key Benchmark Summary

Byte-Level Key Recovery (Desync=0)
- Our BERT-SCA model: 9 traces
- CNN baselines: 22-40 traces (Hoang et al. Byte-2 only, no full-key recovery)
- Other Transformer: EstraNet (Hajra et al., 6 traces (GE metric, not strict rank-0)
- 🔓 Full Key Recovery
  ✅ Our BERT-SCA successfully recovered all 16 key bytes
  ✅ First Transformer-based SCA to achieve full key recovery on ASCAD v1 variable-key, robust across desynchronization (0/50/100)

Model	Model Type	Year	Traces for Byte-2	Full Key Recovery
BERT-SCA (Ours)	BERT Transformer	2025	9 (Rank-0)	True
CNNP (Hoang) [1]	CNN + Plaintext	2020	≈40 (Rank-0)	False
SECNNP (Hoang) [2]	Stacked Ensemble CNN	2022	≈22 (Rank-0)	False
EstraNet (Hajra) [3]	Transformer (Linear Complexity)	2023	≈6 (GE)	False

References

[1] Anh-Tuan Hoang, Neil Hanley, and Maire O'Neill. "Plaintext: A Missing Feature for Enhancing the Power of Deep Learning in Side-Channel Analysis?" In: IACR Transactions on Cryptographic Hardware and Embedded Systems 2020.4 (2020), pp. 49–85. doi: 10.13154/tches.v2020.i4.49-85 (cit. on pp. 17, 85).
[2] Anh Hoang et al. "Stacked Ensemble Model for Enhancing the DL based SCA." In: Jan. 2022, pp. 59–68. doi: 10.5220/0011139700003283 (cit. on pp. 16, 85).
[3] Suvadeep Hajra, Siddhartha Chowdhury, and Debdeep Mukhopadhyay. EstraNet: An Efficient Shift-Invariant Transformer Network for Side-Channel Analysis. Cryptology ePrint Archive, Paper 2023/1860. 2023. url: https://eprint.iacr.org/2023/1860 (cit. on pp. 4, 18, 21).

All images are generated using GPT by A. A. Sakar

*** Source code available at https://github.com/AliAlperSakar/bert_ascad_sca_masterthesis/tree/master/work/BERT/ASCAD_all_latest**