International Conference on Artificial Intelligence and Information Communication (ICAIIC)

MT-MO: Efficient and Robust Non-Profiled Side-Channel Analysis Using Multitask Learning

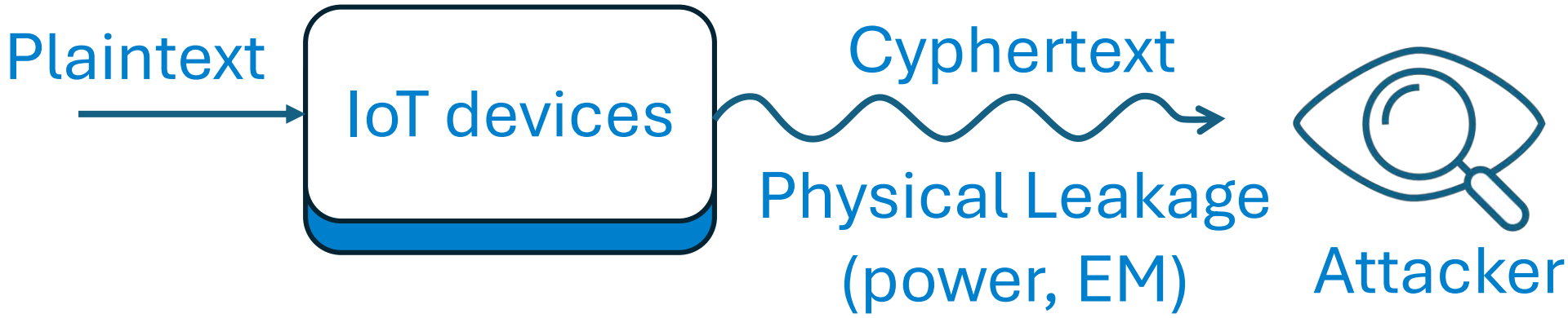
Van-Phuc Hoang et al.
Le Quy Don Technical University, Hanoi, Vietnam

February, 2026

Presentation Outline

- Introduction of Side-Channel Analysis (SCA)
- Proposed method multitask multi-output (MT-MO) SCA
- Results & discussions
- Conclusions

Side-Channel Analysis (SCA): Physical Leakage in Cryptographic Devices Creates a Critical Attack Vector



Profiled Attacks

Require an identical reference device for profiling. Highly accurate but often impractical for closed commercial systems.

Non-Profiled Attacks

Operate directly on the target device without a profiling phase. More practical but traditionally less efficient.

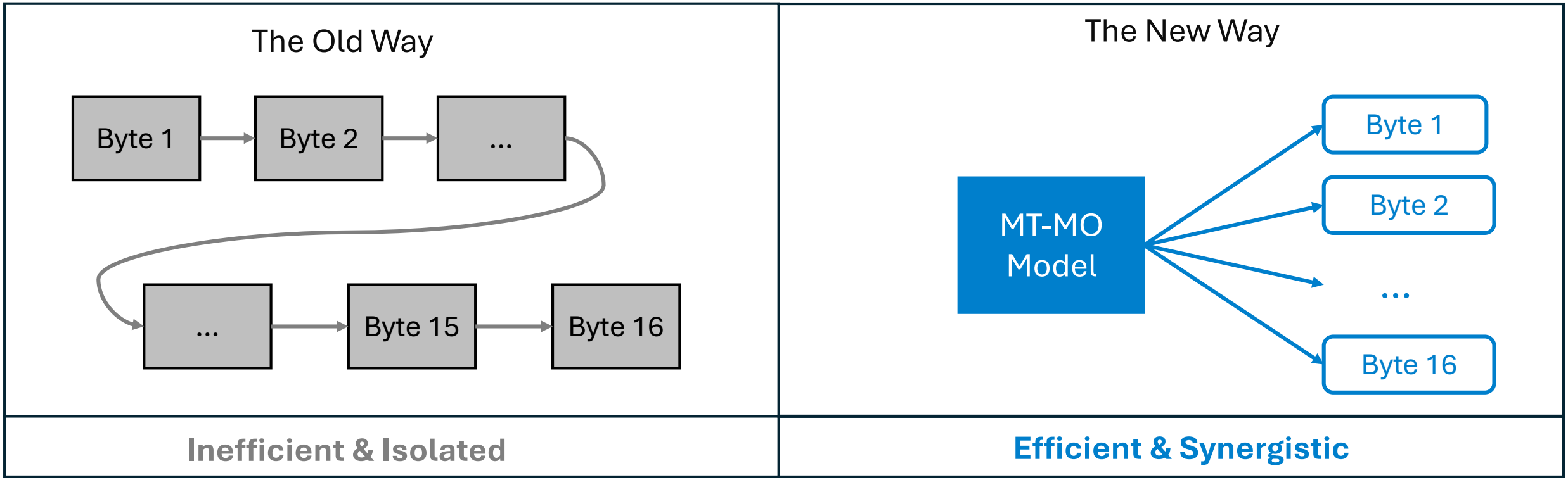
Current Deep Learning SCA Models Attack Key Bytes Sequentially, Creating a Major Bottleneck

Key Concept

Recent advances like Multi-Output Classification (MOC) and Regression (MOR) have improved non-profiled attacks by evaluating all 256 key hypotheses for a single byte at once.

The Problem

To recover a full 16-byte AES key, these models must be trained and run 16 separate times.



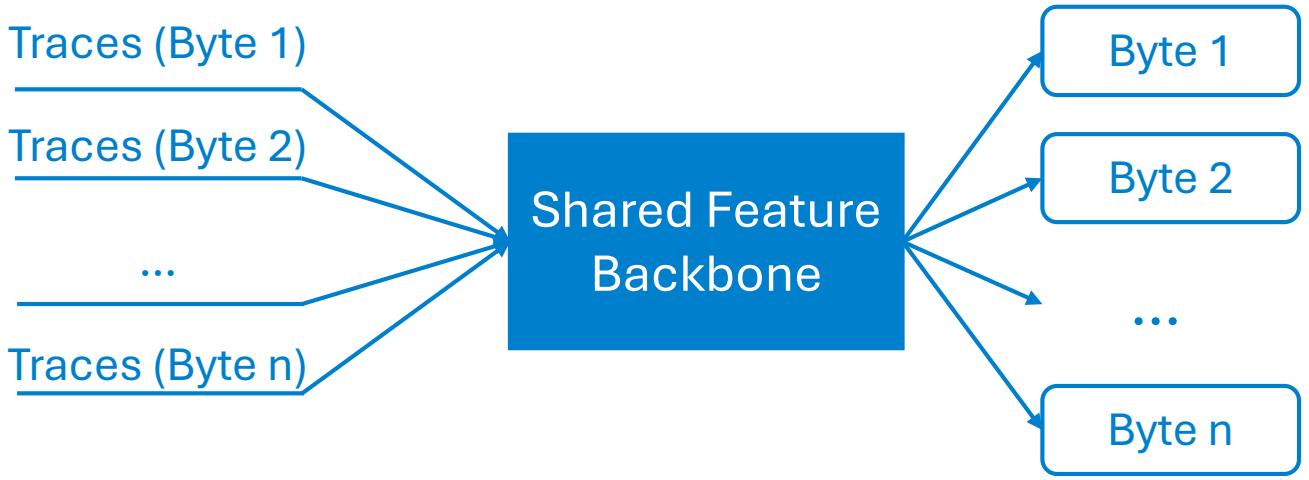
Our Solution Treats Multi-Byte Key Recovery as a Single, Unified Multitask Learning Problem

Core Insight

Instead of treating each key byte as an independent task, we attack multiple bytes simultaneously within a single neural network architecture.

Mechanism

Hard Parameter Sharing: We utilize a shared feature extraction backbone that processes trace data from multiple byte operations.



Eliminates Redundancy

The shared backbone learns generalized leakage features applicable to all target bytes, preventing redundant learning.

Improves Generalization

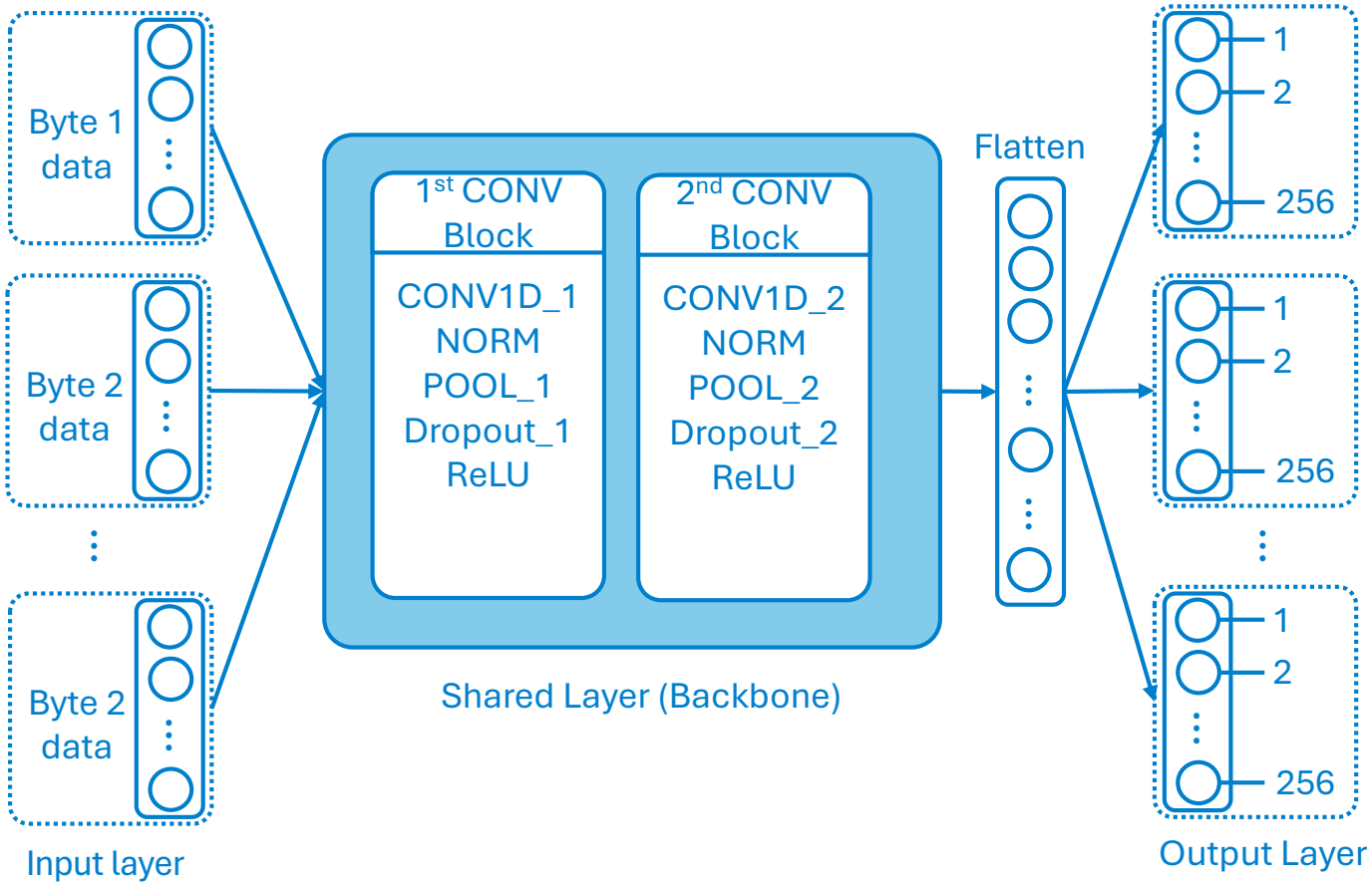
Exposing the model to leakage variations across multiple bytes forces it to learn more robust, fundamental patterns.

Dramatically Reduces Attack Time

A single training session replaces multiple sequential sessions..

Proposed MT-MO Architecture Employs a Shared Backbone and Task-Specific Heads

Multitask Input Layer
Accepts *n* distinct power trace segments simultaneously, each corresponding to a different target key byte.



Task-Specific Output Layer: The shared feature vector is fed into *n* independent output 'heads'. Each head is a fully connected layer with 256 nodes, representing the 256 key hypotheses for its specific byte.

Shared Layer (Backbone)
The core of the model. All inputs are processed by two convolutional blocks. This enforces the learning of generalized features.

MT-MOC Variant Simultaneously Classifies Leakage for Multiple Bytes

Objective: Predict the Least Significant Bit (LSB) of S-Box output for each key hypothesis

Key Technical Choices & Rationale:

Vectorized Output Heads

All 256 key hypotheses are fused into a single dense layer.

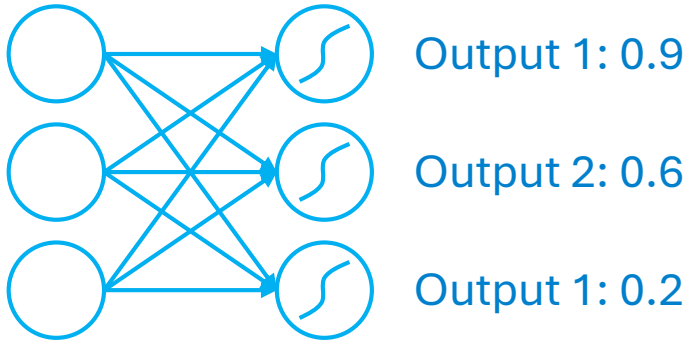
Sigmoid Activation

Allows the model to estimate the probability $P(\mathrm{LSB} = 1|\boldsymbol{t},k)$ for each of the 256 candidates independently, without the mutual suppression imposed by Softmax.

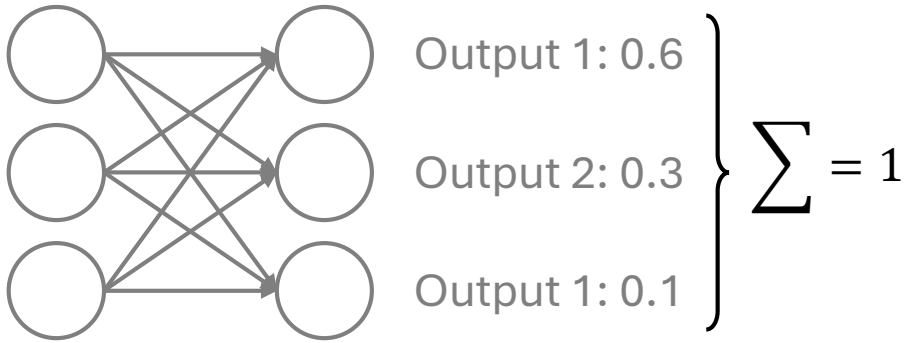
Loss Function

A sum of Binary Cross-Entropy (BCE) losses across all tasks (bytes). This joint optimization forces the shared backbone to learn features beneficial for all bytes simultaneously.

Sigmoid Activation



Softmax Activation



MT-MOR Variant Aligns its Objective with Physical Power Consumption Models

Objective: Directly regress the scalar leakage value associated with the cryptographic operation

Key Technical Choices & Rationale:

Labeling Strategy: Hamming Weight

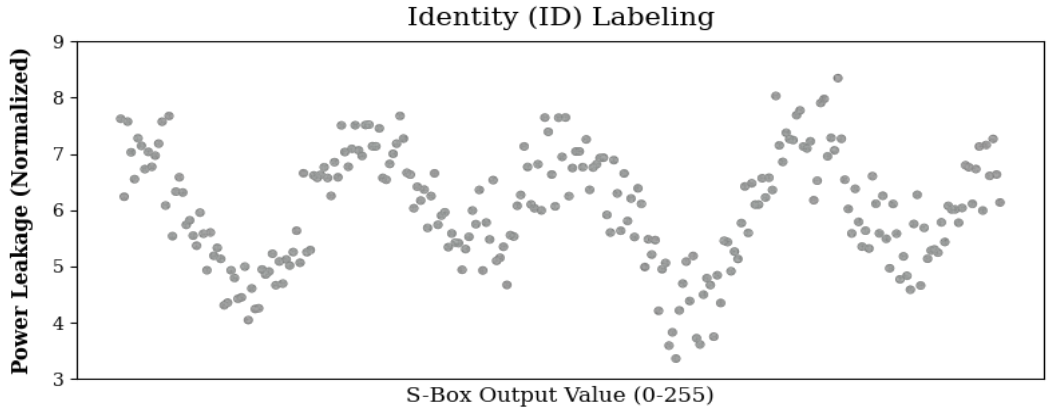
The regression task is aligned with the physical leakage model, simplifying the problem and accelerating convergence.

Linear Activation

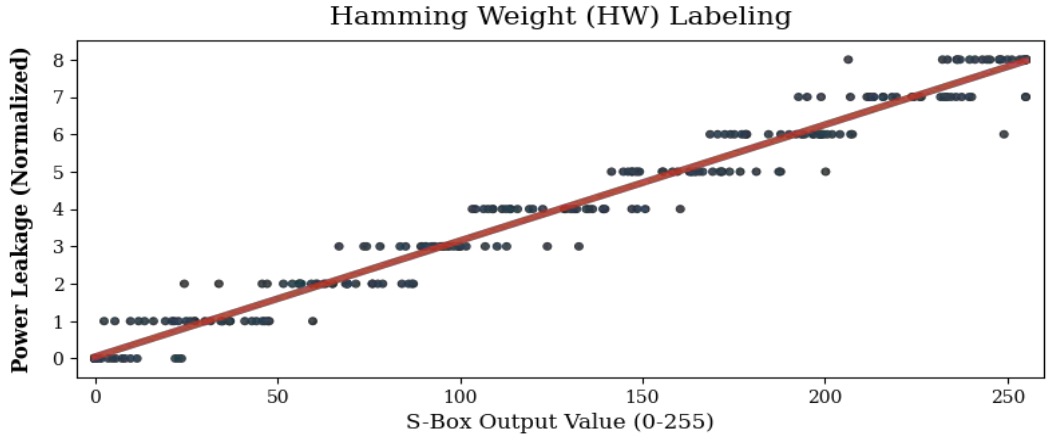
Allows the output heads to predict continuous values representing the estimated leakage.

Loss Function

MSE heavily penalizes large deviations, forcing the model to converge on the correct key hypothesis whose predicted leakage values most closely match the HW-based labels

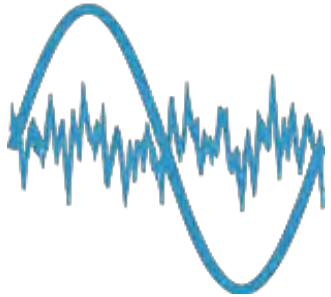


Complex, non-linear relationship. Harder for the model to learn.



Clear, linear relationship.
Aligns with physical reality, simpler to learn.

Testing Proposed MT-MO Architecture Against Modern Countermeasures

Setup Details	Countermeasures		
Hardware Platform Apple M1 with 8GB of unified memory. Experiments were restricted to 3 simultaneous bytes as a proof-of-concept due to memory constraints.			
Evaluation Metrics ***Success Rate (SR)***: Percentage of successful key recoveries over 30 repeated trials. ***Total Attack Time (TA)***: Wall-clocktime in seconds.	**Masking** ASCAD dataset with first-order Boolean masking (15,000 traces).	**Noise Injection** ASCAD dataset with synthetic AWGN (σ = 1.0) added (5,000 traces).	**Desynchronization** ChipWhisperer (CW) dataset with random temporal shifts up to 20 samples (5,000 traces).

Against Masking, MT-MOC Achieves Higher Success Rates, 4.6x Faster

Quantitative Results

Models Compared: Baseline CNN_{MOC} (sequential) vs. Proposed MT-MOC (simultaneous).

Speed

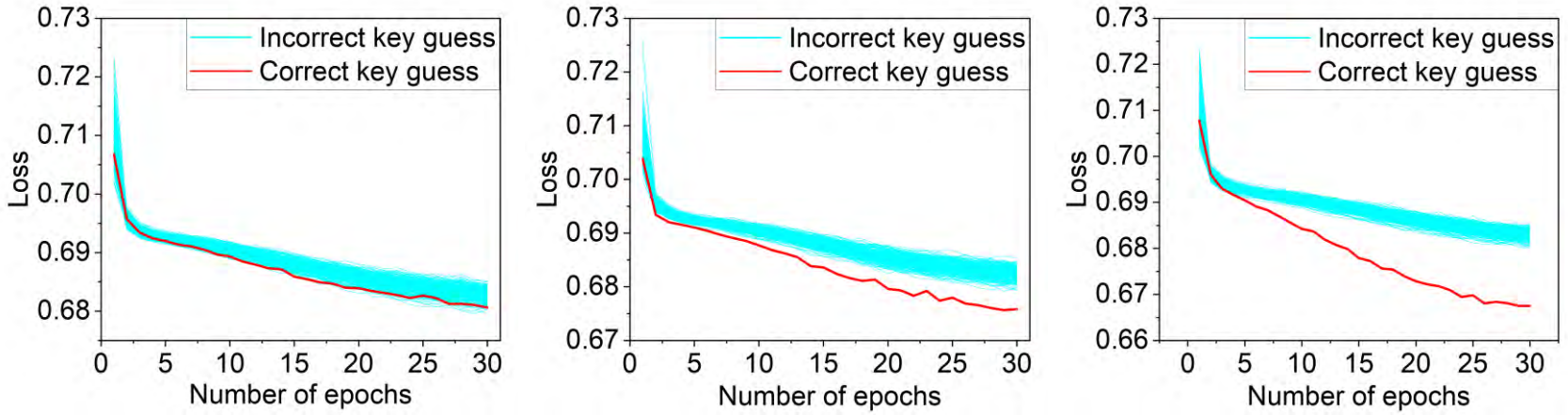
Total time for 3 bytes reduced from 538s to 116s

4.6x speedup

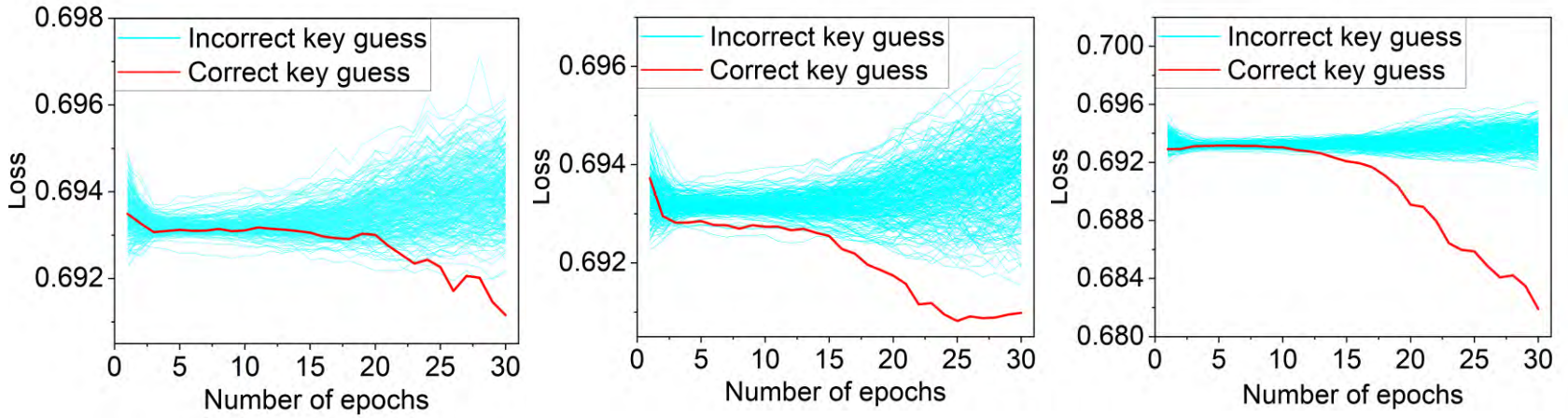
Success Rate

For the challenging Byte 3, SR improved from 50% (CNN_{MOC}) to 73.3% (MT-MOC).
For other bytes, SR was maintained at 100%.

Byte 3 | Byte 4 | Byte 6



Baseline CNN_{MOC}



Proposed MT-MOC

Against Noise Injection, MT-MOR Achieves 11x Speedup

Quantitative Results

Models Compared: Baseline CNN_{MOR} (sequential) vs. Proposed MT-MOR (simultaneous).

Speed

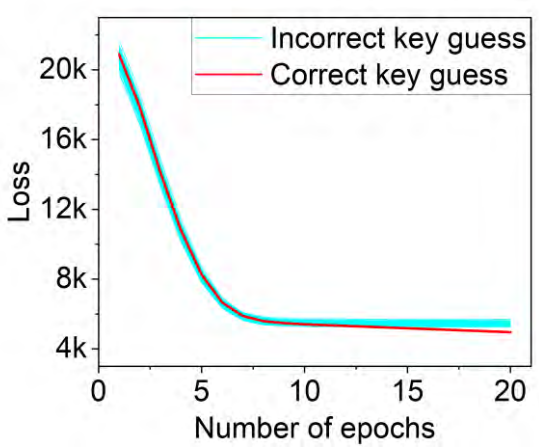
Total time for 3 bytes plummeted from ~317s to just 28.2s

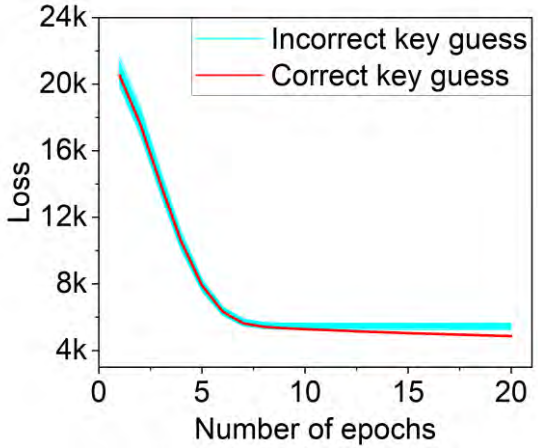
11x speedup

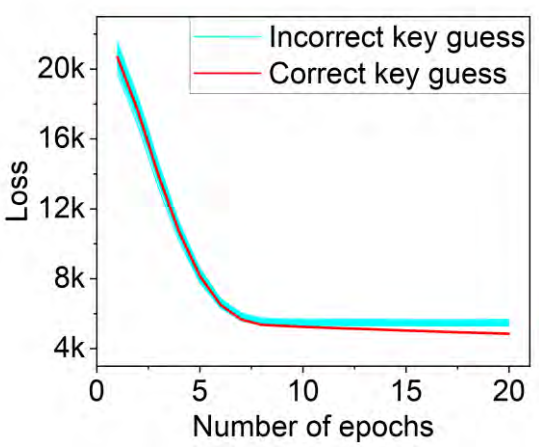
Success Rate

Both models achieved a 100% Success Rate, but MT-MOR did so in a fraction of the time.

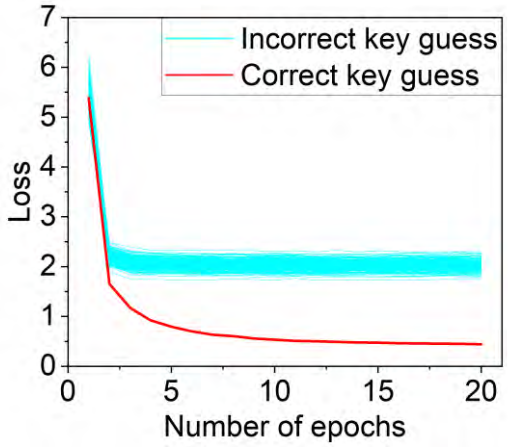
Byte 5 | Byte 6 | Byte 7

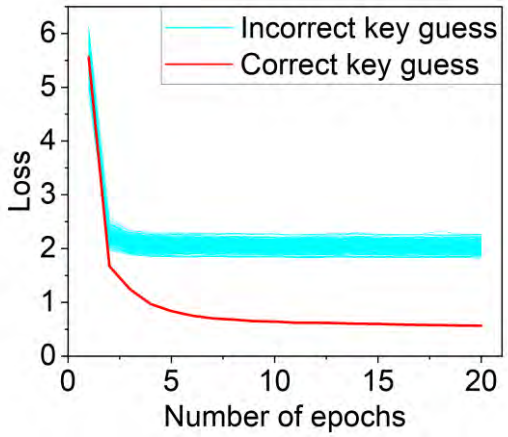


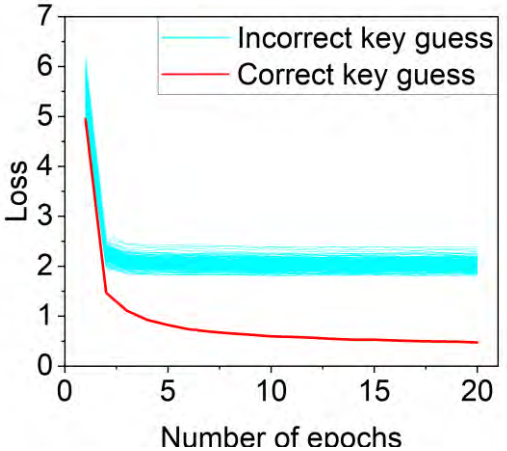




Baseline CNN_{MOR}







Proposed MT-MOR

MT-MOR Demonstrates Remarkable Robustness to Temporal Desynchronization

Quantitative Results

Models Compared: Baseline CNN_{MOR} (sequential) vs. Proposed MT-MOR (simultaneous).

Speed

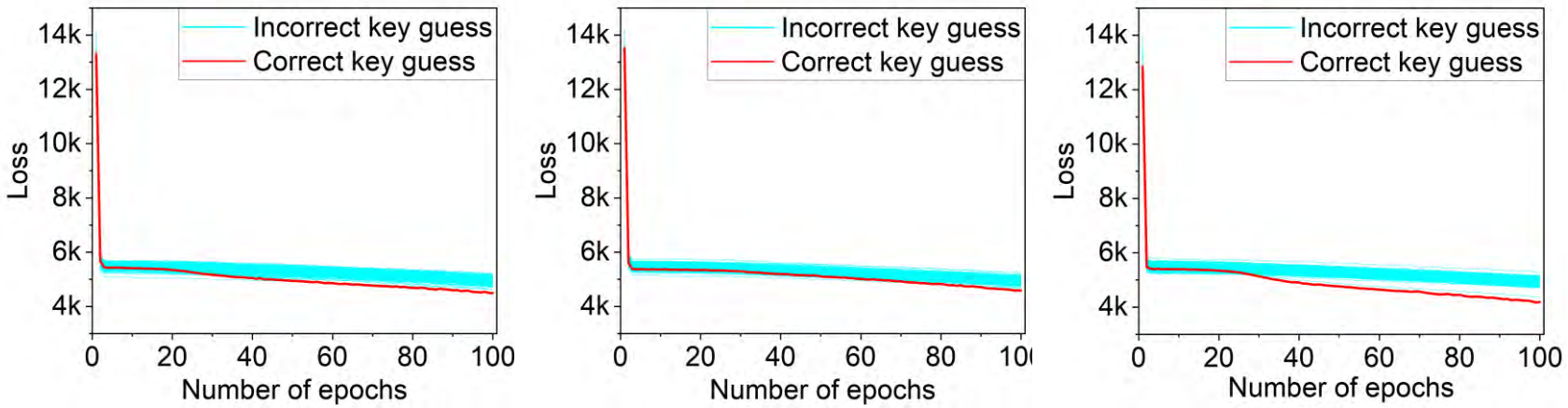
Total time for 3 bytes reduced from 1790.6s to 97.6s

18x speedup

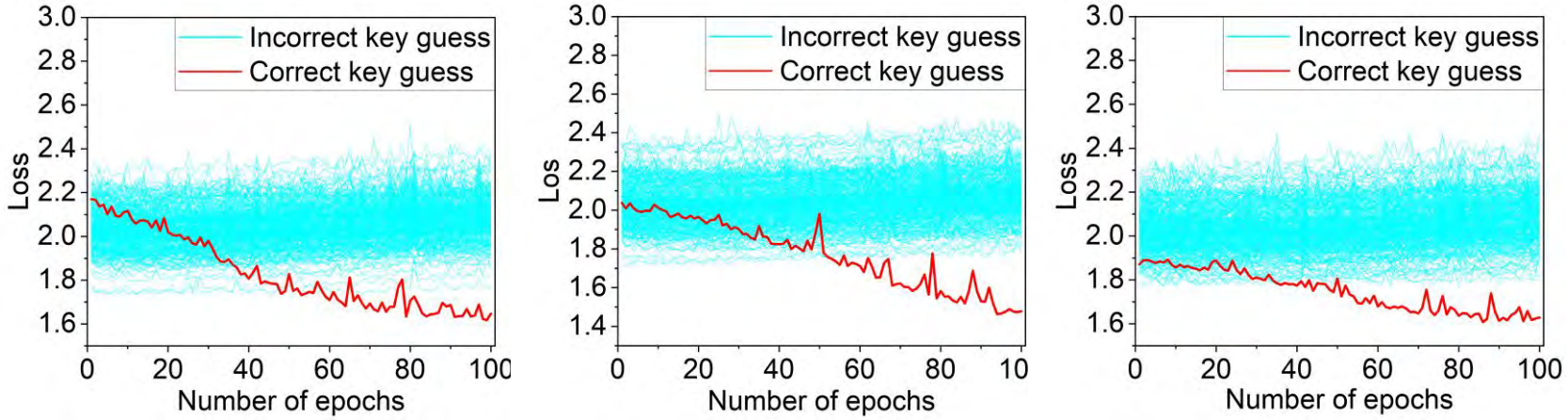
Success Rate

The baseline struggles, with SR as low as 53.3% and 56.7%. The proposed MT-MOR shows a substantial improvement, with SR increasing to 80% and 86.7%

Byte 1 | Byte 5 | Byte 7

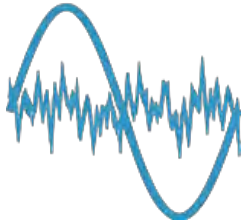


Baseline CNN_{MOR}



Proposed MT-MOR

The Multitask Approach Consistently Outperforms Sequential Attacks in Both Speed and Robustness

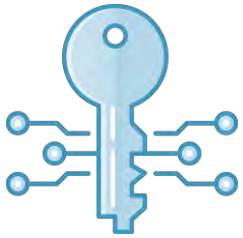
 Vs. Masking (MT-MOC)	 Vs. Noise (MT-MOR)	 Vs. Desynchronization (MT-MOR)
4.6x Faster Execution Time SR increased from 50% → 73.3% on the most difficult byte.	**11x** Faster Execution Time 100% Success Rate maintained.	**18x** Faster Execution Time SR improved from 56.7% → 86.7%

Conclusions

- Simultaneous multi-byte attacks redefine efficiency in non-profiled SCA

Core Contribution

We introduced a multitask multi-output (MT-MO) framework that fundamentally mitigates the computational redundancy inherent in sequential, single-task SCA attacks.



Main Findings

The Hard Parameter Sharing strategy is highly effective. It forces the shared backbone to learn generalized leakage features, leading to:

- Significantly reduced attack times.
- Improved model generalization and robustness against countermeasures.

This work validates that a holistic, multitask approach is a more efficient and powerful paradigm for non-profiled deep learning-based side-channel analysis.

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.