Hessisches Kompetenzzentrum
für Hochleistungsrechnen

Exploring Future Directions of Machine Learning-Based Cryptanalysis

Project Manager
Moritz Huppert

Researchers
Marlon Friedl and Simon Povh

Principal Investigator
Prof. Dr. Marc Fischlin

Project Term
2025 - 2025

Clusters
Lichtenberg II Cluster Darmstadt

Software
PyTorch

Additional Software
CryptoSMT, Mercy-Framework

Institute
Cryptography and Complexity Theory

University
Technische Universität Darmstadt

Introduction

Differential cryptanalysis exploits non-uniformities in a block cipher's output-difference distribution that arise from a known input difference. At CRYPTO 2019, Gohr showed that convolutional neural networks (CNNs) can detect these biases more effectively than traditional distinguishers. Leveraging a CNN-based distinguisher, Gohr achieved state-of-the-art round-key recovery attacks on 11- and 12-round SPECK32/64. Building on insights from the literature and empirical results from earlier experiments on the Lichtenberg Cluster, we explored several promising directions for further improvement. As in many AI-driven approaches, the work is computationally demanding and therefore relies on large-scale experiments on the Lichtenberg Cluster.

Methods

We explored several approaches to improve neural differential cryptanalysis. First, rather than training on a single fixed dataset of 10^7 samples, we generated fresh ciphertext samples continuously throughout training to reduce artifacts caused by a static dataset. Second, we trained networks to detect the specific differential trail induced by a prepended differential, enabling early stopping or ensemble methods during key recovery. Third, we investigated a curriculum strategy in which models were first trained on easier, pre-filtered 8- and 9-round Speck ciphertexts and then gradually adapted to the real, unfiltered distribution. Fourth, we trained distinguishers on

7-round decryption to assess whether backward models could be useful for attacks. Finally, we investigated “neural neutral bits”, i.e., bit flips that leave the distinguisher’s output largely unchanged as they preserve similar differential behaviour, and used them to construct corresponding plaintext structures.

Results

Generating fresh ciphertext samples during training stabilized optimization but yielded only marginal improvements in accuracy. The trail-detection was unsuccessful because, for most trails, the differential bias vanished too quickly for models to learn a reliable (>6 rounds) distinguisher. The curriculum approach produced promising early gains, but these mostly disappeared by the end of training: for 8 rounds, performance became similar to layer-freezing baselines, while for 9 rounds, the advantage was lost during the final unfiltering phase. The 7-round reverse-direction distinguishers were reasonably strong, but still underperformed Gohr’s forward models and, hence, did not suggest a clear attack application. Neural neutral bits gave modest improvements but were insufficient to replace Gohr’s classical prepended differential.

Discussion

On-the-fly generation of ciphertext pairs has the clear advantage of reducing dataset bias, which is particularly important when distinguishing distributions that are already very close to uniform. It also reduces memory requirements, making it especially attractive for multi-pair distinguishers or very large training sets, such as the 10^9 samples used in Gohr’s staged training pipeline. In our experiments, we observed a slight improvement in training stability. However, further research is needed to determine under what conditions on-the-fly data generation can yield substantial gains in distinguisher accuracy. If the prepended differential chosen by Gohr is not fulfilled, most alternative differential trails lead to input differences whose induced bias appears too weak to be exploited by a neural distinguisher. Whether the same limitation also holds for other choices of prepended differentials remains an open question. The curriculum-style training strategy initially helped mitigate the difficulty of training distinguishers for a larger number of rounds. However, in our experiments, the initial gains disappeared once training transitioned to the true, unfiltered ciphertext distribution. We showed that neural distinguishers can also be trained in the decryption direction. How to turn these distinguishers into an effective key-recovery attack remains an open question for future work. We found that classical neutral bits for a prepended differentials provide better practical performance than “neural neutral bits.”

Publications

Reference

Last Update: 2026-04-10 19:14