Offre de stage - Internship Offer

Towards Interpretable Generative Side-Channel Attacks on Desynchronized Traces

6-*months master internship anytime in 2027*

1 Internship Subject

Side-Channel Attacks (SCA) exploit physical leakages produced during the execution of cryptographic algorithms, such as power consumption or electromagnetic emissions, in order to recover sensitive information. In the context of security evaluation, one of the main challenges is to design attacks that are not only efficient, but whose behavior can also be understood and theoretically justified.

Over the last decade, Deep Learning based Side-Channel Analysis (DLSCA) has demonstrated strong performance against complex implementations and countermeasures. In particular, discriminative neural networks such as Convolutional Neural Networks (CNNs) have shown an ability to automatically cope with trace misalignment or desynchronization. Nevertheless, these models are generally difficult to interpret, and their architecture often relies on empirical choices. This black-box behavior is particularly problematic in a security evaluation context, where the evaluator should be able to justify the selected attack methodology and understand its limitations.

To bridge the gap between classical side-channel analysis and deep learning, the community introduced the Conditional Variational AutoEncoder based on Stochastic Attacks (cVAE-SA) [ZBC+23]. This generative approach connects conditional variational autoencoders with the statistical foundations of stochastic side-channel attacks. In contrast with standard discriminative neural networks, the architecture is designed so that its parameters can be related to meaningful quantities of the physical leakage. More recently, this work have been extended in [BCG+25]. The authors introduced a simplified conditional variational autoencoder that integrates an optimal dimensionality reduction directly within the neural network. The resulting architecture is both explainable and interpretable and converges towards the optimal statistical model under the assumptions considered in the work. In particular, its trainable parameters can be associated with quantities such as the deterministic leakage model and the variance of the traces.

An important limitation nevertheless remains: *cVAE-OSM assumes that the input side-channel traces are synchronized*. In practice, this assumption is often unrealistic. Traces acquired from embedded devices may exhibit temporal shifts caused by clock variations, acquisition jitter, interrupt mechanisms, random delays, or intentional hiding countermeasures. Classical statistical attacks therefore usually require an additional alignment or preprocessing stage. In contrast, discriminative deep-learning models can partially learn shift-invariant representations automatically, for example through convolutional layers, but at the cost of reduced interpretability. Handling desynchronization directly inside an interpretable generative model therefore constitutes an important open problem. Both cVAE-SA [ZBC+23] and cVAE-OSM [BCG+25] identify this issue as a promising research direction. More recent extensions of cVAE-OSM to non-profiled side-channel analysis also highlight automatic management of desynchronization while preserving explainability and interpretability as an important remaining challenge [BCG+26].

The objective of this internship is therefore to **extend conditional variational autoencoder based side-channel attacks to desynchronized traces, while preserving as much as possible the theoretical guarantees, explainability, and interpretability of the original approach**.

A natural research direction will consist in explicitly incorporating temporal desynchronization into the generative leakage model. For synchronized traces, the leakage is typically modeled as

$$\mathbf{T} = \Psi(Y) + \mathbf{N},$$

where Y denotes a sensitive intermediate value, Ψ the deterministic leakage model and $\mathbf{N}$ the acquisition noise. In the presence of temporal desynchronization, this model could for example be extended as

$$\mathbf{T} = \mathcal{S}_{\Delta}(\Psi(Y)) + \mathbf{N},$$

where Δ is a random temporal displacement and $\mathcal{S}_{\Delta}$ denotes the corresponding shift operator. From a probabilistic perspective, the temporal shift then becomes an additional nuisance or latent variable. Consequently, the distribution $P(\mathbf{T} \mid Y)$ considered by the original cVAE is no longer necessarily described by a single aligned conditional Gaussian distribution, but can be viewed as resulting from the marginalization of the unknown displacement:

$$P(\mathbf{T} \mid Y) = \sum_{\delta} P(\mathbf{T} \mid Y, \Delta = \delta) P(\Delta = \delta),$$

in the case of discrete temporal shifts.

This observation raises several research questions: how should the temporal displacement be represented in the generative model? Can it be estimated jointly with the leakage model? Can it instead be marginalized during optimization? Is it possible to derive a shift-invariant dimensionality reduction related to the optimal dimensionality reduction of cVAE-OSM? Finally, can convolutional or other shift-equivariant components be introduced without losing the interpretability properties of the original architecture? The internship will investigate these questions from both theoretical and experimental perspectives.

1.1 Objectives

The internship will be structured around several phases in order to progressively extend the existing cVAE framework towards realistic desynchronized side-channel measurements.

1. Understanding and reproducing generative side-channel attacks.

The first phase will focus on studying the statistical foundations of profiled side-channel attacks and their connection with generative neural networks. The intern will study stochastic attacks, Conditional Variational AutoEncoders, cVAE-SA [ZBC+23], and in particular the cVAE-OSM model [BCG+25]. Particular attention will be given to the construction of the

encoder and decoder, the interpretation of their trainable parameters, the customized ELBO loss, and the optimal dimensionality reduction performed by cVAE-OSM. The existing implementation will then be reproduced on simulated traces and selected public side-channel datasets. This phase will provide a reference implementation and establish synchronized baselines against which the extensions developed during the internship can be evaluated.

2. Modeling desynchronization in the generative framework.

The second phase will focus on understanding the impact of desynchronization on the statistical model underlying cVAE-OSM. Different desynchronization models will first be considered, starting from controlled random temporal shifts. Simulated traces will make it possible to precisely control parameters such as the maximum shift, shift distribution, trace dimension, signal-to-noise ratio, and number of informative samples. The intern will then investigate how the presence of a random temporal displacement modifies the conditional probability distribution estimated by cVAE-OSM and the optimal dimensionality-reduction mechanism on which the model relies. An important objective will be to retain a clear interpretation of the quantities learned by the network. Ideally, the resulting model should allow an evaluator to distinguish quantities corresponding to the physical leakage model, acquisition noise, and temporal displacement.

3. Design of an interpretable cVAE for desynchronized traces.

Based on the theoretical analysis, the intern will propose and implement one or several extensions of cVAE-OSM capable of operating directly on desynchronized traces. A central challenge will be to identify the appropriate compromise between shift invariance and interpretability. For example, conventional convolutional layers can provide robustness against translations but introduce additional trainable parameters whose physical or statistical interpretation is less immediate than in the original cVAE-OSM architecture. The internship will therefore investigate whether additional structural constraints can be imposed on such layers or whether the desired invariance can instead be incorporated into the probabilistic model and the loss function. Whenever possible, theoretical results will be sought to characterize the proposed solution. The objective is not necessarily to preserve all the optimality properties of cVAE-OSM under the most general desynchronization model, but to clearly determine which properties remain valid, which assumptions must be introduced, and what is lost when these assumptions are relaxed.

4. Experimental evaluation. The proposed approaches will be evaluated first using simulated traces. Simulations will enable controlled experiments in which the exact leakage model, noise distribution, and temporal shifts are known. The robustness of the different models will be analyzed as a function of the desynchronization amplitude and distribution. Several criteria will be considered, including the quality of leakage-model estimation, key rank or success rate, number of profiling and attack traces, architecture complexity, and training stability. The proposed method will then be evaluated on publicly available side-channel datasets, where controlled artificial temporal shifts can first be introduced in order to reproduce increasingly difficult desynchronization scenarios.

5. Additional contributions.

Depending on the theoretical and experimental results and the remaining time, several extensions may be investigated. A first direction consists in studying more realistic forms of desynchronization than a global temporal translation, such as local clock jitter, random delays inserted at several locations in the computation, or time scaling. A second direction consists in extending the proposed solution from the profiled cVAE-OSM setting to the recently proposed non-profiled NPcVAE-OSM framework [BCG+26]. Such an extension would investigate whether temporal desynchronization can be handled automatically while simultaneously considering all key hypotheses. Finally, the interaction between desynchronization and other countermeasures, particularly masking, could be investigated. Combining these two problems would constitute a longer-term objective towards an interpretable generative model capable of automatically handling several common side-channel countermeasures.

6. Final report and dissemination.

The internship will conclude with the writing of a detailed technical report describing the theoretical analysis, proposed architectures and loss functions, and experimental results. The intern will present the results to the team and, depending on the novelty and quality of the obtained results, the work may provide the basis for a scientific publication in the side-channel and cryptographic community.

2 Candidates

This internship offer is intended for a Master's-level student (or equivalent) interested in applied cryptography, machine learning, and research. The candidate should have a solid background in mathematics and/or computer science. Knowledge of probability, statistics, or cryptography will be particularly relevant for this internship. Good programming skills are required, primarily in Python. Previous experience with a deep-learning framework such as PyTorch or TensorFlow is appreciated but not mandatory. Knowledge of side-channel analysis is an advantage, although the necessary background can be acquired during the first phase of the internship. The candidate should demonstrate autonomy, curiosity, and an interest in understanding both the theoretical and experimental aspects of machine-learning-based cryptanalysis. A good level of English will also be appreciated. After a first interview, selected candidates may be given a technical challenge to carry out at home before a second technical interview.

3 Internship Conditions

The intern will be supervised by Sana Boussam and Gabriel Zaid, both researchers in cryptography. The internship will last six months, with a starting date in March or April 2027. The intern will be based at INRIA Saclay, within the ECLAIR team (1 Rue Honoré d'Estienne d'Orves, 91120 Palaiseau), and/or at Campus Cyber (5–7 Rue Bellini, 92800 Puteaux). Both locations offer a stimulating and collaborative research environment, fostering interac-

tions with researchers, postdoctoral fellows, PhD students, and fellow interns. Remote work arrangements are also possible.

4 Contact

To apply for this internship offer, please send your résumé to

Sana Boussam sana.boussam@inria.fr
Gabriel Zaid gabriel.zaid@cryptoexperts.com

References

[BCG+25] Sana Boussam, Mathieu Carbone, Benoît Gérard, Guénaël Renault, and Gabriel Zaid. Optimal dimensionality reduction using conditional variational autoencoder. *IACR Transactions on Cryptographic Hardware and Embedded Systems*, 2025(3):164–211, Jun. 2025.

[BCG+26] Sana Boussam, Mathieu Carbone, Benoît Gérard, Guénaël Renault, and Gabriel Zaid. From linear regression to generative model for explainable non profiled side-channel attacks. Cryptology ePrint Archive, Paper 2026/274, 2026.

[ZBC+23] Gabriel Zaid, Lilian Bossuet, Mathieu Carbone, Amaury Habrard, and Alexandre Venelli. Conditional variational AutoEncoder based on stochastic attacks. *IACR TCHES*, 2023(2):310–357, 2023.