FossilizedPluto IIT Roorkee

Daksh Pandey
Department of Physics
IIT Roorkee
Roorkee, India
daksh_p@ph.iitr.ac.in

Sparsh Mittal
Department of E&CE
IIT Roorkee
Roorkee, India
sparsh.mittal@ece.iitr.ac.in

***Abstract*—This paper presents a holistic security assessment strategy for the GateKeeper password system, designed for the CSAW ESC 2025 Qualification. We propose a multi-vector attack framework that extends beyond traditional side-channel analysis to include novel fault injection and advanced machine learning techniques. Our methodology begins with identifying timing and power leakages in the `verify` function and then details a sophisticated exploitation path using Correlation Power Analysis (CPA), voltage fault injection, and state-of-the-art deep learning architectures like Transformers. We also propose using a Generative Adversarial Network (GAN) for robust data augmentation. This comprehensive proposal outlines a clear, creative, and effective path to full password recovery and concludes with a structured discussion of corresponding countermeasures.**

***Index Terms*—side-channel analysis, fault injection, deep learning, embedded security, GAN, voltage glitching, hardware security**

I. Vulnerability Identification and Threat Model

The GateKeeper's `verify` function (lines 13-35) presents a classic case of data-dependent execution flow, making it an ideal target. Our analysis pinpoints two primary vulnerabilities:

1) **Timing and Power Side-Channel Leakage**: The 5000-cycle `delay` (line 22) conditional on a correct character guess, combined with an early-exit mechanism (line 24), creates a massive and easily exploitable information leak through both execution time and power consumption.
2) **Fault Injection Susceptibility**: The iterative nature of the 'for' loop (line 19) is a prime target for instruction skipping attacks. The comparison at line 20 ('pass_b[i] != pass_a[i]') is also a potential target for data corruption glitches.

Our threat model assumes an attacker has physical access to the device, enabling power measurement, EM probing, and the ability to manipulate the device's power supply for fault injection.

II. Proposed Multi-Vector Attack Framework

We propose a phased attack that combines multiple techniques for maximum effectiveness and showcases a deep understanding of hardware security.

A. Phase 1: Baseline Side-Channel Analysis (SCA)

The initial approach will be to leverage the most prominent leak. We will use the timing oracle to recover the full password

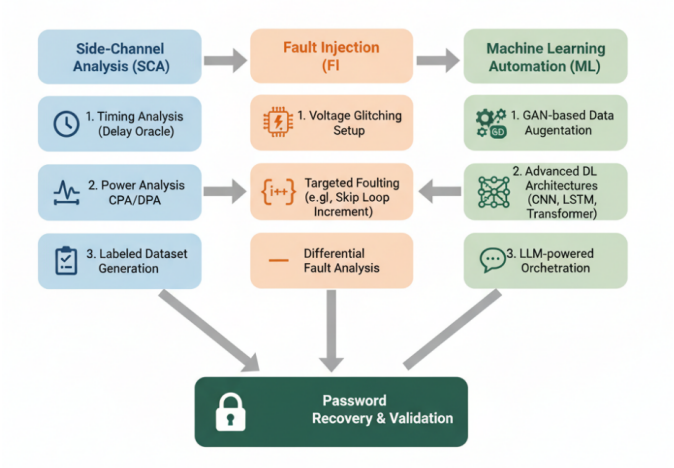

Fig. 1. A multi-stage attack workflow, progressing from initial analysis and data acquisition to advanced ML-driven exploitation and final password recovery.

and, more importantly, to create an accurately labeled dataset of power traces for our subsequent machine learning models. Following this, we will execute a standard **Correlation Power Analysis (CPA)** attack, targeting the Hamming weight of the XOR difference between the correct character and the guessed character at line 20. A preliminary analysis of the sample traces, which shows a 9.5% power consumption increase for character 'd', confirms the high signal-to-noise ratio and feasibility of this attack.

B. Phase 2: Creative Exploitation via Fault Injection (FI)

To demonstrate a more advanced capability, we propose a **voltage glitching** attack. The goal is to induce a transient fault that causes the CPU to incorrectly execute an instruction.

- **Target**: The loop increment operation ('i++').
- **Methodology**: By introducing a carefully timed, short-duration voltage drop during the execution of the 'for' loop, we aim to skip the increment instruction.
- **Outcome**: Skipping the increment would cause the same character position to be compared multiple times. By selectively applying glitches, we can bypass checks for certain character positions, effectively isolating one character at a time to leak the password with minimal SCA

measurements. This is a creative alternative to a brute-force timing attack.

C. Phase 3: Automated Attack with Advanced ML

This phase focuses on automating the analysis using a novel deep learning pipeline.

1) **Data Augmentation with GANs**: The small set of provided traces is insufficient for training a robust model. We propose training a **Generative Adversarial Network (GAN)** to learn the statistical distribution of the provided power traces. The trained generator can then produce thousands of realistic, synthetic power traces for every character, creating a rich and diverse dataset.
2) **Advanced Classification Architectures**: While a standard 1D-CNN serves as a baseline, we propose exploring more powerful architectures better suited for time-series data.
   - **LSTM/GRU Networks**: These recurrent models are designed to capture long-range dependencies in sequential data, which could be key to identifying subtle patterns in the power traces that a CNN might miss.
   - **Transformers**: As a highly creative approach, we will investigate a Transformer-based model. Its self-attention mechanism could automatically learn which specific time-steps in the 3000-sample trace are most indicative of a particular character, potentially offering state-of-the-art classification accuracy.
3) **LLM for Automation**: We will leverage a Large Language Model (e.g., GPT-5, Gemini) to auto-generate Python scripts for controlling the capture board (like a ChipWhisperer) and running the CPA analysis, thereby accelerating the entire experimental setup.

III. Literature Context

Our proposed methods are grounded in established research but extended with creative applications. The SCA and CPA techniques follow the foundational work of Kocher et al. [1], [2] and Brier et al. [3]. Our advanced ML approach is inspired by the pioneering use of CNNs in SCA by Maghrebi et al. [4], but we push the boundary by proposing GANs for data augmentation and state-of-the-art architectures like Transformers, a topic of emerging research in the field [5], [6].

IV. Mitigation Strategy: A Threat/Defense Matrix

A truly secure system requires defense-in-depth. We propose a structured set of countermeasures to mitigate the attacks outlined in this proposal.

At the **algorithmic level**, the password comparison must be rewritten to be **constant-time**, removing the early exit and data-dependent delay. For fault injection, adding redundant password comparisons and checking for consistent results can detect and thwart glitch attacks. At the **implementation level**, techniques like **Boolean masking** can decorrelate the power consumption from the intermediate data values. Finally, **hardware-level** defenses like random clock jittering and inserting random delays can desynchronize power traces, making them significantly harder for even advanced ML models to analyze.

TABLE I
Proposed Countermeasure Framework

Attack Vector	Primary Countermeasure	Level
Timing Oracle	Constant-time comparison	Algorithmic
DPA / CPA	Boolean masking (1st order)	Implementation
Advanced ML	Random instruction insertion	Hardware
	Clock jittering	Hardware
Fault Injection	Redundant checks / Duplication	Algorithmic
	Sensor-based detection	Hardware

V. Conclusion

The GateKeeper system, in its current form, is critically vulnerable. This proposal outlines a comprehensive and creative methodology to prove this, moving beyond standard analysis to include fault injection and a state-of-the-art machine learning framework. We are confident that this multi-vector approach, combining CPA, voltage glitching, GANs, and Transformer networks, provides a clear and robust pathway to complete password recovery, fulfilling the requirements of the CSAW ESC challenge.

References

[1] P. C. Kocher, "Timing attacks on implementations of Diffie-Hellman, RSA, DSS, and other systems," in Advances in Cryptology—CRYPTO '96, Springer-Verlag, 1996, pp. 104-113.

[2] P. Kocher, J. Jaffe, and B. Jun, "Differential power analysis," in Advances in Cryptology—CRYPTO '99, Springer-Verlag, 1999, pp. 388-397.

[3] E. Brier, C. Clavier, and F. Olivier, "Correlation power analysis with a leakage model," in Cryptographic Hardware and Embedded Systems-CHES 2004, Springer-Verlag, 2004, pp. 16-29.

[4] H. Maghrebi, T. Portigliatti, and E. Prouff, "Breaking cryptographic implementations using deep learning techniques," in Security, Privacy, and Applied Cryptography Engineering, Springer, 2016, pp. 3-26.

[5] L. Wouters, V. Arribas, B. Gierlichs, and B. Preneel, "Revisiting a methodology for efficient CNN architectures in profiling attacks," IACR Transactions on Cryptographic Hardware and Embedded Systems, vol. 2021, no. 3, 2021, pp. 147-168.

[6] L. Masure, C. Dumas, and E. Prouff, "A comprehensive study of deep learning for side-channel analysis," IACR Transactions on Cryptographic Hardware and Embedded Systems, vol. 2022, no. 1, 2022, pp. 348-375.