

Secure Attestation of ESP32-S3 IoT Devices Based on Intrinsic SRAM PUFs Using Machine Learning and Multi-Objective Optimisation

Florent Houetchekpo

Faculty of Computer Science and Mathematics
University of Passau
Passau, Germany
houetc01@ads.uni-passau.de

Nikolaos Athanasios Anagnostopoulos

Faculty of Computer Science and Mathematics
University of Passau
Passau, Germany
anagno02@ads.uni-passau.de
ORCID: 0000-0003-0243-8594

Abstract—Internet of Things (IoT) devices require lightweight authentication mechanisms that remain secure even when devices are physically accessible to adversaries. Traditional approaches based on stored cryptographic keys are vulnerable to extraction attacks. In this work, the feasibility of using Static Random-Access Memory (SRAM)-based Physical Unclonable Functions (PUFs) on ESP32-S3 microcontrollers for device attestation is investigated. SRAM startup behaviour was characterised across five ESP32-S3 devices under room-temperature (20–25°C) and elevated-temperature (30–50°C) conditions, with 230,852 memory addresses systematically evaluated against five quality criteria. Experimental results show an average intra-device Hamming distance of 6.16% and an inter-device Hamming distance of 51.22%, indicating acceptable reliability and strong uniqueness. To improve robustness against noise, a windowed Hamming-weight feature extraction method is proposed that converts 256-bit SRAM responses into compact statistical features. A Logistic Regression (LR) classifier achieved 100% classification accuracy on the five-device dataset across all four evaluated temperature scenarios. Non-dominated Sorting Genetic Algorithm II (NSGA-II) multi-objective optimisation was applied to identify feature subsets that balance classification accuracy and communication efficiency; subsequent Out-Of-Distribution (OOD) evaluation selected a 24-feature operating point. The recommended 24-feature solution maintains 100% classification accuracy and complete impostor rejection on the evaluated dataset at a confidence threshold of 95%, while reducing the feature payload by 25%. Finally, a multi-device attestation protocol is demonstrated using Wi-Fi communication, HMAC-SHA256 authentication, and server-side verification. Experimental deployment with five devices achieved 486 completed attestations over 21 minutes of continuous operation; one temporary Wi-Fi disconnection triggered graceful degradation.

Index Terms—Physical unclonable functions, SRAM PUF, ESP32-S3, IoT security, machine learning, NSGA-II, multi-objective optimisation, attestation.

This work has been partially funded by the German Research Foundation – Deutsche Forschungsgemeinschaft (DFG), under the Young Scientist Projects “Security Applications of Printed Electronics” and “Machine-Learning-Augmented Chaotic Memristive PUFs for Printed-Electronics Security” of the Priority Programme – SchwerpunktProgramm (SPP) 2253: “Nano Security: From Nano-Electronics to Secure Systems”.

I. INTRODUCTION

The proliferation of IoT systems has introduced critical security challenges. Modern deployments consist of billions of resource-constrained devices operating in smart homes, healthcare systems, industrial environments, and agricultural monitoring [1], [2]. Since many of these devices are physically accessible to adversaries, traditional authentication based on stored cryptographic keys is vulnerable to extraction attacks such as cold-boot attacks [3].

Physical Unclonable Functions (PUFs) provide an alternative approach by exploiting microscopic manufacturing variations during semiconductor fabrication to create unique hardware fingerprints that are hard to clone [4], [5]. Among PUF implementations, Static Random-Access Memory (SRAM) PUFs are especially attractive since SRAM memory is already present in most microcontrollers without requiring additional hardware [6], [7]. When SRAM cells power on, each cell settles into a logical 0 or 1 depending on transistor mismatches, producing a device-unique startup pattern.

However, SRAM PUF responses are inherently noisy: repeated power-up measurements may exhibit bit errors due to electrical noise and variations in operating conditions such as temperature and supply voltage. Traditionally, fuzzy extractors with error-correcting codes have been employed to handle this noise [8], but these introduce computational overhead and helper data that may leak information about the PUF response. Machine Learning (ML) classifiers offer an alternative by learning to recognise devices despite noisy fingerprints [9], [10]. When combined with multi-objective optimisation algorithms, such as Non-dominated Sorting Genetic Algorithm II (NSGA-II) [11], which systematically balance accuracy and bandwidth, practical authentication systems can be realised.

Despite extensive research on Field-Programmable Gate Array (FPGA)-based [6] and legacy microcontroller [12] SRAM PUFs, to the best of our knowledge, no prior peer-reviewed study has systematically characterised SRAM PUF behaviour on the ESP32-S3 platform. Furthermore, most existing studies focus on single-device authentication rather than

scalable multi-device attestation, and recent PUF-based swarm proposals such as PISA [13] rely on simulation rather than hardware validation.

This work makes the following four main contributions:

- A systematic characterisation of ESP32-S3 SRAM PUF reliability and uniqueness is provided across 230,852 memory addresses and two temperature conditions.
- A lightweight windowed Hamming-weight feature extraction approach is proposed for noise-tolerant authentication.
- Multi-objective optimisation of feature subsets is performed using NSGA-II with classification error and feature count as the optimisation objectives; the resulting candidate subsets are subsequently evaluated for Out-Of-Distribution (OOD) rejection.
- A multi-device attestation prototype is operationally demonstrated on real hardware with five concurrent ESP32-S3 devices.

II. BACKGROUND AND RELATED WORK

In this section, we briefly examine SRAM PUFs, machine-learning-based PUF authentication methods, multi-objective feature selection, and swarm attestation, which constitute some of the basic building blocks of the proposed secure attestation protocol.

A. SRAM PUFs

PUFs exploit intrinsic manufacturing variations to generate device-unique identifiers without dedicated secure key-storage hardware [14], [15]. SRAM PUFs are classified as weak PUFs due to their limited number of challenge-response pairs. PUF quality is evaluated using two key metrics. *Intra-device Hamming distance* (HD_{intra}) measures reliability, and is defined as the percentage of response bits that differ between repeated measurements from the same device. Ideally, this value is 0%; values up to 10% of the overall response are considered acceptable for Commercial Off-The-Shelf (COTS) devices [16]. *Inter-device Hamming distance* (HD_{inter}) measures uniqueness, and is defined as the percentage of response bits that differ between two different devices for the same challenge. The ideal value is 50% of the overall response, indicating statistically uncorrelated responses [17].

It was demonstrated by Holcomb et al. [7] that SRAM startup patterns on discrete chips achieve $HD_{\text{intra}} \approx 3.57\%$ and $HD_{\text{inter}} \approx 50\%$. FPGA-embedded SRAM was found to be noisier (HD_{intra} of 3–4%) by Guajardo et al. [6], and the first microcontroller characterisation on ARM Cortex-M3 devices was provided by Böhm et al. [12]. Temperature variation is a major source of PUF reliability degradation: Cortez et al. [18] demonstrated that temperature variation significantly affects memory-based PUF response noise. Moreover, very low temperatures can induce data-remanence effects that may be exploited to attack intrinsic SRAM PUFs implemented on COTS devices [19], [20].

B. ML-Based PUF Authentication

Machine learning has been applied both to attack strong PUFs [21], [22] and to authenticate weak PUFs. An authentication accuracy of 99.96% was achieved by Yan et al. [9] using Support Vector Machines (SVMs) while tolerating up to 9.4% bit error rates. DeepPUF was proposed by Najafi et al. [23] using Convolutional Neural Networks (CNNs) for Dynamic Random-Access Memory (DRAM) PUF authentication (94.9% accuracy). More recent work has further demonstrated error-resilient SRAM-PUF authentication using lightweight neural networks [24], [25], as well as probabilistic and ensemble learning approaches for DRAM-PUF authentication [26], [27]. The present work follows this lightweight classifier approach for SRAM PUFs, demonstrating that Logistic Regression (LR) suffices when PUF quality provides adequate device separation.

C. Multi-Objective Feature Selection

NSGA-II [11] is a widely used evolutionary algorithm for multi-objective optimisation that identifies Pareto-optimal trade-offs. While evolutionary algorithms have been applied to cryptographic design problems [28], their application to PUF authentication remains unexplored. This gap is addressed in the present work by applying NSGA-II to simultaneously minimise authentication error and feature count.

D. Swarm Attestation

Lightweight hardware-based security architectures leveraging PUFs have been proposed for heterogeneous systems-of-systems environments [29]. Collective attestation protocols such as SEDA [30], SANA [31], and LISA [32] organise devices into networks for efficient group verification. PISA [13] combined PUF-based authentication with swarm attestation but was validated only through simulation. In the present work, hardware-validated PUF authentication is provided with an operational multi-device protocol on real ESP32-S3 devices.

III. SYSTEM DESIGN AND METHODOLOGY

The proposed system operates in two phases. In the *offline phase*, PUF samples are collected, a classifier is trained, and the feature set is optimised. In the *runtime phase*, repeated device attestation is performed via a challenge-response scheme.

A. SRAM PUF Extraction on ESP32-S3

The ESP32-S3 contains 512 KB of internal SRAM and two 8 KB Real-Time Clock (RTC) memory regions [33]. To identify memory locations suitable for PUF extraction, the accessible 32-bit addresses were evaluated across the relevant SRAM and RTC address regions. Specifically, the evaluation covered 106,496 addresses in Internal SRAM 1 (Data Random-Access Memory), 8,192 addresses in Internal SRAM 2 (Data Random-Access Memory (Data RAM)), 112,068 addresses in the Internal SRAM instruction-bus mapping (Instruction Random-Access Memory (Instruction RAM)), and 2,048 addresses in each of RTC FAST and RTC SLOW memory, giving a total of $106,496 + 8,192 + 112,068 + 2,048 + 2,048 = 230,852$

TABLE I
MEMORY REGION EVALUATION RESULTS

Region	Addresses	Accepted	Rate
Internal SRAM 1 (Data RAM)	106,496	53,881	50.6%
Internal SRAM 2 (Data RAM)	8,192	7,445	90.9%
Internal SRAM (Instruction RAM)	112,068	58,889	52.5%
RTC FAST Memory	2,048	1,784	87.1%
RTC SLOW Memory	2,048	1,890	92.3%
Total	230,852	123,889	53.7%

evaluated 32-bit addresses. The Data RAM and Instruction RAM regions represent different bus mappings of the same physical internal SRAM and therefore should not be interpreted as separate physical SRAM capacities.

Each of the 230,852 addresses was systematically assessed using five PUF quality criteria: non-constant behaviour (at least two unique values), uniformity (average Hamming weight between 20% and 80%), reliability ($HD_{\text{intra}} < 15\%$), uniqueness (HD_{inter} between 45% and 55%), and byte aliasing (no byte remaining constant across all devices). For each address, 50 genuine power-up samples were collected from each of five devices, yielding 250 samples per evaluated address, to evaluate each address as part of the overall PUF response.

Standard bit-level aliasing screening rejected 94% of candidate addresses. With only five devices, this criterion is statistically unstable because even an unbiased bit has a $2(\frac{1}{2})^5 = 6.25\%$ probability of taking the same value across all five devices by chance. Consequently, byte-aliasing was adopted instead. The probability that a random byte takes the same value across all five devices is $(\frac{1}{256})^4 \approx 2.3 \times 10^{-10}$, making chance-induced false rejections negligible. This modification increased the address acceptance rate from 6.0% to 53.7%.

Table I summarises acceptance rates by memory region. RTC memories achieved the highest rates (87–92%). Address selection was performed using the five-device enrollment dataset and was fixed before classifier training and evaluation.

A total of 32 bytes (256 bits) were selected from three regions: Internal SRAM 1 Region A (0x3FCBA374, 10 bytes), Internal SRAM 1 Region B (0x3FCC9DF8, 11 bytes), and RTC FAST (0x600FF38C, 11 bytes). This selection provides physical diversity across power domains.

Extraction timing is critical. SRAM startup values were extracted during the early boot stage using a GCC constructor with priority 101, which captures SRAM contents after minimal C runtime setup but before ESP-IDF (Espressif IoT Development Framework) initialisation overwrites memory. Fresh power-on detection was implemented using a magic value (0xDEADBEEF) in the RTC_NOINIT memory, since the ESP32-S3 preserves SRAM across software resets. Each sample required physical power cycling with a 30-second discharge period.

B. Data Collection

Experiments were conducted using five ESP32-S3-DevKitC-1 N16R8 development boards (chip revision v0.2, dual-core Xtensa LX7 at 240 MHz). Data collection was

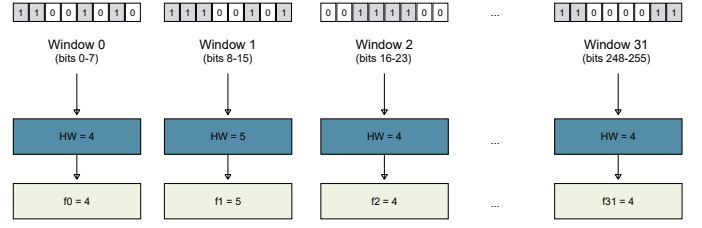


Fig. 1. Windowed Hamming-weight feature extraction. The 256-bit PUF response is partitioned into 32 consecutive 8-bit windows, and the Hamming weight of each window becomes one feature.

performed under two environmental conditions: 50 samples per device were collected at room temperature (20–25°C) and 50 at elevated temperature (30–50°C) in a climate chamber, yielding a total dataset of 500 fingerprint samples.

Authentication was evaluated across four temperature scenarios: Room→Room (baseline), Room→Hot (critical cross-temperature case), Hot→Hot, and Hot→Room. The Room→Hot scenario is considered as the most important for practical deployment, as devices enrolled indoors must authenticate outdoors.

C. Feature Extraction: Windowed Hamming Weight

Rather than using raw 256-bit fingerprints, the PUF response is partitioned into 32 windows of 8 bits and the Hamming weight of each window is computed:

$$\mathbf{f} = [f_0, f_1, \dots, f_{31}] = [HW(b_0 \dots b_7), HW(b_8 \dots b_{15}), \dots, HW(b_{248} \dots b_{255})] \quad (1)$$

where $HW(\cdot)$ counts the Hamming weight, i.e., the number of bits having the logical value 1. For $i \in \{0, \dots, 31\}$, each feature $f_i \in \{0, 1, \dots, 8\}$. As illustrated in Figure 1, this approach provides noise tolerance (a single bit flip changes one feature by ± 1), computational simplicity (under 0.1 ms on ESP32-S3), and privacy preservation (the produced features leak very little information relevant to the original PUF bits).

The Hamming-weight transformation reduces the $2^8 = 256$ possible raw bit patterns of each 8-bit window to only nine possible Hamming-weight values. The resulting nominal feature space for 32 features is therefore $9^{32} \approx 3.4 \times 10^{30}$. This value, however, represents only the theoretical cardinality of the feature space. During PUF-region selection, SRAM addresses whose average Hamming weight fell outside the range of 20%–80% were discarded, thereby excluding strongly biased memory regions. Furthermore, Hamming-weight values are not necessarily uniformly distributed, and correlations may exist between features.

For a more conservative estimate, we assume that each feature effectively occupies only four distinguishable value ranges (low, mid-low, mid-high, and high). This yields a conservative feature space of $4^{32} \approx 1.8 \times 10^{19}$. Assuming statistically independent and uniformly distributed feature vectors, the probability of at least one collision among N devices can be approximated using the birthday bound as $P_{\text{coll}}(N) \approx \frac{\binom{N}{2}}{4^{32}} = \frac{N(N-1)}{2 \cdot 4^{32}}$. This gives $P_{\text{coll}}(5) \approx 5.4 \times 10^{-19}$ for five devices and $P_{\text{coll}}(10^6) \approx 2.7 \times 10^{-8}$ for one million devices. Hence, although the Hamming-weight transformation

substantially compresses the original 256-bit response and the preceding 20–80% Hamming-weight filtering further restricts the responses from which the features are derived, the resulting feature space remains large under these assumptions. Nevertheless, this estimate assumes feature independence, while spatial correlation in weak memory PUFs [34], [35] may reduce the effective feature space and, thus, should be evaluated for the selected ESP32-S3 SRAM regions.

D. Machine Learning Classification

Logistic Regression was employed with L2 regularisation ($C=1.0$), L-BFGS solver, one-vs-rest multi-class strategy, and StandardScaler normalisation. This classifier was selected for its computational efficiency on resource-constrained hardware, interpretability, and a convex optimisation objective. The model outputs class-probability estimates $P(y = k | \mathbf{f})$ for each device class k , which serve as natural confidence scores for authentication decisions. A device is authenticated only if both the predicted class matches the claimed identity and the maximum posterior probability exceeds a predefined confidence threshold τ .

Beyond classification accuracy, Out-Of-Distribution (OOD) detection [36] was evaluated using leave-one-device-out validation. In each fold, the classifier is trained on 4 devices and the held-out device is treated as an unknown impostor. The OOD rejection rate measures the percentage of impostor samples for which the maximum posterior probability falls below τ , correctly triggering rejection. This procedure is critical because in real deployments, the system must not only identify enrolled devices correctly but also reject devices it has never seen before. Nevertheless, because only five devices were available, the OOD evaluation should be considered preliminary and primarily intended to demonstrate feasibility.

During enrollment, labelled PUF feature vectors are used to fit the scaler and train the LR classifier; at runtime, the server applies these stored models to the received features and accepts the claimed identity only if the predicted class matches it and the confidence exceeds τ . The model is fixed after enrollment, and adding new devices requires retraining.

E. NSGA-II Feature Optimisation

Feature selection is formulated as a bi-objective problem:

$$\text{Minimise: } f_1(\mathbf{x}) = 1 - \text{Acc}_{\text{CV}}(\mathbf{x}) \quad (2)$$

$$\text{Minimise: } f_2(\mathbf{x}) = \|\mathbf{x}\|_1 \quad (\text{feature count}) \quad (3)$$

where $\text{Acc}_{\text{CV}}(\mathbf{x})$ is the mean 5-fold cross-validation accuracy obtained using the selected feature subset, and $\|\mathbf{x}\|_1$ denotes the number of selected features. NSGA-II was configured with a population size of 100, for 50 generations, two-point crossover ($p = 0.9$), and bit-flip mutation ($p = 1/32$). Fitness was evaluated via 5-fold cross-validation.

To validate robustness, 8 operator configurations (crossover type, mutation rate, crossover probability) were tested across 30 independent runs each (240 total). Non-parametric bootstrap equivalence testing [37] with Bonferroni correction was

employed, since Shapiro-Wilk tests rejected normality for all configurations.

F. Multi-Device Attestation Protocol

A Wi-Fi station (STA) mode star topology was adopted where each ESP32-S3 connects directly to an access point. The challenge-response protocol proceeds as follows:

- 1) A cryptographic 128-bit nonce with 60-second Time To Live (TTL) is generated by the server.
- 2) PUF features are extracted on-device and a SHA-256 firmware hash is computed.
- 3) The device transmits features, firmware hash, and an HMAC-SHA256 signature computed over all fields using a pre-provisioned 256-bit symmetric key.
- 4) Six verification checks are performed by the server: nonce validity, HMAC verification, firmware hash match, ML classification, confidence threshold (95%), and identity match.

The protocol sequence is illustrated in Figure 2. Individual results are aggregated by the server into a swarm trust status: TRUSTED (all devices pass), DEGRADED (at least one device fails while others pass), COMPROMISED (all devices fail), or UNAVAILABLE (no devices responding). Devices are automatically isolated after three consecutive failures to prevent compromised nodes from disrupting the network. The swarm status provides operators with an immediate overview of fleet health, enabling rapid incident response. Nonce freshness (60-second TTL) prevents replay attacks, while the HMAC-SHA256 signature binds the device identifier, nonce, PUF features, and firmware hash together, preventing selective field manipulation.

IV. EXPERIMENTAL RESULTS

A. PUF Characterisation

PUF quality was evaluated using standard reliability and uniqueness metrics. Table II summarises the results. The average HD_{intra} was measured at 6.16% at room temperature, within the 0–10% range considered acceptable for COTS devices [16]. At elevated temperatures, the noise increased modestly to 6.34%. Only 2.34% of bits (approximately 6 out of 256) were found to exhibit temperature sensitivity across the 20–50°C range.

The average HD_{inter} of 51.22% (51.05% at elevated temperatures) is close to the theoretical ideal of 50%. As shown in Figure 3, HD_{intra} is concentrated at low values, whereas HD_{inter} clusters around the ideal 50%, with clear separation between the distributions. The resulting $8.3\times$ separation indicates that

TABLE II
PUF QUALITY METRICS

Metric	Room Temp	Elevated Temp
Intra-device HD	6.16%	6.34%
Inter-device HD	51.22%	51.05%
Separation ratio	8.3×	
Temp-sensitive bits	2.34% (6/256)	

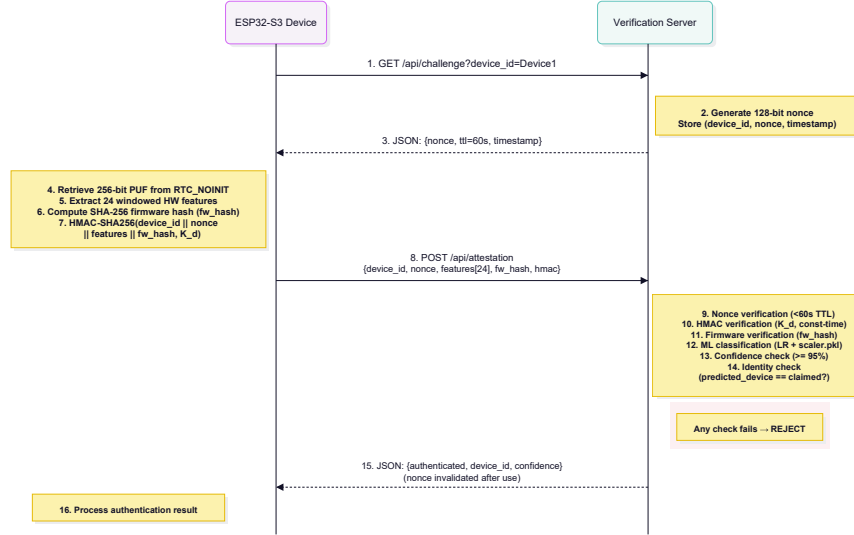


Fig. 2. Challenge-response attestation protocol with six-step server verification including cryptographic authentication, firmware integrity, and ML-based hardware identity verification.

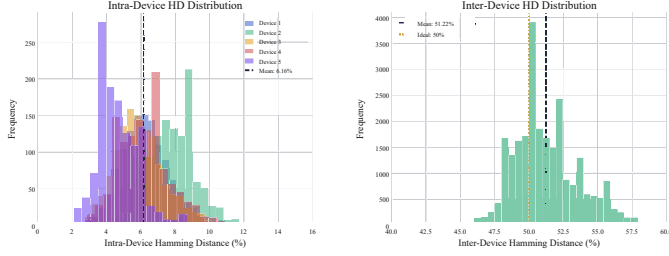


Fig. 3. Hamming distance distributions at room temperature. Left: HD_{intra} (reliability). Right: HD_{inter} (uniqueness). The clear separation confirms suitability for authentication.

within-device noise is substantially smaller than between-device variation, supporting reliable device discrimination under the evaluated conditions.

B. Classification Performance

The LR classifier achieved 100% accuracy on the evaluated dataset in all four temperature scenarios (Table III). The critical Room→Hot scenario, where devices are enrolled indoors and authenticated outdoors, showed zero classification errors. The 95% Confidence Interval (CI), computed using the Wilson score method, quantifies uncertainty in the classification accuracy due to the finite test-set size.

The separability of the device classes is illustrated by Principal Component Analysis (PCA) (Figure 4). The first three principal components, which explain 60.8% of the total variance, reveal five well-separated clusters. The mean temperature-induced displacement of the device centroids in the three-dimensional PCA score space was 0.23, whereas the minimum inter-device centroid distance was 3.54, a dif-

TABLE III
CROSS-TEMPERATURE CLASSIFICATION PERFORMANCE

Scenario	Accuracy	95% CI
Room → Room	100.0%	[98.5%, 100.0%]
Room → Hot	100.0%	[98.5%, 100.0%]
Hot → Hot	100.0%	[98.5%, 100.0%]
Hot → Room	100.0%	[98.5%, 100.0%]

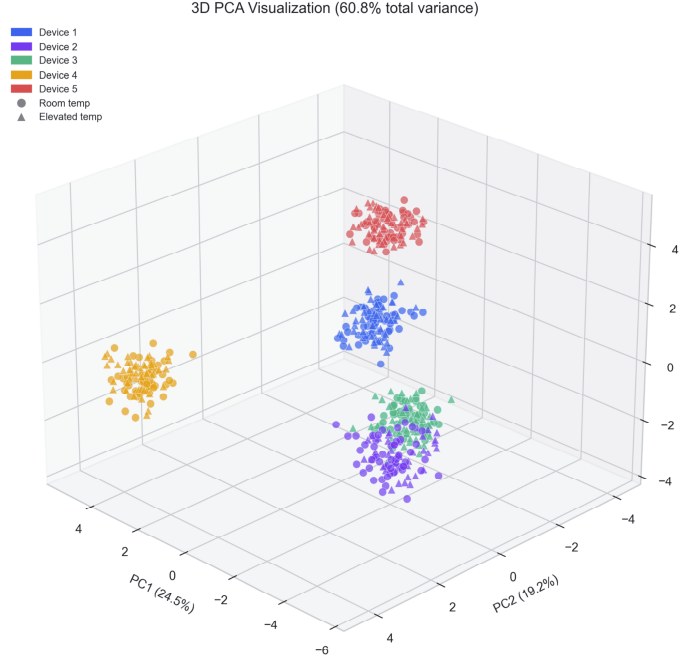


Fig. 4. PCA projection of all 500 samples. Each colour represents a device; circles are room temperature, triangles are elevated temperature. Device clusters are well-separated across both conditions.

ference of approximately $15\times$, indicating that temperature-induced variation is substantially smaller than the separation between devices.

Out-of-distribution detection was evaluated using leave-one-device-out validation, achieving a mean rejection rate of 97.6% at a confidence threshold of $\tau = 0.85$. At the stricter runtime threshold of $\tau = 0.95$, all held-out device samples were successfully rejected.

C. NSGA-II Optimisation

Figure 5 shows the classification performance of representative feature subsets identified during NSGA-II optimisation. Classification accuracy reaches 100% with only three features

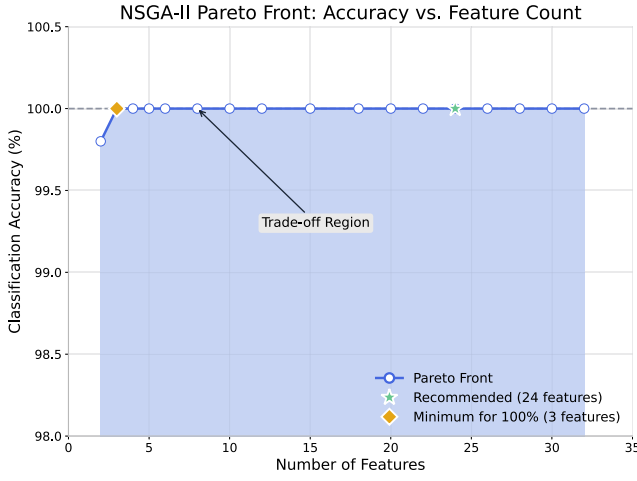


Fig. 5. Classification accuracy of representative feature subsets identified during NSGA-II optimisation as a function of feature count.

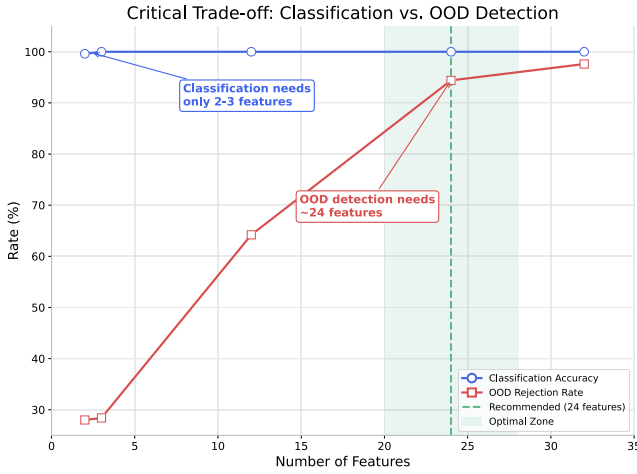


Fig. 6. Classification accuracy and OOD rejection as a function of feature count, highlighting the additional features required for reliable impostor rejection.

and remains saturated for larger subsets. Figure 6 further examines the security implications of this trade-off. Although classification accuracy saturates with only three features, reliable OOD rejection requires substantially more features, motivating the recommended 24-feature solution.

This trade-off is quantified in Table IV. The 2-feature solution achieves 99.6% classification accuracy but only 28% OOD rejection, meaning that 72% of the held-out samples were not rejected by the confidence criterion. The recommended 24-feature solution maintains 100% accuracy and 94.4% OOD rejection at the evaluation threshold ($\tau = 0.85$) (and 100% at $\tau = 0.95$), while reducing the feature payload by 25%.

Robustness was validated across 240 optimisation runs

TABLE IV
NSGA-II SOLUTIONS: CLASSIFICATION VS. OOD DETECTION

Features	Accuracy	OOD Rej.	Feature Payload	Assessment
2	99.6%	28.0%	2 B	Insecure
3	100.0%	28.4%	3 B	Insecure
12	100.0%	64.2%	12 B	Marginal
24	100.0%	94.4%	24 B	Recommended
32	100.0%	97.6%	32 B	Baseline

TABLE V
MULTI-DEVICE ATTESTATION RESULTS (21 MIN)

Device	Successful	Failed	Rate
Device 1	52	0	100%
Device 2	112	0	100%
Device 3	112	0	100%
Device 4	105	0	100%
Device 5	105	0	100%
Total	486	0	100%

(8 configurations $\times$ 30 seeds). Statistical equivalence was confirmed: all 28 pairwise bootstrap equivalence tests passed (Bonferroni-corrected, $\delta = 0.1$), with the largest Cohen's $d = 0.37$. In every run, exactly 3 features were identified as the minimum for perfect classification (standard deviation 0.0), showing reproducibility across all tested runs.

D. Multi-Device Attestation

Five ESP32-S3 devices were operated concurrently for 21 minutes over Wi-Fi using the selected 24-feature solution. The results are presented in Table V. All 486 legitimate attestations passed the six verification steps, corresponding to an observed FRR of 0%. Runtime FAR was not measured because no impostor attempts were included; impostor rejection was evaluated separately using leave-one-device-out validation. Device 1 was attested less frequently (52 vs. 105–112) due to a temporary Wi-Fi disconnection, which triggered a DEGRADED swarm status. Upon reconnection, the swarm recovered to TRUSTED automatically, thereby validating the graceful degradation mechanism. The total bandwidth required per attestation was measured at approximately 177 bytes; for devices attesting hourly, this amounts to around 1.5 MB per device annually.

V. DISCUSSION

The obtained results demonstrate the feasibility of ESP32-S3 SRAM PUF attestation under favourable conditions. The $8.3\times$ separation ratio between HD_{intra} and HD_{inter} provides a substantial margin for classification. In comparison, Holcomb et al. [7] reported $HD_{\text{intra}} \approx 3.57\%$ and $HD_{\text{inter}} \approx 50\%$ on discrete SRAM chips, yielding a separation ratio of approximately $14\times$. The lower separation ratio observed on ESP32-S3 reflects a higher HD_{intra} (6.16% vs. 3.57%), with only 32 bytes extracted compared with Holcomb et al.'s 512 bytes. Nevertheless, the cross-temperature robustness and the absence of failures during the evaluation period (consisting of 486 attestations) are encouraging. However, several limitations constrain generalisability.

Most critically, only five devices were tested. Perfect accuracy on five devices does not guarantee performance on larger populations, where fingerprint clusters may overlap. The collision analysis ($P_{\text{coll}} < 10^{-7}$ at one million devices) is theoretical; empirical large-scale validation (50+ devices from multiple batches) remains essential.

Temperature testing covered 20–50°C, which is suitable for consumer IoT but not for industrial extremes (−40 to +85°C). Laboratory conditions excluded electromagnetic interference,

voltage fluctuations, and humidity. No aging data were collected; transistor drift from Negative-Bias Temperature Instability (NBTI) and Hot Carrier Injection (HCI) could degrade reliability over months [38]. Moreover, experiments on memory-based PUFs have shown that also prolonged thermal stress can alter response reliability and entropy [39], further motivating long-term aging evaluation. An extensive adversarial security evaluation (modelling attacks, side-channel analysis, fault injection) was also not performed.

The trade-off between classification accuracy and OOD detection is considered a key finding with broader implications: evaluating only classification accuracy can produce fundamentally insecure systems. It is recommended that PUF authentication designs include more features than strictly necessary for classification to ensure adequate impostor rejection.

The choice of Logistic Regression (LR) over more complex models (e.g., neural networks [26], random forests [27]) warrants discussion. With only five devices and 32 features, the classification problem is relatively low-dimensional and well-separated. In such settings, simpler models generalise better and are less prone to overfitting. Furthermore, LR provides class-probability estimates that can be used directly as confidence scores for threshold-based OOD rejection without additional post-processing. The model's coefficients are also interpretable, enabling analysis of which features contribute most to device discrimination. Whether LR remains sufficient for larger populations needs to be validated; scaling to hundreds or thousands of devices may require more expressive classifiers.

From a security perspective, the attestation protocol provides protection against replay attacks through nonce freshness, against firmware tampering through hash verification, and against device impersonation through ML-based hardware fingerprinting. However, the reliance on pre-provisioned 256-bit symmetric keys (which can be PUF-based) requires a secure enrollment phase. The protocol does not currently address key revocation or rotation, which would be necessary for production deployment. Finally, adversarial model-extraction and feature-spoofing attacks were not evaluated. Although HMAC verification precedes ML inference and therefore prevents unauthenticated probing of the classifier, attacks based on repeated authorised queries or leaked feature observations remain to be investigated.

VI. CONCLUSION

In this work, SRAM-based device attestation for ESP32-S3 microcontrollers was investigated using machine learning and multi-objective optimisation. Experimental evaluation across 500 fingerprint samples and five devices demonstrated that ESP32-S3 SRAM exhibits strong uniqueness ($HD_{\text{inter}} = 51.22\%$) and acceptable reliability ($HD_{\text{intra}} = 6.16\%$) for authentication purposes. A windowed Hamming-weight feature extraction method was proposed that enabled robust classification despite environmental noise, achieving 100% accuracy across all four temperature scenarios, including the critical cross-temperature case (Room→Hot). NSGA-II optimisation identified that three features provide the best trade-off

between classification accuracy and feature payload efficiency, while post-optimisation OOD evaluation selected 24 features as the operating point. A complete attestation protocol with five concurrent devices achieved 486 successful attestations with zero failures, and graceful degradation under device disconnection was demonstrated.

Future work should address several open challenges. Large-scale validation with 50 or more devices from multiple manufacturing batches is needed to assess scalability and determine at which population size fingerprint clusters begin to overlap. Extended temperature testing across the full industrial range (-40 to $+85^\circ\text{C}$) would establish the boundaries of reliable operation. Longitudinal aging studies lasting 12 months or more are required to quantify fingerprint drift due to transistor aging mechanisms such as NBTI and HCI, and to determine the potentially required re-enrollment frequency. Adversarial security evaluation, including modelling attacks, side-channel analysis, and fault injection, is essential before production deployment. Finally, cross-platform generalisation to other popular microcontrollers (e.g., STM32, nRF52, RP2040) would establish the broader applicability of the proposed methodology.

ACKNOWLEDGMENT

Parts of this work are based on the master's thesis of Florent Houetchequo [40].

REFERENCES

- [1] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of things: A survey on enabling technologies, protocols, and applications," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 4, pp. 2347–2376, 2015. [Online]. Available: <https://doi.org/10.1109/COMST.2015.2444095>
- [2] Statista, "Number of connected Internet of Things (IoT) devices worldwide from 2024 to 2035," 2026. [Online]. Available: <https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/>
- [3] J. A. Halderman, S. D. Schoen, N. Heninger, W. Clarkson, W. Paul, J. A. Calandrino, A. J. Feldman, J. Appelbaum, and E. W. Felten, "Lest we remember: Cold-boot attacks on encryption keys," *Communications of the ACM*, vol. 52, no. 5, pp. 91–98, May 2009. [Online]. Available: <https://doi.org/10.1145/1506409.1506429>
- [4] R. Pappu, B. Recht, J. Taylor, and N. Gershenfeld, "Physical one-way functions," *Science*, vol. 297, no. 5589, pp. 2026–2030, 2002. [Online]. Available: <https://doi.org/10.1126/science.1074376>
- [5] B. Gassend, D. Clarke, M. van Dijk, and S. Devadas, "Silicon physical random functions," in *Proceedings of the 9th ACM Conference on Computer and Communications Security (CCS 2002)*. ACM, 2002, pp. 148–160. [Online]. Available: <https://doi.org/10.1145/586110.586132>
- [6] J. Guajardo, S. S. Kumar, G.-J. Schrijen, and P. Tuyls, "FPGA intrinsic PUFs and their use for IP protection," in *Cryptographic Hardware and Embedded Systems – CHES 2007*, ser. Lecture Notes in Computer Science (LNCS), P. Paillier and I. Verbauwhede, Eds., vol. 4727. Springer, 2007, pp. 63–80. [Online]. Available: https://doi.org/10.1007/978-3-540-74735-2_5
- [7] D. E. Holcomb, W. P. Burleson, and K. Fu, "Power-up SRAM state as an identifying fingerprint and source of true random numbers," *IEEE Transactions on Computers*, vol. 58, no. 9, pp. 1198–1210, 2009. [Online]. Available: <https://doi.org/10.1109/TC.2008.212>
- [8] Y. Dodis, L. Reyzin, and A. Smith, "Fuzzy extractors: How to generate strong keys from biometrics and other noisy data," in *Advances in Cryptology – EUROCRYPT 2004*, ser. Lecture Notes in Computer Science (LNCS), C. Cachin and J. L. Camenisch, Eds., vol. 3027. Springer, 2004, pp. 523–540. [Online]. Available: https://doi.org/10.1007/978-3-540-24676-3_31

- [9] W. Yan, F. Tehranipoor, and J. A. Chandy, "PUF-based fuzzy authentication without error correcting codes," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 36, no. 9, pp. 1445–1457, 2017. [Online]. Available: <https://doi.org/10.1109/TCAD.2016.2638445>
- [10] O. Millwood, J. Miskelly, B. Yang, P. Gope, E. B. Kavun, and C. Lin, "PUF-phenotype: A robust and noise-resilient approach to aid group-based authentication with DRAM-PUFs using machine learning," *IEEE Transactions on Information Forensics and Security*, vol. 18, pp. 2451–2465, 2023. [Online]. Available: <https://doi.org/10.1109/TIFS.2023.3266624>
- [11] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002. [Online]. Available: <https://doi.org/10.1109/4235.996017>
- [12] C. Böhm, M. Hofer, and W. Pribyl, "A microcontroller SRAM-PUF," in *2011 5th International Conference on Network and System Security*. IEEE, 2011, pp. 269–273. [Online]. Available: <https://doi.org/10.1109/ICNSS.2011.6060013>
- [13] S. S. Mehjabin and M. Younis, "PISA: PUF-based IoT swarm attestation protocol," *IEEE Internet of Things Journal*, vol. 12, no. 17, pp. 36 094–36 111, 2025. [Online]. Available: <https://doi.org/10.1109/JIOT.2025.3583269>
- [14] R. Maes, *Physically Unclonable Functions: Constructions, Properties and Applications*. Springer, 2013. [Online]. Available: <https://doi.org/10.1007/978-3-642-41395-7>
- [15] C. Herder, M.-D. Yu, F. Koushanfar, and S. Devadas, "Physical unclonable functions and applications: A tutorial," *Proceedings of the IEEE*, vol. 102, no. 8, pp. 1126–1141, 2014. [Online]. Available: <https://doi.org/10.1109/JPROC.2014.2320516>
- [16] N. A. Anagnostopoulos, "Practical lightweight security: Physical unclonable functions and the internet of things," Ph.D. dissertation, Technical University of Darmstadt, 2022. [Online]. Available: <https://doi.org/10.26083/tuprints-00021494>
- [17] C. Böhm and M. Hofer, *Physical Unclonable Functions in Theory and Practice*. Springer, 2013. [Online]. Available: <https://doi.org/10.1007/978-1-4614-5040-5>
- [18] M. Cortez, S. Hamdioui, V. van der Leest, R. Maes, and G.-J. Schrijen, "Adapting voltage ramp-up time for temperature noise reduction on memory-based PUFs," in *2013 IEEE International Symposium on Hardware-Oriented Security and Trust (HOST)*. IEEE, 2013, pp. 35–40. [Online]. Available: <https://doi.org/10.1109/HST.2013.6581562>
- [19] N. A. Anagnostopoulos, T. Arul, M. Rosenstihl, A. Schaller, S. Gabmeyer, and S. Katzenbeisser, "Attacking SRAM PUFs using very-low-temperature data remanence," *Microprocessors and Microsystems*, vol. 71, p. 102864, 2019. [Online]. Available: <https://doi.org/10.1016/j.micpro.2019.102864>
- [20] —, "Low-temperature data remanence attacks against intrinsic SRAM PUFs," in *2018 21st Euromicro Conference on Digital System Design (DSD)*. IEEE, 2018, pp. 581–585. [Online]. Available: <https://doi.org/10.1109/DSD.2018.00102>
- [21] U. Rührmair, J. Sölter, F. Sehnke, X. Xu, A. Mahmoud, V. Stoyanova, G. Dror, J. Schmidhuber, W. Burleson, and S. Devadas, "PUF modeling attacks on simulated and silicon data," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 11, pp. 1876–1891, 2013. [Online]. Available: <https://doi.org/10.1109/TIFS.2013.2279798>
- [22] G. T. Becker, "The gap between promise and reality: On the insecurity of XOR arbiter PUFs," in *Cryptographic Hardware and Embedded Systems – CHES 2015*, ser. Lecture Notes in Computer Science (LNCS), T. Güneysu and H. Handschuh, Eds., vol. 9293. Springer, 2015, pp. 535–555. [Online]. Available: https://doi.org/10.1007/978-3-662-48324-4_27
- [23] F. Najafi, M. Kaveh, D. Martín, and M. Reza Mosavi, "Deep PUF: A highly reliable DRAM PUF-based authentication for IoT networks using deep convolutional neural networks," *Sensors*, vol. 21, no. 6, 2021. [Online]. Available: <https://doi.org/10.3390/s21062009>
- [24] R. Suragani, E. Nazarenko, N. A. Anagnostopoulos, N. Mexis, and E. B. Kavun, "Identification and classification of corrupted PUF responses via machine learning," in *2022 IEEE International Symposium on Hardware Oriented Security and Trust (HOST)*. IEEE, 2022, pp. 137–140. [Online]. Available: <https://doi.org/10.1109/HOST54066.2022.9839919>
- [25] N. Joshi, N. A. Anagnostopoulos, N. Mexis, S. Neisarian, J. H. Chang, O. Millwood, T. Arul, S. Katzenbeisser, and E. B. Kavun, "Error-resilient PUF-based authentication on IoT edge devices using machine learning," in *2025 IEEE International Conference on Consumer Electronics (ICCE)*. IEEE, 2025, pp. 1–6. [Online]. Available: <https://doi.org/10.1109/ICCE63647.2025.10929847>
- [26] N. Mexis, N. A. Anagnostopoulos, S. Katzenbeisser, E. B. Kavun, S. Tehranipoor, and T. Arul, "Achieving error-free lightweight authentication with DRAM-based physical unclonable functions," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 72, no. 2, pp. 637–646, 2025. [Online]. Available: <https://doi.org/10.1109/TCSI.2024.3480852>
- [27] M. Dudjak, M. Elsharif, S. Neisarian, T. Arul, S. Katzenbeisser, N. A. Anagnostopoulos, and N. Mexis, "Reliable authentication of DRAM PUFs by leveraging random forests in machine learning," in *2025 IEEE 11th World Forum on Internet of Things (WF-IoT)*. IEEE, 2025, pp. 1–7. [Online]. Available: <https://doi.org/10.1109/WF-IoT64238.2025.11270662>
- [28] S. Picek, C. Carlet, S. Guilley, J. F. Miller, and D. Jakobovic, "Evolutionary algorithms for Boolean functions in diverse domains of cryptography," *Evolutionary Computation*, vol. 24, no. 4, pp. 667–694, 2016. [Online]. Available: https://doi.org/10.1162/EVCO_a_00190
- [29] N. Mexis, N. A. Anagnostopoulos, S. Chen, J. Bambach, T. Arul, and S. Katzenbeisser, "A lightweight architecture for hardware-based security in the emerging era of systems of systems," *ACM Journal on Emerging Technologies in Computing Systems*, vol. 17, no. 3, Jun. 2021. [Online]. Available: <https://doi.org/10.1145/3458824>
- [30] N. Asokan, F. Brasser, A. Ibrahim, A.-R. Sadeghi, M. Schunter, G. Tsudik, and C. Wachsmann, "SEDA: Scalable embedded device attestation," in *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security (CCS 2015)*. ACM, 2015, pp. 964–975. [Online]. Available: <https://doi.org/10.1145/2810103.2813670>
- [31] M. Ambrosin, M. Conti, A. Ibrahim, G. Neven, A.-R. Sadeghi, and M. Schunter, "SANA: Secure and scalable aggregate network attestation," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (CCS 2016)*. ACM, 2016, pp. 731–742. [Online]. Available: <https://doi.org/10.1145/2976749.2978335>
- [32] X. Carpent, K. ElDefrawy, N. Rattanavipanon, and G. Tsudik, "Lightweight swarm attestation: A tale of two LISA-s," in *Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security (AsiaCCS 2017)*. ACM, 2017, pp. 86–100. [Online]. Available: <https://doi.org/10.1145/3052973.3053010>
- [33] Espressif Systems, *ESP32-S3 Technical Reference Manual*, 2026, version 1.8. [Online]. Available: https://documentation.espressif.com/esp32-s3_technical_reference_manual_en.pdf
- [34] T. Arul, N. A. Anagnostopoulos, S. Reißig, and S. Katzenbeisser, "A study of the spatial auto-correlation of memory-based physical unclonable functions," in *2020 European Conference on Circuit Theory and Design (ECCTD)*. IEEE, 2020, pp. 1–4. [Online]. Available: <https://doi.org/10.1109/ECCTD49232.2020.9218302>
- [35] N. Mexis, T. Arul, N. A. Anagnostopoulos, F. Frank, S. Böttger, M. Hartmann, S. Hermann, E. B. Kavun, and S. Katzenbeisser, "Spatial correlation in weak physical unclonable functions: A comprehensive overview," in *2023 26th Euromicro Conference on Digital System Design (DSD)*. IEEE, 2023, pp. 70–78. [Online]. Available: <https://doi.org/10.1109/DSD60849.2023.00020>
- [36] D. Hendrycks and K. Gimpel, "A baseline for detecting misclassified and out-of-distribution examples in neural networks," arXiv:1610.02136 [cs.NE], 2018. [Online]. Available: <https://arxiv.org/abs/1610.02136>
- [37] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*, 1st ed. Chapman and Hall/CRC, 1994. [Online]. Available: <https://doi.org/10.1201/9780429246593>
- [38] N. P. Bhatta, H. Singh, A. Ghimire, and F. Amsaad, "Analyzing aging effects on SRAM PUFs: Implications for security and reliability," *Journal of Hardware and Systems Security*, vol. 8, pp. 174–186, 2024. [Online]. Available: <https://doi.org/10.1007/s41635-024-00154-6>
- [39] N. A. Anagnostopoulos, N. Mexis, Y. Fan, S. S. Bartsch, T. Schaiër, S. Katzenbeisser, and T. Arul, "Studying the effects of prolonged thermal stress aiming to induce artificial aging on DRAM retention-based physical unclonable functions," in *2024 IEEE International Conference on Consumer Electronics (ICCE)*. IEEE, 2024, pp. 1–6. [Online]. Available: <https://doi.org/10.1109/ICCE59016.2024.10444494>
- [40] F. Houetcheipo, "Authentication of ESP32-based weak memory PUFs using machine learning and multi-objective evolutionary optimization," Master's thesis, University of Passau, 2026.