Preprints.org

Article

Not peer-reviewed version

Securing IoT Networks Using Machine Learning-Resistant Physical Unclonable Functions (PUFs) on Edge Devices

Abdul Manan Sheikh * , Md. Rafiqul Islam , Mohamed Hadi Habaebi , Suriza Ahmad Zabidi , Athaur Rahman bin Najeeb , Mazhar Baloch

Posted Date: 3 December 2025

doi: 10.20944/preprints202512.0284.v1

Keywords: cryptography; IoT; PUFs; CRPs; machine learning

Article

Securing IoT Networks Using Machine Learning-Resistant Physical Unclonable Functions (PUFs) on Edge Devices

Abdul Manan Sheikh [1,2,*], Md. Rafiqul Islam [2], Mohamed Hadi Habaebi [2], Suriza Ahmad Zabidi [2], Athaur Rahman bin Najeeb [2] and Mazhar Baloch [1]

[1] Department of Electrical Engineering and Computer Science, A'Sharqiyah University, Oman
[2] Department of Electrical and Computer Engineering, International Islamic University Malaysia, Kuala Lumpur, Malaysia
* Correspondence: abdul.manan@asu.edu.om

Abstract

The Internet of Things (IoT) has transformed global connectivity by linking people, smart devices, and data. However, as the number of connected devices continues to grow, ensuring secure data transmission and communication has become increasingly challenging. IoT security threats arise at the device level due to limited computing resources, mobility, and the large diversity of devices, as well as at the network level, where the use of varied protocols by different vendors introduces further vulnerabilities. Physical Unclonable Functions (PUFs) provide a lightweight, hardware-based security primitive that exploits inherent device-specific variations to ensure uniqueness, unpredictability, and enhanced protection of data and user privacy. Additionally, modeling attacks against PUF architectures is difficult to execute due to the random and unpredictable physical variations inherent in their design, making it nearly impossible for attackers to accurately replicate their unique responses. This study collected approximately 80,000 Challenge Response Pairs (CRPs) from a Ring Oscillator (RO) PUF design to evaluate its resilience against modeling attacks. The predictive performance of five machine learning algorithms, i.e., Support Vector Machines, Logistic Regression, Artificial Neural Networks with a Multilayer Perceptron, K-Nearest Neighbors, and Gradient Boosting, was analyzed, and the results showed an average accuracy of approximately 60%, demonstrating the strong resistance of the RO PUF to these attacks. The NIST statistical test suite was applied to the CRP data of the RO PUF to evaluate its randomness quality. The p-values from the 15 statistical tests confirm that the CRP data exhibit true randomness, with most values exceeding the 0.01 threshold and supporting the null hypothesis of randomness.

Keywords: cryptography; IoT; PUFs; CRPs; machine learning

---

1. Introduction

The Internet of Things (IoT) has revolutionized remote monitoring and control of systems by enabling the processing of real-time data from numerous sensing devices. Cisco estimated that approximately 21.3 billion IoT devices were deployed in 2022. Intelligent interfaces allow these devices to interact, collect, communicate, and store data efficiently. A trusted ecosystem built on the principles of authentication, authorization, privacy, confidentiality, availability, and integrity is therefore essential to ensure secure data transactions [2,3]. IoT devices are often constrained by limited storage, processing, sensing, and computational resources [4]. Moreover, the diversity of devices from multiple vendors or manufacturers deployed within IoT networks makes them particularly vulnerable to security threats, with device authentication emerging as a critical factor in maintaining trustworthy communication. Authenticating edge devices is as important as authenticating users to ensure that only legitimate devices gain access to IoT resources. Weakly secured devices can compromise the entire system, leading to severe financial and reputational consequences. The heterogeneous nature of IoT

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 December 2025 doi:10.20944/preprints202512.0284.v1

networks, combined with the resource limitations of edge devices, makes traditional authentication schemes impractical. These constraints underscore the need for innovative, lightweight authentication mechanisms tailored specifically to IoT environments [5].

IoT devices operating at the network edge remain highly vulnerable to cyber threats, including data breaches, hardware tampering, and denial-of-service attacks. Traditional cloud-based, centrally managed infrastructures often struggle to counter advanced risks, including Distributed Denial of Service (DDoS) attacks, Man in the Middle (MiTM) intrusions, eavesdropping, and attacks powered by artificial intelligence [6]. To mitigate these risks more effectively, computational capabilities can be shifted closer to the network edge, enabling real-time threat detection and mitigation directly on devices rather than relying solely on remote cloud infrastructure. Cloud-assisted IoT takes advantage of the large storage capacity and powerful computing resources available in cloud platforms to process and analyze data coming from IoT devices. However, when enormous amounts of data are continuously sent from widely distributed IoT nodes, cloud servers experience processing delays, which leads to higher latency and slower response times for the services delivered to network users.

Edge Computing (EC) enables local processing and storage of sensitive data, thereby preserving privacy and facilitating faster access to critical information during security investigations [7]. In addition, EC offers advantages such as reduced latency, improved bandwidth efficiency, and enhanced privacy and security [8]. The EC architecture, shown in Figure 1, is typically organized into three layers: the edge devices layer, the edge nodes or gateways layer, and the cloud. Edge devices, such as IoT sensors, smartphones, cameras, or industrial machines, are primarily responsible for generating or collecting data. Due to limited processing capabilities, these devices usually perform only basic functions such as data filtering or preliminary processing. At the edge node layer, intermediate devices like routers, edge servers, or micro data centers perform more advanced processing, aggregate data that is further complicated by device heterogeneity and resource limitations, the cloud. The cloud layer, in turn, handles complex computations, long-term data storage, and advanced analytics that exceed the capabilities of edge nodes. Connectivity among edge devices, edge nodes, and the cloud is supported by WiFi, 5G, or wired infrastructure to ensure efficient data transmission. Edge-driven IoT systems distribute heterogeneous resources across the network to achieve flexibility and scalability, supporting applications such as smart cities and autonomous vehicles. However, edge servers and devices remain constrained by limited power and processing capacity, and offloading large volumes of data and tasks can overload the network.

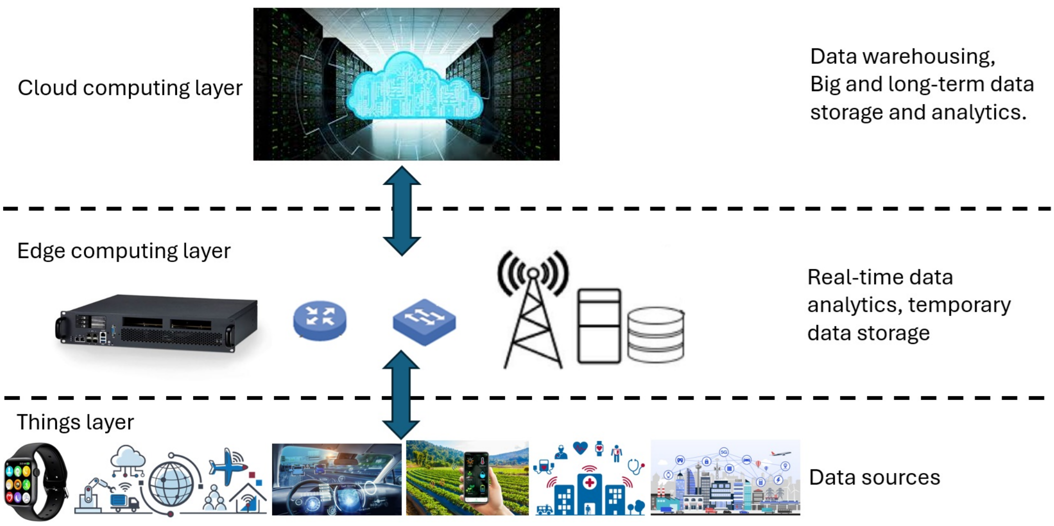


Figure 1. Edge computing architecture

 doi:10.20944/preprints202512.0284.v1

1.1. Motivations

The integration of EC with IoT introduces challenges due to their inherent differences. IoT systems encompass a diverse range of hardware platforms and communication protocols, and their success depends on a unified framework that ensures seamless interoperability between IoT devices and edge nodes. Security and privacy remain primary concerns, complicated further by device heterogeneity and resource limitations such as memory and battery power [9]. As a result, edge devices and servers are highly vulnerable to attacks, and a breach at any point can compromise the entire network. To address these risks, lightweight and robust security mechanisms tailored to the distributed and heterogeneous nature of IoT systems are essential. Several strategies have been proposed to mitigate security challenges in EC. Conventional measures include intrusion detection systems (IDS) to monitor malicious activity, strong access control to regulate permissions, encryption to protect data at rest and in transit, and authentication mechanisms to verify user and device identities. Beyond these methods, advanced technologies such as artificial intelligence (AI) and machine learning (ML) are increasingly used for real-time anomaly detection and adaptive defense against evolving threats [10]. Blockchain utilizes machine learning methods to learn and replicate the behavior of a PUF with high accuracy, thereby undermining sharing, and ensuring transaction integrity [11,12]. In parallel, Physical Unclonable Functions (PUFs) have garnered attention as hardware-based security primitives that provide device authentication, secure key generation, and cryptographic support. PUFs exploit manufacturing variations to generate unique device-specific identifiers that safeguard against cloning and other hardware-based attacks. Field Programmable Gate Arrays (FPGAs), known for their flexibility, reconfigurability, and rapid prototyping, are well-suited for implementing PUFs [13]. Moreover, PUFs can be integrated with intellectual property (IP) cores in FPGA-based systems, making them practical for real-world deployments [14,15].

Meanwhile, ML techniques such as Support Vector Machines, Logistic Regression, and Deep Neural Networks are increasingly used in hardware security to detect hardware Trojans, identify counterfeit integrated circuits (ICs), and evaluate the reliability of PUF [16]. At the edge, ML models are also applied to analyze data streams and detect anomalies, intrusions, and malicious activities in real time. However, ML models deployed on resource-constrained edge devices are themselves vulnerable to adversarial threats, including poisoning attacks that corrupt training data, evasion attacks that deceive models with crafted inputs, inference attacks that extract sensitive information, and exploratory attacks that exploit model weaknesses or replicate behavior. ML-based modeling attacks use machine learning methods to learn and reproduce the behavior of a PUF with very high accuracy, which can undermine its security. An adversary can construct a numerical model of the PUF by obtaining a portion of its challenge response pairs through eavesdropping or any other form of unauthorized access [18]. These risks are exacerbated by the limited storage and computational capabilities of edge devices. To mitigate such vulnerabilities, integrating FPGA-based PUFs with ML frameworks offers a promising approach to strengthening edge security [17]. PUF-derived keys can be employed to encrypt ML models or authenticate devices contributing training data, protecting against tampering and data leakage. Leveraging FPGAs for PUF design provides flexibility, rapid prototyping, and reconfigurability, while seamless integration with other IP cores enables robust and scalable security solutions [19].

1.2. Research Contributions

This research presents a comprehensive investigation into the design, implementation, and security evaluation of a Ring Oscillator Physical Unclonable Function (RO PUF) on an FPGA, with emphasis on machine learning (ML)-based modeling, resistance, and statistical randomness analysis. The main contributions of this work are summarized as follows:

- *FPGA-Based Design and Implementation of RO PUF:* The study successfully implemented a configurable RO PUF architecture on an FPGA platform. A dedicated testbench was developed to

doi:10.20944/preprints202512.0284.v1

acquire large-scale challenge–response datasets under controlled operating conditions, verifying the reproducibility and uniqueness of the PUF behavior across multiple FPGA instances.

- *Comprehensive Evaluation of PUF Metrics:* The implemented RO PUF was analyzed using standard performance metrics such as uniformity, uniqueness, and reliability (intra-Hamming distance). The proposed design demonstrated balanced uniformity near the ideal 50%
- *Machine Learning-Based Attack and Accuracy Estimation:* To assess the resilience of the proposed RO PUF against modeling attacks, various ML algorithms, including Logistic Regression (LR), Support Vector Machine (SVM), Multilayer Perceptron (MLP), and K-Nearest Neighbor (KNN), were employed. Confusion matrix analysis revealed that linear models, such as LR, failed to capture the nonlinear challenge–response relationship, while nonlinear models (SVM, MLP, and KNN) performed moderately better but exhibited trade-offs between precision and recall. The findings confirmed that none of the models achieved strong predictive accuracy, highlighting the robustness and unpredictability of the proposed RO PUF against ML-based cloning attempts.
- *Randomness Validation Using NIST SP 800-22 Tests:* The randomness quality of the generated CRP responses was validated using the NIST statistical test suite, covering tests such as frequency, runs, block frequency, and cumulative sums. The majority of the tests yielded p-values greater than 0.01, confirming that the PUF outputs exhibit strong statistical randomness and are suitable for cryptographic and authentication applications.

2. Related Works

Shen et al. proposed an integrated security framework for EC that combines ML and cryptographic techniques to monitor and detect abnormal activities on the network. The study provides valuable insights into the use of Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Long Short-Term Memory (LSTM) models for time series prediction, evaluated using performance metrics such as Precision, Recall, and F1-Score [20]. In the Collaborative Edge Computing (CEC) model, the edge layer handles data storage and processing in a distributed manner. Given the limited processing capacity of individual edge devices, they cooperate to offload tasks among themselves, a process known as load balancing. The researchers in [21] employed Physically Unclonable Functions (PUFs) to authenticate edge devices during load balancing, eliminating the need to store a database of Challenge-Response Pairs (CRPs) locally. Cheng et al. integrated blockchain technology with certificateless cryptography, elliptic curve cryptography, and pseudonym-based cryptography to enable mutual authentication between edge servers and IoT devices [22]. The study in [23] proposes a privacy-preserving edge computing approach that utilizes federated learning to train a unified deep learning model across multiple end users collaboratively. Instead of sharing raw data, only model parameters (i.e., gradients) are exchanged. The parameters from local deep learning models on various edge nodes are aggregated at the edge and then distributed to all participants. Through several iterations of local training and parameter aggregation, a deep learning model is developed that preserves user privacy without the need to share raw data. Zhang and colleagues introduced a configurable tristate PUF that can operate as an arbiter PUF, a ring oscillator PUF, or a bistable ring PUF. The design uses a bitwise XOR-based obfuscation mechanism to hide the relationship between challenges and responses. As a result, machine learning models fail to build an accurate prediction model, with all attack accuracies remaining around 50% to 60%, which is effectively the same as random guessing [24]. The authors of [25] introduced an authentication method that employs arbiter PUFs together with three protection strategies called challenge splitting, challenge scrambling, and challenge padding. Each strategy disrupts the structure or visibility of the challenge so that machine learning models cannot form an accurate numerical model of the PUF. In [26], a hybrid secure deduplication scheme is introduced that ensures data privacy on the server side while enhancing network performance on the client side. Additionally, an additive homomorphic encryption method is proposed to enable efficient deduplication operations on resource-constrained edge nodes. Intrusion detection systems (IDS) are vital for the security of IoT systems, as they detect traces of known attacks and unusual behavior in IoT

 doi:10.20944/preprints202512.0284.v1

devices and their connecting networks [27]. However, the majority of IDS algorithms use conventional encryption and cryptographic techniques. Unfortunately, these approaches are vulnerable to physical attacks, as they primarily rely on storing secret keys in the device's local memory. Blockchain is considered an ideal approach to store the huge amount of data generated by IoT devices with utmost privacy and security using a distributed ledger. However, the resource-constrained IoT devices can't meet the computational requirements of data mining in blockchain networks. Sarkar et. al. have proposed a robust PUF-based authentication system to replace the popular consensus algorithms [28].

A PUF design based on Generalized Galois Ring Oscillators (GenGARO) has been implemented on Artix-7 FPGAs using both 11-LUT and 3-LUT configurations. The proposed design demonstrates strong resilience against various ML models, including Decision Trees (DT), Random Forests (RF), K-Nearest Neighbors (KNN), Support Vector Machines (SVM), and Multi-Layer Perceptrons (MLP) [29]. Another approach, the Structure-Obfuscated PUF (SO-PUF), is built on a configurable ring oscillator (CRO) framework and dynamically removes NOT gates at each stage based on the challenge input. As reported in [30], SO-PUF was implemented on an Xilinx Spartan-6 FPGA and achieved near-random prediction accuracy. Kareem et al. [31] investigated the vulnerability of three RO-PUF designs to ML attacks and evaluated their prediction accuracy using models such as DT, RF, KNN, and SVM. Abulibdeh et al. introduced the Algorithmically Optimized Configurable Ring Oscillator PUF (AOCRO), which demonstrated strong resistance to machine learning modeling attacks. In their evaluation, five different ML algorithms (SVM, MLP, LR, CNN, and CMA ES) were trained on a dataset of one million challenge response pairs (CRPs), achieving an average prediction accuracy of 61.3%, indicating enhanced robustness compared to conventional CRO PUF designs [32]. Jack Miskelly et al. investigated the impact of ML attacks on Configurable RO PUFs. The study simulated 128-bit CRO PUFs and multi-PUF variants with datasets ranging from 1,000 to 10,000 CRPs. The results showed that conventional CRO PUFs could be accurately modeled using an LR, ML model, achieving prediction accuracies exceeding 98–99% [33]. Laguduva et al. proposed a non-invasive, architecture-independent attack on PUFs using challenge response pairs (CRPs). This method achieved a cloning accuracy of 93.5% without requiring any prior knowledge of the PUF's internal architecture [34]. A summary of the results obtained by above mentioned researchers is shown in Table 1.

Table 1. ML model accuracy against RO PUFs

Article	KNN	SVM	MLP	RF	DT	LR
[29]	56.76 – 63.24	49.76 - 69.71	50.10 -70.86	62.67 – 75.81	57.90 – 72.00	
[30]		49.75	63.13			49.63
[31]	74.6	66.2		72.3	64.6	
[35]		56.64	58.04			51.9
[36]		58.59 – 61.95				

3. Physical Unclonable Functions (PUFS)

Researchers have identified PUFs as robust security primitives that can guarantee the three pillars of security, namely confidentiality, authenticity, and privacy, of IoT data. PUFs extract unique information from the physical characteristics of the IoT device [37]. PUF-based authentication protocols enhance the security of resource-constrained edge devices without requiring them to store credentials in their limited non-volatile memory. Physically Unclonable Functions (PUFs) utilize the inherent random variations introduced during manufacturing to generate secret keys dynamically. They provide essential security functions such as authentication and secret key generation, especially in resource-constrained environments like the Internet of Things (IoT) [38]. The inherent unclonability of PUFs stems from numerous uncontrollable random parameters created during fabrication. When a PUF receives an input, known as a challenge (C), it produces a corresponding output response (R). Figure 2 illustrates that physical variations in the fabrication of integrated circuits (ICs) can result in two ICs yielding different responses to the same challenge. This relationship between the input and output is referred to as a challenge–response pair (CRP) [64]. Traditional authentication methods rely on storing secret credentials in a device's memory, which makes them unsuitable for physically unprotected IoT devices. Attackers can exploit physical vulnerabilities to compromise the entire system. PUFs mitigate such risks in two key ways: first, they generate volatile secrets that are not stored in digital memory but are intrinsically embedded in the hardware structure; second, the uniqueness of each PUF enables it to serve as a unique identifier for individual IoT devices [54].

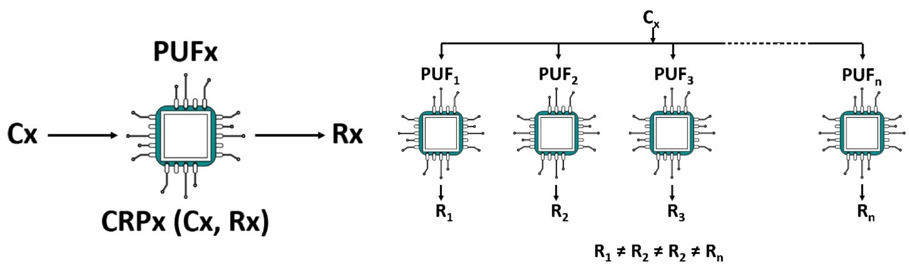


Figure 2. PUFs Challenge & Response Pair (CRP) generation [55]

3.1. PUFs Classification

PUFs can be placed into two groups based on the number of CRPs they can generate, i.e., strong and weak. The limited number of CRPs in a weak PUF are typically proportional to the number of components used in its construction. In contrast, strong PUFs provide a vast number of CRPs, making polynomial-time attacks computationally impractical [56]. Strong PUFs are typically employed in authentication and key establishment protocols, whereas weak PUFs are mainly used for identification and secure key storage applications [57]. Examples of strong PUFs include the Arbiter PUF, the XOR Arbiter PUF, and the Lightweight PUF. In contrast, weak PUFs are represented by designs such as the SRAM PUF, Ring Oscillator (RO) PUF, Anderson PUF, Memristor PUF, Thyristor PUF, and One-Time Programmable (OTP) PUF [58]. Additionally, PUFs are silicon-based and non-silicon PUFs, depending on the manufacturing process adopted. Silicon PUFs count on fabrication mismatches inherent in integrated circuits and can be further divided into delay-based PUFs and memory-based PUFs. In contrast, non-silicon PUFs are based on physical irregularities in systems composed of non-electronic components. Memory-based PUFs utilizes the initial binary states of memory upon power-up, whereas delay-based PUFs exploit variations in signal propagation delays within circuits.

3.2. RO PUFs

Due to inherent random variations in the manufacturing process, two similar ROs do not generate identical oscillation frequencies. The frequency differences between selected RO pairs form the output response of the PUF. An RO PUF consists of an odd number of NOT logic gates arranged in a ring, causing the output to oscillate between logic '1' and '0' at a specific frequency [59]. The basic architecture of an RO PUF, as shown in Figure 3, includes an odd number of inverter gates and an AND gate to enable or disable the feedback loop. The conventional RO-PUF consists of several essential components: n ring oscillators (ROs), two n-to-1 multiplexers (MUXs), two counters, and a comparator circuit. The outputs from each RO are routed to the inputs of both MUXs, whose selected outputs serve as clock signals for the counters. Each counter increments based on the oscillation frequency of the specific RO chosen by its respective MUX. Finally, the comparator compares the values stored in the two counters to produce the RO-PUF's response, corresponding to the applied challenge inputs. The frequency of each RO depends on the delay of the inverters, which is influenced by variations in the manufacturing process.

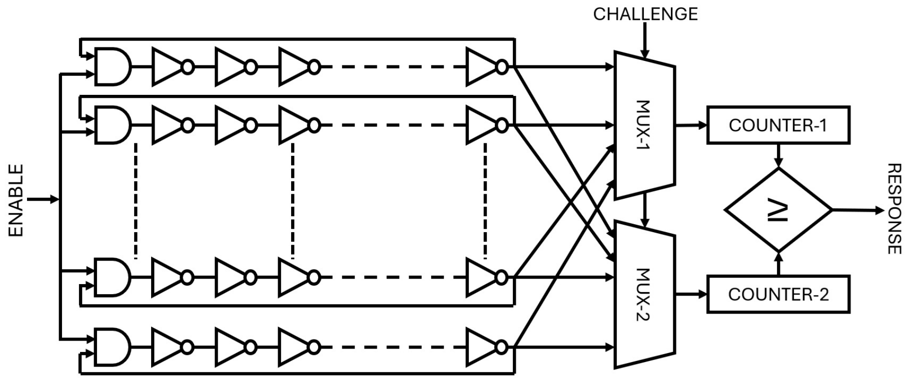


Figure 3. Ring Oscillator (RO) PUF

Algorithm 1 Ring Oscillator PUF (RO-PUF) Algorithm

Require: Challenge input $C = \{Sel\_A, Sel\_B\}$, Time window T
Ensure: Response bit R

```
Initialize N Ring Oscillators: RO[0], RO[1], ..., RO[N − 1]
Initialize two N-to-1 MUXs, two Counters: Counter_A, Counter_B, and a Comparator
Select RO_A = RO[Sel_A] via MUX A
Select RO_B = RO[Sel_B] via MUX B
Connect RO_A output to Counter_A clock
Connect RO_B output to Counter_B clock
Enable Counter_A and Counter_B for fixed time window T
During T, increment counters on each RO oscillation pulse
After T, disable both counters
if Counter_A > Counter_B then
    R ← 1
else
    R ← 0
end if
return R
```

The oscillation frequency of the RO PUF, as shown in eqn. 1 is inversely proportional to both the odd number of gates (n) and their average propagation delay (t_{pd}). Consequently, f_{osc} is sensitive to inherent variations in gate delay, which causes each instance of the RO structure to exhibit a slightly different oscillation frequency [60].

$$f_{osc} = \frac{1}{2nt_{pd}} \tag{1}$$

The delay of each not gate is modelled by eqn. 2, where μ_{pd} is the nominal delay and δ_k is a zero-mean random deviation, often approximated by gaussian, $\mathcal{N}(0, \sigma_{pd}^2)$.

$$t_{pd,k} = \mu_{pd} + \delta_k \tag{2}$$

The fundamental operation of the proposed RO PUF is outlined in algorithm 1. Initially, multiple ROs are instantiated, each consisting of an odd number of inverters connected in a closed loop chain, causing continuous oscillation due to intrinsic gate delays. For every response bit, a corresponding challenge input selects a pair of ROs through multiplexers. The oscillation frequencies of two selected ROs are recorded over a fixed time interval using counters. If the first RO exhibits a higher count, indicating a higher oscillating frequency, a response bit of '1' is generated; otherwise, a '0' is assigned.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 December 2025 doi:10.20944/preprints202512.0284.v1

3.3. PUF Performance Metrices

PUF performance metrics serve as key indicators of functional behavior and security robustness. Hamming distance (HD) serves as a critical metric to measure the degree of dissimilarity between two responses generated by a PUF. HD is further distinguished between intra-PUF and inter-PUF comparisons. The intra-PUF HD is an indicator of the dissimilarity between the responses of a single PUF, highlighting the internal consistency or variability of the PUF. On the other hand, the inter-PUF Hamming distance compares the responses between two different PUFs, offering a gauge of the uniqueness and distinguishability of each PUF's responses. According to the classification presented by Pahlevi et al., the evaluation framework can be grouped into three main categories. The first category, conventional PUF evaluations, includes metrics such as uniqueness, uniformity, and reliability. These metrics assess how distinguishable the responses are between different devices, how evenly distributed the output bits appear, and how consistently the PUF can reproduce the same response under varying environmental or operational conditions. The second category focuses on authentication-oriented metrics, which evaluate the practical suitability of PUF for real-world security applications. Metrics such as the False Acceptance Rate (FAR) and False Rejection Rate (FRR) measure authentication accuracy, while bit aliasing and the Bit Error Rate (BER) capture bit-level stability and possible device bias. Entropy estimation is also included in this category to quantify the randomness present in the response set. The third category comprises machine learning based attack evaluations, which determine how well the PUF can withstand modern modeling attacks aimed at predicting its behavior [62].

The performance of access control mechanisms is evaluated through four fundamental metrics derived from the confusion matrix are defined in equations 3, 4, 5, and 6 are false acceptance rate (FAR), the false rejection rate (FRR), the true acceptance rate (TAR), and the true rejection rate (TRR). The confusion matrix serves as a security evaluation metrics for PUFs, to assess the effectiveness and dependability of an authentication system. FAR indicates instances where unauthorized individuals are mistakenly granted access, while FRR reflects cases where legitimate users are unfairly denied access, highlighting crucial error dimensions that must be minimized [63].

$$\text{FRR} = \frac{\text{Number of False Rejections}}{\text{Total Number of Genuine Attempts}} \times 100\% \tag{3}$$

$$\text{FAR} = \frac{\text{Number of False Acceptances}}{\text{Total Number of Impostor Attempts}} \times 100\% \tag{4}$$

$$\text{TRR} = \frac{\text{Number of True Rejections}}{\text{Total Number of Impostor Attempts}} \times 100\% \tag{5}$$

$$\text{TAR} = \frac{\text{Number of True Acceptances}}{\text{Total Number of Genuine Attempts}} \times 100\% \tag{6}$$

The use of a confusion matrix to evaluate an authentication system highlights its importance in measuring how effectively the system differentiates between legitimate access attempts and potential security breaches.

- Uniqueness: It is used to quantify how different devices respond to the same input challenge. In other words, it is defined as the inter device Hamming Distance (HD) between different devices and its ideal value is 50%. The HD of equation 7 estimates the uniqueness of CRPs, where n is the total number of devices, R_i and R_j are the respective responses of the i^{th} and j^{th} devices under the same challenge, $HD(\cdot,\cdot)$ is the Hamming Distance operator, and m is the bit length of each response.

$$\text{Uniqueness} = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{HD(R_i, R_j)}{m} \tag{7}$$

- Uniformity: It is the probability that the 0s and 1s are uniformly distributed in PUF's response. Uniformity measures the balance between zeros and ones in the responses generated by a PUF. It is obtained by computing the average Hamming weight of the responses, as expressed in equation 8.

$$\text{Uniformity} = \frac{1}{k}\sum_{l=1}^{k} R_l \tag{8}$$

- Reliability: It indicates the ability of a PUF to reproduce the same response bit for a given challenge input even when environmental conditions, such as supply voltage and temperature, vary. An ideal reliability close to 100% means that no bit flips occur across repeated measurements. However, achieving perfect reliability is difficult because PUF outputs are inherently sensitive to these variations.

$$HD_{\text{intra}}(R, R') = \sum_{b=1}^{m} \left| R_b - R'_b \right| \tag{9}$$

  A standard metric for reliability is the intra-class Hamming Distance, is illustrated in equation 9, where R and R' are two responses from the same device under the same challenge.
- Bit aliasing: It complements uniqueness and uniformity by verifying whether a given bit exhibits enough variation. Ideally, each bit appears randomly as 0 or 1 with a typical value of 50%, signifying minimal bias. The aliasing factor for the b^{th} bit is represented in equation 10, where $R_{i,b}$ is the b^{th} bit of the i^{th} device's response.

$$\text{Aliasing}(b) = \frac{1}{n}\sum_{i=1}^{n} R_{i,b} \tag{10}$$

- Bit Error Rate (BER): A BER defined in equation 11, gives an estimates of how often a PUF produces incorrect or flipped bits when the same challenge input is applied multiple times under varying environmental conditions such as temperature and supply voltage.

$$\text{BER} = \frac{\text{Number of flipped bits}}{\text{Total bits measured}} \tag{11}$$

- Entropy: It is used to evaluate the overall randomness of PUF outputs, particularly against advanced modeling attacks and side channel attacks. A higher entropy value reflects a larger and more unpredictable response space. The most widely used metric for this purpose is the Shannon entropy, defined as follows in equation 12:

$$H(R) = -\sum_{\omega} p(\omega)\log_2 p(\omega) \tag{12}$$

  where ω represents every possible output pattern, and $p(\omega)$ denotes the probability associated with each pattern [64].

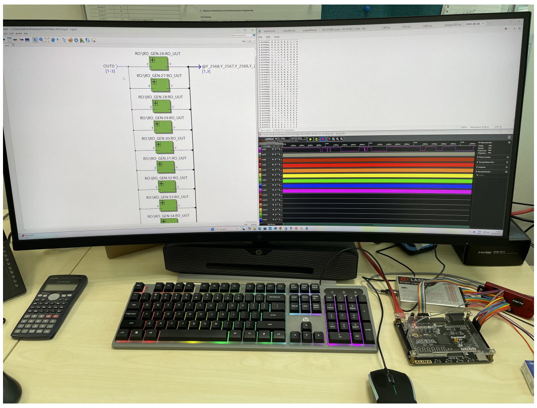
3.4. ML Modelling Attacks

Resistance and reliability to the ML attacks are two major concerns for the overall viability of the PUF-based authentication protocols [42]. ML-based attacks on PUFs can be catalogued as semi-invasive attacks, as adversaries exploit the PUF by intercepting the communication channel between the PUF client and server. The data is then preprocessed, and a parametric numerical model is built using ML algorithms that can successfully predict the PUF responses [40,41]. Adversaries' carryout PUF modeling attacks using deep learning, LR, support vector machine, and evolution strategy, assuming access to CRPs [43]. Countermeasures against ML attacks on PUFs are typically based on increasing the nonlinearity and complexity of the PUF model or integrating additional modules that enable more complex operating modes and communication protocols with the verifier [44].

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 December 2025 doi:10.20944/preprints202512.0284.v1

4. Experimental Setup and Implementation

The experimental setup and synthesized netlist statistics of the proposed RO-PUF is shown in Figure 4. The proposed architecture is implemented and verified using Cyclone IVE FPGA chips (EP4CE10F17C8) in the AX4010 development board. Quartus Prime 18.1 Lite Edition is used to develop and verify the RO-PUF design before configuring the FPGA device. The Cyclone IV E EP4CE10F17C8 is a low-cost, low-power FPGA developed by Intel. It features 10,320 logic elements, 414 Kbits of embedded memory, 179 user I/O pins, and four PLLs for flexible clock management, all packaged in a 256-pin FBGA. The RO-PUF design, implemented in VHDL, is interfaced with the onboard GPIO pins of the AX4010 development board. All control signals, including challenge inputs and response outputs, are connected to a 16-channel logic analyzer through these GPIO pins. The design incorporates 512 ring oscillators (ROs) comprising five inverters. The outputs of the ROs are routed to two sets of 256-to-1 multiplexers (MUXes). Each MUX selects one RO output based on the challenge input, determining which RO signal is passed to the output. The final response is generated by comparing the frequencies of two counters, Counter_A and Counter_B. An 8-bit Linear Feedback Shift Register (LFSR) is used to generate the challenge signals, based on the characteristic polynomial $P(x) = x^8 + x^4 + 1$.



(**a**) Experimental setup

Parameters	#
Total logic elements	1,071 / 10,320 (10 %)
Total registers	40
Total pins	12 / 180 (7 %)
Max LUT depth	8.00
f_{max}	94.23 MHz
Worst case slack	9.388 ns
Nominal core voltage	1.2 V
Low junction temperature	0° C
High junction temperature	85° C

(**b**) Post synthesis netlist statistics

Figure 4. FPGA implementation of proposed RO-PUF

The ML models, such as SVM, LR, MLP, and KNN, require careful parameter tuning to achieve optimal performance. For example, as illustrated in Table **??**, SVM depends on the choice of kernel and gamma values, LR relies on regularization strength and iteration limits, MLP's performance is influenced by hidden layer sizes, learning rates, and solver selection, while KNN requires choosing the appropriate number of neighbors and distance metrics. The training process involves preparing the PUF CRP dataset by separating the features from the responses and then splitting the data into training and testing sets for supervised learning. Validation and evaluation are carried out using metrics such as accuracy and confusion matrices to assess model performance. While train-test splits provide an initial measure of effectiveness, applying cross-validation techniques can further enhance reliability by reducing variance and ensuring the models generalize well to unseen CRPs.

4.1. NIST Randomness Test

Randomness plays a fundamental role in cryptography, as the strength of many security mechanisms relies on the unpredictability of generated sequences. However, generating truly random numbers is inherently challenging, and equally important is the rigorous evaluation of the quality of the generated data. To assess randomness, statistical tests are commonly employed, which yield a p-value. The p-value quantifies the probability that a truly random number generator would produce a sequence with less apparent randomness than the sequence being evaluated. In other words, it provides a statistical measure of how well the tested sequence aligns with the characteristics of an ideal random source.

Most empirical randomness tests, such as those included in the National Institute of Standards and Technology (NIST) Statistical Test Suite (STS), are built on the principles of statistical hypothesis testing. The NIST STS consists of 15 well-defined tests that evaluate binary sequences for signs of non-randomness. These tests analyze both local and global properties of the data. At the local level, they assess features such as the balance between zeros and ones, or the frequency of specific bit patterns within smaller segments of the sequence. At the global level, they examine broader statistical behavior across the entire bitstream to determine overall randomness. To further refine detection, the bitstream is often partitioned into multiple large segments, where each segment is analyzed independently. The results from these segments are then aggregated into final test statistics, which help to identify localized irregularities or systematic deviations from randomness. This multi-layered evaluation ensures that both subtle and significant weaknesses in the sequence can be detected, providing a comprehensive measure of its suitability for cryptographic applications.

In the context of the NIST STS, the interpretation of p-values is crucial. For each test, a significance level (commonly set at 0.01) is defined. If the p-value obtained from a sequence is greater than or equal to 0.01, the sequence is considered to have passed that particular randomness test, indicating no strong evidence of non-random behavior. Conversely, a p-value below 0.01 suggests that the sequence may deviate significantly from randomness. Ideally, when a large number of independent sequences are tested, the distribution of p-values across all tests should be uniform within the interval [0,1]. This uniformity demonstrates that the data behaves consistently with what is expected from a true random source. Therefore, both the proportion of sequences passing each test and the uniformity of their p-value distribution are essential criteria in validating the randomness of cryptographic data.

Each NIST STS test is defined by a specific test statistic, which falls into one of three categories:

- *Bits:* Analyzes characteristics such as proportion of bits, frequency of bit changes, and cumulative sums.
- *m-bit blocks:* Analyzes distribution of m-bit blocks ($m < 30$) within the sequence or its parts.
- *M-bit parts:* Analyzes complex properties of M-bit parts ($M > 1000$), such as matrix rank, sequence spectrum, or linear complexity.

Most tests are parameterized by n (sequence length) and may include a second parameter m or M, depending on the test. Table 2 summarizes the number of sub-tests included in the NIST STS suite. Notably, the non-overlapping template matching test has a variable number of sub-tests determined by m [61].

Table 2. Recommended bitstream parameters

#	Name of the Test	n	M or m	Sub-Test #
1	Frequency	$n \geq 100$	–	1
2	Frequency within a block	$n \geq 100$	$20 \leq M \leq n/100$	1
3	Runs	$n \geq 100$	–	1
4	Longest run of ones	$n \geq 128$	–	1
5	Rank	$n > 38912$	–	1
6	Spectral	$n \geq 1000$	–	1
7	Non-overlapping Template Matching	$n \geq 8m - 8$	$2 \leq m \leq 21$	148
8	Overlapping Template Matching	$n \geq 10^6$	–	1
9	Maurer's Universal	$n > 387840$	–	1
10	Linear Complexity	$n \geq 10^6$	$500 \leq M \leq 5000$	1
11	Serial	–	$2 < m < [\log_2 n] - 2$	2
12	Approximate Entropy	–	$m < [\log_2 n] - 5$	1
13	Cumulative Sums	$n \geq 100$	–	2
14	Random Excursions	$n \geq 10^6$	–	8
15	Random Excursions variant	$n \geq 10^6$	–	18

4.1.1. Brief Description of NIST Randomness Tests

- *Frequency (Monobit) test:* Checks whether the number of ones and zeros are approximately equal.
- *Frequency within a block test:* Evaluates proportion of zeros and ones in M-bit blocks; expected frequency of ones is $M/2$.
- *Runs test:* Measures consecutive runs of zeros and ones; checks if transitions occur at expected frequencies.
- *Longest run of ones in a block test:* Examines if the longest run of ones (and zeros) in M-bit blocks matches the expected distribution.
- *Random binary matrix rank test:* Evaluates the rank of sub-matrices to detect linear dependencies in the sequence.
- *Discrete Fourier Transform (Spectral) test:* Detects periodic features using DFT peak heights.
- *Non-overlapping template matching test:* Detects excessive occurrences of aperiodic m-bit patterns using a sliding window that resets after each match.
- *Overlapping template matching test:* Counts occurrences of target substrings; window slides by one bit to allow overlaps.
- *Maurer's Universal Statistical test:* Measures compressibility of the sequence; overly compressible sequences indicate non-randomness.
- *Linear complexity test:* Estimates the length of the feedback register required to reproduce the sequence; shorter lengths indicate predictability.
- *Serial test:* Examines frequency of all overlapping m-bit patterns.
- *Approximate Entropy test:* Compares frequencies of overlapping m-bit and $(m+1)$-bit patterns to detect regularity.
- *Cumulative Sum (Cusum) test:* Evaluates maximal deviation from zero in the cumulative sum of bits mapped to $\{-1, +1\}$.
- *Random Excursions test:* Measures the number of cycles with exactly K visits in cumulative sum random walks.
- *Random Excursions Variant test:* Analyzes frequency of visits to specific states in cumulative sum random walks to detect non-random patterns.

4.1.2. Calibration

Calibration plays a critical role in ensuring the accuracy, reliability, and reproducibility of CRP acquisition when implementing PUFs on FPGAs. Since PUF responses are highly sensitive to environmental conditions such as temperature, voltage fluctuations, and aging, as well as measurement artifacts like timing misalignment, noise, and signal distortions, calibration methods are employed to minimize errors before feeding the data into ML-based security models.

- *Timing Calibration:* Logic analyzers capture digital signals at high sampling rates ranging from hundreds of MHz to several GHz. However, any misalignment between challenge signals, response outputs, and control clocks can result in corrupted CRPs. The calibration strategies are discussed below.
  - i) Use a known reference signal, such as the FPGA internal clock or a test pattern generator, to align acquisition channels.
  - ii) Apply trigger-based synchronization in the logic analyzer to ensure consistent alignment of challenge vectors with corresponding responses.
- *Voltage and Signal Level Calibration:*FPGA output signals may degrade due to voltage drop, temperature variations, or I/O mismatches. This can cause the logic analyzer to misinterpret logical '0' and '1' levels. The calibration techniques adopted are,
  - i) Adjust threshold voltage levels on the logic analyzer to match FPGA I/O standards (e.g., LVTTL, LVCMOS).

   ii) Periodically recalibrate using known test vectors to verify that digital transitions are accurately captured.

- *Environmental Calibration:* PUF responses are known to vary with temperature, supply voltage, and device aging. Environmental calibration ensures that CRPs remain stable and consistent under varying conditions. Calibration strategies include,
   i) Use environmental profiling, where CRPs are collected across controlled temperature and voltage ranges, followed by applying corrective models.
   ii) Apply ML-based preprocessing such as normalization or majority voting to compensate for environmental drift.
- *Noise Filtering and Signal Cleaning:*High-frequency noise or transient glitches can distort CRP acquisition and lead to unstable datasets. Techniques used in noise filtering and signal conditioning are,
   i) Apply digital filtering techniques (e.g., glitch removal, debouncing) during data preprocessing.
   ii) Perform repeated measurements followed by majority voting to ensure transient noise does not bias the dataset.
- *Data Alignment and Synchronization:* During multi-channel CRP acquisition, timing skew between channels can lead to incorrect challenge-response mapping. The calibration methods used for data alignment and synchronization include,
   i) Perform multi-channel skew calibration by applying the same known signal to all acquisition channels and adjusting offsets accordingly.
   ii) Use post-processing alignment algorithms to re-synchronize challenge-response mapping before ML training.
- *Statistical Calibration for ML Training:* Before feeding CRPs into machine learning models, statistical calibration ensures data integrity and uniformity for reliable analysis. The calibration techniques include,
   i) Compute intra-class and inter-class metrics to evaluate the reliability and uniqueness of CRPs.
   ii) Apply whitening techniques (e.g., Linear Feedback Shift Register (LFSR) or hash-based methods) to eliminate bias in raw PUF data.
   iii) Normalize datasets to prevent ML models from being influenced by imbalanced or skewed response distributions.

5. Results and Discussion

PUFs exploit the devices inherent physical randomness to generate unique and repeatable responses corresponding to challenge inputs. The key metric for measuring randomness, i.e., the balance between 0s and 1s in the output, is the relative frequency of bit 1 across all generated responses. This frequency provides a mathematical means to assess how uniformly bit 1 appears, which is critical in determining the PUF's effectiveness for security purposes. In security applications, high unpredictability and an even distribution of binary values are essential. As expressed in the equation 13, relative frequency is calculated by dividing the total number of 1s by the overall number of response bits, thus an absolute indicator of the PUF's randomness, and thereby its reliability and security suitability.

$$p = \frac{1}{KL} \sum_{k=1}^{K} \sum_{l=1}^{L} b_{k,l} \tag{13}$$

where p and K represent the relative frequency of bit 1 in all responses and the total number of responses, while L denotes the response length. In contrast, k and l refer to the response k^{th} and the position l^{th} bit in the response output, respectively.

 doi:10.20944/preprints202512.0284.v1

$$H = -\log_2 \max(p, 1-p) \tag{14}$$

Table 3 lists the uniqueness and uniformity values, which fall within the range of published RO PUF designs, but its reliability is noticeably lower than that of most prior works. However, lower reliability strengthen resilience against ML attacks as it introduces label noise into the challenge response pairs, making the PUF's input–output mapping harder to learn.

Table 3. Comparison of RO PUF metrics reported in the literature.

Reference	Uniqueness	Uniformity	Reliability	Description
[45]	47.64%, 45.15%	49.8%, 48%	98.5%, 96%	RO PUF using three and five stage oscillators on Artix seven FPGA with XOR and inverter logic.
[46]	50.1%	49.45%	98.33%	CLU-based design using XOR and XNOR to create a low hardware CRO PUF.
[47]	49.23%	49.76%	98.05%	A lightweight configurable RO PUF that combines RRAM with CMOS inverters.
[51]	48.64%	46.78%	86%	Strong RO PUF (BST RPUF) designed for improved CRP count and stable responses.
[49]	49.2%	49.8%	97.6%	RO PUF design used as a hardware security primitive for IoT applications.
[50]	44.46%, 47.33%, 47.48%	59.61%, 60.62%, 62.89%	97.96%, 98.09%, 99.16%	Uses one hundred RO blocks with five, eleven, and twenty stages for response generation.
[51]	48.64%	46.78%	BER < 10^{-9}	Highly reliable BST RPUF robust against ML-based modeling attacks.
This work	49.1%	56.86%	60.13%	A configurable RO PUF architecture is implemented on an FPGA platform and its resilience against ML attacks is measured.

To assess the randomness of a bit sequence, the equation 14 defines H as the minimum entropy anticipated from PUF outputs to exhibit uniform distribution. H peaks at 1 when $p = 0.5$ and hits its lowest at 0 when $p = 0$ or $p = 1$. The best p is 0.5 for a binary system, as it produces the maximum entropy of 1 bit and represents the maximum uncertainty or randomness in a system. It also provides the strongest unpredictability and ensures resilience against cloning and prediction. The bitwise probability 'p' and entropy 'H' are plotted against varying CRPs in Figure 5. This study also conducts a vulnerability assessment of the proposed FPGA-based RO-PUF design against machine learning (ML) modeling attacks. The challenge-response pair (CRP) data is split into training and testing sets in an 80% to 20% ratio to build the attack model. This setup enables a clear and meaningful evaluation of the prediction accuracy of various ML models, with CRP lengths scaling up to 80,000. The graph shown in Figure 6 illustrates that the security of the PUF decreases as more CRP become available. LR and MLP achieve the highest prediction accuracy, meaning they pose the most effective modeling threat, while the SVM also shows moderate attack capability. In contrast, kNN performs poorly with low and unstable accuracy, indicating strong resistance against such attacks. Accuracy values rise with increasing challenge response pairs and peak around thirty-two thousand to thirty-three thousand, after which improvements saturate or slightly decline. Since lower accuracy corresponds to stronger

 doi:10.20944/preprints202512.0284.v1

security, the physical unclonable function remains most robust against attacks with limited challenge response pairs and when weaker models like kNN are used.

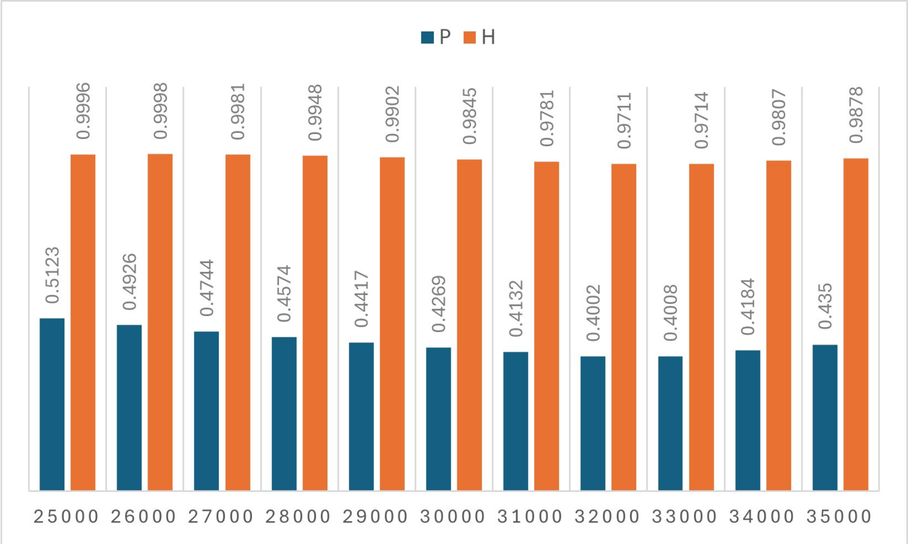


Figure 5. Bitwise probability (p) and entropy (H) against CRP numbers

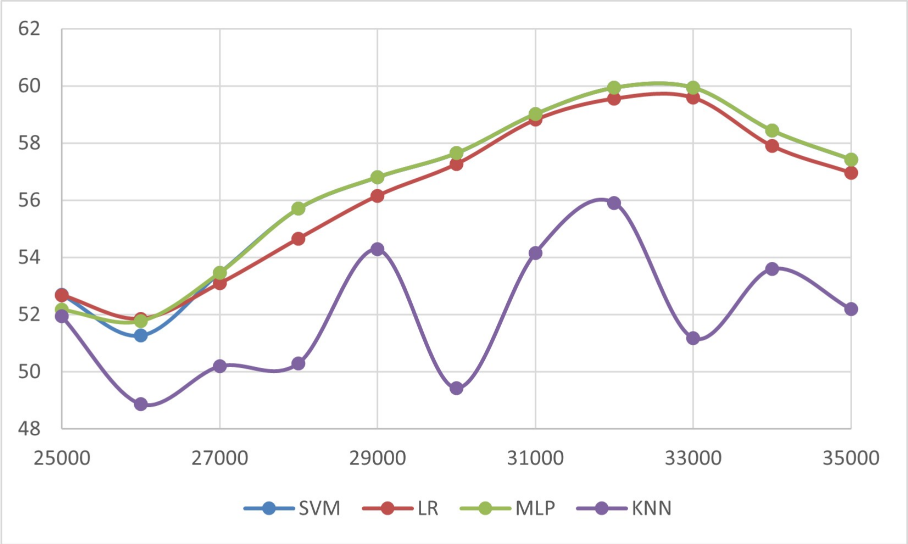


Figure 6. ML accuracy versus varying number of CRP's

Table 4. ML model accuracy against varying CRPs

Algorithm	25K	26K	27K	28K	29K	30K	31K	32K	33K	34K	35K
SVM	52.7	51.27	53.46	55.71	56.81	57.65	59.02	59.94	59.94	58.44	57.43
LR	52.68	51.85	53.09	54.66	56.16	57.27	58.82	59.56	59.91	57.9	56.96
MLP	52.17	51.77	53.46	55.71	56.81	57.65	59.02	59.94	59.94	58.44	57.43
KNN	51.94	48.87	50.19	50.30	54.28	49.43	54.16	55.91	51.18	53.6	52.19

The confusion matrix shown in Figure 7 evaluates the performance of a classification model. The top-left cell represents true positives (TP), indicating the positive instances correctly identified by the model. The bottom-left cell shows false positives (FP), which are negative cases incorrectly classified as positive, and also known as type I errors. The top-right cell displays the number of false negatives (FN), referring to positive instances that were mistakenly predicted as negative. Finally, the bottom-right cell indicates true negatives (TN), where the model correctly identified negative instances. The sum of these four values gives the total number of predictions made by the model. Metrics such as accuracy, precision, recall, and F1-score can be derived from the confusion matrix to provide deeper insight into the model's performance. Equation 15 presents an estimation of model accuracy based on the values in the confusion matrix.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{15}$$

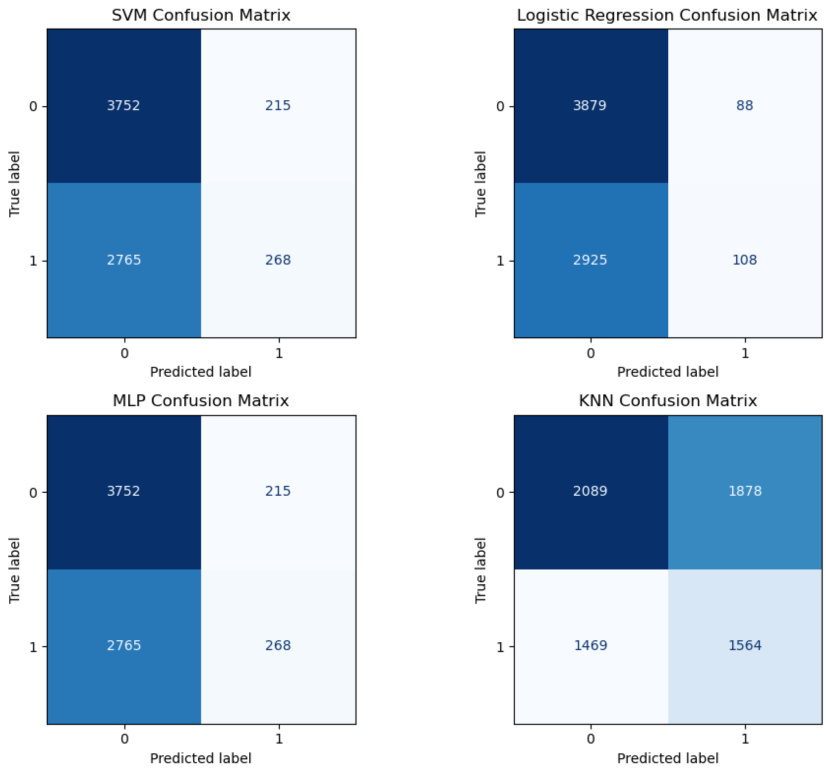


Figure 7. Confusion matrix of ML models (trained using 35K CRPs)

Modeling attacks is becoming a significant security concern, aiming to imitate the challenge–response behavior of PUFs and replicate their secret authentication keys. In the context of lightweight PUF-based systems, these attacks are generally classified into three main types: (a) machine learning (ML)-based software attacks, (b) side-channel attacks, and (c) hybrid attacks that combine ML techniques with side-channel analysis. A typical generic model, typically polynomial, is used in ML-based modeling attacks for approximating CRP datasets. Adversaries divide the compromised CRP dataset into training and testing subsets. An iteration of inputting a challenge into the developed PUF model and its predicted response is carried out. The error between the predicted and actual response estimates a loss function; subsequently, the model parameters are updated using optimization strategies such as gradient descent in logistic regression (LR) or maximum likelihood estimation in support vector machines (SVMs). Finally, the model's ability to replicate the actual PUF behavior is validated using the test set [65]. However, the effectiveness and precision of ML-based attacks decline as the structural complexity of the PUF increases. In particular, strong PUFs, such as Ring Oscillator (RO) and Arbiter PUFs, that incorporate a high degree of nonlinear logic are more resistant to such modeling efforts.

5.1. Randomness Results

Table 5 summarizes the results of the NIST randomness tests conducted on the RO PUF CRP data. The P-value represents the likelihood that the observed bit sequence could have been produced by an ideal random source under the null hypothesis. In general, a sequence is considered random if its P-value exceeds the predefined significance level, typically 0.01, whereas values below this threshold indicate potential non-random behavior. It is important to emphasize that no single P-value alone confirms perfect randomness. Instead, a robust PUF should yield P-values that are uniformly distributed between 0 and 1 across various tests and sequences. Such a uniform distribution, along with the majority of P-values surpassing the 0.01 threshold, indicates that the CRP data exhibit statistical characteristics consistent with true randomness. Thus, it is concluded that FPGA-based RO-PUF can effectively resist ML modeling attacks, as shown in Table 4. The ML models perform no better than random guessing (50% accuracy for binary responses) and the RO-PUF successfully prevents learning patterns, as the highest accuracy achieved is approximately 60%.

Table 5. RO PUF NIST statistical test results

Tests	p-values
Block Frequency	0.444570
Cumulative sums	0.907298
FFT	0.561658
Frequency Test	0.583604
Runs	0.677681
Longest run of ones	0.164698
Rank	0.945607
Non overlapping Template matching	0.51082702 (Average)
Overlapping template matching	0.711526
Universal statistical	0.829717
Approximate entropy	0.120839
Random excursions	0.332701
Random excursions variant	0.447202222
Serial test	0.2013255
Linear complexity	0.420000

5.2. Correlation Matrix

The correlation matrix of the RO PUF CRP data in Figure 8 illustrates the degree of dependency between different output bits generated by the RO PUF. Each cell in the matrix represents a correlation coefficient between two response bits, ranging from –1 to +1. The diagonal elements show perfect correlation (value of 1), indicating that each bit is fully correlated with itself.

The colors transition from red along the diagonal to blue in the off-diagonal regions, reflecting varying levels of inter-bit relationships. From the heatmap, it can be observed that adjacent bits exhibit moderate correlation, while distant bits tend to be weakly correlated or nearly independent. This pattern suggests that neighboring ring oscillators may share certain environmental or structural influences, such as local routing, power distribution, or temperature variations, which introduce partial dependency among nearby bits. However, as the distance between bits increases, the influence diminishes, and their correlation approaches zero. The overall low inter-bit correlation indicates that the RO PUF produces outputs that are largely independent and random, fulfilling one of the key requirements of a strong PUF design. The slight correlation observed between adjacent bits is expected in practical implementations due to physical proximity effects. To further enhance independence, post-processing techniques such as XOR-based bit mixing or improved oscillator placement strategies can be employed. In summary, the correlation matrix demonstrates that the RO PUF achieves good randomness and uniqueness characteristics, making it suitable for secure identification and authentication in hardware security applications.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 December 2025 doi:10.20944/preprints202512.0284.v1

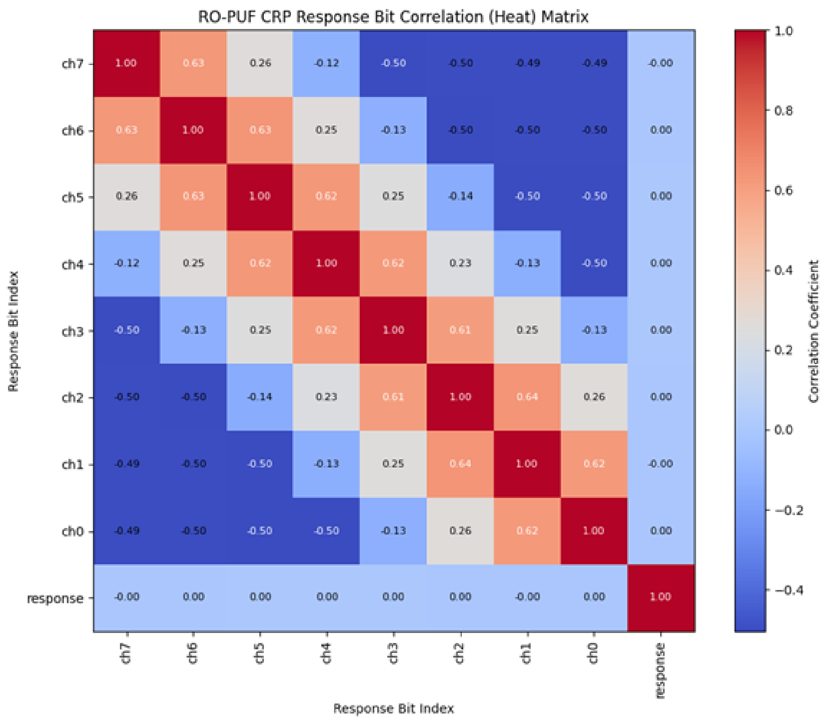


Figure 8. RO PUF Correlation Matrix

5.3. *Receiver Operating Characteristic (ROC)*

The ROC analysis for RO-PUF in Figure 9, shows that all models—Logistic regression, Random Forest, Gradient Boosting, SVM, and MLP—produced Area Under the Curve (AUC) values around 0.49, with ROC curves nearly overlapping the diagonal reference line. This behavior indicates that none of the models could distinguish the RO-PUF responses from random guessing, confirming its strong resistance to ML-based prediction. The intrinsic randomness and frequency-based variations of the oscillators introduce complex, nonlinear behavior that conventional ML algorithms cannot capture effectively.

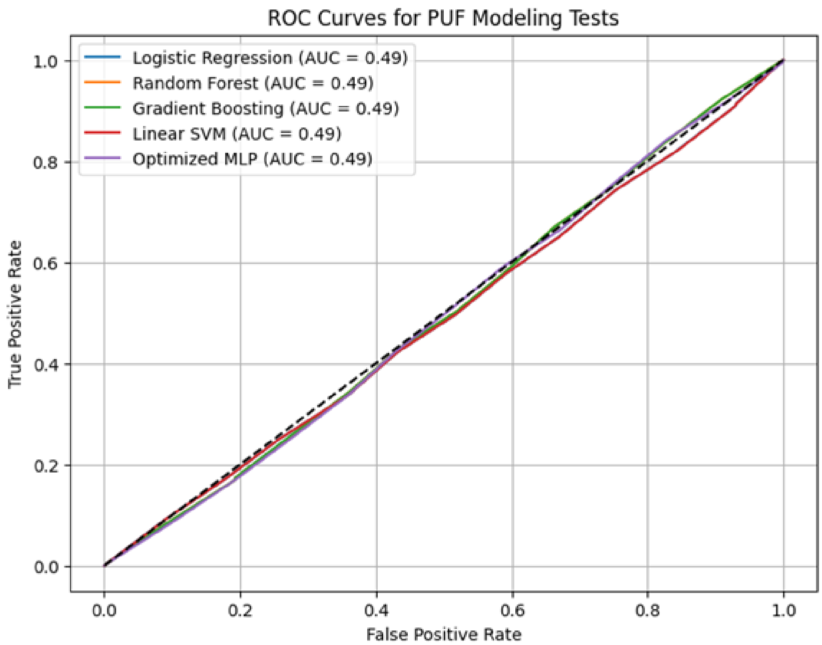


Figure 9. Receiver Operating Characteristic of RO PUF

 doi:10.20944/preprints202512.0284.v1

6. Conclusion

Physically Unclonable Functions (PUFs) offer a cost-effective and reliable method for authenticating IoT devices due to their unique physical characteristics and ease of implementation. Nevertheless, despite being labeled as unclonable, PUFs can be vulnerable to modeling attacks if an adversary gains access to a portion of their challenge–response pairs (CRPs). In this work, we present a machine learning (ML)-resistant strong Ring Oscillator PUF (RO-PUF) architecture implemented on an FPGA. The design utilizes 512 RO chains, each consisting of five inverters, and employs an eight-bit challenge to generate a response bit. Experimental results demonstrate that the proposed RO-PUF significantly improves resilience against ML-based modeling attacks, effectively hindering adversaries from constructing accurate predictive models.

Author Contributions: Conceptualization, A.M.S., M.R.I. and M.H.H.; methodology, A.M.S.; software, A.M.S. and M.H.H.; validation, A.M.S., M.R.I. and M.H.H.; formal analysis, A.M.S., M.H.H., S.A.Z., A.R. and A.K.; investigation, A.M.S. and M.R.I.; resources, M.R.I., M.H.H. and A.K.; data curation, A.M.S. and M.R.I.; writing—original draft preparation, A.M.S; writing—review and editing, M.R.I., M.H.H., S.A.Z., A.R. and A.K.; visualization, M.H.H., S.A.Z. and A.R.; supervision, M.R.I. and M.H.H.; project administration, A.K.; funding acquisition, A.M.S. and A.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by A'Sharqiyah University, Oman-Internal Research Grant (IRG-16), 2024-26, "Intrusion detection in an IoT network through Machine Learning (ML) of hardware characteristics" and "The APC was funded by A'Sharqiyah University, Oman.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
AUC	Area Under the Curve
CEC	Collaborative Edge Computing
CRPs	Challenge Response Pairs
DDoS	Distributed Denial of Service
DT	Decision Trees
EC	Edge Computing
FN	False Negatives
FP	False Positives
FPGAs	Field Programmable Gate Arrays
ICs	Integrated Circuits
IDS	Intrusion Detection Systems
IoT	Internet of Things
IP	Intellectual Property
KNN	K-Nearest Neighbor
LR	Logistic Regression
LSTM	Long Short-Term Memory
MiTM	Man in the Middle
ML	Machine Learning
MLP	Multilayer Perceptron
NIST	National Institute of Standards and Technology
PUF	Physical Unclonable Function
RF	Random Forests
RO	Ring Oscillator
STS	Statistical Test Suite

SVM Support Vector Machine
TN True Negatives
TP True Positives

References

1. Albreem, M. A., Sheikh, A. M., Alsharif, M. H., Jusoh, M., and Yasin, M. N. M., "Green Internet of Things (GIoT): applications, practices, awareness, and challenges," *IEEE Access*, vol. 9, pp. 38833–38858, 2021.
2. Kokila, M, and Srinivasa Reddy, "Authentication, access control and scalability models in internet of things security-A review," *Cyber Security and Applications*, pp. 100057, 2024.
3. Albreem, M. A., Sheikh, A. M., Bashir, M. J., and El-Saleh, A. A., "Towards green Internet of Things (IoT) for a sustainable future in Gulf Cooperation Council countries: Current practices, challenges and future prospective", *Wireless Networks*, vol. 29(2), pp. 539–567, 2023.
4. Gupta, Divya, Shalli Rani, Saleem Raza, Nawab Muhammad Faseeh Qureshi, Romany F Mansour, and Mahmoud Ragab, "Security paradigm for remote health monitoring edge devices in internet of things," *Journal of King Saud University-Computer and Information Sciences*, vol. 35, no. 6, pp. 101478, 2023.
5. Badhib, Arwa, Suhair Alshehri, and Asma Cherif, "A robust device-to-device continuous authentication protocol for the internet of things," *IEEE Access*, vol. 9, pp. 124768–124792, 2021.
6. Giguère, François, "Edge Computing in IoT: Transforming Security Paradigms Through Advanced Forensics," 2024.
7. Thomas, Chips, "IoT and Edge Computing Convergence: Revolutionizing Security and Forensic Applications," 2024.
8. Sheikh, Abdul Manan, Md Rafiqul Islam, Mohamed Hadi Habaebi, Adnan Kabbani, Suriza Ahmad Zabidi, and Athaur Rahman bin Najeeb, "Securing the IoT Edge Devices Using Advanced Digital Technologies," *Asian Journal of Electrical and Electronic Engineering*, vol. 4, no. 2, pp. 52–60, 2024.
9. Kong, Linghe, Jinlin Tan, Junqin Huang, Guihai Chen, Shuaitian Wang, Xi Jin, Peng Zeng, Muhammad Khan, and Sajal K Das, "Edge-computing-driven internet of things: A survey," *ACM Computing Surveys*, vol. 55, no. 8, pp. 1–41, 2022.
10. Raza, A., Yusoff, M. Z., Khan, M., Baloch, M., Shaikh, A. M., and Chauhdary, S. T., "Solar Energy Optimal Grid Integration Through Machine Learning Techniques", *International Journal on Energy Conversion*, vol. 12(2), 2024.
11. R. A. Alnuaimi, R. K. Almasalmeh, E. A. Alhammadi, G. E. B. M. E. Hassan and H. Zia, "Optimizing Edge Security: Comprehensive Analysis and Mitigation Strategies for Securing Edge Computing," *2023 9th International Conference on Optimization and Applications (ICOA)*, Abu Dhabi, United Arab Emirates, 2023, pp. 1–7, doi: 10.1109/ICOA58279.2023.10308853.
12. Sheikh, A. M., Islam, M. R., Habaebi, M. H., Kabbani, A., Zabidi, S. A., and bin Najeeb, A. R., "Securing the IoT Edge Devices Using Advanced Digital Technologies", *Asian Journal of Electrical and Electronic Engineering*, vol. 4(2), pp. 52–60, 2024.
13. Manan, A., "Efficient 16 nm SRAM Design for FPGA's", *5th International Conference on Signal Processing and Integrated Networks (SPIN)*, pp. 457–461,IEEE, 2018.
14. Manan, A., "Implementation of image processing algorithm on FPGA", *Akgec Journal of Technology*, vol. 2(1), pp. 25–28, 2006.
15. K. Lata and L. R. Cenkeramaddi, "FPGA-based PUF designs: A comprehensive review and comparative analysis," *Cryptography*, vol. 7, no. 4, p. 55, 2023.
16. Sheikh, A. M., Islam, M. R., Habaebi, M. H., Zabidi, S. A., Najeeb, A. R. B., and Basahel, A., "Machine Learning (ML) assisted Edge security framework on FPGAs", *9th International Conference on Computer and Communication Engineering (ICCCE)*, pp. 155–160, IEEE, 2023.
17. Sheikh, A. M., Islam, M. R., Habaebi, M. H., Zabidi, S. A., Bin Najeeb, A. R., and Kabbani, A., "Integrating Physical Unclonable Functions with Machine Learning for the Authentication of Edge Devices in IoT Networks", *Future Internet*, vol. 17(7), no. 275, 2025.
18. Yu, Sungjin, Kisung Park, and Youngho Park. "A Machine Learning Attack-Resistant PUF-based Robust and Efficient Mutual Authentication Scheme in Fog-enabled IoT Environments." IEEE Internet of Things Journal (2025).
19. N. N. Anandakumar, M. S. Hashmi and S. K. Sanadhya, "Design and analysis of FPGA-based PUFs with enhanced performance for hardware-oriented security," *ACM Journal on Emerging Technologies in Computing Systems (JETC)*, vol. 18, no. 4, pp. 1–26, 2022.

20. T. Shen, L. Ding, J. Sun, C. Jing, F. Guo, and C. Wu, "Edge Computing for IoT Security: Integrating Machine Learning with Key Agreement," in *Proc. 2023 3rd Int. Conf. on Consumer Electronics and Computer Engineering (ICCECE)*, pp. 474–483, 2023.
21. S. G. Aarella, S. P. Mohanty, E. Kougianos, and D. Puthal, "Fortified-edge: Secure PUF certificate authentication mechanism for edge data centers in collaborative edge computing," in *Proc. Great Lakes Symposium on VLSI 2023*, pp. 249–254, 2023.
22. G. Cheng, Y. Chen, S. Deng, H. Gao, and J. Yin, "A Blockchain-Based Mutual Authentication Scheme for Collaborative Edge Computing," *IEEE Trans. Comput. Social Syst.*, vol. 9, no. 1, pp. 146–158, 2022.
23. G. Liu, C. Wang, X. Ma, and Y. Yang, "Keep your data locally: Federated-learning-based data privacy preservation in edge computing," *IEEE Network*, vol. 35, no. 2, pp. 60–66, 2021.
24. Zhang, Jiliang, Chaoqun Shen, Zhiyang Guo, Qiang Wu, and Wanli Chang. "CT PUF: Configurable tristate PUF against machine learning attacks for IoT security." IEEE Internet of Things Journal 9, no. 16 (2021): 14452-14462.
25. Ebrahimabadi, Mohammad, Mohamed Younis, and Naghmeh Karimi. "A PUF-based modeling-attack resilient authentication protocol for IoT devices." IEEE Internet of Things Journal 9, no. 5 (2021): 3684-3703.
26. H. Shin, D. Koo, and J. Hur, "Secure and efficient hybrid data deduplication in edge computing," *ACM Trans. Internet Technol.*, vol. 22, no. 3, pp. 1–25, 2022.
27. A. Rullo, E. Bertino, and K. Ren, "Guest editorial special issue on intrusion detection for the Internet of Things," *IEEE Internet of Things Journal*, vol. 10, no. 10, pp. 8327–8330, 2023.
28. A. Sarkar, S. Ganguly, P. S. Sarkar, and S. R. Chatterjee, "PUF-Based Authentication System with Resilience against Multi-Faceted Attacks for Blockchain-based IoT Networks," in *Proc. 2024 IEEE 3rd World Conference on Applied Intelligence and Computing (AIC)*, pp. 1279–1284, 2024. doi:10.1109/AIC61668.2024.10731113.
29. R. Aparicio-Téllez, M. Garcia-Bosque, G. Díez-Señorans, and S. Celma, "Novel machine learning-resistant RO-based PUF optimized for IoT device authentication," *IEEE Access*, vol. 13, pp. 46147–46160, 2025.
30. L. Fan, Z. Huang, J. Wang, L. Zhou, Y. Zhu, and Q. Wang, "A Novel Configurable RO-Obfuscated PUF Design with Machine Learning Immunity," in *Proceedings of the 2023 International Conference on Networking and Network Applications (NaNA)*, pp. 680–685, 2023, doi: 10.1109/NaNA60121.2023.00117.
31. H. Kareem and D. Dunaev, "Machine Learning Vulnerability Assessment of Ring Oscillator Physical Unclonable Functions," in *Proceedings of the 2023 International Conference on Control, Automation and Diagnosis (ICCAD)*, pp. 1–5, 2023, doi: 10.1109/ICCAD57653.2023.10152413.
32. Abulibdeh, E., Saleh, H., Mohammad, B., Alqutayri, M. and Santikellur, P., "Algorithmically Optimized Configurable Ring Oscillator Puf for Iot Devices". Available at SSRN: https://ssrn.com/abstract=5327693
33. Miskelly, J., Gu, C., Ma, Q., Cui, Y., Liu, W. and O'Neill, M., "Modelling attack analysis of configurable ring oscillator (CRO) PUF designs,"*IEEE 23rd international conference on digital signal processing (DSP)*, pp. 1–5, 2018.
34. Laguduva, V., Islam, S.A., Aakur, S., Katkoori, S. and Karam, " Machine learning based iot edge node security attack and countermeasures " *IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*, pp. 670–675, 2019.
35. E. E. Abulibdeh, *Reliable and Efficient Hardware Implementation of PUFs for Secure IoT Applications*, Ph.D. dissertation, Khalifa University of Science, 2024.
36. J. H. L. Teo, N. A. N. Hashim, A. Ghazali, and F. Hamid, "Ring oscillator physically unclonable function using sequential ring oscillator pairs for more challenge-response-pairs," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 13, no. 3, pp. 892–901, 2019.
37. F. Zerrouki, S. Ouchani, and H. Bouarfa, "PUF-based mutual authentication and session key establishment protocol for IoT devices," *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 9, pp. 12575–12593, 2023.
38. Y. Gao, S. F. Al-Sarawi, and D. Abbott, "Physical unclonable functions," *Nature Electronics*, vol. 3, no. 2, pp. 81–91, 2020.
39. A. Al-Meer and S. Al-Kuwari, "Physical unclonable functions (PUF) for IoT devices," *ACM Computing Surveys*, vol. 55, no. 14s, pp. 1–31, 2023.
40. Ali, Rashid, Haoyuan Ma, Zhengyi Hou, Deming Zhang, Erya Deng, and You Wang. "A reconfigurable arbiter MPUF with high resistance against machine learning attack." IEEE Transactions on Magnetics 57, no. 10 (2021): 1-7.
41. Anupama, A., Immanuel Raja, Deepu Roy, and K. Padmakumar. "A Ring Oscillator-based Strong Physical Unclonable Function with Excellent Resilience to Machine Learning Attacks." IEEE Internet of Things Journal (2025).

doi:10.20944/preprints202512.0284.v1

42. Sajadi, Abolfazl, Ahmad Shabani, and Bijan Alizadeh. "DC-PUF: Machine learning-resistant PUF-based authentication protocol using dependency chain for resource-constraint IoT devices." Journal of Network and Computer Applications 217 (2023): 103693.
43. Rajput, Shailesh, and Jaya Dofe. "Counteracting modeling attacks using hardware-based dynamic physical unclonable function." In 2023 IEEE International Conference on Cyber Security and Resilience (CSR), pp. 586-591. IEEE, 2023.
44. Ferens, Mieszko, Edlira Dushku, and Sokol Kosta. "When Random is Bad: Selective CRPs for Protecting PUFs against Modeling Attacks." IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (2024).
45. Hazari, Noor Ahmad, Faris Alsulami, Ahmed Oun, and Mohammed Niamat. "Performance analysis of XOR-inverter based ring oscillator PUF for hardware security." In 2019 IEEE National Aerospace and Electronics Conference (NAECON), pp. 253-256. IEEE, 2019.
46. Kareem, Husam, and Dmitriy Dunaev. "A novel low hardware configurable ring oscillator (CRO) PUF for lightweight security applications." Microprocessors and Microsystems 104 (2024): 104989.
47. Cui, Yijun, Chenghua Wang, Weiqiang Liu, Chongyan Gu, Máire O'Neill, and Fabrizio Lombardi. "Lightweight configurable ring oscillator PUF based on RRAM/CMOS hybrid circuits." IEEE Open Journal of Nanotechnology 1 (2020): 128-134.
48. He, Zhangqing, Chen Wang, Tao Ke, Yuejiao Zhang, Wenjun Cao, and Jiuchun Jiang. "A highly reliable FPGA-based RO PUF with enhanced challenge response pairs resilient to modeling attacks." IEICE Electronics Express 18, no. 20 (2021): 20210350-20210350.
49. Zulfikar, Zulfikar, Norhayati Soin, Sharifah Fatmadiana Wan Muhamad Hatta, Mohamad Sofian Abu Talip, and Anuar Jaafar. "Routing density analysis of area-efficient ring oscillator physically unclonable functions." Applied Sciences 11, no. 20 (2021): 9730.
50. Zulfikar, Zulfikar, Norhayati Soin, Sharifah Fatmadiana Wan Muhamad Hatta, and Mohamad Sofian Abu Talip. "Runtime analysis of area-efficient uniform RO-PUF for uniqueness and reliability balancing." Electronics 10, no. 20 (2021): 2504.
51. He, Zhangqing, Chen Wang, Tao Ke, Yuejiao Zhang, Wenjun Cao, and Jiuchun Jiang. "A highly reliable FPGA-based RO PUF with enhanced challenge response pairs resilient to modeling attacks." IEICE Electronics Express 18, no. 20 (2021): 20210350-20210350.
52. Maiti, Abhranil, Jeff Casarona, Luke McHale, and Patrick Schaumont. "A large scale characterization of RO-PUF." In 2010 IEEE International Symposium on Hardware-Oriented Security and Trust (HOST), pp. 94-99. IEEE, 2010.
53. Suh, G. Edward, and Srinivas Devadas. "Physical unclonable functions for device authentication and secret key generation." In Proceedings of the 44th annual design automation conference, pp. 9-14. 2007.
54. A. M. Naveed, K. C. Chua, and B. Sikdar, "Physical unclonable functions for IoT security," in *Proc. 2nd ACM International Workshop on IoT Privacy, Trust, and Security*, 2016.
55. A. M. Sheikh, M. R. Islam, M. H. Habaebi, S. A. Zabidi, A. R. Bin Najeeb, and A. Kabbani, "A Survey on Edge Computing (EC) Security Challenges: Classification, Threats, and Mitigation Strategies," *Future Internet*, vol. 17, no. 4, p. 175, 2025.
56. B. Chatterjee, D. Das, S. Maity, and S. Sen, "RF-PUF: Enhancing IoT security through authentication of wireless nodes using in-situ machine learning," *IEEE Internet of Things Journal*, vol. 6, no. 1, pp. 388–398, 2018.
57. A. M. A. Modarres, N. S. Anzabi-Nezhad, and M. Zare, "A new PUF-based protocol for mutual authentication and key agreement between three layers of entities in cloud-based IoMT networks," *IEEE Access*, vol. 12, pp. 21807–21824, 2024.
58. C. Yehoshuva, R. R. Adhithan, and N. N. Anandakumar, "A survey of security attacks on silicon based weak PUF architectures," in *Proc. International Symposium on Security in Computing and Communication*, pp. 107–122, 2020.
59. H. Kareem and D. Dunaev, "A robust architecture of ring oscillator PUF: Enhancing cryptographic security with configurability," *Microelectronics Journal*, vol. 143, p. 106022, 2024.
60. M. Budnik, "Design and evaluation of a Ring Oscillator based Physically Unclonable Function," B.S. thesis, University of Twente, 2023.
61. Marton, K., and Suciu, A., "On the interpretation of results from the NIST statistical test suite,"*Science and Technology*, vol. 18, no. 1, pp. 18–32, 2015.

 doi:10.20944/preprints202512.0284.v1

62. Pahlevi, Rizka Reza, Hirokazu Hasegawa, Yukiko Yamaguchi, and Hajime Shimada. "A Pre-Selection–Enhanced Arbiter PUF for Strengthening PUF-Based Authentication." IEEE Access (2025).
63. Sukarno, Parman, and Fachrul Reiza Medina. "Enhancing IoT Security: Optimizing PUF Responses through Pre-Processing Techniques." JURNAL INFOTEL 17, no. 2 (2025): 210-228.
64. Al-Meer, Abdulaziz, and Saif Al-Kuwari. "Physical unclonable functions (PUF) for IoT devices." ACM Computing Surveys 55, no. 14s (2023): 1-31.
65. A. Bhatia, S. Bitragunta, and K. Tiwari, "PUF-AQKD: A Hardware-Assisted Quantum Key Distribution Protocol for Man-in-the-Middle Attack Mitigation," *IEEE Open Journal of the Communications Society*, pp. 1–1, 2025. doi:10.1109/OJCOMS.2025.3575206.