Machine Learning Attacks on Optical Physical Unclonable Functions

By

Iskandar Atakhodjaev

A dissertation submitted to The Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy.

Baltimore, Maryland

August, 2018

Abstract

Traditional security algorithms for authentication and encryption rely heavily on the digital storage of secret information (e.g. cryptographic key), which is vulnerable to copying and destruction. An attractive alternative to digital storage is the storage of this secret information in the intrinsic, unpredictable, and non-reproducible features of a physical object. Such devices are termed physical unclonable functions (PUFs), and recent research proves that PUFs can resolve the vulnerabilities associated with digital key storage while otherwise maintaining the same level of security as traditional methods. Modern cryptographic algorithms rest on the shoulders of this one-way principle in certain mathematical algorithms (e.g. RSA or Rabin functions). However, a key difference between PUFs and traditional one-way algorithms is that conventional algorithms can be duplicated.

Here, we investigate a silicon photonic PUF a novel cryptographic device based on ultrafast and nonlinear optical interactions within an integrated silicon photonic cavity. This work reviews the important properties of this device including high complexity of light interaction with the material, unpredictability of the response and ultrafast generation of private information. We further explore the resistance of silicon photonic PUFs against numerical modeling attacks and demonstrate the influence of cavity's inherent nonlinear optical properties on the success of such attacks. Finally, we demonstrate encrypted data storage and

compare the results of decryption using a genuine silicon PUF device the “clone” generated by the numerical algorithm.

Finally, we provide similar analysis of modeling attacks on another well-known type of optical PUF, called the optical scattering PUF (OSPUF). While not as compatible with integration as the silicon photonic PUF, the OSPUF system is known to be extremely strong and resistant to adversarial attacks. By attacking a simulated model of OSPUF, we attempt to present the underlying reasons behind the strong security of this given device and how this security scales with the OSPUFs physical parameters.

Primary Advisor: N. Peter Armitage

Secondary Advisor: Mark A. Foster

Acknowledgement

The first couple of years in the United States were quite challenging for me. I had to adapt to a new country, make new friends, speak and think in a new language. But I was quite lucky to meet people that made my transition comfortable and smooth. First of all, I would like to thank my academic advisors: Peter Armitage and Mark Foster. I am deeply grateful for Prof. Foster's encouragement at the times when I really needed it. I appreciate his both material and moral support, brilliant expertise and great ideas that significantly helped me throughout my program. I am also thankful to Prof. Armitage since he is the one who introduced me to Prof. Foster back in 2013. His support and advice will always remain invaluable to me. I would also like to thank Prof. Amy Foster for being my instructor at one of her classes and for the smooth transition to the research of the lab.

I appreciate Ms. Kelley Key and Prof. Oleg Tchernyshyov who supported me in the early beginning of my Ph.D. program. Kelley helped me to navigate the educational issues and program related procedures, for which I am grateful for.

I am thankful to my parents Alisher Atakhodjaev and Iroda Vafokulova for their support, patience for my regular visits, personal sacrifices, and encouragement. I would not have achieved anything without their countless lessons and life experience. I am thankful to my lovely sister Dildora Atakhodjaeva for her sense of humor and limitless support during my life in Baltimore. I thank my aunt Abdalimova Guzal, my cousins,

relatives and closest friends in Moscow who have always supported me over the years and always waited for my visits.

I deeply thank my colleagues in the lab, Dr. Bryan Bosworth, Milad Alemohammad, Jasper Stroud, Jeff Shin, Hongcheng Sun, Kangmei Li for all the assistance they provided me, for countless questions I asked them and for teaching me the necessary skills in order to survive in the program. My deep respect goes to Bryan Bosworth who was patient enough to listen to my arguments, to explain the same things many times, and for spending with me zillion hours in the lab assisting me to align my experimental setup. I am grateful for Milad's great ideas he always suggested, his sincere willingness to help, and his kindness to listen to my concerns.

My life would have been more boring without my friends I met in Baltimore. My special thanks go to Prasenjit Bose. He was the first friend at JHU who bought me a sandwich when I even didn't know where to buy it. He helped me out countless times and we had great fun times together. I thank Mikhail Osanov, Subranshu Mishra, Ankur Gupta, Daniil Pakhomov, Polina Koroleva, Chris Sapsanis, Nikita Ivkin, Robert Dipietro, Nurgissa Umatay for great times we shared at JHU. I thank Diyora Khasanova who makes my life brighter by occasional visits to Baltimore. Her support and love always helped me to stay afloat in the busiest period of times at work.

On a final note, I would like to say that I will owe everyone mentioned here for the rest of my life, as their actions have played a critical role in the shaping of my personal life and professional career.

Table of Contents

List of Figures

List of Tables

Chapter 1 : Introduction and Overview

In this dissertation, we investigate a novel approach of generating and reliably storing vast amounts of digital key material within a complex physical object. Such objects are known as physical unclonable functions (PUFs) and have potential applications in secure authentication, anti-counterfeiting, and data encryption. Specifically, the extraction of private information from a PUF can be used for encryption/decryption of stored material or of a communication channel or also as a unique signature for granting the access to a system or verifying the authenticity of an object. The development and implementation of the photonic PUFs investigated here require knowledge from multiple disciplines including integrated optics, dynamical systems, nonlinear and ultrafast optics, information theory, and cryptography. Specifically, in this dissertation, we study a cryptographic device based on the ultrafast optical interactions in an integrated silicon photonic cavity. The cavity's behavior results from a mixture of fundamental physical processes underlying the extremely complex light-matter interaction achieved at the operating conditions. In detail, the characterization of the device requires the consideration of a number of phenomena in the areas of nonlinear optics, ultrafast optics, and semiconductor physics. In practice, the nonlinear optical behavior is typically avoided in other optical systems since it often distorts signals of interests and can introduce instability and noise. On the contrary, in this work, we exploit a range of nonlinear optical processes and benefit from the additional

complexity and signal distortion during the device's operation. The research here includes the description of these optical processes and their impact on the reliability and security of optical PUFs as cryptographic devices.

1.1 Internet of Things and Digital Era

The concept of the Internet of Things (IoT) was first introduced back in early 1999 by K. Ashton who was a brand manager at Procter & Gamble [1]. Studying the data of supply chains led him to deploy Radio Frequency Identification (RFID) tags on shop inventory, thus allowing early *"talk"* of *Things* between each other. With the progress of technology, the definition of *Things* has changed, although the main goal of keeping *Things* interconnected still remains the same.

Ever since the phrase *Internet of Things* was born, it received a variety of definitions, including the *Network of Everything* or *Network of Objects* or *Internet of People* [2]. Despite all these descriptions, what remains truly undoubted is the fact that IoT is approaching our day to day lives inevitably and silently. Presently, one of the popular perceptions of the IoT concept is the global network of digital devices interconnected with standard communications protocols. In order to clarify this concept, let me give some examples and applications of IoT:

Smart environment application domains.

	Smart home/office	Smart retail	Smart city	Smart agriculture/forest	Smart water	Smart transportation
Network size	Small	Small	Medium	Medium/large	Large	Large
Users	Very few, family members	Few, community level	Many, policy makers, general public	Few, landowners, policy makers	Few, government	Large, general public
Energy	Rechargeable battery	Rechargeable battery	Rechargeable battery, energy harvesting	Energy harvesting	Energy harvesting	Rechargeable battery, Energy harvesting
Internet connectivity	Wifi, 3G, 4G LTE backbone	Wifi, 3G, 4G LTE backbone	Wifi, 3G, 4G LTE backbone	Wifi, satellite communication	Satellite communication, microwave links	Wifi, satellite communication
Data management	Local server	Local server	Shared server	Local server, shared server	Shared server	Shared server
IoT devices	RFID, WSN	RFID, WSN	RFID, WSN	WSN	Single sensors	RFID, WSN, single sensors
Bandwidth requirement	Small	Small	Large	Medium	Medium	Medium/large

Table 1.1: Current and potential applications of Internet of Things [3]

Figure 1.1 summarizes that with IoT anything/anybody in the world will be able to connect to internet from any place in the world.

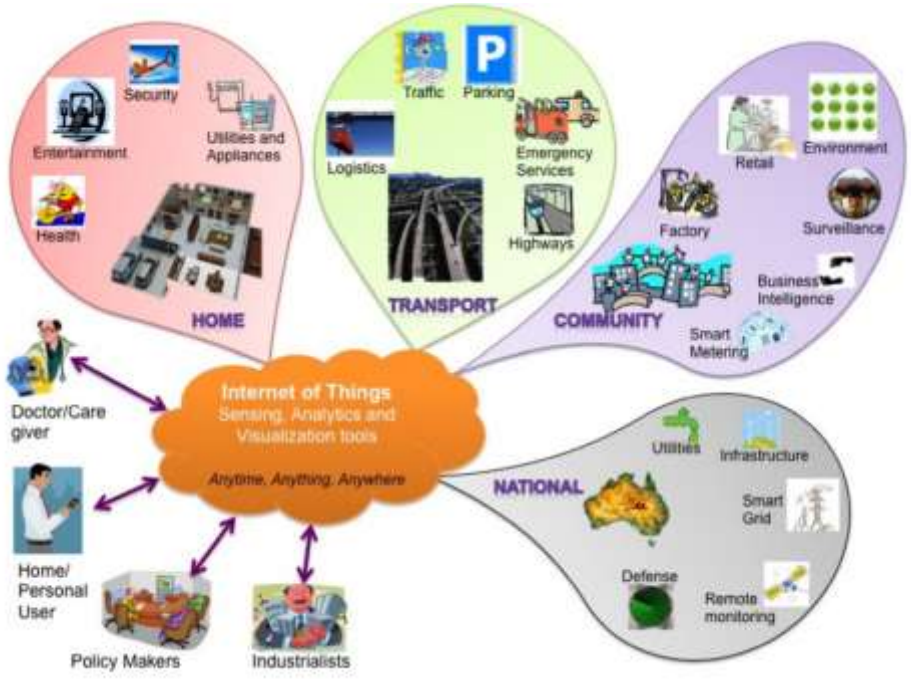


Figure 1.1: Schematic of Internet of Things concept indicating the end users and applications in various aspects of life [3].

It is clear now that IoT is the next revolution of technology leading to the transformation of our society at all levels. The current state of IoT is still in the emerging

phase, consisting of 25 billion devices connected to each other [3]. According to Cisco's network growth forecasts (**Figure 1.2**) the number of connected devices on the internet will exceed 50 billion by 2020 and by 2022 the world will be drowning in 1 trillion sensors [4]. One of the most important consequences of this trend is the rapid increase in the amount of data that will be generated by each device in the network. Unprecedented amounts of data will have to be managed, stored and protected requiring novel technologies. The proliferation of such amounts of data will inevitably magnify security threats of the network, lead to authentication problems, access control, privacy of the data and its confidentiality. Therefore, it is critically important to understand whether we are prepared for such dramatic changes at this level of pace and how we are planning to solve the problems mentioned above.

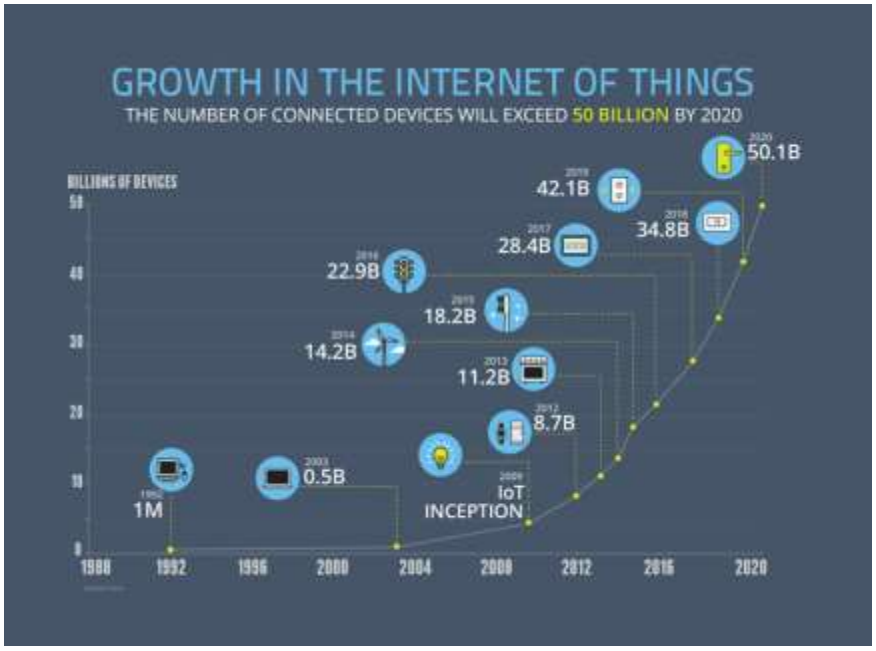


Figure 1.2: Growth of digital devices in the Internet of Things [4].

Keeping data private and secure has always been a challenging task. The need for providing secured information has existed for centuries. However, in the last couple of decades, given the exponential growth of the amount of data generated in the world, this need has also exponentially grown. According to [5], in 2012 the digital world of data was expanded to 2.7 Zettabytes (10^{21} bytes) and this amount is predicted to double every two years [6]. Everyday 2.5 Exabytes (10^{18} bytes) is created along with the fact that 90% of the total amount of current data has been generated in the last 5 years [7]. This situation is astonishing and frightening at the same time. Mass media and academia are overfilled with numerous articles and research papers about this so-called *Digital Era.* Some people call it the *Era of Big Data*, others refer to it as *Industry 4.0* [8]. The latter term encompasses the global digitalization of society, IoT, smart environment and manufacturing, cyber-physical systems and etc. Big IT corporations such as Google, Amazon, Apple, IBM, Microsoft and etc. are already in a race to achieve dominance in this emerging market providing a variety of services and products to efficiently store data, operate on it and securely transfer it between two parties.

As we can see, there is a vast amount of problems coming in the near future and in this dissertation, I focus on one of them: information security of data.

1.2 Information Security in a Modern World

The main goals of information security can be classified in the following way [9]:

- Confidentiality of data, i.e. keeping the information protected from unauthorized parties

- Authentication. Proper protocols to provide the proof of the identity as well as entity one is interacting with. For example, if someone wants to access a university library, he/she must provide a student card in order to be allowed to enter
- Data Integrity. The communication of two parties must not be altered by an adversary
- Non-repudiation, i.e. one party of a transaction cannot deny having received a transaction, nor can the other party deny having sent a transaction [10]

Most of these requirements are addressed by fields such as Cryptography, which is providing the range of solutions being presented here. The most widespread family of security schemes are Symmetric Key Algorithms and Public Key Algorithms (Asymmetric Scheme).

1.2.1 Symmetric Key Algorithms

If two parties communicate with each other using symmetric key algorithms, they use cryptographic keys (exchanged apriori via secured channel or during physical meeting) for encryption of plaintext and decryption of ciphertext that are known for both sides. Those keys are shared between two parties and must be kept in secret from any external malicious attackers. Most of the time, the encryption and decryption keys are identical. In **Figure 1.3**, the principle of symmetric key algorithm is demonstrated [11]. The best-known examples of the given security scheme are DES, AES, VERNAM and One -Time - Pad algorithms [12-15].

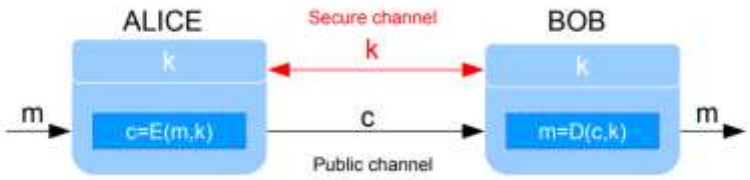


Figure 1.3: Communication of two parties with symmetric key algorithm. m – plaintext, c – ciphertext, k – cryptographic key, E – encryption function and D – decryption function [11].

1.2.2 Asymmetric Key Algorithms

Asymmetric Key or Public Key Algorithms revolutionized cryptography in the 1970s and remain the most popular cryptographic system in modern security [16]. The system still uses a pair of keys: a public key, that can be spread widely without any restrictions and a private key which is known only to the owner of encryption of plaintext. The private key must be kept in a strict secret, otherwise the security of the whole system is at risk.

In public key cryptography any person can encrypt a message using the receiver's public key, however the encrypted message can only be decrypted with the receiver's private key. Before the communication starts, typically the pair of public and private keys are generated in a fast and efficient way. The whole idea of public cryptography consists of the fact that it is almost impossible to computationally derive the private key knowing its paired public key. Systems using public key algorithms rests on mathematical problems that currently don't have an efficient solution such as elliptic curves relationships, discrete algorithms etc. **Figure 1.4** shows the general principle of communication between two parties utilizing public key algorithm [11].

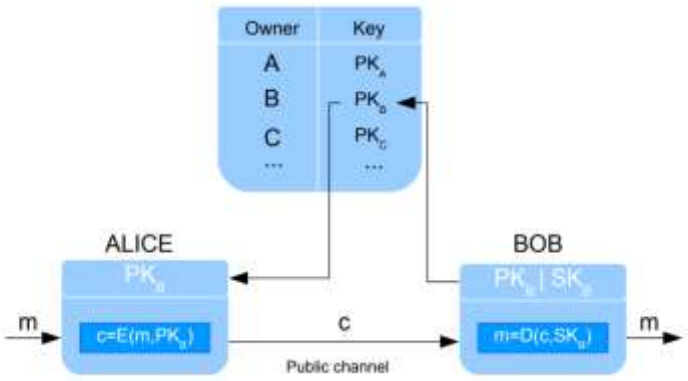


Figure 1.4: Communication of two parties with asymmetric key algorithm. *m* – plaintext, *c* – ciphertext, PK_x – public key of user *X*, SK_x – secret key of user *X*, *E* – encryption function and *D* – decryption function [11].

Before Alice sends the message, she fetches Bob's public key and encrypts the plaintext with it. The encrypted message (ciphertext) is transmitted to Bob via a public channel, where on the other end Bob, using his private key decrypts the ciphertext. Public key cryptography is the essential ingredient in modern cryptosystems, communication protocols, and other applications. The most widely used RSA (Rivest-Shamir-Adleman) system is one of the earliest public key cryptosystems and is extensively deployed in secure data transmission [17].

1.3 Current Challenges in Information Security

In the last section, I discussed that modern information security rests on computational asymmetry, i.e. algorithms that are easy to compute, but difficult to invert. This type of "one-way" property, however is not exhaustively studied and according to [18, 19], the full security of one-way algorithms is not yet proven. For example, the SHA-1

cryptographic hash function was reliable until 2005, when Rijmen, et al. proposed an attack against it [20]. Since then many organizations have switched to more secure versions of hash functions (SHA-2, SHA-256). Further, with an increase of parallel computing power and supercomputers, asymmetric algorithms can be potentially jeopardized in the future. At this point, we cannot just rely on protective mechanisms ensured by algorithmic or mathematical security scheme. Many of the key-based mathematical algorithms are implemented in electronic circuits to prevent counterfeiting, frauds, and theft. However, the latter methods cause other issues such as insecure key storage and generation, complex and costly physical anti-tampering mechanisms, and cumbersome heavy designs. The secret key, for example, in those physical primitives must be stored in non-reliable and non-volatile electronic memory occupying large portions of the integrated circuit (IC). Moreover, cryptographic devices often have to be powered on in order to keep the memory active thereby increasing the cost of device manufacturing and decreasing their operation speed. Since the secret key is stored permanently in the digital memory in most of the cryptographic devices, they are highly vulnerable to adversarial attacks causing a rise of counterfeit market. Several reports from last year Frontier Economics indicate that the global market of counterfeiting and piracy could potentially reach US$2.3 trillion by 2022 [21].

Complexity and insecure behavior of asymmetric cryptosystems mentioned above motivated much of the research to develop new alternative approaches that are reliable, secure, lightweight, easy to use in authentication and key storage applications and cheap to manufacture. Physical Unclonable Functions (PUF) offer a promising and innovative solution to the issues of reliable private key storage, secure authentication schemes and

easy operation. In the next section, I introduce the concept of a PUF and the state-of-the-art techniques related to this approach.

1.4 Physical Unclonable Functions (PUFs)

1.4.1 Concept

The concept of a Physical Unclonable Function is closely related to the old idea of using intrinsic physical features to uniquely identify objects in the world. Biometric authentication has been known for centuries going back to early Egyptian times when traders were distinguished by their physical characteristics. Human fingerprints are the best example of such physical features which were first collected in 1891 in Argentina to track criminals [22]. Fingerprint authentication is a widespread security scheme nowadays for many reasons. Firstly, they are specific to one person and only one, thereby human fingerprints possess *individualism.* Besides this, a fingerprint is also *inherent,* so that every human on the Earth has this physical feature, unlike other identification features like a hand signature or name. Finally, and the most importantly, fingerprints are *unclonable,* hence it is difficult to generate identical copies of fingerprints through artificial or biological processes. Since times when people started using fingerprints, many other biometric technologies were born including voice authentication, face recognition, infrared thermogram, DNA and etc.

Besides physiological features, unique identification can be implemented using other properties of objects. For example, at the end of the twentieth century, random patterns in paper fibers and optical tokens were used in unique identification of currency

notes and strategic arms [23, 24]. The formalization of such approaches began later in the early twenty-first century and initially it was introduced as a physical one-way function and finally as *physical unclonable function* or PUF.

Mathematically, the concept of PUF is akin to the concept of an algorithmic one-way function. These are the *functions that are easy to compute, but hard to invert*. A function that is easy to compute means that there is a polynomial computational time to produce the output given the input, while the difficulty of inversion indicates the negligible probability of finding any algorithm that finds the input given the output. Asymmetric key algorithms, as I discussed in section 1.2, are exactly based on this definition where it is almost impossible in the finite time to derive the private key from only knowing its public key.

There is no strict definition of a PUF, but the one that is most frequently mentioned in the literature is the following. A PUF is a *physical object which produces the output signal to the input which is dependent on physical structures which are impossible to clone* [25]. To further accentuate, three keywords in the PUF abbreviation corresponds to specific requirements that have to be satisfied in each PUF instance (**Figure 1.5**).

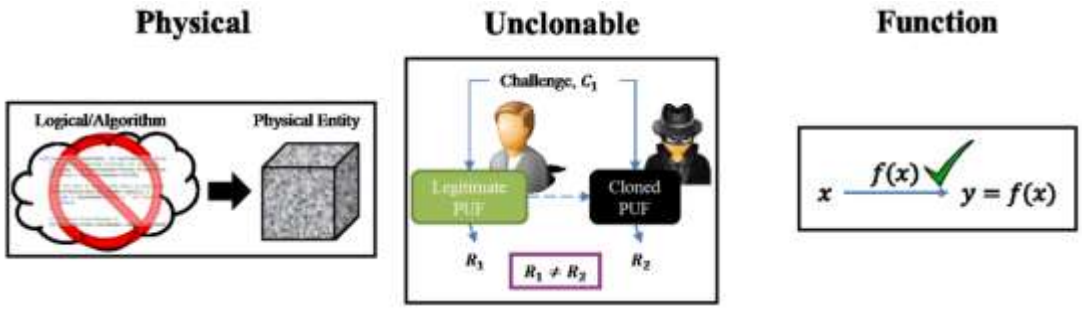


Figure 1.5: Physical Unclonable Function Model [26].

a) **Physical**, i.e. real physical entity (not an algorithm or logical procedure).

b) **Unclonable**, i.e. cannot be replicated by any means with any infinite resources even by the original creator of the entity.

c) **Function**, i.e. it has to perform some operation on the given input signal and generate the output. In other words, it is a function in an engineering sense.

Interest in PUFs has risen significantly over recent years leading to an increase of published works in this information security area. The majority of conventional information security approaches discussed earlier, rely on the concept of a piece of information that must be kept in secret in storage permanently. If an attacker finds a way to steal this piece of information, the whole security of the system is compromised. PUFs, on the other hand, suggest a novel approach to building security. The key idea in PUF is to leverage the small-scale random disorders that are inevitable during the manufacturing process of the device. Those disorders are unavoidable and uncontrollable effect during the fabrication of the system, making the device truly unclonable even for the original manufacturer that has a complete knowledge of the design. The physical structure of the device plays a role analogous to a fingerprint or DNA.

PUFs offer a variety of advantages over the traditional cryptographic systems. For example, instead of a permanent storage of the secret key in non-volatile memory, PUFs derive the secret key from their behavior, which is sensitive to the unique physical random structure. In other words, all private information is placed in the device's physical structure and can be accessed only by the holder of the device. Moreover, in PUF systems there is no need to keep the memory powered in order to access the private key, since the key is

obtained only at the time of the external stimuli, typically called *challenges.* On top of that, physical intrusion in the PUF device would inevitably and irreversibly modify the original PUF behavior making PUF systems tamper-evident.

Physical Unclonable Functions prove to be a lightweight cryptographic primitive with a range of potential applications including low-cost authentication and secret key generation. In the next subsection, I consider the terminology and basic definitions of components associated with PUF systems that subsequently will be used throughout this dissertation.

1.4.2 PUF Terminology

Any PUF device queried with a specific input produces a measurable output, i.e. a PUF performs a functional operation. Typically, an input to a PUF is called a *challenge* and the output is called the *response.* An applied challenge and corresponding response are referred to as a *challenge-response pair (CRP)*. The relationship between challenges and responses is generally called *CRP behavior.* The process of collection of CRPs is referred to as the e*nrollment process* and the collection itself is called a *CRP database* or *CRP library.* I want to stress the fact that the CRP behavior is unique to one particular PUF token and only one. In order to be authenticated the client, who physically possesses a PUF device, queries it with one of the challenges from CRP library. The response produced by PUF is then compared to the response from CRP library. This is a standard challenge-response authentication protocol based on PUF systems that is also demonstrated in **Figure 1.6**. Another set of terms is associated with PUF security metrics that would also be

considered in detail in section 1.4.3 and 1.4.4. In order to evaluate PUF performance I introduce two important concepts: *inter-distance* and *intra-distance.*

- Applying a particular challenge twice to the same PUF instantiation at different times, the *intra-distance* μ_{intra} indicates the distance between two corresponding responses, produced at these different times by the same PUF.

- Applying a particular challenge to two different PUF instantiations at the same time, the *inter-distance* μ_{inter} indicates the distance between two corresponding responses, produced by these two different PUFs.

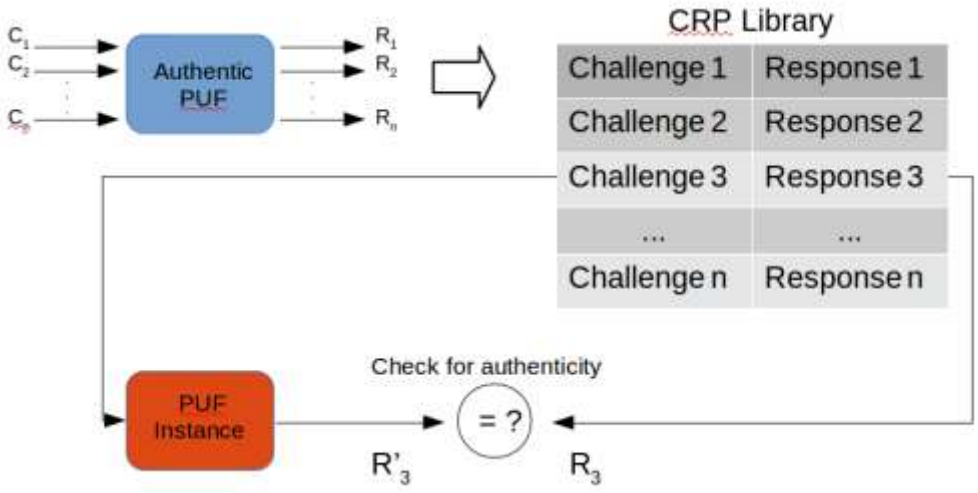


Figure 1.6: Typical authentication scheme using PUF based systems.

The intra-distance of PUF token measures the reproducibility of a response with respect to a previous response produced from the same challenge. Unlike mathematical algorithms, which always produce the same output, PUFs are physical processes that exhibit noise and other variation. “Good” PUFs are consistent, meaning that the same

challenge should correspond to the same response up to a certain noise level. It is clear that μ_{intra} should be as small as possible since this indicates a very reliable and repeatable PUF.

The inter-distance, on the other hand, demonstrates the degree of uniqueness of a PUF device, resulting in good differentiation of two systems. As a result, the value of μ_{inter} should be as high as possible, depending on the metrics and challenge-response representation.

Overall, uniqueness and repeatability are the most important properties that an ideal PUF should exhibit along with the others that would be described in the next section.

1.4.3 Main Properties of PUFs

In this section, I outline the most important and frequent properties encountered across a variety of PUF instantiations. Physical Unclonable Functions are still an emerging area in hardware security, so the following list is not exhaustive and can be complemented by various other properties depending on the specific PUF architecture.

Ideally, a PUF should exhibit six regularly occurring properties (**Figure 1.7**):

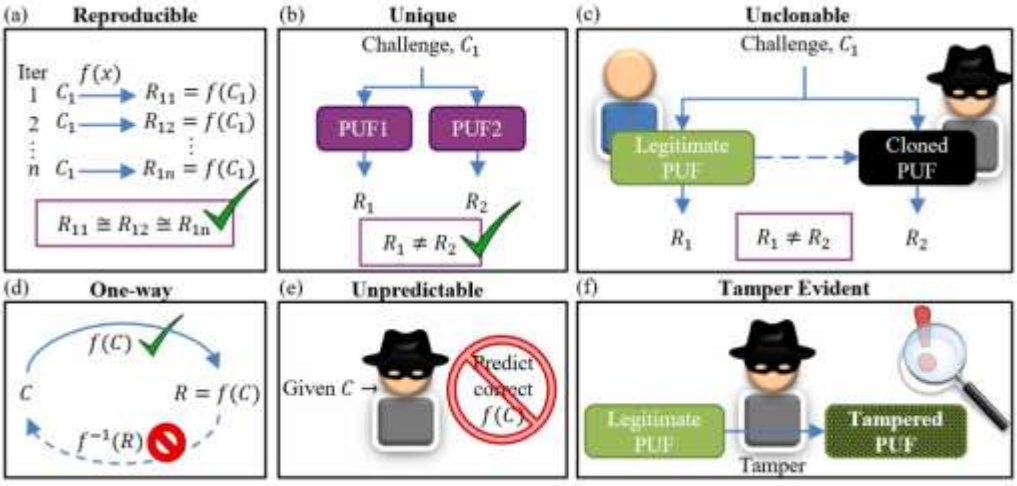


Figure 1.7: Most common properties of PUF systems [67].

a) *Reproducible:* The responses to different evaluations of the same challenge on the same PUF device should be close to each other (up to some threshold level) in the distance metric chosen apriori. Normally, reproducibility is measured by intra-distance value μ_{intra}

b) Unique: This property is self-explanatory and is derived straight from the definition of PUF. Different PUF devices should be unique, such that the same challenge given to two different devices produces significantly different responses. Normally, uniqueness is measured by inter-distance value μ_{inter} mentioned earlier

c) Unclonable: This is the core property of any PUF. PUF is truly unclonable if it is *mathematically unclonable* and *physically unclonable.* By physical unclonability, I mean that it is very hard to come up with the physical design of entity which would emulate the same behavior as the genuine design. If it is challenging to construct a mathematical model or algorithm that mimics original PUF, then PUF is claimed to be mathematically unclonable. Importantly, the latter gives rise to research in the direction of *modeling attacks* against PUFs, which is the focus of this dissertation.

d) One – Way: Classical property coming from asymmetric cryptography describing the fact that it is infeasible to invert the functional operation of PUF.

e) Unpredictable: This property is very similar to unclonability. If an adversary can predict the outcome of PUF for a chosen challenge, then PUF is considered to be spoofed. One way to do that is to observe the subset of CRP library and build a

mathematical model that learns PUF behavior based only on this subset. In the next chapter, I will go over attacks against PUFs in a detail.

f) *Tamper – Evident:* Tampering is the process of making changes to the integrity of physical entity. Since the whole point of PUF is in random physical idiosyncrasies it is clear that tampering the device would change PUF behavior forever.

Another important thing to consider while dealing with PUF systems is related to environment influence on the operation of PUF. During the physical measurement and response generation, there are a number of unwanted physical effects that could interfere with the final results. For example, the repeatability property of specific PUF instance can be affected by fluctuations of temperature or input power. Most of these factors carry the systematic effects, so there is a range of approaches developed to reduce their influence:

- One technique is called compensation, where instead of measuring the absolute values of responses, the differences between those are measured. Following this way, the influence of the environment is reduced, and the system is considered to be more robust.

- Another simple approach is to manually select those responses that turned out to be stable and robust and ignore other responses influenced by the environment the most.

Environmental effects are highly dependent on the specific implementation details of a given PUF. Certain designs do not require any of the approaches mentioned above and this will be observed in section 1.4.6 where I present a variety of PUF implementations.

1.4.4 PUF: metrics and evaluations

Given all the definitions and properties of PUFs, one needs to raise a question of the evaluation of PUF design quality. The most common set of evaluation metrics includes robustness, uniqueness, and unclonability.

Let me consider a PUF as a function $f(\bullet)$ with some provided challenge c_i and a produced response $r_i = f(c_i)$. In an ideal noiseless system, the repeated challenges applied to a PUF provide the same response. However, practically due to various phenomena such as noise, some misalignments in the system, optical fiber-to-waveguide conversion losses, surface impurities and etc., PUFs have small variations in the measured responses. Therefore, it is important to define the "distance" between the responses and evaluate the repeatability of the system based on it. Typically, in the research community, Fractional Hamming Distance (FHD) is one popular metric and is defined as the number of positions in which two binary sequences of the same length differ [25]. The FHD value is confined to the range of [0, 1], where FHD = 1 means that two sequences are different at all positions (i.e. identical but inverted), while FHD = 0 means that two sequences are identical. Two ideally unique and random sequences will thus have an FHD = 0.5.

To estimate how close the evaluations of challenges are, we compute FHD between all the subsequent responses, corresponding to the same interrogations applied to PUF. Then, plotting FHD values on a histogram plot, we obtain the distribution that is, normally, referred to as *same* or sometimes *like* distribution. The mean and width of the FHD distribution indicate the repeatability and error rate of the system. The mean of the distribution is also the intra-distance μ_{intra} mentioned earlier. PUFs that are robust and

repeatable have the *like* distribution centered around 0 with ideally a small standard deviation. Therefore, FHD distribution is a convenient mathematical tool to estimate the repeatability of PUF.

Now, let me consider two different PUF devices $f(\bullet)$ and $g(\bullet)$. Applying one specific challenge c_i, two PUFs generate corresponding responses $r_i^1 = f(c_i)$ and $r_i^2 = g(c_i)$. Then I calculate FHD (r_i^1, r_i^2) $\forall$ $i = 1..N$, where N – is the total number of challenge-response pairs. The distribution of FHD values, in this case, is called *different* or *unlike* distribution, where the mean of it indicates the uniqueness of the certain PUF. Therefore, if two PUFs generate uncorrelated responses to the same applied challenge, then the mean of unlike distribution should be centered around 0.5. The mean of the unlike distribution is also the inter-distance μ_{inter} mentioned earlier.

Lastly, to estimate the unclonability property of certain PUF implementation, we follow the same procedure of calculating FHD values, but instead of comparing responses from different devices we compare the responses from clones of one PUF. The closer the mean of obtained FHD distribution to 0.5 the harder to clone PUF.

In a typical authentication scenario, as was mentioned in section 1.4.2, expected responses from CRP library are compared to measured responses during authentication. Since the means of the like and unlike distributions are not exactly centered around 0 and 0.5 (due to noise effects, environmental factors, etc.) a specific threshold value α is set. If FHD value between the measured response and the expected response is below α then PUF is deemed authentic, otherwise it is unauthentic. In **Figure 1.8** [26] like and unlike distributions are presented and it is shown how to define the authentication threshold.

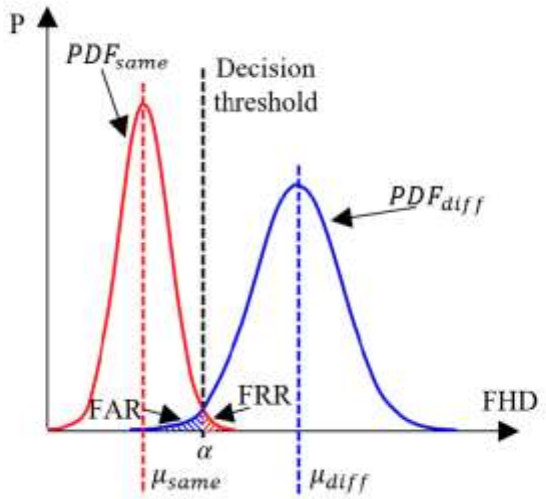


Figure 1.8: Authentication error using like and unlike distributions. FAR – False Acceptance Rate, FRR – False Rejection Rate. α – authentication decision threshold [26].

Two other important characteristics the false acceptance ratio (*FAR*) and false rejection ratio (*FRR*) of PUF are defined as the probability of accepting the wrong responses and the probability of rejecting the correct responses respectively.

To complete this section, I conclude that ideal PUF performance is reflected in terms of the characteristics mentioned above. High-quality PUFs have a significant separation of like and unlike FHD distributions and correspondingly a very low FAR and FRR. For uniqueness and unclonability, it is typically preferred that unlike distributions are centered around 0.5 with a small width of distribution so that responses are well uncorrelated. For repeatability and robustness of PUF system, it is typically preferred to have like distributions centered around 0 with a small width of distribution.

1.4.5 Types of PUFs

Given the huge number of PUF proposals and architectures, there is a need to systematically categorize their designs. One-way PUFs can be classified as based on the material used to manufacture the device. Early PUF systems such as random fiber structure of paper or optical scattering medium for reflection of light were based on non-electronic technologies. On the other side, there is a big class of PUFs containing electronic circuits integrated on a chip. Therefore, all the designs can be further categorized as *electronic* and *non-electronic.* I present the overview of the best examples of each class in the section 1.4.6.

Another popular way of classification of PUF systems is based on the source of randomness. As Guajardo et al. initially suggested, PUFs that satisfy the following two conditions [9, 27]: *i)* evaluations are performed internally within the PUF setup *ii)* random physical idiosyncrasies are implicitly introduced during the manufacturing process, are called *intrinsic* PUFs. In e*xtrinsic* designs, responses are typically evaluated externally, and random features are introduced explicitly. For example, one of the popular PUF systems is called coating PUF, where the surface of an electronic chip is sprayed with randomly distributed dielectric particles [28], so the randomness of the system is explicitly generated during the manufacturing.

The last classification approach and the most relevant to this dissertation is based on security parameters of CRP behavior. In this regard, PUFs can be distinguished as a *weak* or *strong* PUFs. The main difference between those two types is in the domain of CRP library or the total number of possible CRPs. A weak PUF generally have a very small number (sometimes even one) of CRPs whereas strong PUFs are characterized by a huge

set of challenge-response pairs. This distinction has an immediate effect on usage scenarios and adversarial attack scenarios against weak and strong PUFs. To "hack" weak PUF one simply needs to know all the CRPs and there are very few of them, whereas attacking strong PUFs it is much more challenging as characterizing the CRP space within a limited timeframe is exceedingly difficult due to the large CRP dataset. Clearly, strong PUFs exhibits higher security than weak PUFs, hence they are generally used in different application areas.

1.4.6 PUF Implementations

In this subsection, I review a non-exhaustive list of various PUF proposals and designs each of which has their own advantages and drawbacks. This list includes optical [29, 30], electronic [4, 28], acoustic [28] and coating PUFs [28]. As it is shown later, optical systems promise higher resistance to adversarial attacks and exhibit stronger security in general, while electronic PUFs are still vulnerable to cloning. On the other hand, electronic PUFs are much easier to implement and integrate on silicon chip allowing for mass production and inexpensive manufacturing. Generally, optical approaches are much more complex systems requiring high-precision mechanisms and bulky setups.

1.4.6.1 Electronic PUFs

Electronic PUFs with innate randomness coming from delayed measurements of a signal are called *delay – based intrinsic PUFs*. One the most popular approaches using this

technique are Arbiter PUF [31, 32] and Ring Oscillator PUF [4, 33]. In **Figure 1.9** schematic diagrams for both of architectures are presented:

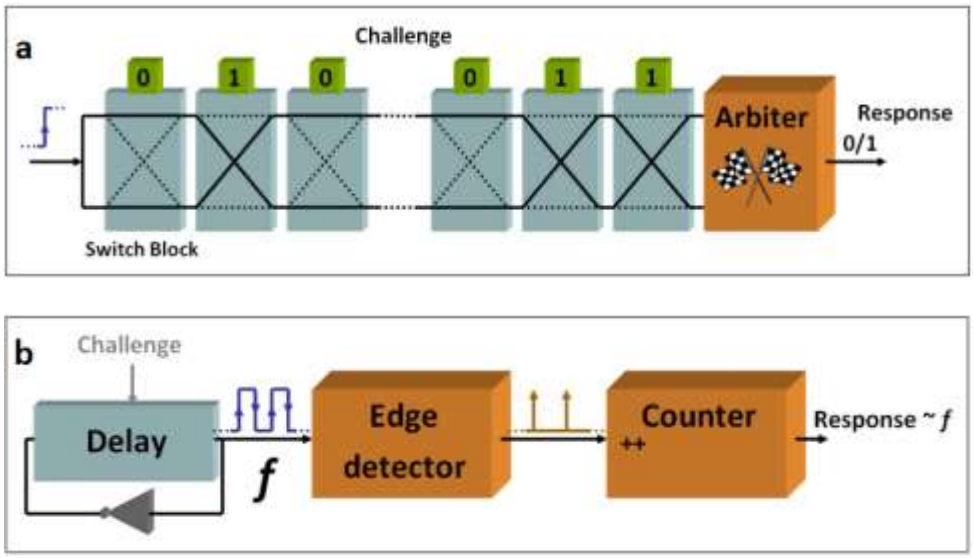


Figure 1.9: a) Arbiter PUF b) Ring Oscillator PUF [34].

The basic idea of Arbiter PUF is a race condition on two electronic signals propagating through two paths on a chip. The arbiter module indicates which of the two paths won the race by detecting the signal that comes earliest to the module. The intrinsic randomness of Arbiter PUF comes from the fact that during the chip manufacturing process it is impossible to fabricate two paths with zero delay between them. Therefore, there is always a small random offset between the two delays. Moreover, Arbiter PUF consists of so-called switch blocks in the initial design [31, 32]. By choosing the way of connecting inputs to outputs (straight or crossed), the challenge signal is set up. The number of unique challenges in Arbiter PUF is exponential in the number of those switch blocks making Arbiter PUF a strong PUF.

Another approach based on delays in electronic signals is Ring Oscillator PUF. According to the diagram above, the output of delay block is fed back to its input making this block as an oscillating loop. The frequency of this oscillator depends on the amount of the delay introduced in each round of signal propagation. Random manufacturing variations make this frequency also random and unpredictable. Typically, the delay is parametrized by the external challenge and in the initial proposal of Ring Oscillators they were used in pairs in parallel so that the counter blocks are compared between two delay blocks. The number of ways to choose two oscillators out of total number N oscillators is proportional to N^2 thereby making Ring Oscillators correspond to the class of weak PUFs.

Two PUF constructions described above were based on the random delays during the fabrication of chips. Besides that, there is a plenty of other electronic PUFs based on fluctuations and instabilities of digital memory such as SRAM (static random-access memory) [27], Butterfly [35], Latch PUFs [36] etc. On top of that, extrinsic PUFs such as Coating PUF (**Figure 1.10**) is also considered an electronic PUF as discussed earlier [28]:

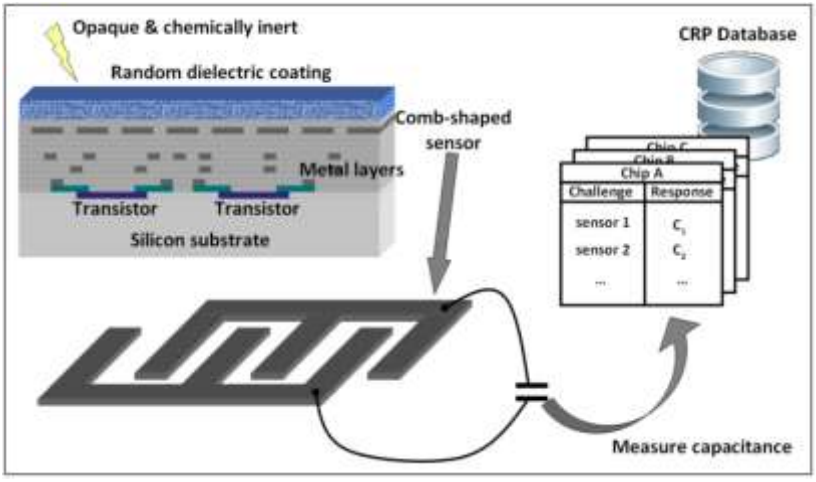


Figure 1.10: Principle of Coating PUF operation [34].

In Coating PUF random features are introduced not during the manufacturing but by explicit spraying of the dielectric coating on top of the sensors on the chip. After that, the capacitance of surfaces of comb-shaped sensors is measured and written down in the CRP library as a response. The challenge in this approach plays the role as the sensor itself.

Here, I presented a concise list of the most common electronic PUF architectures so that the reader can have a general idea of details of techniques and implementations. Although electronic PUFs remain the most common approach, due to easy and cheap production and the ability to integrate them on silicon chips, many research papers have found them to be susceptible to cloning, model building attacks, and invasive attacks. For example, given the linear structure of Arbiter PUF, mathematical algorithms have been developed to be able to predict the response to a certain challenge [37, 38]. To leverage the security of Arbiter PUFs many other modifications were proposed such as Feed-Forward, XOR, Lightweight architectures, but even these countermeasures failed to resist simple machine learning attacks using logistic regression or support vector machines [38 – 40]. In addition, electronic weak PUFs with finite CRP library size are easy targets for adversaries to read out all CRPs and fully characterize the device. Given all the weaknesses of electronic PUFs, there is significant recent interest in developing other PUF constructions based on different materials or on a completely different scheme.

1.4.6.2 Optical PUFs

Pappu et al. first proposed an optical PUF in 2001 where he gave a comprehensive overview of Physical One – Way Functions (POWF) [29]. The basic idea of Pappu's approach was to exploit an optical inhomogeneous medium filled with glass spheres (500

μm). This token was illuminated by a continuous beam of a HeNe laser and due to multiple scattering of the light the speckle pattern was detected on the CCD camera ready for further computational processing. The variation of the orientation of laser results in different speckle patterns thereby making the angle of illumination a challenge and Gabor filtered (the special case of Fourier transform) speckle image a response. The scattering medium in this approach is impossible to copy even for the original manufacturer of the token due to inherent random small-scale fluctuations of the microsphere positions. It is also tamper evident since the physical intrusion of an attacker will destroy the original design of the medium. The basic operation of optical PUF is depicted in **Figure 1.11** [34]. Even though Pappu's PUF was never successfully attacked and remains to be very secure, there are plenty of shortcomings associated with this system. First, optical scattering setup requires a free space optics which is hard to align and stabilize.

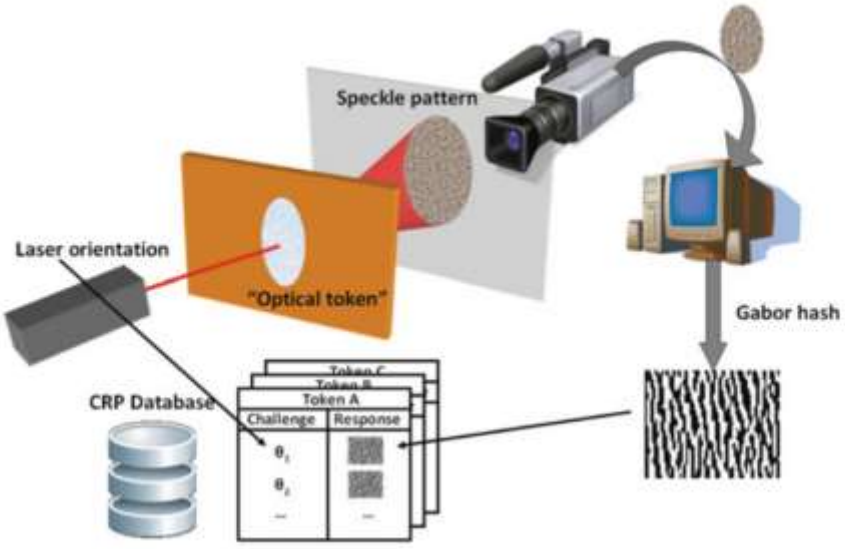


Figure 1.11: Basic operation of optical scattering PUF [34].

This results in unstable and error-prone readout of speckle patterns harming the repeatability and consistency of the PUF. Further, the large setup with the precise mechanical positioning of laser and slow readout speed make the whole system complex and expensive to build. The latter motivated researchers to find another way to miniaturize Pappu's method onto a single device and provide more integrated and more robust designs such as the Optical Scattering PUF probed via a spatial light modulator (SLM) developed by Horstmeyer et al. [41]. In this technique rather than changing the position of the laser source, Horstmeyer et al. utilized an SLM to modulate the phase of an optical wavefront that is further focused on the scattering token. The results turned out to be slightly more stable and robust, therefore, more repeatable. Later in 2013, the same idea of exploiting a volumetric scattering medium was refined by Rührmair et al. who proposed an integratable optical PUF [42]. Instead of a single laser source, he suggested two theoretical approaches where in the first scheme he proposed an immobile array of phase-locked laser diodes to illuminate the disordered scattering medium (**Figure 1.12a**). Each of diodes can be independently switched on and off leading to 2^k challenges, where k is the number of diodes in the array. In the second approach, he proposed a single laser source that passes a light modulator array as it is shown in **Figure 1.12b**. Lastly, Rührmair built a real experimental prototype (**Figure1.12c**) where he investigated the security of the system.

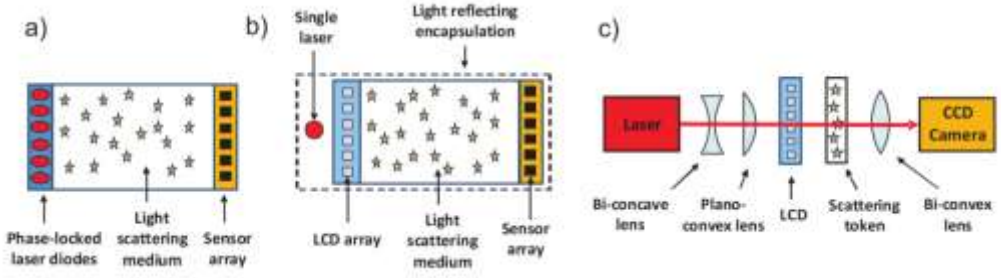


Figure 1.12: a) and b) are the theoretical types of integrated PUF systems. c) schematic illustration of the prototype [42].

Intriguing recent optical approaches are based on exploiting a scattering medium with *quantum* challenges consisting of several photons state. The main idea of the method developed by Goorden et al. is in quantum secure authentication (QSA) where adversarial attacks are excluded by the quantum mechanical properties where exact determination the quantum state of photons is prohibited given the small number of photons [43]. QSA exhibits greater security, but building the aforementioned system is still expensive and laborious with limited readout time of responses (100 ms).

During the last two decades since the first optical PUF was proposed there is a variety of other techniques such as laser surface authentication [44], authentication schemes using CDs [45] and etc. The main barrier for widespread practical usage of those techniques is the slow operation of the setup, their size, and complexity as well as the cost not to mention a high vulnerability to adversarial attacks in many implementations. In general, optical PUFs are proven to be more secure and resistant to third party attacks but this comes at the cost of complexity of the system. Given all the aforementioned practical and theoretical challenges, it is highly important to develop more advanced optical PUF technology that is easy to build, cheap and compatible with electronic circuits, fast in terms of readout and key generation, resistant to adversaries and possessing large information content.

In Chapter 3 I present a new class of PUF systems called *photonic* PUFs developed by our group [26]. Photonic PUFs benefit from a range of factors such as usage of CMOS-compatible silicon materials, operation in telecommunications spectrum bands, compatibility with electronic integration, large optical nonlinearity, and large information

density. Those benefits make photonic PUFs an attractive and desirable choice for information security applications.

1.5 Applications of PUFs

PUFs are mainly used in two areas of hardware cryptography: system identification and secure key generation.

1.5.1 Low-Cost Authentication

The challenge-response protocol allows PUFs to be a secure and inexpensive way to authenticate the objects. The CRP library that is collected during the enrollment phase can be securely stored on a server and when the client wishes to, for example, access his/her bank account, some set of CRPs are chosen from the library and applied to the client's PUF circuit in a secure terminal. Measured responses are compared with the expected responses to determine authenticity. It is very important that challenges should be chosen at random to prevent malicious attacks.

This standard protocol was implemented on RFID tags [46, 47]. According to this paper, utilizing 128-bit response results in few parts per billion of FAR and FRR values. The use of PUFs has also been suggested in anti-counterfeiting of integrated circuits by embedding the PUF within the integrated circuit [48]. Furthermore, PUFs are well suited for a variety of other fields such as IP protection

and tracking [27], smart credit cards with built-in PUF chips [49], wireless sensor network security [50] etc.

1.5.2 Key Generation

In any encryption and secure information communication scheme a secret key is required. PUFs are found to be useful sources of such keys since the functional operation of PUF devices is ideally random and can't be duplicated. The first proposal for a secret key generation was by Suh et al. [51] and then later was studied by O'Donnell et al. [52] in the context of random number generation. As I mentioned earlier, the responses of PUF systems are typically noisy and highly dependent on the environmental conditions preventing the direct usage of a PUFs output as a source of a reliable secret key. Generally, after the responses are measured the implementation of error correction techniques such as *fuzzy extractors* are adopted in order to increase the privacy and security of generated key [53, 54]. The main idea of fuzzy extractors is to provide the secure, uniformly random and reliably reproducible random output from the noisy response generated by certain cryptographic primitive.

1.6 Dissertation Outline

To summarize this introductory chapter, I presented an overview of the concept of Physical Unclonable Functions, their main goals and challenges encountered in the existing cryptographic primitives. I discussed how PUFs can potentially solve the aforementioned

limitations such as non-reliable storage of secret keys, susceptibility to cloning and duplication, non-compatibility with electronic circuits.

The rest of this dissertation is structured in the following way:

In Chapter 2, I present an overview of adversarial attacks against PUF systems using various approaches such as Machine Learning attacks, Side Channel attacks, Invasive and Non- invasive attacks depending on the type of PUF. I investigate the trade-offs between attacking Strong and Weak PUFs. I provide the most up to date results of spoofing Electronic PUFs, perspectives and ideas of attacking Optical PUFs and describe the main challenges associated with it. On a final note, I give a brief motivation for developing novel optical PUF system that is resistant to Numerical Attacks and compatible with CMOS platforms.

In Chapter 3, I present the silicon photonic PUF, which is the primary focus of this dissertation. I describe the main characteristics of this PUF device: repeatability and uniqueness, information capacity and key generation rates. I briefly describe the physical processes that influence light propagation inside the cavity and in particular the influence of nonlinear optical phenomena. I conclude Chapter 3 by raising a question regarding the unclonability of silicon photonic PUF, hence motivating the research work of constructing a numerical modeling attack on it.

In Chapter 4 theoretical investigations of attacking a simplified silicon photonic PUF is presented. I simulate the propagation of light in the device under different scenarios introducing a step by step nonlinear optical phenomena: dispersion, self-phase modulation, etc. In each scenario I conduct the set of machine learning attacks against challenge-response behavior gathered throughout the simulation codes.

In Chapter 5 I construct machine learning attacks against an experimental silicon photonic PUF using experimentally acquired data. I utilize a Deep Neural Network as a machine learning method since it was acknowledged as the state-of-the-art technique outperforming other traditional solutions such as Support Vector Machines or Random Forests. I demonstrate that unlike the simplified simulated processes, silicon photonic PUFs are resistant to two possible attack scenarios. I conclude that this resistance is rooted in the true complexity of the optical nonlinearity and the sensitivity to precise conditions resulting from the device's ray chaotic design. Lastly, I investigate the application to encrypted data storage and compare the results of decryption using genuine PUF device and machine learning "clone".

In Chapter 6 I investigate eavesdropper (Eve) attacks against a simulated optical scattering PUF. Similar to silicon photonic PUF attacks, I first generate the CRP library via speckle simulation codes. Having this artificial dataset, I provide machine learning algorithms attempting to emulate the challenge-response behavior. Further, being in an attacker role, I explore the amount of total CRP dataset the adversary has to possess in order to successfully spoof a scattering PUF as a function of device size. In other words, I answer the important question of the minimum subset of CRP database an adversary should have in order to successfully predict the response to any given challenge.

In Chapter 7, I conclude this work with an overview of future steps in this direction. Artificial Intelligence and Machine Learning are fast-growing areas with novel approaches introduced from year to year. I demonstrate that silicon photonic PUF is resistant to the state-of-the art family of algorithms referred to as Deep Learning, but it is more important to ensure its resistance to future algorithms. Regarding the optical scattering PUF, there

are still several challenges that exist including the stability of the system in general as well as the study of its security. Therefore, the next steps in this direction would be to apply ML attacks against experimental data and investigate the dependence of the results on a range of parameters such as the size of the scattering volume, the dimension of the CRP space, the speed of response readout, etc.

Chapter 2 : Attacks on PUF systems

2.1 Introduction

Physically Unclonable Functions are a promising security technology that has a strong identification capability and can be applied in authentication as well as secure communication. Theoretical works on applying PUFs presume that PUFs are reliable and not susceptible to adversary cryptographic primitives due to their unclonability, uniqueness and repeatability properties. However, it has been shown that most electronic PUFs can be emulated with software algorithms. On top of that, there is no comprehensive study on the full unclonability of electronic PUFs making such security analysis an ongoing area of research.

In this chapter, I present an overview of different attack methods targeted against various PUF systems. Specifically, it was shown that certain PUF constructions are susceptible to attacks ranging from invasive to non-invasive attacks [55]. Invasive attacks are typically accompanied by physical modification of the PUF to study its structure and understand its implementation. Non-invasive attacks, on the other hand, are implemented without any physical interaction with PUF. This last type of attack is generally applicable solely to Strong PUFs due to their huge domain of CRP database.

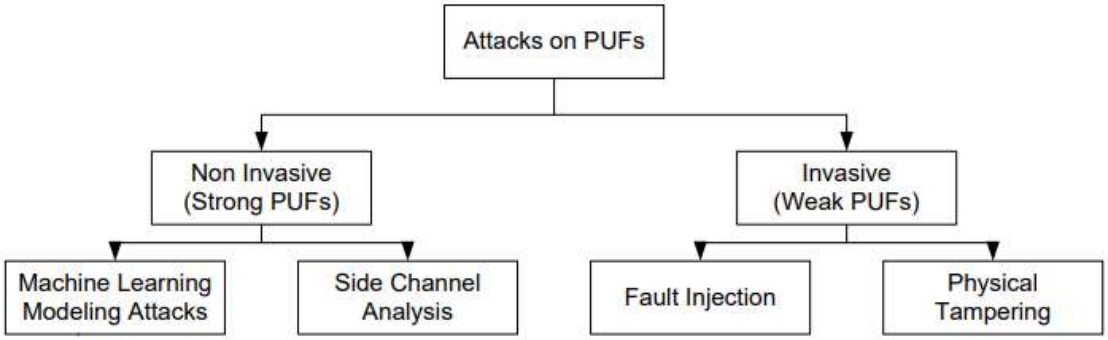


Figure 2.1: Type of attacks on PUFs [55].

According to **Figure 2.1**, weak PUFs and strong PUFs are generally attacked in different ways. Numerical Attacks are not relevant for Weak PUFs due to the insignificant size of the CRP library. To characterize the weak PUF system it is sufficient to read out all the challenge-response pairs within a reasonable time window and thus with complete knowledge of the Weak PUF's behavior numerical attacks are irrelevant. Therefore, to keep Weak PUF secure all CRPs must be kept secret.

2.2 Attacks on Weak PUFs

Generally, to break the security of Weak PUFs fault injection and physical tampering methods are adopted. According to [56], fault injection attacks aim to introduce erroneous behavior in a device by manipulating it in some way leading to the faulty recovery of the key. These manipulations could be introduced in many ways, e.g., exposing the device to extreme environmental conditions or injecting a transient fault to specific components of the device. The Ring Oscillator PUF, that was described in section 1.4.6.1, can be broken by increasing the fraction of unstable CRPs during the enrollment phase

under change of environmental factors such as temperature, DC supply voltage [57]. Memory-based PUF systems are also shown to be attacked successfully exploiting fault injection method where Oren at al. cloned SRAM (static random-access memory) PUFs using the remanence effect of memory decay, where the data, written in a volatile memory, is typically not immediately lost after power-off [58]. In addition, SRAM PUFs are susceptible to physical tampering attacks and side-channel analysis [59]. Rührmair et al. demonstrated the emulation attacks on Ring Oscillator PUF, where he selected the most straightforward way of reading out all CRPs in a database, size of which is $O(k^2)$, where k – is the total number of oscillators [37]. Applying the quicksort algorithm to all the frequency outputs of PUF, he was able to obtain prediction results surpassing 99% accuracy with a different number of oscillators in the system. Importantly, the whole procedure of attacking takes a polynomial time since it is based on sorting algorithms.

Relative to Weak PUFs, Strong PUFs exhibit higher security potential in terms of resistance to adversary leading to the development of other attacking methods that are covered in the next section.

2.3 Attacks on Strong PUFs

Currently, the most relevant attacks on Strong PUFs are called modeling attacks [37, 60, 40]. Unlike Weak PUFs, Strong PUFs possess a sufficiently large CRP space that it is presumed to be ineffective to capture the entire CRP space. In modeling attacks, an adversary (Eve) presumably has acquired a subset of the CRP space associated with the attacked PUF. Then Eve uses this subset to derive a numerical model, which is an algorithm

or computational procedure that mimics the behavior of the authentic PUF by providing the correct responses to an arbitrary challenge with relatively high accuracy. If one finds such an algorithm, then the security of PUF is compromised. One of the most popular and powerful tools among modeling attacks is based on artificial intelligence and machine learning [62]. I would like to stress again that modeling attacks on Weak PUFs are inappropriate since the core idea of machine learning model is to learn PUF behavior based on the subset of the extremely large CRP database only. Weak PUFs have a few CRPs or sometimes even one. Therefore, no emulation of weak PUF based on known CRPs is needed.

Considerable effort has taken place to successfully attack electronic PUFs including Arbiter PUF and its variations. Rührmair et al. in his survey presented the extensive existing research in attacking Arbiter, XOR, Feed – Forward and Lightweight Arbiter PUFs using Logistic Regression (LR), Evolution Strategies (ES) and Support Vector Machines (SVM) techniques that are studied in ML community [37, 62]. The simplest form of an Arbiter PUF that consists of a sequence of k stages was successfully attacked by LR where the subset of CRPs was used for training step and the rest of CRPs was used as a test set. Mathematically, Arbiter PUF can be represented as linear delay model, where the final delay $\Delta = \vec{w}^T\vec{\Phi}$ is the sum of intermediate delays at each stage, where $\vec{w}$ is the parameter vector of corresponded delays and $\vec{\Phi}$ is a feature vector of applied challenge both with $k+1$ dimension (including bias). [63]. Taking the sign function of the final delay, $sgn(\Delta)$ the response of Arbiter PUF can be calculated. This response serves as a label for supervised learning during the training step. Hence, given the described mathematical model of Arbiter PUF, I obtain a typical classification task in machine learning. Logistic regression

is a natural method to tackle this problem where LR determines the decision boundary by learning a parameter vector via maximum likelihood optimization steps. The whole optimization is carried out using the known to adversary CRPs and after the parameter vector is optimized the final prediction results of LR are evaluated on unknown CRPs. The best results among other ML algorithms (SVM and ES) were achieved by LR and are shown in the table below [37]:

ML Method	Number of stages	Prediction Rate	CRPs
LR	64	95% 99% 99.9%	640 2555 18050
LR	128	95% 99% 99.9%	1350 5570 39200

Table 2.1: Logistic Regression attack on Arbiter PUF with 64 and 128 linear stages.

It is clear that the Arbiter PUF is easily emulated by a mathematical model with the accuracy of > 99% using the certain size of CRP database. It is important to point out that a PUF is considered to be attacked successfully if the computational complexity of the adversary's model is a low-degree polynomial in terms of Arbiter PUF stages. Rührmair et al. demonstrated that this condition is satisfied in his experiments showing that the total number of computational steps is equal to $O\left(\frac{k^2}{\varepsilon}\ log\frac{k}{\varepsilon}\right)$, where k is the number of stages in arbiter PUF and ε is the classification error of LR. Thus, the Arbiter PUF in its simplest form is vulnerable to mathematical clonability, which has led to the development of modifications to strengthen the resilience against machine learning. One well-known possible way to reinforce the Arbiter PUF is to use the XOR Arbiter PUF architecture, where l number of Arbiter PUFs are used in parallel each with k stages. The same challenge

is applied to all *l* Arbiter PUFs and the global response is an XOR operation of individual outputs. Another approach to make Arbiter PUF more secure is a Lightweight PUF that has similar to XOR architecture but each individual Arbiter PUF is interrogated with its own challenge bit sequence [64]. Lastly, the most resilient Arbiter PUF variant against ML attacks is called Feed-Forward Arbiter PUF [31, 63]. In this approach, some of the stages are not subjected to external challenge bits but rather dependent on the delay values accumulated at stages before. According to Rührmair et al. even these new modifications can still be broken with Logistic Regression or Evolution Strategies. The main results are shown in the table below [37]:

PUF Type	XORs/ Loops	ML Method	Number of stages	Prediction Rate	CRPs ($x10^3$)	Training Time
Arbiter	N/A	LR	128	99.9%	39.2	2.10 sec
XOR	5	LR	128	99.0%	500	16:36 hrs
Lightweight	5	LR	128	99.0%	1000	267 days
FF	8	ES	64	95.5%	50	46 days

Table 2.2: Summary of ML attacks against Arbiter, XOR, Lightweight and Feed-Forward PUFs.

On a final note, it has been found that all Arbiter-based PUFs are successfully attacked with both low-degree polynomial training time and the number of internal parameters. Besides the vulnerability to attacks, the instability of XOR-based approaches increases with the number of stages harming the essential repeatability properties of an ideal PUF system.

For the sake of a complete picture, there is a vast number of other machine learning techniques such as Bagging & Boosting, Ensemble Learning, Unsupervised learning that are also used to perform modeling attacks against Strong PUFs [61, 65]. With the advent

of massive computation power and data parallelism, Deep Learning (DL) techniques have rapidly become the dominant and most powerful tool outperforming conventional machine learning algorithms on benchmark tests in artificial intelligence research. Yashiro et al. showed that DL can also be used as a successful modeling attack against PUF primitives [66].

So far, I have discussed attacks against electronic Strong PUFs. Surprisingly, very few attacks have been reported against optical PUFs. Specifically, the original scattering PUF designed by Pappu et al. has never been successfully broken due to multiple reasons including extremely large information content (2.37×10^{10} challenges), complex optical refractions within the scattering medium and practical limitations such as confined illumination area of the medium [30, 37]. Given these factors, non-integrated scattering optical PUFs demonstrate an unprecedented level of security, although unfortunately at the cost of laborious, expensive, impractical implementation. However, the more practical integrated optical PUF prototypes developed by Rührmair et al. turned out to be broken by Linear Regression models [42]. The main assumptions of the attacks in the given paper are: *i)* linear scattering medium, *ii)* an adversary has a direct access to raw speckle images. Moreover, as it has been claimed in [42] to remedy the failure of resistance to attacks, Rührmair et al. suggested exploiting nonlinear scattering materials that is still an essential open problem in the research.

2.4 Security vs Practicality

After a review of electronic and optical PUF constructions, there is one clear trade-off that can be identified. Optical PUFs are shown to be much more robust to malicious attacks and unclonable, but the construction of devices, as well as compatibility with electronic chips, are very challenging. At the same time, electronic PUFs are widespread in usage and easily integrated with chips, but they are easily threatened by modeling and invasive attacks. For clearer demonstration, I show different PUF devices on a 2D plane where the y-axis indicates the compatibility with electronic circuits, while x-axis measures the strength of PUF in terms of challenge domain size. **Figure 2.1** is rather qualitative than quantitative to give a better general view of the most up to date trends in PUF research area. Scattering based and Quantum PUFs, as it was mentioned earlier, are very robust systems with low perspective of embedding them on chips. The brightest examples of electronic PUFs such as Arbiter, SRAM and Ring Oscillators were initially and inherently in nature designed on silicon platforms.

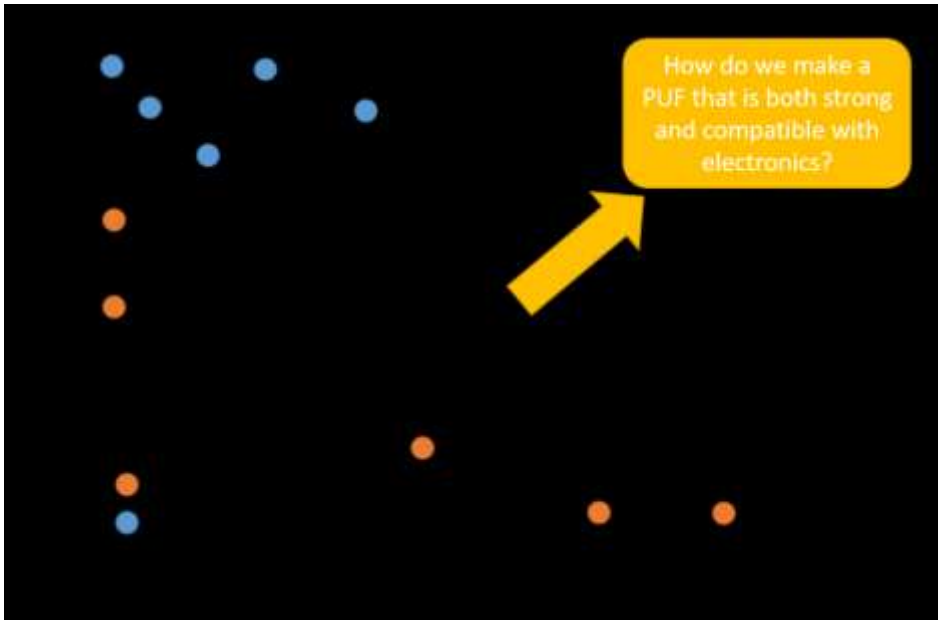


Figure 2.2: PUFs strength and electronic compatibility trade-off [67].

Given this landscape, it is natural to raise the question of developing a new approach that potentially can possess advantages from both types of PUFs. Specifically, a new PUF technology should satisfy the following requirements:

a) easy integration with semiconductor electronic circuits

b) simple and cheap to produce

c) robust during its operation and reliable in usage

d) complex enough to resist machine learning attacks

The answer to this question, I believe, can be found in integrated silicon photonics with the introduction of a new class of devices termed as *silicon photonic PUFs*. Photonic PUFs benefit from a complexity of optical interactions and at the same time are easily compatible with well-developed silicon microelectronics platforms. The optical operating wavelength of silicon photonic PUFs allows them to be compatible with telecommunications infrastructure thus minimizing the additional costs of future deployment of these devices.

Further, the silicon as a core material of photonic PUFs is well-known for offering a variety of complex nonlinear optical effects under specific external conditions. This can be used for maximizing the complexity of device operation as well as the security enhancement. As a result, photonic PUFs demonstrate a promising alternative for developing highly secure, robust and repeatable, low size, and cost source of private information for potential applications in a hardware integrity and information security. In the next chapter, I present an overview of silicon photonic PUF prototype developed by Grubel et al. [26, 67].

Chapter 3 : Silicon Photonic Physical Unclonable Function

The original work was developed by Grubel et al. [67, 68, 69] and here I review the main properties that are the most relevant to the rest of this dissertation.

3.1 Introduction

The Silicon Photonic PUF was first introduced by Grubel et al in 2017 [26, 67, 68, 69]. It is the novel CMOS-compatible cryptographic device based on ultrafast nonlinear optical interactions in a silicon microcavity that is designed as a disk-shaped resonator with a chamfer, which generally exhibits chaotic behavior. An electron microscopy image of an example microcavity is shown in **Figure 3.1** with 30-μm diameter.

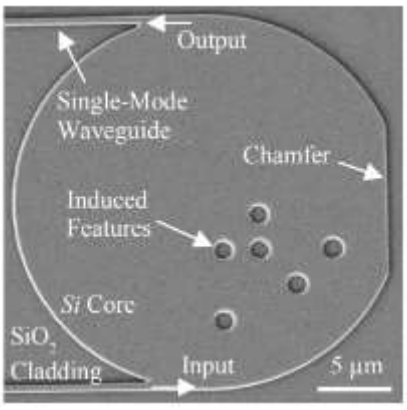


Figure 3.1: Scanning electron microscopy (SEM) image of an example PUF cavity [26].

Random physical idiosyncrasies of the cavity such as film thickness, material impurities, surface roughness, artificially induced features, and the precise geometry shape enhance the device's resilience against cloning and other adversarial attacks. Even with the complete knowledge of the design, these physical structures are impossible to replicate exactly, even for the original manufacturer, thus making the devices unique. In addition, at high optical power levels the non-linear behavior of silicon increases the complexity of the challenge-response relationship, thus providing necessary properties such as one – wayness, unclonability, and unpredictability. Robust optical coupling via add and drop single-mode waveguides facilitates the high repeatability of the device behavior.

The certain shape and the geometry characteristics of the micro-cavity are the result of computational optimizations and finite difference time-domain simulations over diameter, chamfer orientation, and chamfer size. As a result of these optimizations, a general trade-off between the input-to-output loss and cavity lifetime was observed. Specifically, large cavity geometry results in a longer photon lifetime and increased optical losses, whereas the smaller diameters exhibit decreased optical loss and photon lifetime thus less complexity of input-to-output relationship but greater signal strength which improves repeatability. Given this trade-off, an optimal point was found and a device with 30-μm diameter was selected as a baseline device.

In [26, 67, 68] Grubel et al. provide an extensive work on silicon photonic PUFs beginning from the details of the ray chaotic design and fabrication processes of the cavity and finishing with the applications of photonic PUFs in an encrypted communications and system authentication. However, one of the future steps indicated in the previous work is

the evaluation of unpredictability through measuring resistance to emerging machine learning attacks. The analysis of such attacks is the subject of this dissertation. Therefore, in this chapter, I provide an overview of the previous work that is the most relevant to adversarial attacks on silicon photonic PUFs. In particular, I will describe the nonlinear properties of the photonic PUF that increase the complexity of the optical interactions within the cavity thereby enhancing its security (section 3.2). These properties, as we will see later in Chapter 5, play a crucial role in the performance of modeling attacks on silicon photonic PUFs. Further, Grubel et al. investigate photonic PUF as an authentication token in a challenge-response protocol with thorough description of the challenge and response signals, their generation procedures and further post-processing algorithms (section 3.3). Lastly, the physical unclonability of silicon photonic PUF is proven with the set of FHD distributions for the set of physical clones of the device, leaving the mathematical unclonability or vulnerability to ML attacks as an open question. It is important to recall that PUF is truly unclonable if it is both mathematically and physically unclonable.

3.2 Nonlinear Properties

In this section, I give a very brief background on the nonlinear optics and cover a few of the nonlinear processes that are relevant to silicon photonic PUF operation.

3.2.1 Nonlinear optics

The first strides towards the studying the nonlinear optical processes are dated back to 1870 when John Kerr discovered the change in the refractive index of solids and liquids

under the strong external DC field [77]. This now well-known phenomenon is called Kerr effect. Further substantial progress in this direction was made with the advent of intense light sources, i.e. lasers, in 1960 and since then nonlinear optics continues as an active field of research with innumerable applications.

For many years until the Kerr effect was discovered it was considered that optical materials respond linearly to an applied external electric field. In linear conditions, beams of light do not interact (superposition principle). However, later it was realized that this assumed linear response is valid only at small field strengths. From a theoretical point of view, the linearity of the medium's response can be described by the relationship between the induced polarization density field and the electric field

$$\vec{P}(t) = \varepsilon_0 \chi^{(1)} \vec{E}(t),$$

where ε_0 is the permittivity of free space and $\chi^{(1)}$ is known as the linear susceptibility. In nonlinear optics this relationship is generalized by presenting the polarization field as a power series of electric field strengths

$$\vec{P}(t) = \varepsilon_0 \left[\chi^{(1)} \vec{E}(t) + \chi^{(2)} \vec{E}^2(t) + \chi^{(3)} \vec{E}^3(t) + \ldots\right]$$

The terms $\chi^{(2)}$ and $\chi^{(3)}$ are the second- and the third-order nonlinear optical susceptibilities and the terms proportional to $\vec{E}^2(t)$ and $\vec{E}^3(t)$ represent the second- and third-order nonlinear polarization effects, respectively. For simplicity, I take optical susceptibilities as constants whereas in the general case they depend on the frequencies of the external field. In the simple case (and ignoring the vector direction), the external electric field $E(t) = A\ cos\ (wt)$ creates a dipole moment *p* per atom aligned with the applied field, or $P = Np$, where *N* is the atomic number density. Therefore, the polarization field up to second order is aligned with the applied field and its magnitude is given by,

$$P(t) = \varepsilon_0 \left[\chi^{(1)} A \cos(wt) + \frac{1}{2}\chi^{(2)} A^2 (1 + \cos(2wt))\right].$$

The second order term of polarization consists of a contribution at zero frequency $\sim \chi^{(2)} A^2$ and a contribution at the frequency of $2w$. The latter leads to the generation of the second harmonic radiation which was observed experimentally by a team led by Peter Franken [78]. In his experiment, a slab of crystalline quartz was illuminated by a ruby laser with intense radiation and $\lambda = 694.3$ nm resulting in a detectable second harmonic radiation at $\lambda = 347.15$ nm.

Such nonlinear optical effects are the root of a range of fundamental physical mechanisms at the atomic and molecular level of the material. The classical approach of treating optical nonlinearities is based on the extended Lorentz model of the atom with additional quadratic displacement terms under the external force [79, 80]. The main shortcoming of this model is that this approach is based on the single resonance frequency of the atom. In addition, it is not suitable for the external fields with the frequencies much lower than the resonance frequency of the material system. Therefore, the quantum mechanical theory of nonlinear susceptibility was developed to describe the atom with many energy eigenvalues, hence with many frequencies [79, 80]. According to this theory, the underlying origin of the nonlinearity is hidden inside $\chi^{(n)}$ coefficients that are calculated with the time-dependent perturbation theory. Moreover, the values of these coefficients are directly related to the symmetry properties of the material. The consequence is that $\chi^{(2)}$ is non-zero only for materials that are non-centrosymmetric, whereas $\chi^{(3)}$ is non-zero for all media including those with $\chi^{(2)} = 0$. Therefore, second harmonic generation (SHG), for example, is prohibited in gases, amorphous materials, and centrosymmetric crystalline materials (e.g. diamond lattice) etc. The third order term

$\chi^{(3)}\vec{E}^3(t)$ gives rise to a variety of phenomena including third harmonic generation (THG), the Kerr effect, four-wave mixing (FWM), self-phase modulation (SPM), and two-photon absorption (TPA). Due to the vastness of nonlinear optical effects and theory, in the remainder of this sections I will cover only those effects that are relevant to the silicon photonic PUF.

The four-wave mixing (FWM) process, as it is demonstrated later in the silicon cavity, is one of the third-order nonlinear processes based on the mutual interaction between two or three lightwaves to produce light at new wavelengths. The basic idea of this phenomenon consists in the response of the medium to the propagation of two strong waves of angular frequencies w_1, w_2 that are typically referred to as *pumps.* As a result, two new frequencies are produced $w_3 = 2\ w_1 - w_2$ and $w_4 = 2\ w_2 - w_1$. When the process is seeded by either of the new frequencies this seed frequency is called the *signal* and the unseeded frequency is called the *idler*. This case of four different frequency components interacting with each other is called *non-degenerate* FWM. However, there is more commonly a case of *pump-degenerate* FWM where the two pump lightwaves are at the same frequency. For example, a single pump can provide the amplification for signal and idler components. In the next section, FWM along with other nonlinear processes were experimentally observed in the silicon resonator by observing the output spectrum and identifying the new frequencies generated from two input laser pulses centered around two different wavelengths.

Another third-order effect is caused by the variation of the refractive index at high intensities of the external fields. It is shown that:

$$\Delta n \approx \left(\frac{3\chi^{(3)}}{4c\varepsilon_0\, n^2}\right) I \;=\; n_2\, I,$$

$$n(I) = n + n_2 I,$$

where I is the intensity of the field. This effect is known as the optical Kerr effect. The order of the magnitudes for coefficient n_2 is 10^{-16} in glasses to 10^{-2} in semiconductors. As a result of the optical Kerr effect, a strong lightwave experiences a self-induced phase shift known as self-phase modulation (SPM) during propagation in the medium. The phase shift due to the change in the refractive index can be calculated as follows:

$$\varphi = -n(I)k_0 L = -2\pi(n + n_2 I)L/\lambda_0 ,$$

$$\Delta\varphi = -2\pi n_2 \frac{L}{\lambda_0} I$$

Generally, the nonlinear phase shift caused by SPM is rapidly time-varying due to the rapid fluctuations in optical intensity typical of optical pulses. This time-varying phase shift results in the frequency shift of the lightwave and thus a change of the optical spectrum. Specifically, it often leads to a spectral broadening of the pulse. As it is experimentally demonstrated in the next section, strong SPM effects also lead to the spectral distortion of the signal at very high input powers. Self-phase modulation (SPM) plays a central role in laser pulse propagation in a nonlinear medium and a notable result is the formation of a stable type of optical pulses so-called *solitons* in suitably dispersive media.

Nonlinear processes are generally avoided in conventional optical systems (e.g. fiber optic communications) due to various distortions of the signal and resultant impaired performance of the system. However, I demonstrate that nonlinearity is a highly beneficial property for enhancing the security of photonic PUFs, through enhancing both their unpredictability and unclonability.

3.2.1 Nonlinear processes in the silicon cavity

Silicon material is well-known to exhibit a centrosymmetric property, thus third-order optical nonlinear effects are most relevant and, in this section, the presence of these phenomena is demonstrated.

There are several sources of nonlinearity in silicon devices: self-phase modulation (SPM), two-photon absorption (TPA), four-wave mixing (FWM), stimulated Raman scattering (SRS), etc. Each process is individually well understood and mathematically described, but most optical systems permit nonlinear effects in a collective manner. Generally, systems dealing with ultrashort pulses in nonlinear and dispersive media are described by the nonlinear Schrödinger equation (NLSE), due to the similarity of Schrödinger equation with a nonlinear potential term [79, 80]. The evolution of the pulse's amplitude *A(t, z)* which propagates in lossless medium with dispersion and SPM effect, is described as:

$$i\frac{\partial A(t,z)}{\partial z} = \frac{\beta_2}{2}\frac{\partial^2 A(t,z)}{\partial t^2} - i\gamma|A(t,z)|^2 A(t,z)$$

where β_2 – second-order chromatic dispersion and γ – the coefficient corresponding to Kerr nonlinearity. The NLSE above is in the simplest form since it doesn't reflect high-order dispersion and the other third-order nonlinear processes such as TPA, Raman Scattering, and linear loss of the pulse. NLSE is difficult to solve analytically in practice. Therefore, the NLSE is typically solved numerically with several well-known methods including split-step Fourier analysis [71]. We use this method for simulation of models of the photonic PUF in a presence of SPM and chromatic dispersion in the cavity (Chapter 5).

As already mentioned, nonlinearity is observed at very high light intensities or electric fields, so to demonstrate the presence of nonlinearity behavior of photonic cavity the output spectrum as a function of input power level is first examined on an optical spectrum analyzer (OSA). In particular, one of the spectral response to a certain challenge pulse is observed at three different input pulse energies. The combination of SPM, FWM, and TPA results in the distinct variations of the spectrum profile indicating that the photonic PUF is functioning in a nonlinear regime. In addition, FWM is demonstrated in one of the PUF device by inputting two 6.7 ps pulses at different wavelengths (**Figure 3.2**).

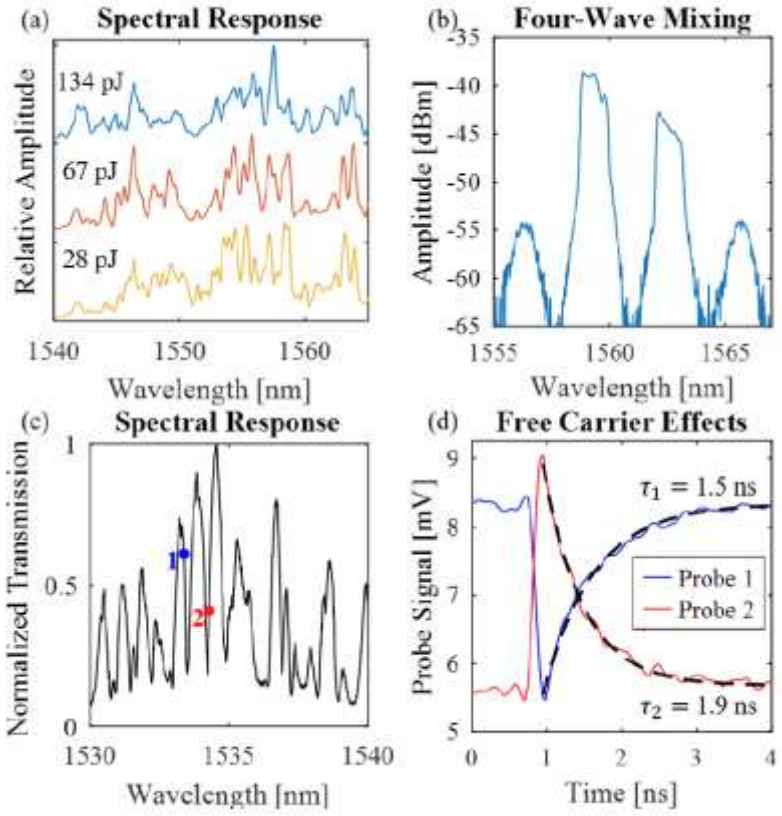


Figure 3.2: Nonlinear processes in silicon photonic PUF. a) Variations of spectral density in a response at different input laser pulse energies. b) Demonstration of FWM effect in a cavity by inputting two 6.7 ps pulses centered at υ_1 = 191.94 THz and υ_2 = 192.43 THz. Observed sidebands are centered at υ_3 = 191.57 THz and υ_4 = 192.80 THz. c) Spectral response of the cavity and two probe measurements. d) Temporal response of the two probes demonstrating the showing free-carrier dispersion effects.

By pump-probe measurement the presence of free-carrier dispersion (FCD) and free-carrier absorption was experimentally observed [83]. The pump of 3.5ps 300-pJ laser pulse from 90MHz mode-locked laser sent through a bandpass filter, whereas the probe is a continuous wave source. During the pulse propagation in the cavity, free-carriers are generated which introduce the absorption and the shift in the cavity's resonance. This can be observed by placing the probe at two locations of the transmission spectrum of the cavity and detecting the inverted temporal response. From the **Figure 3.6d**, the free-carrier lifetime can be determined to be approximately 1.9ns.

Spectral distortion of the signal, generation of new frequencies, the generation of free carriers and their impact on the semiconductor loss and refractive index all contribute to an extremely complex and unpredictable output response. Thus, nonlinearity is playing a critical role in constructing reliable photonic PUF.

3.3 Challenge – Response Authentication

Any strong PUF proposal is typically used in a system authentication. In [67] a silicon photonic PUF is exploited as an authentication token in a challenge-response authentication protocol. The sequence of spectrally-encoded ultrashort optical pulses constitutes a challenge signal whereas the post-processed optical output from the cavity is considered a response. The experimental setup for the authentication scheme is presented in **Figure 3.3**.

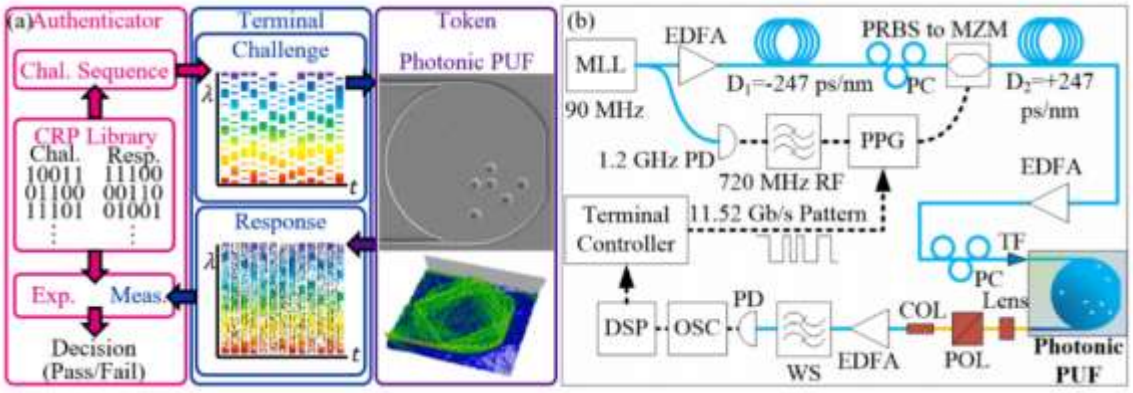


Figure 3.3: An experimental setup for testing an authentication protocol. a) An authentication protocol where the measured response is compared to expected response from CRP library associated with certain PUF token. b) Using Mach-Zender Modulator (MZM) a sequence of ultrafast pulses sourced from mode-locked laser (MLL) are encoded with binary sequences from a pulse pattern generator (PPG). After a series of compression and amplification of pulses they are sent to photonic cavity and the measured analog response is detected with photodetector (PD) [67].

The generation of challenge pulses is implemented in the following way. 300-fs mode-locked laser (MLL) pulses are stretched to 11 ns by dispersion compensating fiber (DCF) and the temporally broadened spectrum is modulated with a length of 128 pseudorandom binary sequence (PRBS) that is generated with pulse pattern generator (PPG). After encoding the optical pulses are compressed to 6-ps duration using the standard single mode fiber, amplified with an EDFA to a certain power level and sent to photonic PUF device. The amplification of light can be controlled on the EDFA and depending on the power level, the PUF functions in a varyingly nonlinear regime. For this experiment, the total number of unique challenge pulses generated during the enrollment of PUF is 8550. Each challenge pulse sent to the PUF results in analog response sequence that is further converted to digital power samples using an analog-to-digital converter (ADC).

For the sake of PUF quality evaluations, it is typically more convenient to operate with challenges and responses in binary form. Therefore, to extract a binary representation

of each response, a post-processing algorithm is applied to the digital power samples obtained in the experiment (**Figure 3.3**). The post-processing procedure is algorithmic in its nature and has no relation to the hardware operation. To derive the binary sequence from power samples, the probability density function (PDF) is estimated for response energies and an equalization procedure is implemented such that the probability of choosing any power value becomes equal. Using a Gray code conversion each power sample is converted to binary with specified number of resampling bits. Then, an XOR operation is performed on adjacent binary sequences. In the end, the results of XOR operation are appended together to create a single bit sequence that constitutes to a binary response. The total length of binary sequence per response depends on a number of resampling bits and the number of least significant bits (LSB) kept per sample.

As for authentication itself, the authenticator selects randomly a specific CRP from CRL, encodes the binary challenge via spectral patterning described above, sends the encoded optical pulse to PUF device and measures the post-processed analog response. The acquired response is compared to the expected response from CRL calculated during the enrollment process. The comparison is based on the FHD calculation. If two responses match up to a certain threshold value, the authentication is successful. The threshold for authentication is predetermined by the designer of the protocol to optimize system security and usability.

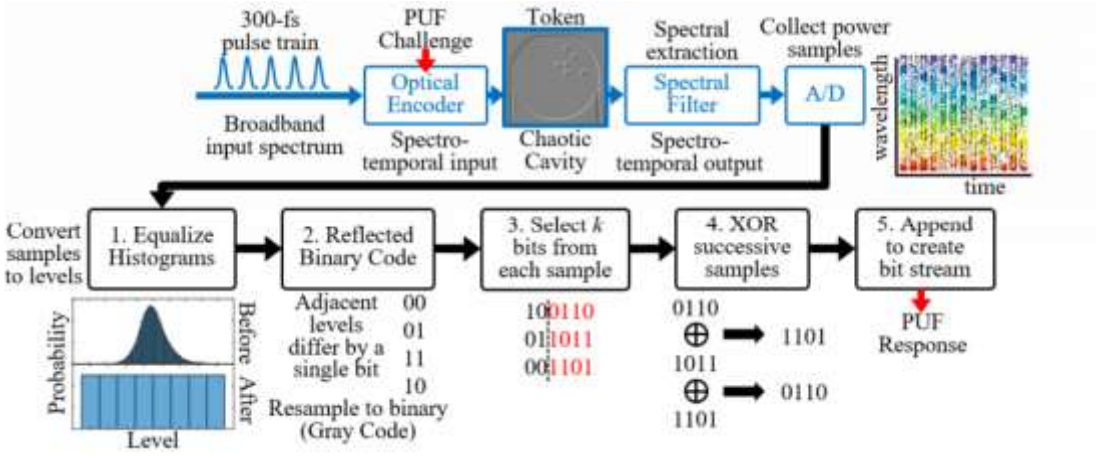


Figure 3.4: Post-processing algorithm for binary sequence derivation from analog response [67].

3.4 Experimental Results

3.4.1 Physical Unclonability

To investigate the silicon PUF's unclonability and repeatability properties, six prototypes were fabricated. For each design a CRL is built by averaging 460 analog response corresponding to the same challenge sequence. This CRL was utilized during the FHD histogram calculation where an individual measured response is compared to the averaged response from CRL associated with the certain device. The resulted FHD distributions are plotted for six different cavities. The set of histograms on the left are "same" or "like" distributions whereas the histograms on the right are "different" or "unlike" distributions (**Figure 3.4**). The threshold for authentication error is determined by the distance between the "same" and "different" distributions.

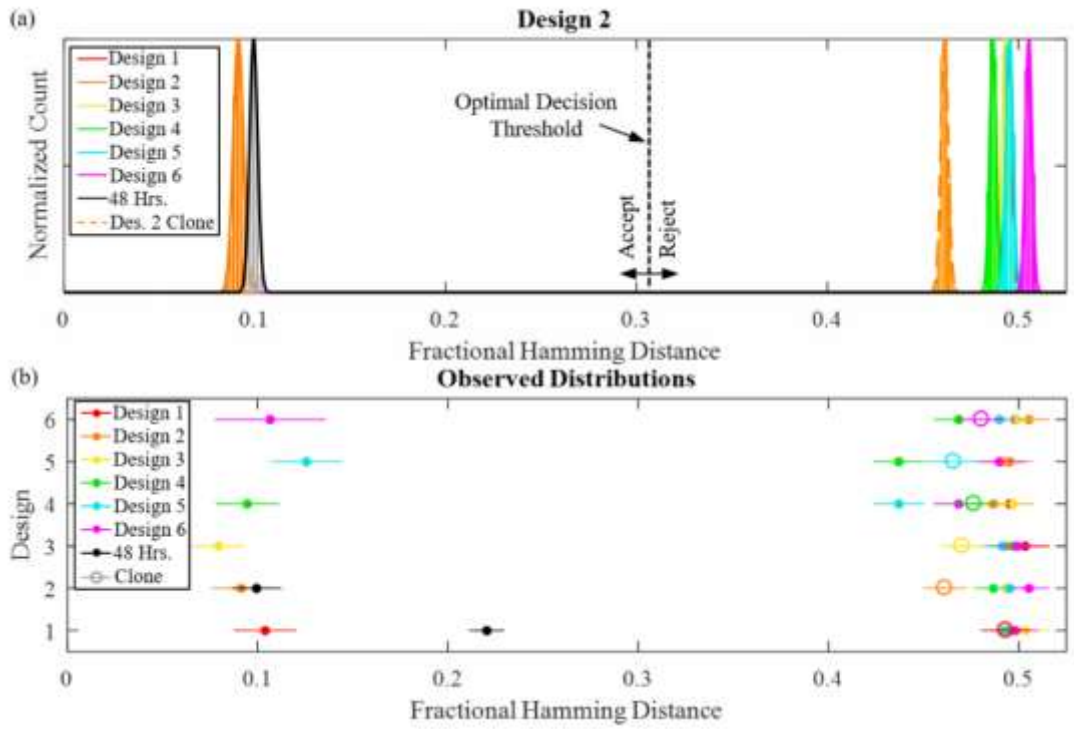


Figure 3.5: Authentication results. a) FHD histograms for each cavity calculated against design 2 along with 2 additional FHD histograms corresponding to the clone of design 2 and to the same design 48 hours later. b) Normalized FHD histograms for each design against every other cavity. Error bars represent ± standard deviations [67].

The total authentication error of the protocol is the sum of false acceptance rate (FAR) and false rejection rate (FRR) that is optimized under the variation of the number of bits kept at post-processing algorithm. It is shown that authentication error can be minimized to 10^{-21} of false accepting or false rejection for a key material length of 17.1 kb. The results in the **Figure 3.4** indicate that the mean value and standard deviation of FHD values for the "same" distribution reflects the reproducibility of PUF to identical challenges. The width of "same" distribution comes from a variety of noise sources and environmental factors. In this case, the "same" distribution for design 2 is centered around 0.1 meaning that the number of positions in response sequences differs by 10% of the total length on average. At the same time, "different" FHD distributions are centered around 0.5

indicating the degree of uniqueness of the device. Besides different cavity designs, the FHD distributions of the clone are calculated where the responses of design 2 were compared with CRL of identical design fabricated at the same conditions and at the same time. The distribution for the clone is located very close to "different" distribution indicating the high degree of distinguishability between the genuine device and the clone.

Lastly, the repeatability of the system over time is estimated by plotting FHD values corresponding to design 2 at the certain time and subsequently plotting FHD values corresponding to the same device 48 hours later. As shown in **Figure 3.4**, there is a clear repeatability of PUF system with the small drift of the mean value that is accounted for the temperature variations in the laboratory room.

Thus, the experimental results above directly demonstrated the reproducibility, uniqueness and physical unclonability properties of photonic PUFs making this approach highly desirable in system identification applications and other areas related to hardware security.

3.4.2 Information Content Metrics

Since the silicon photonic PUF is envisioned as a source of random private key material it is useful to estimate the information capacity of the device, i.e. the number of unique random bits of information that can be derived from a single PUF device. PUFs with high information content are harder to fully characterize by an adversary. Further, knowing the specifications related to information content allows one to answer critically important questions. For example, what is the maximum possible key length can be generated by photonic PUF? What is the maximum amount of information can be

encrypted or decrypted using photonic PUF? How fast is key material generation and how much information will an adversary need to gather in order to successfully break the device's security? To answer these questions, experimental and theoretical investigations of information capacity of photonic PUFs are presented [69]. Specifically, the nonlinear properties of the silicon photonic PUF make a great impact on the security enhancement and susceptibility to modeling attacks. On top of that, the nonlinearity allows significant improvements in PUF's information capacity. Finally, the key generation rate is estimated and because of the use of lightwave signals, it is shown to be around 200 Mbps outperforming the best optical scattering PUFs rates.

To derive the information capacity limits for silicon photonic PUF, the spectro-temporal information mapping model is utilized [26, 69] (**Figure 3.6**). The input laser pulse is composed of multiple encoded spectral features each of which with the spectral width Δf_{in} and temporal width Δt_{in}. The total number of spectral features and time slots in the pulse are $m_f = \Omega/\Delta f_{in}$ and $m_f = \tau/\Delta t_{in}$ respectively, where Ω - is the spectral bandwidth of the light source and τ – is the cavity lifetime. The input feature sizes may be arbitrarily chosen and in order to maximize the number of possible symbols encoded in the input map, $\Delta f_{in} = 1/\Delta t_{in}$ was chosen. Therefore, the total number of input symbols is $m_f\, m_t = \Omega\, \tau$. In case of a linear system, the mapping from input symbol s to output symbol r may be represented as m_i x m_j transmission matrix T. The maximum number of independent rows of this matrix is equal to its rank, thus the challenge space is linear in terms of the total number of symbols. To attack this linear system, it is enough to calculate the inverse matrix of T and derive the input given the output. In a nonlinear system, the transmission function is a combination of nonlinear equations, thus the inversion of T no longer exists.

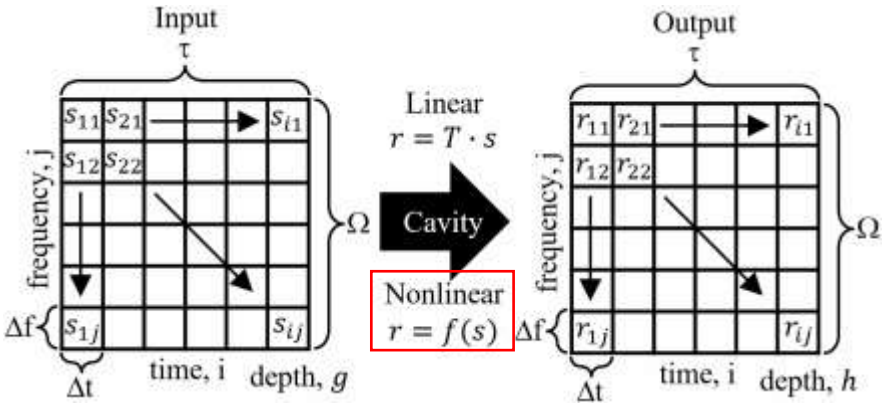


Figure 3.6: Spectro-temporal input and output mapping model [69].

The calculation of the information capacity of silicon PUF is based on the product of the number of bits per spectral response β and the total number of uncorrelated spectral responses n per cavity: $N \leq \beta n$. The total number of possible spectral patterns is bound to the spectral width of the source and the cavity lifetime. Specifically, the total number of input symbols with binary modulation yields $n = 2^{\Omega\tau}$. In the original experiment, the spectral width of the 300-fs laser pulse 4.2THz and the average feature size Δf_{in} = 44.2 GHz were measured. Hence, the upper bound of the independent number of responses is given by $n = 2^{\frac{\Omega}{\Delta f_{in}}} = 2^{94}$. The number of bits per response β can be estimated via calculation entropy rate of the spectral output of the device. In [69] the spectral probability mass function is determined from the spectral responses of the device and the number of bits per spectral feature is estimated to 5.2 for one of the polarization states. Thus, $\beta = 5.2 * \Omega/\Delta f_{in}$ = 286 bits leading to the upper bound of the total number of nits per device $N \leq \beta n \leq 10^{22}$Gbits. Given the area of the cavity 707 μm^2 this yields to the information density of 10^{22} Tbits/mm^2 that is much greater than any information density of current digital storage devices.

For clarity, I present the comparison of silicon PUF against various OSPUF designs in key characteristics such as information generation rate, information content and information density are presented in the table below:

PUF Type	PUF Performance Metrics				
	Year	*Info. Rate [bps]*	*Info. Limit [Gbits]*	*Volume [mm^3]*	*Info. Density [$Tbit/mm^3$]*
OSPUF with Probe [7]	2002	233	5522	254	0.022
OSPUF with SLM	2013	2×10^6	151	0.151	1
Integrated OSPUF	2013	1.3×10^5	?[b]	250[b]	?[b]
Photonic PUF	2016	1.8×10^8	10^{22}	1.6×10^{-7}	6.3×10^{25}

[a] Used for secure communications and encryption

[b] Estimated/data not available

Table 3.1: PUF performance metrics comparison [69].

3.4.3 Security Evaluation

A photonic PUF's susceptibility to an adversarial attack is a result of the chaotic behavior of the cavity, nonlinearity, physical limits, information density, and ultrafast operation speed. As it was shown before, the structure of the cavity with induced nano-scale features and complex nonlinear optical interactions prevents direct cloning of the device. But besides physical duplication, there is another factor prohibiting cloning photonic PUF: the ultrafast response of the cavity (~20 ps). Let me assume that an adversary may attempt to construct CRP library associated with legitimate photonic PUF device. To be able to do that, an adversary has to build the setup with required optical equipment with necessary components and store the collected CRP values in a non-volatile

computer memory that subsequently would be used to provide the correct responses to the client for a given challenge. For the successful attack, the whole system would need to respond to the user faster than 20 ps. In other words, the whole time of looking up the correct response for a certain challenge should take less than the ultrafast response of the physical cavity. Any current memory technologies and CPUs are incapable of performing these extremely fast search algorithms, thus making this attack infeasible and completely unforeseeable in near future [70].

3.5 Summary

The original work described above was developed by Grubel et al. [67, 68, 69] and here I review the main properties that are the most relevant to the rest of this dissertation. In this chapter, silicon photonic PUF is introduced with its key properties and characteristics as well as application in an authentication protocol. Reproducibility, uniqueness, unclonability, and low authentication error of the device are directly demonstrated. From the information capacity perspective, it is demonstrated that photonic PUF outperforms previous optical PUF implementations in many key parameters such as information content, information density, and speed of private key generation. From the security perspective photonic PUFs are proven to be robust and resistant to adversarial attacks such as direct physical duplication and CRL optoelectronic characterization of the device. All these benefits with the compactness of the device, easy integration with electronic circuits and design simplicity make photonic PUFs extremely attractive in a range of technologies including smart tokens, secure data storage devices, and smart

authentication systems. Despite comprehensive works on photonic PUFs, there is one important question that remains unanswered. Specifically, unclonability means that the PUF must be both physically and mathematically unclonable, where the latter implies the infeasibility of any computational algorithm to emulate PUF's behavior. In the original work, computational means of cloning were not investigated. Therefore, the next two chapters are focused on measuring the resistance to machine learning attacks and proving the unpredictability of the photonic PUF device.

Chapter 4 : Deep Learning Attacks on Simulation Models of Silicon PUF

4.1 Introduction

The main focus of this and the next chapter is the study of the resistance of silicon photonic PUFs to machine learning attacks. For better understanding of the cavity operation I design a set of computational models, where I provide attempts to learn the propagation of optical challenge pulses in a simulated photonic PUF. Specifically, I start with the simplest model of deriving the analog power values via random spectral filtering of the binary challenges and applying the same post-processing algorithm to produce the binary responses. Then I proceed to create more sophisticated simulation models taking into account the nonlinear optical interactions and other features specific to the silicon PUF device. Every simulation model produces the set of analog power sequences that is eventually post-processed for binary response extraction. As a result, I obtain an artificially generated CRP database with binary challenges uniquely mapped to binary responses. To emulate the CRP behavior, I aim to design a set of machine learning attacks against all the simulated models of the photonic PUF.

Machine learning tasks are typically classified into two broad types: *supervised learning* and *unsupervised learning*. The former type is referred to the family of algorithms

with the goal of mapping input to the corresponding outputs (labels), whereas in the latter the learning process is based on data with no labels. Since the supervised learning approach fits our problem, I choose a Deep Neural Networks (DNN) model that is acknowledged to be the most powerful algorithm, outperforming all the previous conventional methods in ML area [73]. In every simulation model, I generate a set of 80,000 challenge-response pairs, 70% of which is used for training process of DNN and the rest is used for evaluating the performance of trained neural network. Each binary challenge and binary response consist of 128 and 186 bits respectively since each of 31 channels in the post-processing algorithm is digitized to 6 bits. The attack is successful if DNN model predicts the binary response to a given challenge with sufficiently high accuracy. I present the prediction accuracies for every simulation model and study the convergence of prediction curves with respect to the size of the training subset that is fed to the DNN model. Further, I investigate the performance of DNN attacks with respect to the number of resampling bits kept per channel during the post-processing procedure. Variation of the number of LSBs yields different lengths of binary responses and affects the final performance of ML attacks as well as the repeatability of the PUF system. Therefore, it is important to study the performance of the DNN across the bit number in the channel.

4.2 Simulation models

4.2.1 Linear Spectral Filter PUF

We start with the trivial simulation model of a linear spectral filter that I call *Linear Spectral Filter PUF*. Specifically, we generate 80,000 128-bit spectrally encoded random

challenges and a filter of size 128 with uniformly distributed random numbers in the interval (0,1). Each challenge is multiplied in the spectral domain by the filter in an element-wise manner and integrated across the spectrum to obtain the analog power samples. These power samples are fed to post-processing algorithm to extract the binary response sequences. After collecting all 80,000 challenge-response pairs, the attacking procedure by DNN is performed. Due to the extremely simplicity of the given approach, we expect it to be the easiest task for DNN to correctly map the challenges to responses

4.2.2 Nonlinear PUF with a Single Spatial Mode

A more realistic model of the light propagation in the cavity includes a range of optical effects such as chromatic dispersion and self-phase modulation (SPM). In this model I assume that the optical pulse propagates non-resonantly as single spatial mode, thus there are no optical interactions between multiple modes that could potentially enhance the security of the device. Similar to the previous model, I create 80,000 128-bit challenge sequences each of which I encode onto 5-THz bandwidth optical pulse where each bit occupies 25 GHz of spectrum. To simulate the propagation of the optical signal in a presence of nonlinear effects and dispersion, I exploit the well-known *split-step Fourier method* [71]. This method is extensively used as a numerical approach to solve the pulse propagation problem in a nonlinear dispersive medium. The core idea of the split-step Fourier technique is straightforward. To model the nonlinearity effects and chromatic dispersion the medium is typically divided into a large number of segments, where at each segment the nonlinear and dispersion terms are applied in time and frequency domain

respectively. For example, the propagation of laser pulse with amplitude *E(t, z)* in optical fiber is dictated by nonlinear Schrödinger equation:

$$\frac{\partial E}{\partial z} = -\frac{i\beta}{2}\frac{\partial^2 E}{\partial t^2} + i\gamma|E|^2 E$$

We dropped the term corresponding to the loss and high order nonlinear effects for the sake of simplicity. The first term $\frac{i\beta}{2}\frac{\partial^2 E}{\partial t^2}$ governs the effect of dispersion, whereas $i\gamma|E|^2 E$ is responsible for Kerr nonlinearity. According to the split-step Fourier method, the equation can be split into linear part:

$$\frac{\partial E_D}{\partial z} = -\frac{i\beta}{2}\frac{\partial^2 E}{\partial t^2} = \widehat{D}\, E\,,$$

and nonlinear part:

$$\frac{\partial E_N}{\partial z} = i\gamma|E|^2 E = \widehat{N}\, E$$

Both parts have analytical solutions separately, but the NLSE does not have a general analytical solution. However, if the propagation path is divided into many small segments, then the two parts can be treated separately with a minor numerical error. Typically, the dispersion step has an analytical solution in the frequency domain, so it is convenient to Fourier transform the signal and convert it back to the time domain where the nonlinear step can be applied.

In the case of the silicon cavity, I model every segment to be one roundtrip propagation distance. At the first half of this distance the pulse accumulates a certain nonlinear phase shift in time domain, whereas at the second half the dispersion effect is applied in the frequency domain. We specify cavity's material parameters such as group velocity dispersion β = -21.7 fs^2/mm and nonlinear parameter $\gamma = 10^3$ $W^{-1}m^{-1}$ that is typically used for nonlinear phase shift calculation. The amount of nonlinearity in the

cavity is controlled via the peak amplitude of the input laser pulse. Since the cavity has an inherent loss and the coupling waveguides tangent to the disk, I incorporate these effects as well. After each round of propagation, the spectral and temporal profile of the pulse are detected. The sequence of power values is obtained by calculating the intensity profile of the output signal after 100 roundtrips. Power samples corresponding to a given challenge pulse are post-processed and stored in a CRP table for the subsequent attacking process.

4.2.3 Nonlinear PUF with Multiple Spatial Modes

The last modification made to the developed simulation model is the excitement of multiple spatial modes of the cavity. Here, I set 100 modes propagating in the silicon resonator each of which has its own optical loss, roundtrip time and coupling ratio of the power. By introducing 100 spatial modes I allow the complex intermodal optical interactions in the cavity. I expect that a modeling attack against this simulation would yield the worst performance due to the greater complexity of the nonlinear interactions between modes.

4.3 Results

In this section, I present the results of attacking simulated photonic PUF via Deep Neural Networks. The implementation details of DNN and all parameters associated with the design of the network are given in section 5.4.

By performing DNN attacks against the aforementioned simulation models of the PUF I address two important questions. First, is DNN capable of emulating every

simulation model? Second, what is the minimum number of CRPs the attacker needs in order to successfully break the simulated PUF? To answer the first question, it is enough to train DNN on 70% of the total 80,000 CRPs and evaluate the prediction results on the rest 30% of the CRP library. To answer the second question, I run DNN for different sizes of CRP database and identify the least number of samples at which DNN gives the lowest prediction error. In **Figure 4.1** results of DNN attacks are presented.

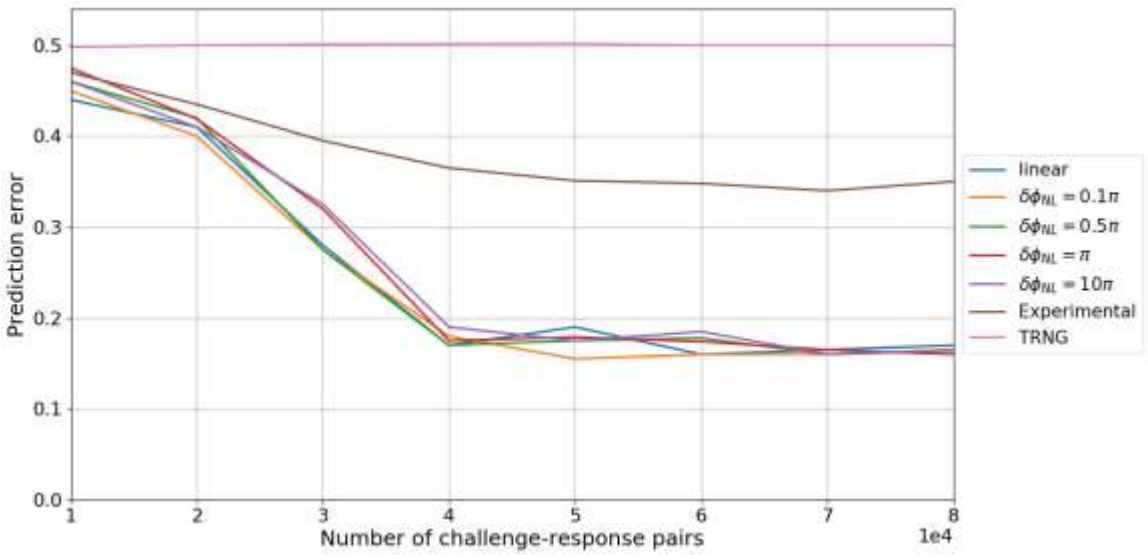


Figure 4.1: DNN attack on simulated photonic PUF. Prediction results are obtained on 30% of CRPs (test set) after training process on 70% of CRPs (train set). Linear (blue), nonlinear PUF with dispersion and single spatial mode (orange) and nonlinear PUF with multiple spatial modes at three different input energy pulses (green, red, purple) are presented. For a comparison, performance of DNN on experimental dataset is also demonstrated (brown). Purple curve represents the accuracy of random guessing of every response generated by TRNG.

Surprisingly, we do not observe any differences in the performance of DNN across all simulation models. Even the most complex model of the cavity with 100 spatial modes and with the highest dominance of SPM effect ($\Delta\varphi_{NL} = 10\pi$) is attacked by DNN yielding the same prediction accuracy of ~86%. For benchmark comparison we run the DNN on

two additional CRP datasets. I plot the prediction curves corresponding to experimental dataset of 80,000 CRPs collected in the original work. This dataset was gathered at a relatively low energy of the input pulse with minimum nonlinear effects. As it can be seen, the prediction accuracy of DNN, in this case, is ~60%. In addition, we create 80,000 challenges where each challenge is associated with random binary sequence generated by a true random number generator (TRNG). In **Figure 4.1**, it was confirmed that the DNN is incapable of learning anything from this dataset. In other words, DNN is equivalent to random guessing of correct binary responses where the prediction rate is 50%. Hence, we answered the first question regarding the ability of the ML algorithm to emulate the simulated PUF behavior.

An additional observation we make from the prediction results is the dependence of DNN's performance on the size of CRP database. Interestingly, starting from 40,000 CRPs the prediction accuracy of DNN hits a plateau with insignificant fluctuations. This answers the second question regarding the minimum number of samples the attacker needs for the successful design of the ML model. In a real scenario **Figure 4.1** indicates that if an Eve steals 70% of the 40,000 samples to train the Neural Network then this model would extrapolate PUF behavior on any new challenge-response pairs with 86% accuracy.

Lastly, we study the prediction accuracies of DNN as a function of bit number that is kept during the post-processing of raw analog power samples. In all simulation models above, I keep 6 bits per channel leading to 186-bits responses and I assume that most significant bit (MSB) is the most repeatable and easy learnable bit representing the largest fluctuations in the power value, whereas the least significant bit (LSB) is the most sensitive and the hardest bit to learn for DNN. For this reason, I perform six DNN attacks against

80,000 CRPs, where the length of every response is 31 bits (one bit per channel). In every attack I keep only one bit starting from the MSB (bit number 1) and finishing with LSB (bit number 6) (**Figure 4.2**)

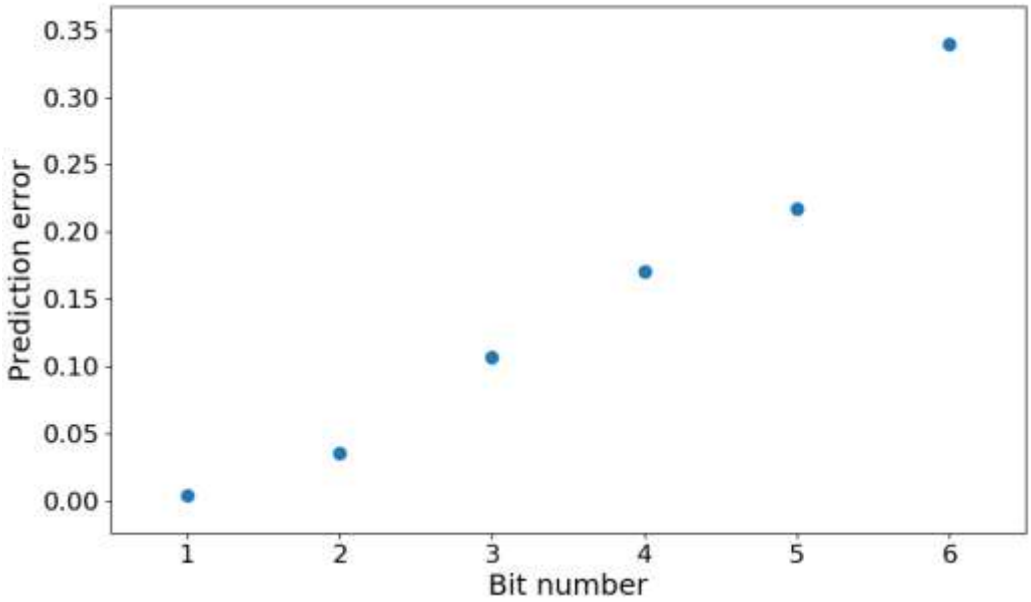


Figure 4.2: DNN performance as a function of bit number kept in digitized channel. Bits are ordered from the most significant bits (MSB) to the least significant (LSB) ones. Notably, the average of prediction errors for 6 bits matches to the overall prediction error of DNN against CRP with 186 bits responses (86%)

As expected, the worst prediction accuracy is shown for bit number 5 and 6, thus leading to the trade-off between the ML resistance and repeatability of PUF. Keeping only the most sensitive (and noisiest in the experiment) bits harms the reproducibility but enhances the security of the PUF. On the other hand, if I keep, for example, the first two bits in each channel, then PUF is more repeatable but at the cost of the higher vulnerability to modeling attacks. Given this trade-off, we choose 4 bits that I consider an optimal number of bits.

4.4 Conclusion

In this chapter, we designed various computer simulations of the PUF cavity that, I believe, is a natural step for the next modeling attacks against large experimental CRP dataset. In these models, we attempted to capture several significant complex nonlinear interactions and other details specific to the photonic cavity. However, on the security side, simulations demonstrated to be relatively weak against Deep Neural Networks. As shown, DNN easily characterizes the CRP relationship produced by all type of simulations with high accuracy (86%). In addition, the resistance to ML algorithms of PUF models can be managed via the converting power samples to binary representations in the post-processing algorithm. Specifically, it is shown that keeping a different number of bits in the digitized form of the response makes an impact on two important properties such as repeatability and mathematical unclonability. As the future step, it is important to extend simulation models by including other high order nonlinear effects such as TPA, Raman scattering, self-steepening.

Given these results, I proceed to the next set of DNN attacks on the silicon photonic PUF that was exploited in the original experiment.

Chapter 5 : Deep Learning Attacks on Silicon Photonic Physical Unclonable Function

5.1 Introduction

The focus of this chapter is to provide a set of Machine Learning Attacks against a true experimental silicon photonic PUF, a recent approach based on ultrafast nonlinear optical interactions in silicon microcavity. I demonstrate that in practice nonlinear silicon PUFs are resistant to two possible ML attack scenarios. I find that this resistance is rooted in the optical nonlinearity of the silicon photonic PUF token in tandem with its complex ray chaotic structure. Finally, I investigate encrypted data storage and compare the results of decryption using genuine PUF device and ML "clone".

To quantify the degree of protection offered by a PUF, it is important to quantify the difficulty of determining the input, given the observed output. The security of a PUF rests on the inability to duplicate the physical device or to accurately model its behavior, so that only the device holder can extract the CRP database. In recent years artificial intelligence and machine learning have made great strides in learning the behavior of a physical process or device via training, without the need for a physical model. If ML can learn the behavior of the device after exposure to a subset of the CRP database, then the

security of the PUF (as well as of the resource that it protects) is at risk, as the machine can then generate the entire CRP database at any time. Notably, ML attacks have been very successful against electronic strong PUFs, including Arbiter PUFs, Ring Oscillators, XOR Arbiter PUFs and other electronic-based cryptographic devices [37, 42, 72]. It is known that only optical scattering-based PUFs continue to resist ML attacks, e.g., the first non-integrated optical PUF implemented by Pappu et al. has, as of this writing, not yet been successfully attacked by ML algorithms [29, 30, 42]. However, this resistance is trivially achieved in this bulk approach as a different spatial region of a random material is probed with each new challenge and thus there can be no way to learn the behavior based on previous observations. This is akin to having a large number of unique PUF devices and using each only once. However, this greater security comes at the cost of large device size, a lack of electronic integration, and extremely poor reproducibility of behavior. Notably, attempts to integrate scattering PUFs robustly into CMOS circuits have been shown vulnerable to modeling and ML attacks due to the reuse of the scattering volume and the linear nature of the scattering process [41, 42].

Recently our research group demonstrated a novel photonic PUF that harnesses nonlinear optical behavior in an integrated silicon photonic device that maintains high compatibility with electronics [67, 68]. This silicon photonic PUF can be easily integrated with both CMOS electronics and telecommunications hardware in particular due to the recent emergence of silicon photonic chip foundries [69]. However, as with other PUF technologies, a remaining crucial factor is the device's unpredictability and resistance to ML attacks, which is the focus of this chapter.

In this investigation, I construct ML-attacks against a silicon photonic PUF as follows. If an adversary Eve can obtain access to the photonic PUF or otherwise steal a subset of CRPs, I aim to determine whether Eve, within a limited time frame, can derive a mathematical model that can correctly emulate the full device behavior and thus generate the full challenge-response space by predicting unobserved CRPs (**Figure 5.1**).

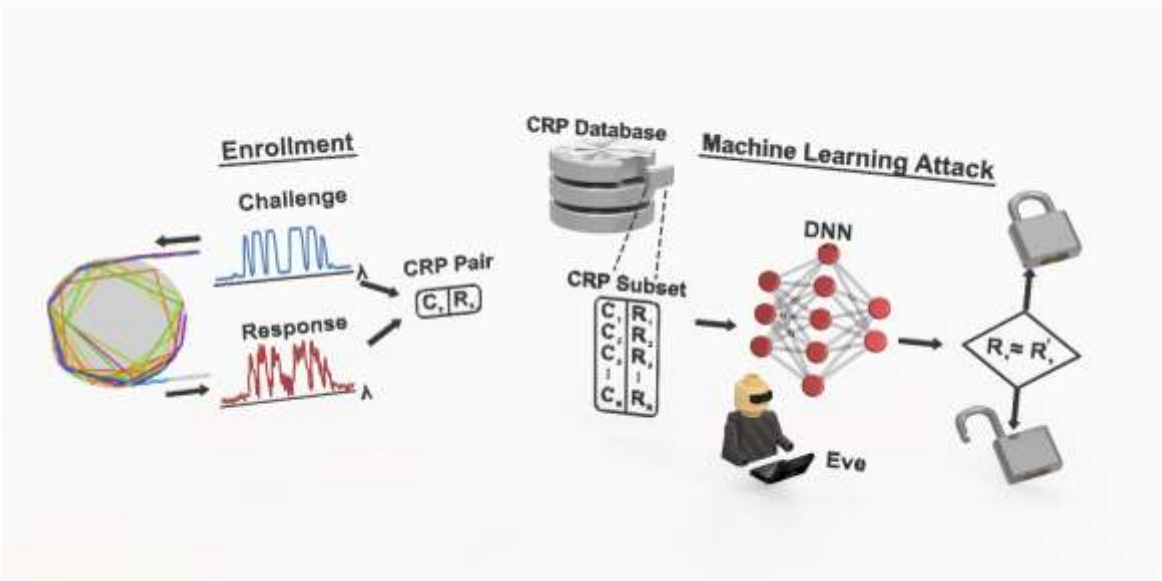


Figure 5.1: An adversary attack procedure. Having a subset of CRPs from the full challenge-response space, Eve has a limited time to design the machine learning algorithm in order to obtain the approximate behavior of a PUF device. Specifically, Eve trains a Deep Neural Network (DNN) on the stolen set of CRPs, feeds the DNN with new challenges and attempts to predict unobserved CRPs. If the DNN predicts the correct responses up to some error threshold, then PUF is considered to be compromised.

Previous successful attacks on electronic PUFs were conducted using Support Vector Machines (SVM), Boosting, Logistic Regression, and Evolution Strategies [37, 40, 55]. Here, I present ML-based attacks against the silicon photonic PUF in both authentication and encryption scenarios using Deep Learning on a Deep Neural Network (DNN) and demonstrate the PUF's high resistance to learning due to the complexity of its nonlinear optical behavior. In the ML community, Deep Learning is acknowledged as the

state-of-the-art technique and outperforms other solutions in multiple fields such as computer vision, image and speech recognition, classification and machine translation [73]. A major advantage of the Deep Learning framework is that it can model nonlinearity and can be easily adapted to new problems. For these reasons, I chose Deep Learning to investigate ML attacks against silicon photonic PUF. Notably, for completeness I also investigated other methods (e.g. SVM, Logistic Regression) and all performed inferiorly to a DNN and thus only the Deep Learning results are presented here.

5.2 Results

5.2.1 Data Collection

I employ the methods described in [67] for token authentication and methods based on [69] for encryption. Ultrafast 300-fs laser pulses with 90-MHz repetition rate undergo frequency-to-time mapping in dispersion compensating fiber (DCF) and are encoded with 128-bit random binary amplitude sequences generated by a pulse pattern generator at 11.52 Gbit/s. These spectrally-encoded challenge pulses are compressed with complementary dispersion single-mode fiber (SMF) to near their transform-limited duration and coupled into the PUF. The sequence of response pulses emanating from the device is amplified, filtered with a set of spectral masks, and recorded with a photodiode (PD) and synchronized analog-to-digital converter (ADC) providing one 16-bit sample at the peak of each pulse. A post-processing algorithm derives the response binary sequence from the analog samples using probability equalization and resampling to a selectable number of significant bits per

sample. Each 128-bit random sequence serves as a single challenge; 32 output spectral filters, resampled to 4 bits with an XOR applied to successive spectral filter outputs, yields 124 bits per response (31 *channels* of 4 bits). During the enrollment phase, I collect averaged responses to 960,000 challenges to create the CRP database for both the training and test datasets for the Deep Learning attacks. I performed the enrollment process three times at different optical power levels of the challenge pulses to study the effect of optical nonlinearity on the success of the ML-attack. Training and test data were generated according to a 60/20/20 partition: 60% of the data was used for *training*, 20% of the data was used for *cross-validation* and tuning hyperparameters of the neural nets, and the remaining 20% of the data was used as a *test* set to evaluate the ML performance to unobserved CRPs.

5.2.2 Machine Learning Attacks Scenarios

I investigate two possible ML attack scenarios based on the point at which an eavesdropper (Eve) manages to observe the output of the device (**Figure 5.2a**). First, Eve might attempt to emulate the device by observing the input and output binary sequences and training an ML algorithm to predict all of the CRPs in the database. I refer to this scenario as a *direct attack* (**Figure 5.2b**). Because the bit extraction algorithm does not attempt to add security beyond the device itself, I also consider the scenario in which an eavesdropper can probe the output waveguide and record unprocessed power samples directly. I refer to this as a *side-channel attack* (**Figure 5.2c**), where instead of predicting the binary responses, Eve attempts to model the analog optical power transfer of the device. I consider the side-channel attack to be the best possible chance to model the device.

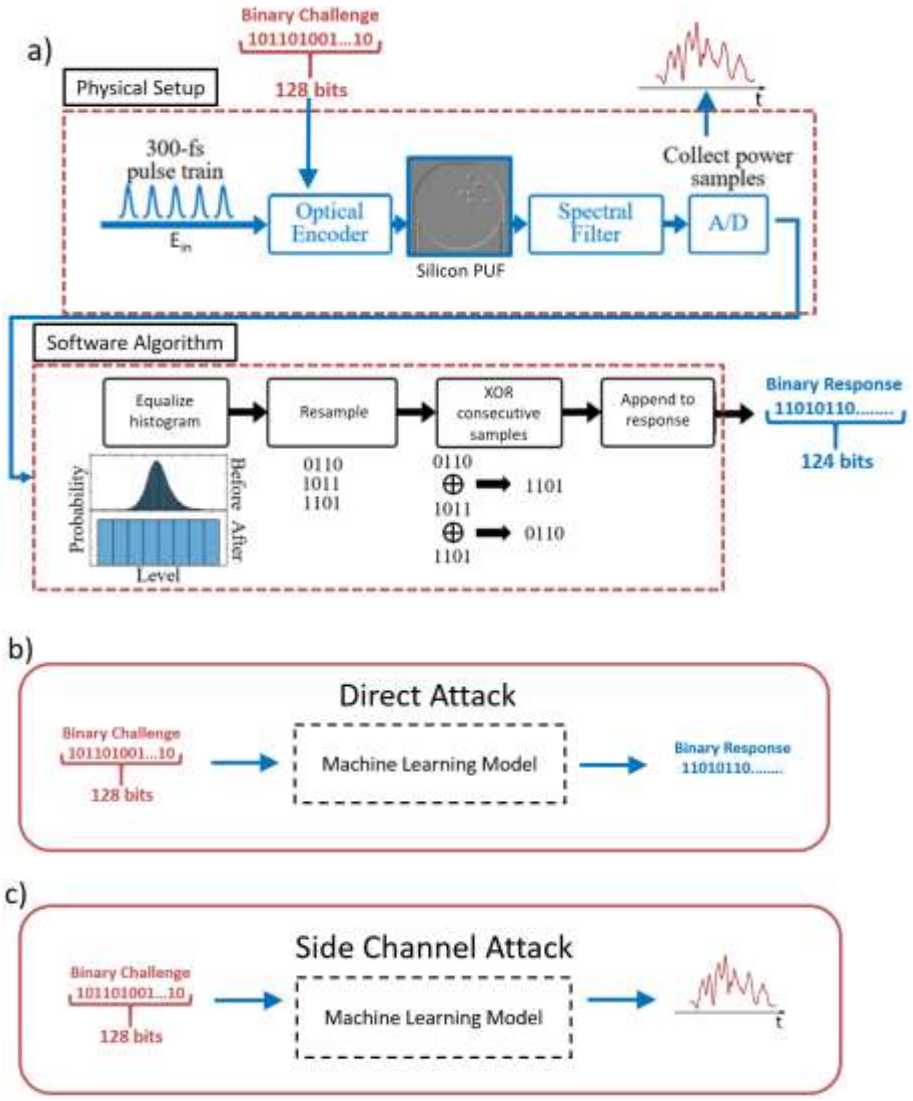
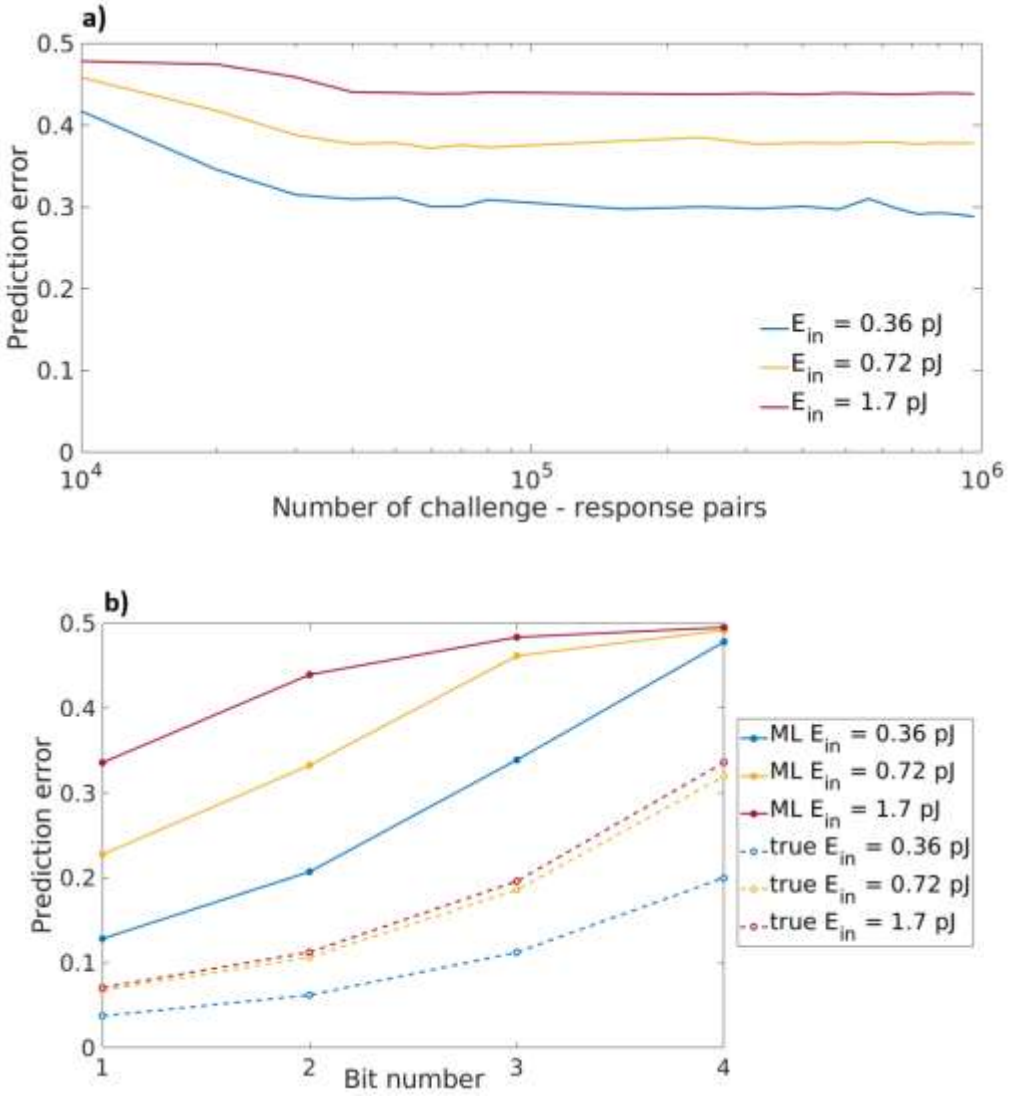


Figure 5.2: Machine Learning Attacks scenarios. a) General setup of challenge–response generation with hardware setup producing analog power samples response and post-processing algorithm producing the binary version of the response b) Direct attack with ML model mapping binary-to-binary relationship c) Side-channel attack with ML model mapping binary-to-real relationship.

5.2.3 Direct Attack

I first studied the convergence of the DNN to a stable prediction of the entire 124-bit binary response to each challenge with increasing size of the total training dataset (i.e.,

employing an increasing percentage of the total 960k to create the training, validation, and test data). In **Figure 5.3a**, three learning curves are presented for different input optical power levels. In all cases, I observe that the performance of the DNN plateaus after a training set size of roughly 10^5 and in all cases the DNN fails to accurately reproduce the behavior of the PUF. Notably, the performance of the DNN shows a clear dependence on the amount of optical nonlinearity in the PUF. I observe that increasing the optical nonlinearity by increasing the optical power significantly increases the prediction error.



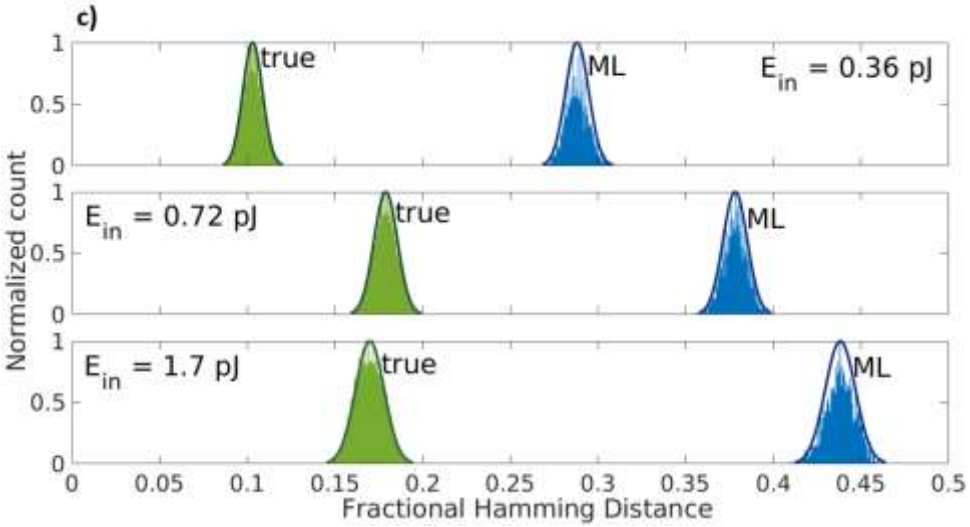


Figure 5.3: ML direct attack results. a) Convergence of NN generalization errors with respect to amount of the dataset at average pulse energy 0.36 pJ (blue), 0.72 pJ (yellow) and 1.7pJ (red) b) NN prediction error of each bit in channel at maximum number of samples used for training phase. c)Normalized FHD distributions and histograms calculated against CRP of legitimate PUF token at different power levels in the setup: "like" distribution (green) represents the FHD values between repetitions and the response sequence from CRP of the legitimate PUF, ML "clone" distribution (blue) represents the FHD values between ML predicted response sequences and the response from CRP of legitimate PUF.

Notably, for a given challenge each response channel is digitized to 4 bits. With the most significant bit (MSB) representing the largest scale fluctuations and the least significant bit (LSB) representing the finest scale fluctuations. We expect the MSB (bit number 1) to be the easiest to learn and the LSB (bit number 4) to be the hardest to learn. For this reason, I also study the performance of the DNN as a function of bit number. As shown in **Figure 5.3b**, I find that lesser bits of each 4-bit channel are the most difficult to predict, as expected. Notably, the probability density function (PDF) of the analog power samples has an entropy of 6 bits [67], but during bit extraction I downsample to 4 bits by discarding the least significant levels to improve repeatability (limited by the signal to noise

ratio). Thus, the ML performance as a function of bit number should also be considered to determine the optimal trade-off between repeatability and ML-resistance.

In the experiments, the authentication threshold is optimized based on the FHD distribution generated by repeated probing of the device and that generated by non-authentic devices. In practice, there is no unique optimal threshold; the threshold is set to the value best suited to the user's needs. In this context, the ML performance is relevant to determine an optimized threshold. Here, I computed the set of *like* histograms at different power levels in the experiment as well as a set of machine learning clone histograms, depicted in **Figure 5.3c**. Notably, I observe excellent separability between the genuine PUF device and the ML clone PUF using 9.9-kbit keys generated from concatenated responses. This occurs even at the lowest power level when the effects of optical nonlinearity are the weakest. However, at higher pulse energies the mean of the ML clone distribution moves closer to 0.5, consistent with the observation that the optical nonlinearity in the device enhances its unpredictability. Notably, the ML clone performs markedly better than the actual physical clones on the same chip, which typically exhibit an FHD mean > 0.45 [67], underscoring the importance of these ML-resistance studies.

5.2.4 Side-channel Attack

To investigate a side-channel attack, I assume that Eve has temporary access to the raw power measurements after challenges are presented to the cavity. Thus, I train a second DNN to minimize the mean squared error (MSE) between the genuine device and the ML predictions. A comparison (using MSE) of the analog signals generated by Neural Networks with power samples obtained during the experiment is presented in **Figure 5.4a**.

In addition, I apply the bit extraction process to the ML predicted power samples to obtain a set of binary responses to generate an ML clone histogram (**Figure 5.4b**).

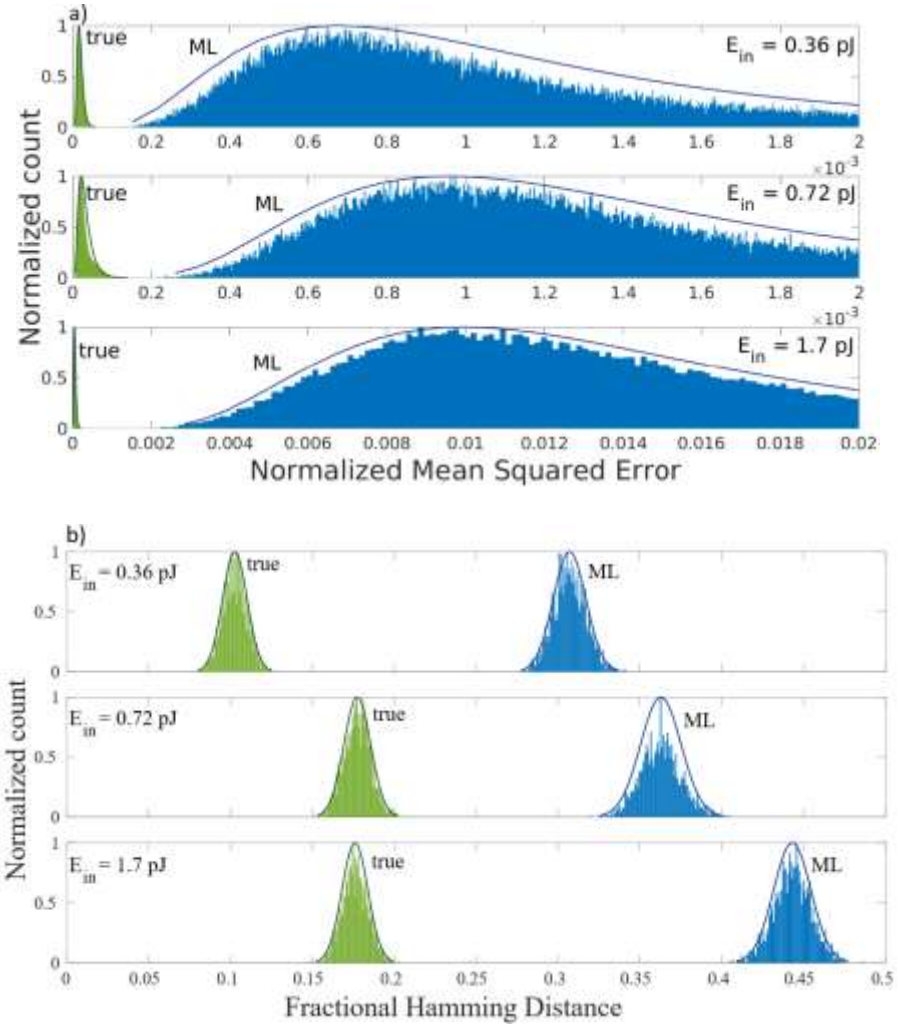


Figure 5.4: Side-channel attacks results. a) Normalized MSE distributions based on comparison between power repetitions and averaged power sample of PUF device(green) and comparison on averaged power samples of PUF device and ML predicted power samples (blue). Note that the scale in the last figure is different from the previous two. b) Normalized FHD distributions of binary response sequences obtained after post-processing algorithm on analog power samples. Both charts are presented at different power of optical signal in the system

From **Figure 5.4**, it is evident that even in the case of a side-channel attack, when Eve has a chance to extract raw analog data without its digital post-processing it is still not

possible to emulate the device. The FHD analysis leads to nearly the same results as in case of direct attacks (**Figure 5.4b**) with distinct separability of "clone" device and legitimate one and increased ML-resistance at high optical power levels.

5.2.5 Encryption Results

The exponentially large challenge-response space permitted with strong PUFs becomes most interesting for cryptography schemes that require extremely large key lengths, such as one-time pad (OTP) encryption [41]. Notable, a genuine PUF will always have some bits that differ from the CRP database because of noise. Forward error correction and fuzzy extraction of the usable cryptographic key material has been successfully employed to eliminate errors for secure communication with PUFs [61, 69]. Statistically, the ML clone can correctly predict a portion of the response bits, so it is necessary to test OTP encryption using key material generated from fuzzy extraction with a genuine PUF and consequent decryption (by Eve) of the message with key material extracted from the ML "clone" PUF using the same fuzzy extractor. To investigate the encryption performance under such an ML attack at a range of error correction code rates, the fuzzy extractor is applied to the response bits from the PUF to produce blocks of reliable and strong key material to encrypt a message. The message is XORed with this encryption to form the ciphertext. In **Figure 5.5**, I decrypt the message using the subsequent output of the genuine PUF and the trained ML clone to compare the success of decryption at different code rates. The performance is quantified based on the bit-error rate (BER) of the decrypted message. In the low power case, for example, code rates < 0.1 yielded no errors for the 19.1-Mbit message upon decryption with the legitimate device. By contrast, the ML

"clone" was unable to accurately reconstruct any of the original message (BER ~0.5) at all code rates.

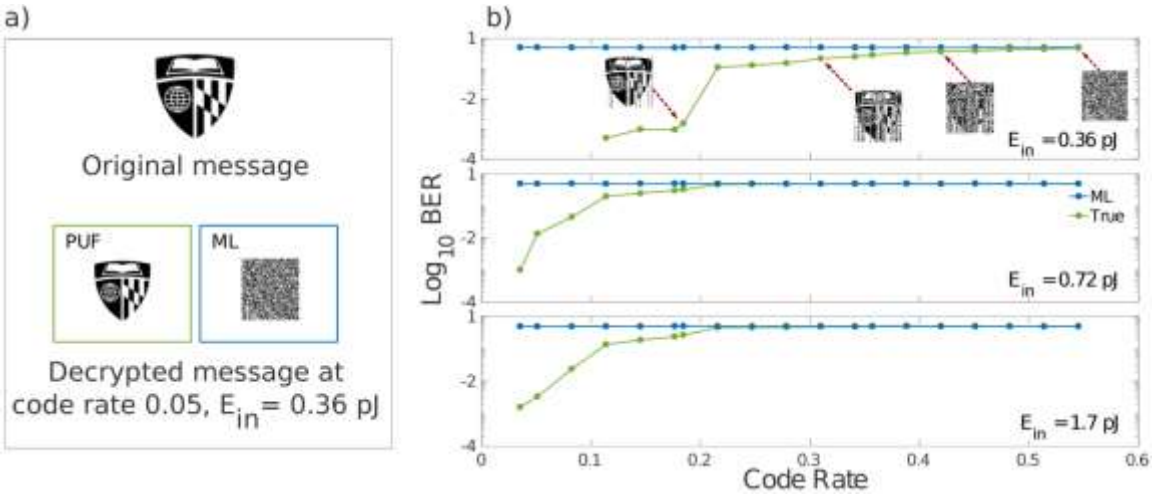


Figure 5.5: a) Original message used and corresponding decryption results for ML clone and genuine PUF. b) The mean BER for the message decryption using ML clone and legitimate PUF CRL responses at different average power levels in the system. Inset pictures show the quality of decryption at various code rates. ML clone is unable to reconstruct the original image even at the lowest code rates.

5.3 Neural Network Design

For this study, the machine learning algorithm was designed using the open-source Keras library [74] with Theano backend [75], implemented in Python and CUDA. The input layer of the NN consists of 128 nodes corresponding to the 128-bit challenge sequence length. The total number of hidden layers was kept to 2, each of which consists of 500 nodes, and, depending on the attack scenario, the NN and its parameters were adjusted to be consistent with the output format of the data. In direct attacks, where the NN should learn how to derive binary responses given the input challenge, the output layer nodes used a "sigmoid" activation function that is typically used for classification problems and the hidden layer nodes used the ReLU function. In the side-channel attacks, the output

layer of the NN should give the analog or continuous pulse energy values. Therefore, a linear activation function was used. I tested different configuration of Neural Network designs and concluded that deeper and wider networks yield the same out-of-sample errors, although it is more time consuming to train them.

It is well-known, that NNs possess a huge set of hyperparameters that need to be tuned at cross-validation phase to achieve the greatest accuracy. Using the 60/20/20 splitting schema of the whole dataset, I tuned such parameters as batch size, weight decay of regularization techniques, learning rate of optimization process and its momentum. In addition, dropout regularization was utilized to reduce overfitting and improve the generalization error. The optimization method was also varied, but generally set to "Adam" [76] which is a popular technique in the state-of-the-art neural networks configurations.

5.4 Conclusion

In conclusion, I have demonstrated the strong resistance of silicon photonic PUFs to state-of-the-art machine learning attacks. Neither a direct attack, attempting to reproduce the PUF's extracted binary response to binary input challenges, nor a side-channel attack, granted access to the raw optical output from the PUF, succeeds in replicating the behavior of a legitimate photonic PUF. The optical nonlinearity is clearly shown to have critical importance in the resistance to such machine learning attacks. The demonstrated combination of device robustness and machine learning resistance is superior to any PUF yet developed.

Chapter 6 : Deep Learning Attacks on Simulation Model of Optical Scattering Physical Unclonable Function

6.1 Introduction

The goal of this chapter is to study the security aspects of an optical scattering PUF (OSPUF). OSPUFs have been shown to exhibit the unprecedented level of security and the resistance to modeling attacks. The optical scattering system originally proposed by Pappu et al. was one of the first scattering PUFs [29] (**Figure 1.10**). In the original work, he presented a non-integrated PUF system that possesses a number of advantages including low cost of equipment piece, extremely high output complexity, great resistance to adversarial attacks such as modeling attacks as well as physical cloning. However, on the downside, the whole OSPUF setup requires many moveable components and high precision mechanisms for stable read out of the responses. Therefore, the implementation of Pappu's setup is laborious, expensive, and error-prone, thus motivating the research goal of embedding optical PUFs into electronic chips.

Despite many advantages of OSPUF, very few attempts have been made in integrating OSPUF with electronic microcircuits. For example, one of the first miniaturized version of Pappu's setup uses expensive and slow piezo positioners [81]. Later Rührmair et al. presented the prototypes of integrated OSPUF where LCDs and phase locked arrays

were utilized as spatial light modulators (SLM) [42]. However, his approaches were studied more for security analysis but not for physical implementation.

Even less activity has been observed in the attacking OSPUF systems. Surprisingly, no machine learning attacks have ever been reported on scattering PUFs despite the fact that a linear scattering medium is typically exploited in the setup. Due to this fact, we decided to investigate the level of complexity of this problem by constructing an integrated PUF experimental system similar to [42]. After multiple attempts to attack the experimental OSPUF, all our efforts remained unsuccessful. To better understand why we study the security of a simplified optical scattering PUF in the simulation. We believe that breaking the security of the simulated scattering system can give us the useful insight about the underlying reasons why OSPUF's remain robust to modeling attacks.

6.2 Simulation Model

In practice when coherent light (e.g. a laser beam) illuminates a rough surface a speckle image is formed because of multiple interference of a set of wavefronts. Mathematically, speckle generation is described as a random walk where each wave experiences the random phase during the scattering. If each wave is modeled as a vector with random angles (phases) then the length of a resultant vector is distributed from zero to the sum of individual wave vectors. From diffraction theory, each point on a scattering surface acts as a source of secondary spherical waves. The light at any point in scattered space is the sum of amplitudes of each ray resulting in the exponential distribution of intensity values.

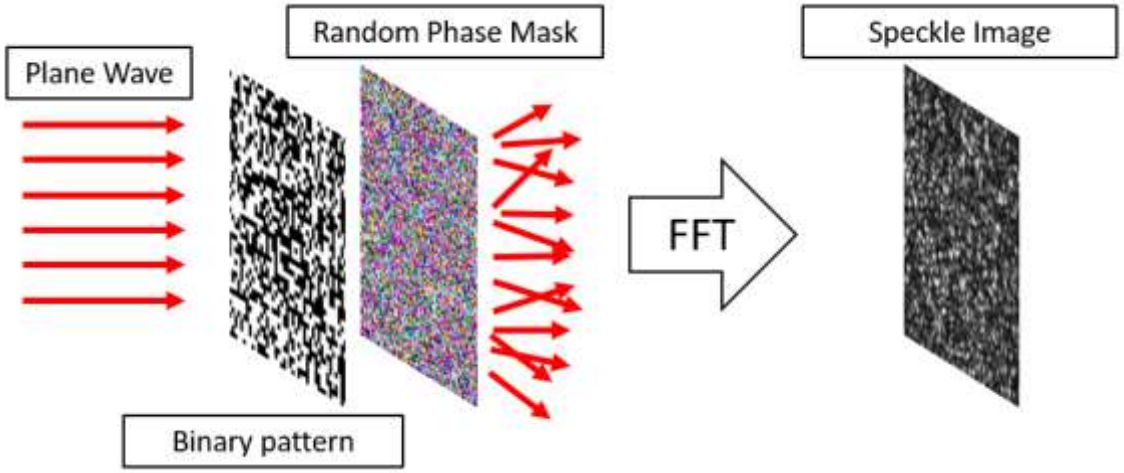


Figure 6.1: Single surface scattering of the modulated plane wave using the random phase mask.

Based on this, we model in Matlab a single surface scattering of a plane wave using a random spatial phase mask. The whole computation procedure of the speckle image consists of the following steps (**Figure 6.1**). First, we apply an "on-off" scheme to the plane wave to obtain the binary illumination pattern with sizes that can be varied further. This binary pattern plays a role of a challenge in the scattering PUF model. The challenge spatial field is then multiplied by the random spatial phase matrix in an element-wise manner. The random phase matrix consists of Gaussian distributed numbers with mean value equal to the width of the surface and possesses the same number of features as the challenge. Therefore, each wavefront's propagation length is random leading to the random accumulation of phase. In the end, I apply the Fourier transform to the result of element-wise multiplication in order to obtain the far-field spatial intensity resulting in the speckle image, which plays a role of a response of the OSPUF. The Fourier step is implemented based on the assumption that the detection of speckle images is observed at much longer

distances than the wavelength of the source, leading to the far-field diffraction. For the sake of clarity, I present a pseudocode of essential steps for speckle generation:

```
S = 32; % size of binary pattern
N = 10000; % number of CRPs
for i = 1:N
      % binary challenge matrix
      challenge_matrix = randi([0,1,S);

      % random phase mask with Gaussian random numbers
      r_mask = randn(S);

      field = challenge_matrix.*exp(2j*pi*r_mask);
      FFT = fft2(field);
      response = FFT.*conj(FFT); % real valued intensity
```

Following the procedure above, we collect 100,000 challenge-response pairs for different sizes of binary patterns: 8x8, 16x16, 32x32, and 64x64 pixels. The example of binary challenge and the corresponding speckle image is shown below for the size 32x32 (**Figure 6.2**).

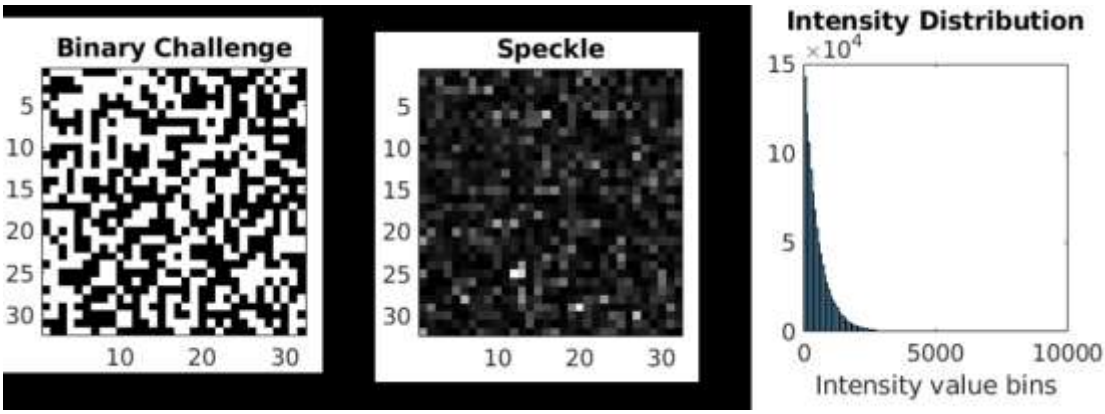


Figure 6.2: Example of 32x32 binary pattern and corresponding obtained speckle image via the procedure described above. Exponential distribution of intensity values of all 100,000 speckle images plotted for sanity checks.

As it can be seen, the distribution of intensity values is exponential and in the next section, this fact is used for benchmark analysis of the ML attacks performance. After the CRP collection, I perform the set of machine learning attacks based on the deep neural network algorithm. Essentially, the goal of these attacks is to explore the capability of the attacker to break the simulation model of OSPUF in its simplest form without using a physical model. Similar to the attacking procedure on silicon photonic PUF, we are interested in the minimum amount of CRP information that can be revealed to the attacker to accurately predict the responses given the unseen challenges.

6.3 Simulation Results

In this section, I present the results of the DNN performance against all sets of 100,000 CRPs. The specific details of DNN structure, tuned hyperparameters, and necessary infrastructure are given in the section 6.5.

The set of 100,000 CRPs is divided into three subsets: training data (70%) for model learning, validation data (20%) for tuning the model's hyperparameters and test data (10%) for model's performance evaluation. Before the training process, we preprocess the speckle images by normalizing the intensity values to the global maximum of the value across all speckle images. Therefore, the pixel intensities of the processed speckle images range between 0 and 1. As a metric for comparison between the correct and predicted responses, we choose the root-mean-square error (RMSE) that is calculated as follows:

$$RMSE = \sqrt{\frac{\sum_i^N (Y_i - \hat{Y}_i)^2}{N}}$$

where N – is the total number of CRPs, Y_i – vector corresponding to the ground truth speckle image, $\hat{Y}_i$ – vector corresponding to the DNN predicted speckle image. The next set of figures (**Figure 6.3 – 6.6**) shows the results of DNN for each of the four of pattern sizes from 8x8 to 64x64 pixels. For comparison, we evaluate the DNN prediction accuracy against the accuracy of random guessing. For example, in 8x8 speckle pattern, for each of 64 pixels we randomly draw samples from the exponential distribution obtained from all simulated 100,000 8x8 speckle images. Then, we calculate the RMSE between the ground truth speckle image and the speckle image generated by the sampling procedure. This RMSE corresponds to the accuracy of random guessing. This is repeated for all pattern sizes.

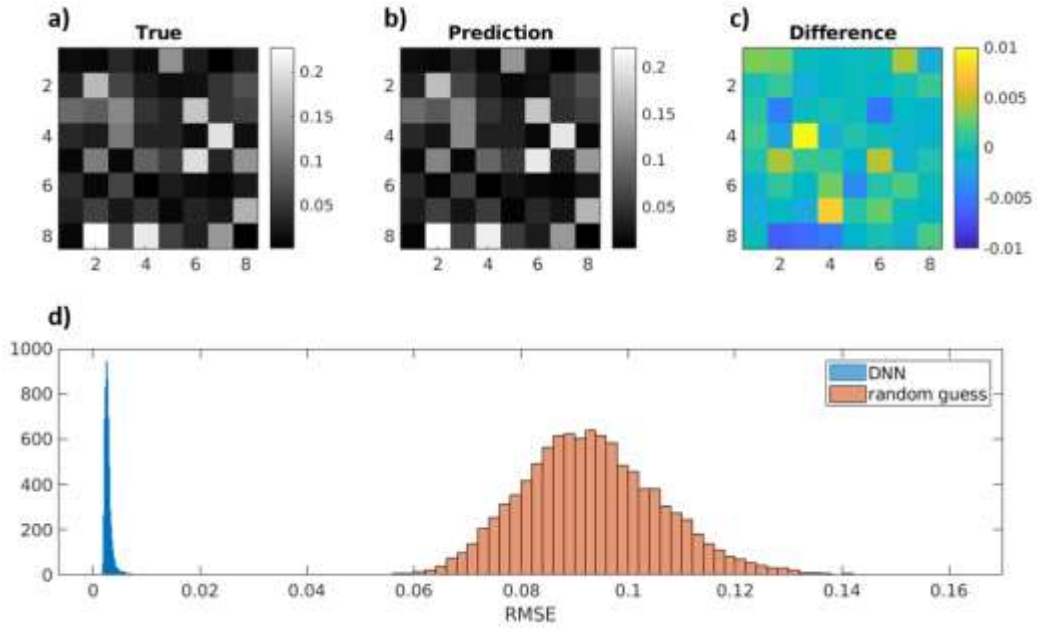


Figure 6.3: DNN performance on the set of 100,000 8x8 binary patterns and corresponding 8x8 normalized speckle images. a) Speckle image generated in simulation code. b) Speckle image predicted by DNN c) Difference map between true and prediction speckles. d) RMSE distributions for DNN (centered around 0.004) and for random guessing algorithm (centered around 0.09).

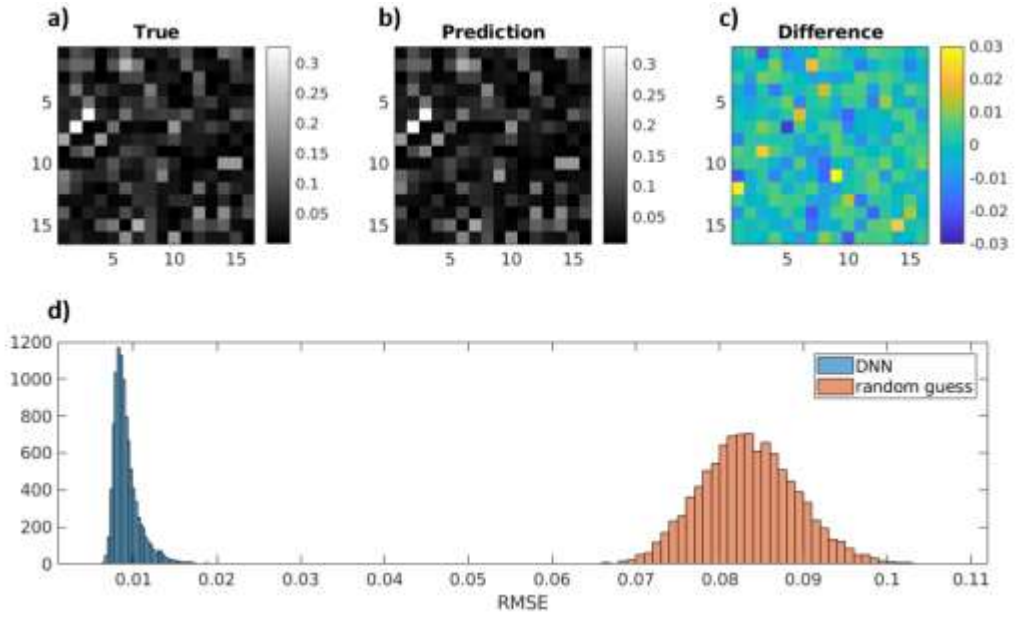


Figure 6.4: DNN performance on the set of 100,000 16x16 binary patterns and corresponding 16x16 normalized speckle images. a) Speckle image generated in simulation code. b) Speckle image predicted by DNN c) Difference map between true and prediction speckles. d) RMSE distributions for DNN (centered around 0.01) and for random guessing algorithm (centered around 0.085).

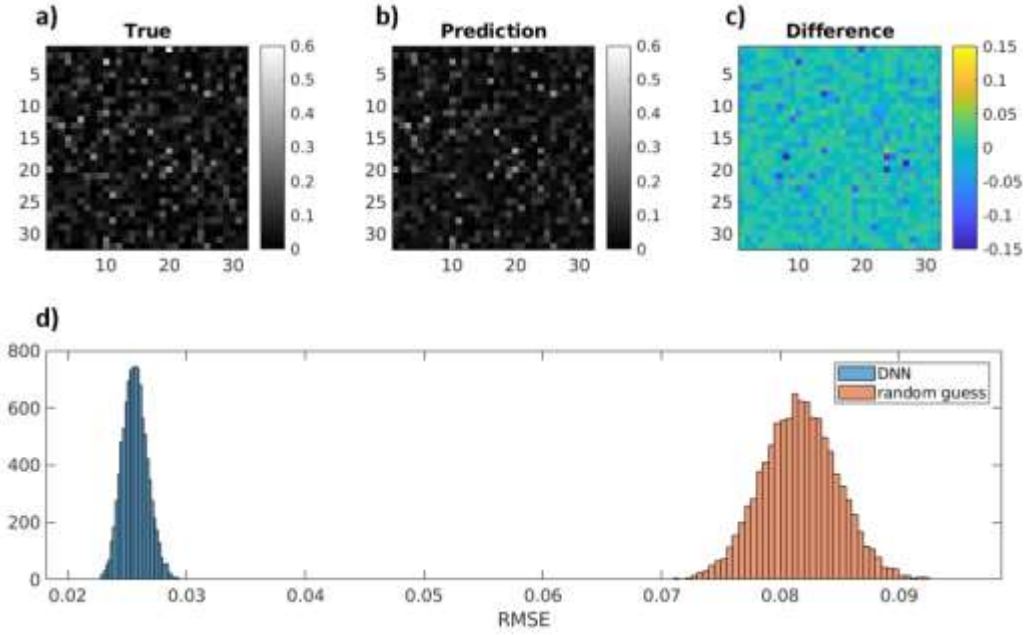


Figure 6.5: DNN performance on the set of 100,000 32x32 binary patterns and corresponding 32x32 normalized speckle images. a) Speckle image generated in simulation code. b) Speckle image predicted by DNN c) Difference map between true and prediction speckles. d) RMSE distributions for DNN (centered around 0.026) and for random guessing algorithm (centered around 0.082).

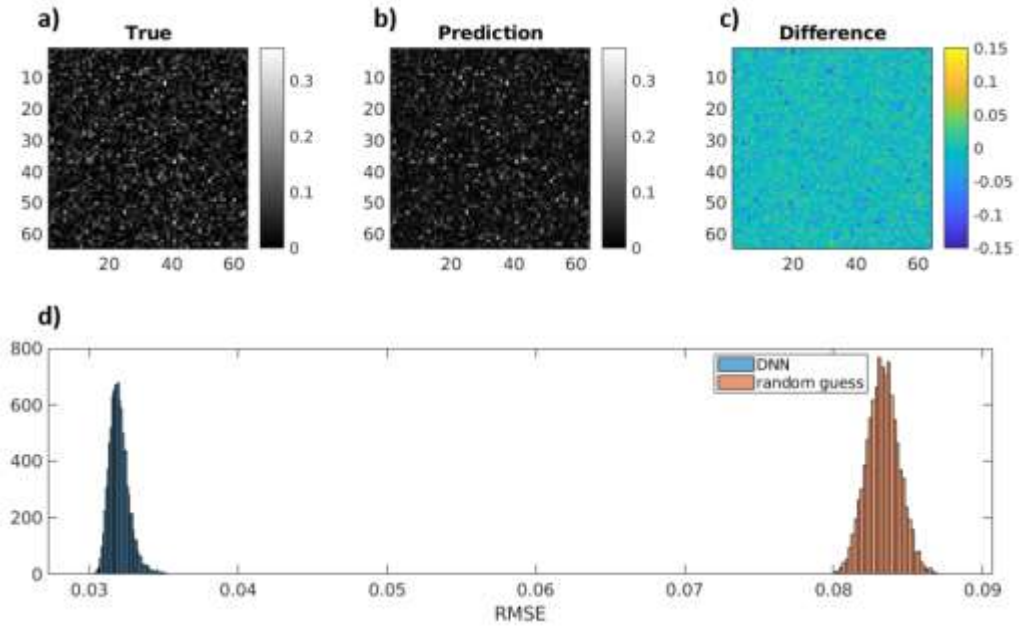


Figure 6.6: DNN performance on the set of 100,000 64x64 binary patterns and corresponding 64x64 normalized speckle images. a) Speckle image generated in simulation code. b) Speckle image predicted by DNN c) Difference map between true and prediction speckles. d) RMSE distributions for DNN (centered around 0.033) and for random guessing algorithm (centered around 0.084).

At 8x8 CRPs the DNN's prediction accuracy is ~20 times better than random guessing, whereas at 64x64 DNN's prediction is only ~2.5 times better than random guessing. As expected, with the increase of pattern size and thus system complexity the DNN's performance is deteriorated. The mean of RMSE distribution shifts from $4e^{-3}$ to $3.3e^{-2}$ at 8x8 and 64x64 pattern sizes respectively. In addition, the width of RMSE distribution is increasing meaning that at large pattern sizes the DNN's performance is noisier, so the reliability of the DNN model becomes worse.

Lastly, we studied the performance of neural networks against the training size of CRPs available to the attacker across all the pattern sizes. The results of this study would give the minimum number of CRPs the attacker should steal in order to train the DNN properly.

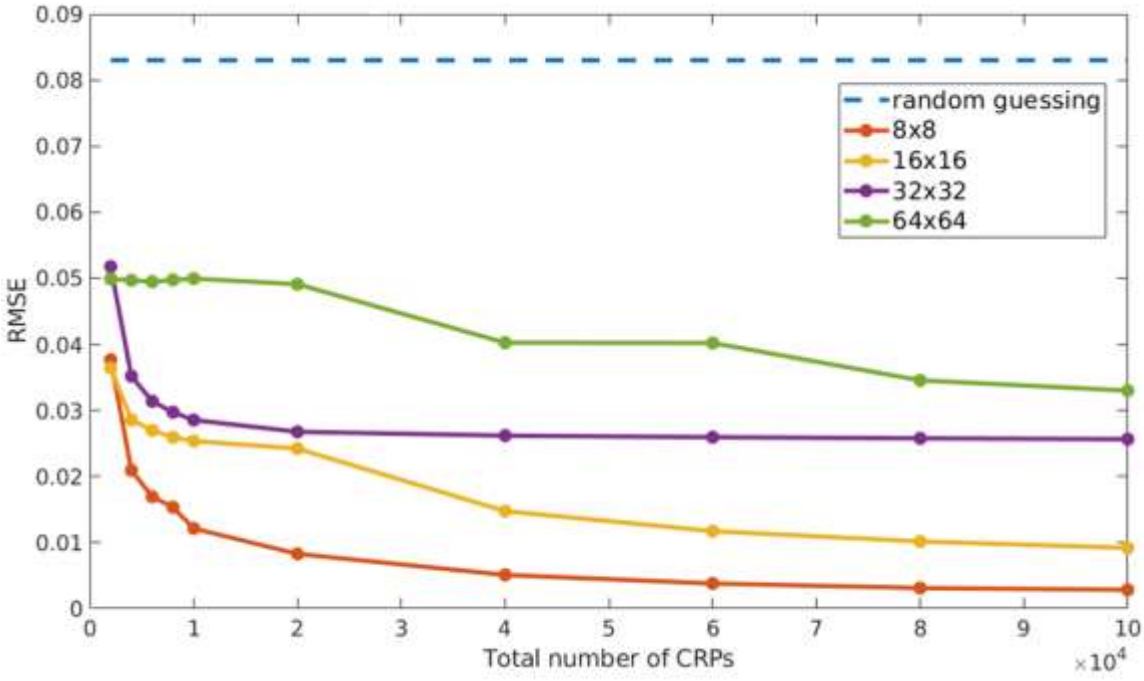


Figure 6.7: DNN root mean squared error on test data for all pattern sizes and random guessing prediction.

In **Figure 6.7** four learning curves are presented along with the performance curve corresponding to random guessing. On the x-axis I measured the total size of the database, 70% of which is used for training DNN. Notably, **Figure 6.7** shows the clear dependence of the prediction accuracy on the input size pattern. In cases for 8x8 and 16x16 patterns, I observe that the performance of the DNN plateaus after a total size of CRPs 40,000, whereas for 32x32 it plateaus after 20,000 CRPs. Hence, for 32x32 patterns an acquirement of 20,000-30,000 CRPs would allow the fraud to achieve relatively high prediction accuracy at the same level as at 100,000 CRPs. In other words, if an eavesdropper steals

70% of the total minimum CRP size which is 14,000-21,000 CRPs, he/she can use this subset to train DNN and emulate the whole behavior of simulated OSPUF. Further, the absence of the significant drop of prediction error at 64x64 case indicates that it is not clear how much of the CRP data required for breaking the simulated OSPUF.

6.4 Neural Networks Architectures

We find that the optimal configuration of the Neural Networks (NN) is highly dependent on the size of the input data, i.e. pattern size. For all NNs, we use Keras library with Theano/Tensorflow backend for the GPU-accelerated training of neural network [74, 75, 82]. The computer we use for training is running under Linux-Ubuntu 16.04 operating system with Nvidia GTX 1080 possessing 2560 CUDA cores and 8GB memory GDDR5X.

For the 8x8 pattern size, the best performing NN consists of one input layer with 64 neurons matching the input pattern size, 2 hidden layers with 64 and 4096 neurons each of them followed by ReLU activation layer. Since the speckle images are normalized and ranged between 0 and 1, the output layer consists of 64 neurons followed by the sigmoid activation layer. We train this NN for 2,000 epochs with batch size 256 and learning rate of $1e^{-3}$ using Adam optimization algorithm. The total number of trainable parameters is ~530,000 with the total training time of 10 minutes.

For the 16x16 pattern size, the best performing NN consists of one input layer with 256 neurons matching the input pattern size, 2 hidden layers with 256 and 10,000 neurons each of them followed by ReLU activation layer. Since the speckle images are normalized and ranged between 0 and 1, the output layer consists of 256 neurons followed by the sigmoid activation layer. We train this NN for 2,000 epochs with batch size 64 and learning

rate of $1e^{-3}$ using Adam optimization algorithm. The total number of trainable parameters is ~5M with total training time ~1.5 hour.

For the 32x32 pattern size, the best performing NN consists of one input layer with 1024 neurons matching the input pattern size, 2 hidden layers with 1024 and 8192 neurons respectively each of them followed by ReLU activation layer. Since the speckle images are normalized and ranged between 0 and 1, the output layer consists of 1024 neurons followed by the sigmoid activation layer. We train this NN for 2,000 epochs with batch size 128 and learning rate of $1e^{-4}$ using Adam optimization algorithm. The total number of trainable parameters is ~17M with total training time ~3 hours.

For the 64x64 pattern size, the best performing NN consists of one input layer with 4096 neurons matching the input pattern size, 3 hidden layers with 4096 neurons each of them followed by ReLU activation layer. Since the speckle images are normalized and ranged between 0 and 1, the output layer consists of 4096 neurons followed by the sigmoid activation layer. We train this NN for 2,000 epochs with batch size 64 and learning rate of $1e^{-3}$ using Adam optimization algorithm. The total number of trainable parameters is ~50M with total training time ~11 hours.

Besides the fully connected NN architectures, other configurations based on the convolutional neural networks and residual networks were investigated. However, the results for these types of networks were equivalent to the architectures above, thus I do not present them here.

Notably, increasing the input data dimension drastically increases the training time as well as the number of trainable parameters causing an extreme increase in computational

overhead in terms of required GPU memory. This fact can be clearly observed in **Figure 6.8.**

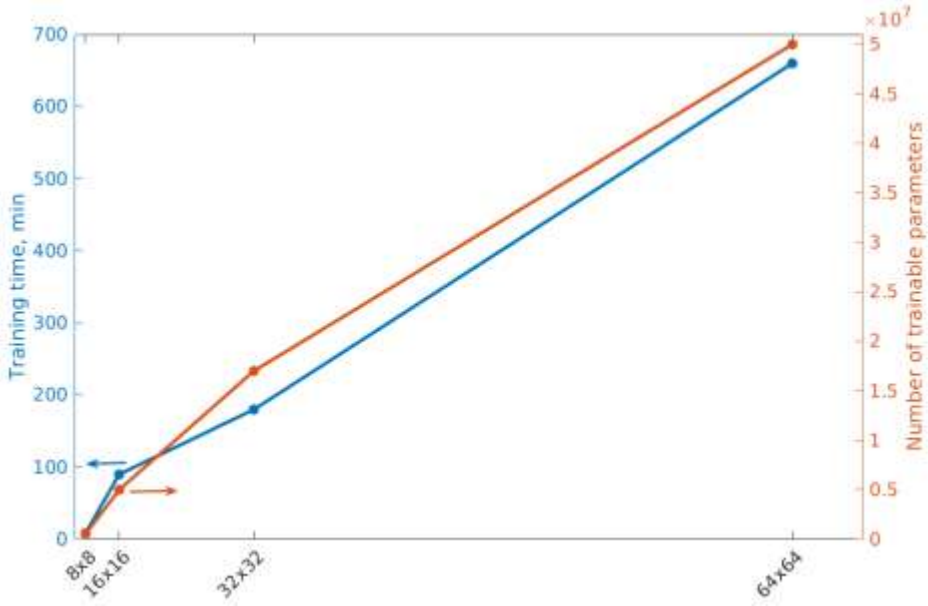


Figure 6.8: Training time and the complexity of NNs in terms of the number of parameters across the input pattern sizes.

Specifically, the time required for training and the complexity of the Neural Networks are drastically increasing with the size of the input pattern. Therefore, for larger input sizes (that is the case for real experimental data), an attacker needs an access to the significant amount of computational resources.

6.5 Conclusion

Our work on silicon photonic PUFs found that the resistance to ML attacks is rooted in the presence of optical nonlinearity in the cavity. In comparison, OSPUF's resistance is rooted in the vast information content and extremely high complexity and unpredictability

of the scattering process for large dimensions of the data, even though the scattering medium is linear.

In this chapter, I demonstrated machine learning attacks against the simplest simulation model of optical scattering PUF. As shown, for small dimensions of input data, Neural Networks are capable to reproduce the model's behavior with high accuracy, which is not the case for the larger pattern sizes. In a real scenario, attacking the experimental OSPUF is a much more challenging process for several factors. First, volumetric scattering is typically observed in the experiment, thus leading to the higher unpredictability of the system. Second, typical sizes of speckle images detected on CCD cameras are ranged from 128x128 to 512x512. At these dimensions, it would be extremely hard to train a Neural Network with such a huge set of parameters, in addition to requiring a significant amount of computational and time resources. Following the results from the simulation model, the DNN would be incapable to correctly predict the speckles given the binary patterns. Despite the fact, that the model presented here is simplified, the obtained results provide baseline requirements for anybody who is interested in attacking a true experimental OSPUF.

Chapter 7 : Conclusion and Future Directions

In this dissertation, I presented the concept of Physically Unclonable Function as a promising alternative hardware solution to existing cryptographic primitives. Conventional security mechanisms are based on the idea of digital storage of secret information which is vulnerable to copying, stealing, and destruction. The idea of Physical One-Way Functions, developed by Pappu et al., allowed us to extract the benefits of using the random disordered media in information security [29]. Pappu's work became a cornerstone for the next couple of decades of PUF research. Since then, there is a vast number of PUF implementations has been developed and applied as an alternative protection mechanism of the secret information. In Chapter 3, as one of the optical PUFs, I presented Silicon Photonic PUF, originally developed by Grubel et. al. [26, 67]. According to the original work, this optical PUF exhibits a number of advantages over the existing optical PUF systems. First, it was demonstrated that silicon PUF can be directly integrated onto electronic circuits and easily deployed with telecommunications infrastructure. Second, silicon photonic PUF is the first PUF that harnesses the chaotic nature of the cavity and nonlinearity of the silicon material that significantly increases the information content and the complexity of the output signal. Third, silicon PUF device exploits the ultrafast response of the cavity as one of the main protection mechanism from adversarial cloning or emulation process. In addition, low-cost

production, simplicity, and compactness make silicon photonic PUF an attractive technology in a range of potential authentication protocols including smart credit cards, mobile and desktop devices.

As an important extension on silicon photonic PUF, in the work presented here, I thoroughly investigated the security of the device and robustness to adversarial attacks. Specifically, I examined the resistance of photonic cavity to the state-of-the-art machine learning techniques. After performing sets of different machine learning attacks under various scenarios, I demonstrated clearly that the optical nonlinearity of the silicon material plays a crucial role in the device's resistance. Therefore, based on the results from [26] and Chapter 4,5, the true unclonability of this PUF is established.

Then, I return to the roots of PUF devices by exploring the possibility of attacks on Pappu's original Optical Scattering PUF. Surprisingly, no machine learning attacks have ever been reported on scattering PUFs. This motivated our group to study the underlying reasons for strong resistance of OSPUFs to modeling attacks. In Chapter 6, I constructed a simple model of single surface scattering to collect the required dataset that would be used in attacking procedure. I demonstrated that even at the simplest level of OSPUF representation, the dimension of input data plays a critical role in the resilience against ML attacks.

The current results for both of optical PUFs open the new avenues for potential research directions. For silicon Photonic PUF, there is a plethora of ways and directions for continued research. Regarding the device itself, a number of future steps are mentioned in B. Grubel Ph.D. thesis [26] including the optimization of coupling efficiency, improvement of ray-tracing models, exploring other techniques for post-processing of raw

responses, and etc. In addition to that, I would like to add, that it would be interesting to implement photonic PUF based on different materials such as amorphous silicon. Potentially, this could enhance the security of the device and overcome the problem of high optical loss of the cavity (~30 dB). Another future step is to implement a set of optimization techniques based on the size of the cavity and the amount of nonlinearity for different shapes and geometric configurations. Regarding the resistance to ML attacks, it is highly important to track the resistance to rapidly growing ML area. It is well-known fact, that Deep Learning is one of the "hottest" areas of Artificial Intelligence area with novel approaches introduced from year to year. Therefore, it is very important to ensure the protection of silicon PUF from future ML models.

Regarding the scattering PUF, I suggest constructing the advanced simulation models of OSPUFs. In Chapter 6, the scattering was estimated using one random phase mask. Hence, it would be interesting to model the volumetric scattering by introducing several phase masks, thus incorporating Fresnel diffraction theory for the light propagation between these masks. Volumetric scattering model would be a more realistic representation of physical scattering token. Similarly, one should investigate the resistance against machine learning attacks in the given case. Since the DNN emulates the simulation model of OSPUF at small pattern sizes, I would also suggest building the experimental setup where thin scattering token is illuminated by small binary patterns. Susceptibility of the given system to ML attacks and comparison of the results to the results from Chapter 6 would be an interesting analysis.

On a final note, I hope that this dissertation would be useful and interesting for people who found themselves in a various research area such as cryptography, information

and hardware security, information theory, nonlinear optics, machine learning and artificial intelligence.

Bibliography

1. K. Ashton, That ‘‘Internet of Things’’ thing, RFiD Journal (2009)
2. Edewede Oriwoh , Marc Conrad , ‘Things’ in the Internet of Things: Towards a Definition, *International Journal of Internet of Things*, Vol. 4 No. 1, 2015, pp. 1-5. doi: 10.5923/j.ijit.20150401.01.
3. Gubbi, Jayavardhana & Buyya, Rajkumar & Marusic, Slaven & Palaniswami, Marimuthu. (2012). Internet of Things (IoT): A Vision, Architectural Elements, and Future Directions. Future Generation Computer Systems. 29. 10.1016/j.future.2013.01.010.
4. B. Gassend, D. Clarke, M. van Dijk, and S. Devadas, “Silicon physical random functions,” Proc. 9th ACM Conf. Comput. Commun. Secur. - CCS ’02, p. 148, 2002.
5. S. Sagiroglu and D. Sinanc, "Big data: A review," *2013 International Conference on Collaboration Technologies and Systems (CTS)*, San Diego, CA, 2013, pp. 42-47.
6. Intel IT Center, "Planning Guide: Getting Started with Hadoop", Steps IT Managers Can Take to Move Forward with Big Data Analytics, June 2012
7. S. Singh and N. Singh, "Big Data Analytics", 2012 International Conference on Communication, Information & Computing Technology Mumbai India, IEEE, October 2011
8. G. J. Cheng, L. T. Liu, X. J. Qiang and Y. Liu, "Industry 4.0 Development and Application of Intelligent Manufacturing," *2016 International Conference on Information System and Artificial Intelligence (ISAI)*, Hong Kong, 2016, pp. 407-410.
9. Maes, Roel. (2013). Physically Unclonable Functions: Constructions, Properties and Applications. 10.1007/978-3-642-41395-7.
10. McCarthy, C. (2006). "Digital Libraries: Security and Preservation Considerations". In Bidgoli, H. *Handbook of Information Security, Threats, Vulnerabilities, Prevention, Detection, and Management*. **3**. John Wiley & Sons. pp. 49–76.
11. Cryptographic Applications with Physically Unclonable Functions, M. Deutschmann, Master Thesis, 2010
12. It was created by IBM's (International Business Machines) *Walter Tuchman (1997). "A brief history of the data encryption standard". Internet besieged: countering cyberspace scofflaws. ACM Press/Addison-Wesley Publishing Co. New York, NY, USA. pp. 275–280.*
13. Daemen, J., Rijmen, V.: The Design of Rijndael: AES— The Advanced Encryption Standard. Springer, Heidelberg (2002)

14. S. Hatkar, S & K. Pawar, Bhagyashri. (2016). Symmetric key algorithm using vernam cipher: VSA. 1-5. 10.1109/INVENTIVE.2016.7830196.
15. M. Bellovin, Steven. (2011). Frank Miller: Inventor of the One-Time Pad. Cryptologia. 35. 203-222. 10.1080/01611194.2011.583711.
16. Stallings, William (1990-05-03). . Prentice Hall. p. 165. .
17. R.L. Rivest, A. Shamir, and L. Adleman, A method for obtaining digital signatures and public-key cryptosystems, Commun. ACM, Feb. 1978, 21(2): 120-126.
18. A. Ekert, "From quantum-codemaking to quantum code-breaking," Geom. Issues Found. Sci. Oxford Univ., pp. 1–21, 1997.
19. R. Anderson, Security Engineering, Second Edition. Indianapolis: Wiley Publishing, Inc., 2008.
20. Rijmen, Vincent and Elisabeth Oswald. "Update on SHA-1." *IACR Cryptology ePrint Archive* 2005 (2005): 10.
21. "The Economic Impacts of Counterfeiting and Piracy: Report prepared for BASCAP and INTA," in Frontier Economics (2016), p. 61.
22. Josh Ellenbogen, Reasoned and Unreasoned Images: The Photography of Bertillon, Galton, and Marey (University Park, PA, 2012)
23. D. Bauder, An Anti-Counterfeiting Concept for Currency Systems. Technical Report PTK- 11990, Sandia National Labs, Albuquerque, NM, 1983
24. Commission on Engineering and Technical Systems (CETS), Counterfeit Deterrent Features for the Next-Generation Currency Design, Appendix E (The National Academic Press, Washington, DC, 1993)
25. C. Bohm and M. Hofer, Physical Unclonable Functions in Theory and Practice. Graz, Austria: Springer, 2013.
26. Brian Grubel, Silicon Photonic Physical Unclonable Functions. PhD Thesis, 2017
27. 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. LNCS. Springer, to appear 2007.
28. Tuyls, P., Schrijen, G.-J., Škorić, B., van Geloven, J., Verhaegh, N., and Wolters, R. 2006. Read-Proof Hardware from Protective Coatings. In Workshop on Cryptographic Hardware and Embedded Systems – CHES 2006. Lecture Notes in Computer Science (LNCS), vol. 4249. Springer, 369–383. pages xvi, 29, 68, 74, 211
29. R. Pappu, B. Recht, J. Taylor, and N. Gershenfeld, "Physical One-Way Functions," Science, vol. 297, no. 5589, Cambridge, pp. 2026–2030, 2002.
30. R. Pappu, "Physical One-Way Functions," Massachusetts Institute of Technology, 2001.
31. J.W. Lee, D. Lim, B. Gassend, G.E. Suh, M. van Dijk, S. Devadas, in A Technique to Build a Secret Key in Integrated Circuits for Identification and Authentication Application. Proceedings of the Symposium on VLSI Circuits, 2004, pp. 176–159
32. D. Lim, Extracting Secret Keys from Integrated Circuits. Master's thesis, MIT, MA, USA, 2004
33. B. Gassend, Physical Random Functions. Master's thesis, MIT, MA, USA, 2003

34. Maes, Roel & Verbauwhede, Ingrid. (2010). Physically Unclonable Functions: A Study on the State of the Art and Future Research Directions. Towards Hardware-Intrinsic Security. 3-37. 10.1007/978-3-642-14452-3_1.
35. S. Kumar, J. Guajardo, R. Maes, G.J. Schrijen, P. Tuyls, in Extended Abstract: The Butterfly PUF Protecting IP on Every FPGA. IEEE International Workshop on Hardware-Oriented Security and Trust, 2008, HOST 2008, Anaheim, CA, USA, 2008, pp. 67–70
36. Y. Su, J. Holleman, B. Otis, in A 1.6pj/bit 96% Stable Chip-ID Generating Circuit Using Process Variations. IEEE International Solid-State Circuits Conference, ISSCC 2007. Digest of Technical Papers (IEEE Computer Society, Washington, DC, 2007), pp. 406–611
37. Rührmair, U., Sehnke, F., Sölter, J., Dror, G., Devadas, S., & Schmidhuber, J. (2010). Modeling attacks on physical unclonable functions. *IACR Cryptology ePrint Archive, 2010*, 251.
38. U. Rührmair *et al*., "PUF Modeling Attacks on Simulated and Silicon Data," in *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 11, pp. 1876-1891, Nov. 2013. doi: 10.1109/TIFS.2013.2279798
39. C. Wachsmann and A.-R. Sadeghi, Physically Unclonable Functions (PUFs): Applications, Models, and Future Directions, vol. 9, no. 1. Morgan and Claypool Publishers, 2014.
40. A. Vijayakumar, V. C. Patil, C. B. Prado and S. Kundu, "Machine learning resistant strong PUF: Possible or a pipe dream?" *2016 IEEE International Symposium on Hardware Oriented Security and Trust (HOST)*, McLean, VA, 2016, pp. 19-24. doi: 10.1109/HST.2016.7495550
41. R. Horstmeyer, B. Judkewitz, I. M. Vellekoop, S. Assawaworrarit, and C. Yang, “Physical key-protected one-time pad,” Sci. Rep., vol. 3, no. x, 2013.
42. U. Rührmair, C. Hilgers, and S. Urban, “Optical PUFs Reloaded,” Eprint.Iacr. Org, 2013.
43. Sebastianus A. Goorden, Marcel Horstmann, Allard P. Mosk, Boris Škorić, and Pepijn W. H. Pinkse, "Quantum-secure authentication of a physical unclonable key," Optica 1, 421-424 (2014)
44. P. R. Seem, J. D. R. Buchanan, and R. P. Cowburn, “Impact of surface roughness on laser surface authentication signatures under linear and rotational displacements.,” Opt. Lett., vol. 34, no. 20, pp. 3175–7, 2009.
45. Hammouri G., Dana A., Sunar B. (2009) CDs Have Fingerprints Too. In: Clavier C., Gaj K. (eds) Cryptographic Hardware and Embedded Systems - CHES 2009. CHES 2009. Lecture Notes in Computer Science, vol 5747. Springer, Berlin, Heidelberg
46. Devadas, Srinivas & Suh, Edward & Paral, Sid & Sowell, Richard & Ziola, Tom & Khandelwal, Vivek. (2008). Design and Implementation of PUF-Based Unclonable RFID ICs for Anti-Counterfeiting and Security Applications. 2008 IEEE International Conference on RFID (Frequency Identification), IEEE RFID 2008. 58 - 64. 10.1109/RFID.2008.4519377.
47. Tuyls P., Škorić B. (2006) Physical Unclonable Functions for enhanced security of tokens and tags. In: ISSE 2006 — Securing Electronic Busines Processes. Vieweg

48. J. A. Roy, F. Koushanfar and I. L. Markov, "EPIC: Ending Piracy of Integrated Circuits," *2008 Design, Automation and Test in Europe*, Munich, 2008, pp. 1069-1074. doi: 10.1109/DATE.2008.4484823
49. Kish, Laszlo B. et al. "Unconditionally secure credit/debit card chip scheme and physical unclonable function." *CoRR* abs/1605.02355 (2016): n. pag.
50. Liu, Wenchao et al. "A Trustworthy Key Generation Prototype Based on DDR3 PUF for Wireless Sensor Networks." *2014 International Symposium on Computer, Consumer and Control* (2014).
51. Suh, G. Edward and Srinivas Devadas. "Physical Unclonable Functions for Device Authentication and Secret Key Generation." *2007 44th ACM/IEEE Design Automation Conference* (2007): 9-14.
52. W O 'donnell, Charles & Edward Suh, G & Devadas, Srinivas. (2018). PUF-based random number generation.
53. Y. Dodis, L. Reyzin, A. D. Smith, C. Cachin, J. Camenisch, "Fuzzy extractors: How to generate strong keys from biometrics and other noisy data", *Proceedings of EUROCRYPT 2004 ser. LNCS*, Springer, vol. 3027, pp. 523-540, 2004.
54. Y. Dodis, R. Ostrovsky, L. Reyzin, A. D. Smith, "Fuzzy extractors: How to generate strong keys from biometrics and other noisy data", *SIAM J. Comput.*, vol. 38, no. 1, pp. 97-139, 2008.
55. Noor, Nur Qamarina Mohd et al. "Defense Mechanisms against Machine Learning Modeling Attacks on Strong Physical Unclonable Functions for IOT Authentication: A Review." (2017).
56. C. Wachsmann and A-R Sadeghi, Physically Unclonable Functions (PUFs). Synthesis Lectures on Information Security, Privacy, and Trust. Morgan & Claypool Publishers. 2004. https://doi.org/10.2200/s00622ed1v01y201412spt012
57. J. Delvaux and I. Verbauwhede, "Fault Injection Modeling Attacks on 65 nm Arbiter and RO Sum PUFs via Environmental Changes," IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 61, no. 6, pp. 1701-1713, June 2014. DOI: 10.1109/TCSI.2013.2290845.
58. Y. Oren, A.-R. Sadeghi dan C. Wachsmann, On the Effectiveness of the Remanence Decay Side-Channel to Clone Memory-Based PUFs. In: Bertoni G., Coron JS. (eds) Cryptographic Hardware and Embedded Systems - CHES 2013. CHES 2013. Lecture Notes in Computer Science, vol 8086. Springer, Berlin, Heidelberg. 2013. DOI: https://doi.org/10.1007/978-3-642-40349-1-7
59. C. Helfmeier, C. Boit, D. Nedospasov and J. P. Seifert, "Cloning Physically Unclonable Functions," 2013 IEEE International Symposium on Hardware-Oriented Security and Trust (HOST), Austin, TX, 2013, pp. 1-6. DOI: 10.1109/HST.2013.6581556.
60. U. Rührmair and D. E. Holcomb, "PUFs at a glance," *2014 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, Dresden, 2014, pp. 1-6.doi: 10.7873/DATE.2014.360
61. Yu MD.., M'Raihi D., Sowell R., Devadas S. (2011) Lightweight and Secure PUF Key Storage Using Limits of Machine Learning. In: Preneel B., Takagi T. (eds) Cryptographic Hardware and Embedded Systems – CHES 2011. CHES 2011. Lecture Notes in Computer Science, vol 6917. Springer, Berlin, Heidelberg

62. Christopher M. Bishop. 2006. Pattern Recognition and Machine Learning (Information Science and Statistics). Springer-Verlag, Berlin, Heidelberg.
63. Daihyun Lim. Extracting Secret Keys from Integrated Circuits. Msc thesis, MIT, 2004.
64. M. Majzoobi, F. Koushanfar and M. Potkonjak, "Lightweight secure PUFs," *2008 IEEE/ACM International Conference on Computer-Aided Design*, San Jose, CA, 2008, pp. 670-673. doi: 10.1109/ICCAD.2008.4681648
65. R. E. Schapire, The Boosting Approach to Machine Learning: An Overview. In: Denison D.D., Hansen M.H., Holmes C.C., Mallick B., Yu B. (eds) Nonlinear Estimation and Classification. Lecture Notes in Statistics, vol 171. Springer, New York, NY. DOI: https://doi.org/10.1007/978-0-387-21579-2_9
66. R. Yashiro, T. Machida, M. Iwamoto, and K. Sakiyama, ―DeepLearning-Based Security Evaluation on Authentication Systems Using Arbiter PUF, ‖ In: Ogawa K., Yoshioka K. (eds) Advances in Information and Computer Security. IWSEC 2016. Lecture Notes in Computer Science, vol 9836. Springer, Cham. https://doi.org/10.1007/978-3-319-44524-3_16.
67. Brian C. Grubel, Bryan T. Bosworth, Michael R. Kossey, Hongcheng Sun, A. Brinton Cooper, Mark A. Foster, and Amy C. Foster, "Silicon photonic physical unclonable function," Opt. Express 25, 12710-12721 (2017)
68. B. C. Grubel, D. S. Vresilovic, B. T. Bosworth, M. Kossey, A. C. Foster, M. A. Foster, and A. B. Cooper, "Light transport through ultrafast chaotic micro-cavities for photonic physical unclonable functions," in Conf. Inf. Sci. Syst. (CISS, 2017).
69. B. C. Grubel, B. T. Bosworth, M. R. Kossey, A. B. Cooper, M. A. Foster, A. C. Foster, Information-Dense Nonlinear Photonic Physical Unclonable Function, arXiv:1711.02222
70. P. D. Fisher and R. Nesbitt, "The test of time. Clock-cycle estimation and test challenges for future microprocessors," IEEE Circuits Devices Mag. 14(2), 37–44 (1998)
71. G. P. Agrawal, "Nonlinear Fiber Optics," 3rd Edition, Academic Press, San Diego, 2001.
72. 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, Aug. 2014.
73. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. Nature 521, 436–444 (2015)
74. Chollet, F. (2015) Keras, GitHub. https://github.com/fchollet/keras
75. Theano Development Team, Al-Rfou et. al Theano: A Python framework for fast computation of mathematical expressions.
76. Kingma and J. Ba. Adam: A method for stochastic optimization. In ICLR, 2015.
77. John Kerr LL.D. (1875) XL. *A new relation between electricity and light: Dielectrified media birefringent*, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 50:332, 337-348
78. Franken, P. A., Hill, A. E., Peters, C. W. & Weinreich, G. Generation of optical harmonics. *Phys. Rev. Lett.* **7**, 118–119 (1961)
79. Robert W. Boyd. 2008. Nonlinear Optics, Third Edition (3rd ed.). Academic Press, Inc., Orlando, FL, USA.

80. New, G. (2011). *Introduction to Nonlinear Optics*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511975851
81. Skoric, Boris & Schrijen, Geert-Jan & Ophey, Wil & Wolters, Rob & Verhaegh, Nynke & Geloven, Jan. (2007). Experimental Hardware for Coating PUFs and Optical PUFs. 255-268. 10.1007/978-1-84628-984-2_15.
82. Abadi, M. et al. TensorFlow: Large-scale machine learning on heterogeneous systems (2015). URL http: //tensorflow.org/.
83. A. C. Turner-Foster *et al.*, "Ultrashort free-carrier lifetime in low-loss silicon nanowaveguides," *Opt. Express*, vol. 18, no. 4, p. 3582, 2010

Vita

Iskandar Atakhodjaev was born in January 4th, 1991 in Samarkand, Uzbekistan to Mr. Atakhodjaev Alisher and Mrs. Vafokulova Iroda. He finished 9 grades of high school № 17 in Samarkand in 2005. The same year he transitioned to mathematics school in Troitsk №3, Russia to finish the last 3 grades. He was accepted to Moscow Institute of Physics and Technology (MIPT) for his undergraduate studies, where he was awarded the Full Scholarship for the excellent performance on Physical and Mathematical Contest arranged by MIPT. He graduated with a B.S. in Applied Mathematics and Physics in June 2012. He attended Johns Hopkins University for his graduate studies where he received his M.A. in Physics in May 2015, the M.S.E in Computer Science in May 2018, and Ph.D. in Physics in August 2018.