

Security of Quantum-Readout PUFs against quadrature based challenge estimation attacks

Boris Škorić, Allard P. Mosk, Pepijn W.H. Pinkse

Abstract

The concept of quantum-secure readout of Physical Unclonable Functions (PUFs) has recently been realized experimentally in an optical PUF system. We analyze the security of this system under the strongest type of classical attack: the challenge estimation attack. The adversary performs a measurement on the challenge quantum state in order to learn as much about it as he can. Using this knowledge he then tries to reconstruct the challenge and to emulate the PUF. We consider quadrature measurements, which are the most informative practical measurements known to us. We prove that even under this attack the expected number of photons detected in the verification mechanism is approximately a factor $S + 1$ too low; here S is the Quantum Security Parameter, defined as the number of modes in the optical system divided by the number of photons in the challenge. The photon count allows for a reliable distinction between an authentic PUF and a challenge estimation attack.

1 Introduction

1.1 Physical Unclonable Functions

Authentication plays an important role in society, providing the trust without which people and automated systems are unwilling to engage in transactions. Authentication is usually based on either “something that you know” or “something that you possess”. In the second case it is highly desirable to possess a token that is difficult to clone, even for the manufacturer of the token. With the advent of Physical Unclonable Functions (PUFs), physical systems have been identified which satisfy strong uniqueness and unclonability properties, e.g. phenomena such as laser speckle based on multiple scattering. A PUF is a complex piece of material that is difficult to reproduce accurately because its manufacture inherently contains uncontrollable steps [16, 8, 3, 18, 11, 5, 25, 19, 17]. A stimulus can be applied to the PUF (‘challenge’), leading to observable behavior (the ‘response’) that depends in a complex way on the challenge and the minute details of the PUF’s structure. The combination of a challenge and the corresponding response is called a Challenge-Response Pair (CRP).

A good example of a physical system satisfying the abstract requirements above are the so-called Optical PUFs. These are three-dimensional diffusive structures containing optical scatterers at random positions. When an Optical PUF is illuminated by a laser, the transmitted and reflected light shows a random-looking pattern of dark and bright spots known as speckle. The properties of the laser beam (such as wavelength, angle, focus) constitute the challenge; the speckle pattern is the response. It depends strongly on the challenge as well as on the exact positions of the scatterers. Optical PUFs support a large number of independent CRPs.[20, 12, 23]

1.2 “Hands-off” verification of PUFs; emulation attacks

A PUF-based authentication or anti-counterfeiting system typically has two phases: enrollment and verification. In the enrollment phase the Verifier applies a limited number of random challenges to a PUF and records the CRPs in a database, e.g. coupled to an identification number. Later, in the verification phase, the Verifier has to decide whether a PUF with a given identifier is authentic.

He looks up the CRPs listed under the given identifier, and by challenging the PUF anew checks if it produces the listed responses. The procedure sketched above is extremely reliable when the Verifier has full control over the PUF, e.g. he holds the PUF during the verification phase. There are many cases, however, where the PUF owner is unwilling or unable to hand over his PUF. He may not trust the Verifier, or he is too far away from him. In such situations the Verifier must do verification without having full control. We call this “hands-off” or “remote” verification. Achieving a high level of security is far more difficult in this setting. There is a serious danger of emulation attacks (‘spoofing’).

For most PUFs the number of supported independent CRPs is ‘finite’, in the sense that anyone holding the PUF can, in a feasible amount of time, extract enough information from the PUF to be able to compute (or look up) the response to any future PUF challenge without having to use the PUF any more. In other words, in practice most PUFs can be *emulated*. This also holds for Optical PUFs, though the emulation may require quite a large database of CRPs. In general, the stricter the robustness requirements (i.e. reproducibility of responses), the smaller the challenge space and hence the bigger the danger of emulation attacks.

Given these considerations, it is prudent to assume that for every PUF that has ever been handed out a fast¹ emulation program is publicly available. The traditional way to retain any control in the “hands-off” setting is to have some means to ensure that no spoofing is going on, e.g. a trusted measurement device in the field or extra sensors for detecting specific kinds of spoofing. There is an important drawbacks to this approach: The extra anti-spoofing means add cost to the verification hardware, while it is difficult to ascertain how secure the system really is. For instance, remote trusted devices need to be tamper-proofed, but hardware attacks improve with time. Similarly, new techniques are continuously developed to spoof sensors. Thus, as often the case in the field of hardware security, it is an arms race between attackers and defenders.

1.3 Quantum readout of PUFs

An elegant way out of this expensive arms race was proposed by Škorić [24]: Quantum Readout (QR) of PUFs. The physical challenge is a quantum state. The PUF interacts with the challenge state via unitary evolution and produces a response that is also a quantum state. The Verifier, knowing from the enrollment phase what the response state is supposed to be, is able to verify if the response is correct. All this can be done without a trusted remote device, because of the inherent tamper-resistant properties of single quanta. The No Cloning Theorem [26, 6] states that an unknown single quantum cannot be copied onto another particle. (It has been exploited spectacularly in Quantum Key Distribution schemes [1]). One of the implications is that the state of an unknown quantum challenge cannot be fully determined, preventing the emulation: *If the attacker cannot be sure what the challenge is, he cannot reliably run his emulation program.*² (In fact, the challenge does not have to consist of a single quantum; it is allowed to consist of multiple quanta, provided that the attacker cannot accurately determine the challenge.) By repeatedly sending random challenges, the verifier ensures that the probability of successful spoofing is brought down exponentially. One of the nice aspects of the QR-PUF technique is that the challenge space does not have to be large, and that the scheme is still secure if the list of responses is publicly known. Apart from solving the “hands-off” authentication problem, the QR-PUF concept can also be used to create an authenticated quantum channel without the need for pre-shared entangled particles; instead the authentication is based on public data. Such a channel could be employed e.g. for Quantum Key Distribution, removing the need for pre-sharing a secret (e.g. a classical key or an entangled quantum state).

Quantum Readout of PUFs was first experimentally realized by Goorden et al. [10]. An optical PUF was used, consisting of a layer of zinc oxide nanoparticles on a substrate. The challenge was

¹Authentication protocols have been proposed for PUFs that can be emulated, but only *slowly*. [13]

²It is assumed that quantum computers do not exist or, if they do, that the emulation on a quantum computer is not feasible, e.g. too expensive or not fast enough.

implemented as a weak n -photon coherent light pulse³ with a randomly chosen wavefront. The scattering in the PUF scrambles the wavefront. The response is the scrambled n -photon light pulse. The problem of testing the correctness of the few-photon speckle pattern responses was solved using recently developed wavefront shaping techniques [21, 7, 15]. The response passes through a Spatial Light Modulator (SLM2 in Fig. 1), an array of switchable phase-rotation pixels. The SLM is programmed to match the phases contained in the wavefront of the expected response. If the response is correct, the SLM aligns all the pixels in the wavefront to have the same phase, leading to an essentially parallel beam that can be focused onto a small area. A sensitive detector measures the number of photons arriving in this area. Only wavefronts that are close to the enrolled response lead to a significant number of detected photons.

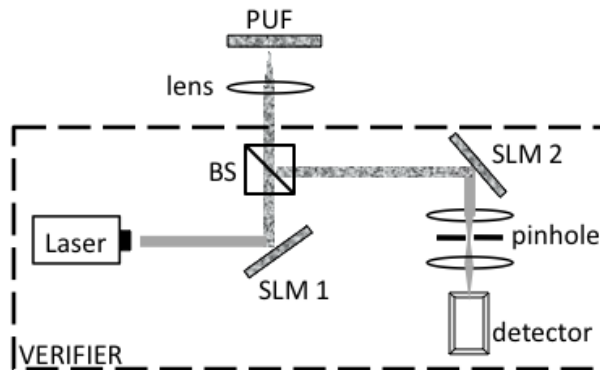


Figure 1: *Schematic overview of the setup used in [10]. The components in the dashed box are under the verifier’s control. The first SLM shapes the wavefront to create the challenge. BS is a polarizing beam splitter. SLM2 is tuned to ‘decode’ the correct response wavefront into a parallel beam. The detector counts how many photons pass through the pinhole.*

Let K denote the number of momentum modes in the challenge wavefront. The amount of information contained in the challenge is of order K . By keeping the number of photons (n) well below K , it is ensured that an attacker who intercepts the photons cannot learn enough about the challenge to fully reconstruct the wavefront. The ratio $S = K/n$ was dubbed the Quantum Security Parameter. A heuristic argument showed that the number of detected photons that an attacker can achieve is approximately a fraction $1/S$ of what the correct response speckle pattern would yield.

1.4 Contributions

We analyze the security of the optical QR-PUF system described in Ref. [10] against *classical* attacks. We consider the strongest class of classical attacks, ‘challenge estimation’ attacks, in which the adversary performs a measurement on the challenge state and then constructs a response state based on his measurement outcome and public information about the PUF. We always assume that the attacker has perfect optical equipment and perfect detectors.

- We model the action of the Spatial Light Modulator and derive a result for the fraction of photons that arrives in the detector when the response wavefront is correct. Our result is consistent with the experiments of Ref. [10].
- We derive a general formula for the number of photons that arrive in the detector given that the challenge estimation attack takes place, with an arbitrary choice of measurement by the attacker.

³A coherent pulse with $n > 1$ was chosen in order to avoid the ‘fragility’ of single-photon states, i.e. the difficulty and cost of handling them.

- We specialize to the case where the attacker uses beam splitters and does quadrature measurements. We prove that the best obtainable result for the attacker occurs when the beam splitting is uniform, i.e. into equal parts. For this attack, the number of expected photons at the detector is roughly a factor $S + 1$ lower than for the correct response wavefront. This discrepancy allows for reliable distinction between an authentic PUF and an attack.

2 Preliminaries

2.1 Notation

Quantum states are represented as vectors in a Hilbert space. We adopt the usual Dirac ‘bra’ and ‘ket’ notation; $|\psi\rangle$ stands for a quantum state labelled by some description ψ which summarizes all the knowable information about the state. The Hermitian conjugate is denoted as $\langle\psi|$. The notation for the inner product between two states is $\langle\psi_1|\psi_2\rangle$. We will consider only normalized states, i.e. satisfying $\langle\psi|\psi\rangle = 1$. Real-valued observables are represented by Hermitian operators acting on the Hilbert space. The expectation value of an operator A , given state $|\psi\rangle$, is denoted as $\langle\psi|A|\psi\rangle$, or in shorthand notation $\langle A\rangle$ when it is clear from the context what the state is. The commutator of two operators is $[A, B] = AB - BA$. The Hermitian conjugate of A is A^\dagger .

Our description of wavefronts follows the standard approach[9, 4] in terms of discrete ‘modes’ labeled by their transverse momentum (wave number). The set of modes in the challenge is denoted as \mathcal{K} , and the modes themselves as two-dimensional wave vectors (in boldface notation) $\mathbf{k} \in \mathcal{K}$. We define $K = |\mathcal{K}|$. We consider only a single wavelength of light. The set of modes in the response is denoted as \mathcal{K}' , with $K' = |\mathcal{K}'|$. It holds that $K' \geq K$, since the diffusion in the PUF causes the area from which light exits to be slightly larger than the illuminated spot.

The creation operator for a photon with wave vector \mathbf{k} is written as $a_{\mathbf{k}}^\dagger$. We have the commutation relation $[a_{\mathbf{k}}, a_{\mathbf{k}'}^\dagger] = \delta_{\mathbf{k}\mathbf{k}'}$. The vacuum is denoted as $|0\rangle$, with $a_{\mathbf{k}}|0\rangle = 0$. The photon counting operator for mode \mathbf{k} is $N_{\mathbf{k}} = a_{\mathbf{k}}^\dagger a_{\mathbf{k}}$. The two ‘quadrature’ operators are defined as $X_{\mathbf{k}} = (a_{\mathbf{k}} + a_{\mathbf{k}}^\dagger)/2$ and $Y_{\mathbf{k}} = (a_{\mathbf{k}} - a_{\mathbf{k}}^\dagger)/(2i)$.

The challenge wavefront is fully characterized by a complex vector $\mathbf{c} = (c_{\mathbf{k}})_{\mathbf{k} \in \mathcal{K}}$ satisfying the normalization $\sum_{\mathbf{k} \in \mathcal{K}} |c_{\mathbf{k}}|^2 = 1$. We write $c_{\mathbf{k}} = c_{\mathbf{k}}^{\text{re}} + ic_{\mathbf{k}}^{\text{im}}$. We model the action of the PUF as a complex K by K' matrix M , in general non-square, which satisfies $MM^\dagger = \mathbf{1}_{K \times K}$ (‘semi-unitary’). Without loss of generality, we consider either pure reflection or pure transmission of the light. The response speckle pattern is characterized by a normalized complex vector $\mathbf{d} = (d_{\mathbf{k}})_{\mathbf{k} \in \mathcal{K}'}$, with $\mathbf{d} = M\mathbf{c}$, and $\sum_{\mathbf{k} \in \mathcal{K}'} |d_{\mathbf{k}}|^2 = 1$.

2.2 Coherent challenge state and response state

Of all possible quantum states of light, coherent light comes closest to a classical state. We denote our coherent challenge state as $|n, \mathbf{c}\rangle$, where \mathbf{c} is the challenge wavefront and n the expected number of photons.

$$|n, \mathbf{c}\rangle = \prod_{\mathbf{k} \in \mathcal{K}} \exp \left[\sqrt{n} (c_{\mathbf{k}} a_{\mathbf{k}}^\dagger - c_{\mathbf{k}}^* a_{\mathbf{k}}) \right] |0\rangle. \quad (1)$$

This state has the following properties:

- If the photon counting operator $N_{\mathbf{k}}$ is measured, the result follows the Poisson distribution with expectation value $\langle N_{\mathbf{k}} \rangle = n |c_{\mathbf{k}}|^2$, i.e. $\Pr[N_{\mathbf{k}} = b] = e^{-n |c_{\mathbf{k}}|^2} (n |c_{\mathbf{k}}|^2)^b / b!$.
- The quadratures $X_{\mathbf{k}}$ and $Y_{\mathbf{k}}$ are Gaussian-distributed, with $\langle X_{\mathbf{k}} \rangle = \sqrt{n} c_{\mathbf{k}}^{\text{re}}$, $\langle Y_{\mathbf{k}} \rangle = \sqrt{n} c_{\mathbf{k}}^{\text{im}}$, $\langle X_{\mathbf{k}}^2 \rangle - \langle X_{\mathbf{k}} \rangle^2 = 1/4$ and $\langle Y_{\mathbf{k}}^2 \rangle - \langle Y_{\mathbf{k}} \rangle^2 = 1/4$. Note that the standard deviations are typically much bigger than the averages: for a typical speckle pattern they differ by an order $\sqrt{K/n}$. This hides the $c_{\mathbf{k}}$ values in the noise.

The PUF produces a response state $|n, \mathbf{d}\rangle = |n, M\mathbf{c}\rangle$, of the same coherent form as (1), but with K' modes over which the photons are spread out instead of K .

2.3 Spatial Light Modulation for response verification

The first contribution of this paper is a description of SLM2's operation in the response verification mechanism. The SLM applies a transform on the response wavefront \mathbf{d} , finely tuned with the aim of creating a pure $\mathbf{k} = \mathbf{0}$ wavefront. The SLM must have a number of pixels at least equal to K' , in order to be able to address all the important degrees of freedom. Without loss of generality, we will model the SLM as having exactly K' pixels.

The discrete Fourier transform between the wavefront parameters $d_{\mathbf{k}}$ in the wave vector domain and amplitudes $d(\mathbf{x})$ in the spatial domain is given by

$$d(\mathbf{x}) = \sum_{\mathbf{k} \in \mathcal{K}'} e^{i\mathbf{k} \cdot \mathbf{x}} d_{\mathbf{k}} \quad ; \quad d_{\mathbf{k}} = \frac{1}{K'} \sum_{\mathbf{x}} e^{-i\mathbf{k} \cdot \mathbf{x}} d(\mathbf{x}). \quad (2)$$

We have $\frac{1}{K'} \sum_{\mathbf{x}} |d(\mathbf{x})|^2 = 1$. Each SLM pixel causes a phase rotation $d(\mathbf{x}) \mapsto d'(\mathbf{x}) = \Lambda(\mathbf{x})d(\mathbf{x})$ where $\Lambda(\mathbf{x}) = \exp i\lambda(\mathbf{x})$, with $\lambda(\mathbf{x}) \in [0, 2\pi)$ freely chosen by the verifier. In the wave vector domain the multiplication $\Lambda(\mathbf{x})d(\mathbf{x})$ becomes a convolution sum,

$$d_{\mathbf{k}} \xrightarrow{\text{SLM}} d'_{\mathbf{k}} = \sum_{\mathbf{p} \in \mathcal{K}'} \tilde{\Lambda}(\mathbf{k} - \mathbf{p}) d_{\mathbf{p}}, \quad (3)$$

where $\tilde{\Lambda}$ is the Fourier transform of Λ . The transformation (3) can be written as a matrix product,

$$\mathbf{d} \xrightarrow{\text{SLM}} \mathbf{d}' = L\mathbf{d} \quad (4)$$

where L is a $K' \times K'$ unitary matrix with the double constraint that it has Toeplitz form, namely $L_{\mathbf{k}\mathbf{p}} = \tilde{\Lambda}(\mathbf{k} - \mathbf{p})$, and that the Fourier transform of $L_{\mathbf{k}+\mathbf{p},\mathbf{k}}$ with respect to \mathbf{p} is a pure phase, namely $\Lambda(\mathbf{x})$. After passing through the SLM, the quantum state is

$$|n, LM\mathbf{c}\rangle, \quad (5)$$

with expected mode occupations $\langle N_{\mathbf{k}} \rangle / n = |(LM\mathbf{c})_{\mathbf{k}}|^2$.

Lemma 1 *Let L be the matrix representing an SLM setting. For all $\mathbf{k} \in \mathcal{K}'$ it then holds that*

$$\sum_{\mathbf{p} \in \mathcal{K}'} (LM)_{\mathbf{k}\mathbf{p}} (LM)_{\mathbf{p}\mathbf{k}}^\dagger = 1. \quad (6)$$

Proof. The left-hand side is equal to $[(LM)(LM)^\dagger]_{\mathbf{k}\mathbf{k}}$. Since L is unitary and M semi-unitary, the product LM is semi-unitary; it satisfies $(LM)(LM)^\dagger = \mathbf{1}_{K' \times K'}$. \square

The optimal concentration of light into $\mathbf{k} = \mathbf{0}$ is achieved when $\Lambda(\mathbf{x})$ orients the phases of the $d'(\mathbf{x})$ to be the same for all \mathbf{x} . Without loss of generality we consider phase zero. Thus, in the optimal case the SLM has $\lambda_{\text{opt}}(\mathbf{x}) = -\arg d(\mathbf{x})$, yielding $d'_{\text{opt}}(\mathbf{x}) = |d(\mathbf{x})|$. The amount of concentration into the $\mathbf{k} = \mathbf{0}$ mode is then given by

$$(d'_{\text{opt}})_{\mathbf{0}} = \frac{1}{K'} \sum_{\mathbf{x}} d'_{\text{opt}}(\mathbf{x}) = \frac{1}{K'} \sum_{\mathbf{x}} |d(\mathbf{x})|. \quad (7)$$

The response \mathbf{d} is a random speckle pattern. The right-hand side of (7), which has the form of a spatial average, becomes very close to the expectation value of $|d(\mathbf{x})|$ over the ensemble of speckle patterns. The light intensity in the challenge wavefront is given by $I(\mathbf{x}) = |d(\mathbf{x})|^2 / K'$. The light intensity in a speckle pattern, integrated over one speckle area, is known to obey a Gamma distribution[9] with $\langle \sqrt{I(\mathbf{x})} \rangle = \frac{1}{2} \sqrt{\pi K'}$. Substitution into (7) yields $(d'_{\text{opt}})_{\mathbf{0}}^2 = \pi/4 \approx 0.79$. Thus, with perfect optics approximately 79% of the light can be concentrated into one mode.⁴

⁴Our derivation of the result $\pi/4$ is different from the derivation in [21], but it is based on the same ingredients.

We briefly comment on the case where the challenge, the PUF and the SLM are not perfectly matched to each other. We model this situation as an imperfectly configured challenge \mathbf{c}_{imp} ,

$$\mathbf{c}_{\text{imp}} = \mathbf{c}\sqrt{1-\varepsilon} + \mathbf{e}\sqrt{\varepsilon}, \quad (8)$$

where $\varepsilon \in [0, 1]$ is a number parametrizing the imperfection, and \mathbf{e} is a random speckle pattern orthogonal to \mathbf{c} , i.e. $\sum_{\mathbf{k}} c_{\mathbf{k}}^* e_{\mathbf{k}} = 0$. Note that $\mathbf{c}_{\text{imp}} \cdot \mathbf{c} = \sqrt{1-\varepsilon}$. We find

$$\frac{\langle n, LM\mathbf{c}_{\text{imp}} | N_{\mathbf{0}} | n, LM\mathbf{c}_{\text{imp}} \rangle}{n} = |(LM\mathbf{c}_{\text{imp}})_{\mathbf{0}}|^2 = (1-\varepsilon)|(LM\mathbf{c})_{\mathbf{0}}|^2 + \mathcal{O}(\sqrt{\varepsilon/K'}). \quad (9)$$

Here we have used that $LM\mathbf{e}$ is a random speckle pattern, i.e. $(LM\mathbf{e})_{\mathbf{k}}$ is of order $1/\sqrt{K'}$. Equation (9) is consistent with the behavior $\langle N_{\mathbf{0}} \rangle / n \approx |\mathbf{c}_{\text{imp}} \cdot \mathbf{c}|^2 |(LM\mathbf{c})_{\mathbf{0}}|^2$ observed in Ref. [10].

3 Security analysis

3.1 Attacker model: the challenge estimation attack

During enrollment a sufficient number of CRPs is measured to completely characterize the PUF. For instance, after measuring a set of $\mathcal{O}(K)$ response speckle patterns *including phase information*, any new challenge can be written as a linear superposition of the enrolled challenges, and the response is the corresponding superposition of enrolled complex-valued speckle patterns.

At authentication time, the verifier chooses a challenge wavefront \mathbf{c} from the space $\{\mathbf{c} : \sum_{\mathbf{k} \in \mathcal{K}} |c_{\mathbf{k}}|^2 = 1\}$ uniformly at random. The attacker performs a measurement on the challenge state $|n, \mathbf{c}\rangle$. The measurement is represented by a Hermitian operator. (We do not consider actions that are tantamount to quantum computing. As mentioned in Ref. [24], a sufficiently fast combination of quantum teleportation [2, 14] and quantum computing can break the security of QR-PUFs. Note, however, that a quantum computer would need K -qubit registers for an attack on the optical QR-PUF under consideration.)

Based on the obtained information, the attacker computes an estimate of \mathbf{c} . We will denote this estimate as $\mathbf{f} = (f_{\mathbf{k}})_{\mathbf{k} \in \mathcal{K}}$, with $\sum_{\mathbf{k} \in \mathcal{K}} |f_{\mathbf{k}}|^2 = 1$. We consider \mathbf{f} to be an operator on the challenge Hilbert space, since it is a function of the attacker's measurement operator. The attacker prepares the state $|n, M\mathbf{f}\rangle$ and sends it to the verifier.

The above scenario is the strongest attack that can be performed without the use of a quantum computer. (A quantum computer would emulate the mapping $\mathbf{c} \mapsto M\mathbf{c}$ based on the public information M without having to perform a measurement on the challenge state.) We will assume that the attacker does not have access to an efficient quantum computer.

3.2 General formula for the mode occupation after SLM2

In Ref. [24] a general⁵ result was derived for the case of a *single quantum* and a Hilbert size of dimension K , independent of the physical system. We have not yet been able to obtain a similar generic bound for the case of n quanta in the challenge. This is left for future work.

First we derive a general formula for the number of photons that arrives in the detector given that the challenge estimation attack takes place, with an *arbitrary* measurement chosen by the attacker. Then (Section 3.3) we specialize to the case where the attacker is somewhat restricted in his choice: only quadrature measurements are allowed, but the attacker has an unrestricted number of ideal beamsplitters and ideal detectors. For this restricted attacker model we prove an upper bound on the probability of passing one round.

SLM2 transforms the fake response into $|n, LM\mathbf{f}\rangle$, where L is tuned to concentrate the front $M\mathbf{c}$ into the $\mathbf{k} = \mathbf{0}$ mode. The mismatch between \mathbf{f} and \mathbf{c} will cause a lack of focusing.

⁵The verifier is allowed to prepare any state in the Hilbert space as his challenge state, and the attacker is allowed to choose any Hermitian operator as his measurement.

Lemma 2 *Let the attacker perform the challenge estimation attack as described in Section 3.1. After SLM2 the modes have the following expected number of photons*

$$\frac{\langle n, LM\mathbf{f}|N_{\mathbf{k}}|n, LM\mathbf{f}\rangle}{n} = \sum_{\mathbf{p}, \mathbf{p}' \in \mathcal{K}} (LM)_{\mathbf{k}\mathbf{p}}(LM)_{\mathbf{p}'\mathbf{k}}^\dagger \langle n, \mathbf{c}|f_{\mathbf{p}'}^* f_{\mathbf{p}}|n, \mathbf{c}\rangle. \quad (10)$$

Proof: By the properties of the coherent state (Section 2.2), the number of photons in mode \mathbf{k} is Poisson-distributed with average $n|(LM\mathbf{f})_{\mathbf{k}}|^2 = n \sum_{\mathbf{p}, \mathbf{p}'} (LM)_{\mathbf{k}\mathbf{p}}(LM)_{\mathbf{p}'\mathbf{k}}^\dagger f_{\mathbf{p}'}^* f_{\mathbf{p}}$. Since \mathbf{f} is an operator on the challenge Hilbert space, we finally need to take the expectation value with respect to the challenge state $|n, \mathbf{c}\rangle$. \square

Lemma 2 with \mathbf{k} set to $\mathbf{0}$ gives us the fraction of light focused into the detector.

3.3 Challenge estimation by quadrature measurements

In this section we restrict the attacker's choice of measurement. He is allowed to use lossless beam splitters. Furthermore, he has an unrestricted number of ideal detectors that can measure quadratures (in X or Y direction, or at any angle), chosen at will for every mode separately. These freedoms lead to the following generic setup.

- The attacker splits $|n, \mathbf{c}\rangle$ into ℓ parts using his lossless beam splitter(s). This results in ℓ copies of the wavefront, which we will label with a Greek index, $\alpha \in \{0, \dots, \ell - 1\}$. The splitting does not have to be equal. We denote the fraction of light in copy α as r_α , where $\sum_\alpha r_\alpha = 1$. The photon number in copy α is Poisson-distributed with mean nr_α . In order to parametrize the non-uniformity of the splitting, we introduce a constant δ as follows,

$$\left(\frac{1}{\ell^2} \sum_{\beta=0}^{\ell-1} \frac{1}{r_\beta}\right)^{-1} = 1 - \delta^2. \quad (11)$$

For uniform splitting we have $r_\alpha = 1/\ell$ for all α , and $\delta = 0$.

- In copy α the attacker does a quadrature measurement at angle φ_α in every mode $\mathbf{k} \in \mathcal{K}$ separately. The corresponding operators are

$$Q_{\alpha\mathbf{k}} = X_{\alpha\mathbf{k}} \cos \varphi_\alpha + Y_{\alpha\mathbf{k}} \sin \varphi_\alpha. \quad (12)$$

Note that $Q_{\alpha\mathbf{k}}$ and $Q_{\beta\mathbf{k}}$ commute, and that $\langle Q_{\alpha\mathbf{k}} \rangle = \sqrt{nr_\alpha}(c_{\mathbf{k}}^{\text{re}} \cos \varphi_\alpha + c_{\mathbf{k}}^{\text{im}} \sin \varphi_\alpha)$, and $\langle \Delta Q_{\alpha\mathbf{k}}^2 \rangle = 1/4$, where $\Delta Q_{\alpha\mathbf{k}} := Q_{\alpha\mathbf{k}} - \langle Q_{\alpha\mathbf{k}} \rangle$.

- The attacker combines the information from all the copies using a weighted sum,

$$\Omega_{\mathbf{k}} = \frac{2}{\ell} \sum_{\alpha=0}^{\ell-1} \frac{1}{\sqrt{nr_\alpha}} e^{i\varphi_\alpha} Q_{\alpha\mathbf{k}}. \quad (13)$$

Here the expression $e^{i\varphi_\alpha} Q_{\alpha\mathbf{k}}/\sqrt{nr_\alpha}$ represents the best guess for $c_{\mathbf{k}}$ based on the information from $Q_{\alpha\mathbf{k}}$. Note that the expectation value of $\Omega_{\mathbf{k}}$ is given by $\langle \Omega_{\mathbf{k}} \rangle = c_{\mathbf{k}} + (c_{\mathbf{k}}^*/\ell) \sum_\alpha e^{2i\varphi_\alpha}$. The second term can be eliminated, if $\ell \geq 2$, by setting the angles φ_α maximally apart (see the Appendix),

$$\varphi_\alpha = \alpha\pi/\ell. \quad (14)$$

From this point on we will consider only $\ell \geq 2$ and use φ_α angles as defined by (14), yielding $\langle \Omega_{\mathbf{k}} \rangle = c_{\mathbf{k}}$.

- Even though the operators $\Omega_{\mathbf{k}}$ have expectation value $c_{\mathbf{k}}$, the attacker cannot directly use them as his estimators for the wavefront shape, since the normalization is incorrect. In fact, the $\Omega_{\mathbf{k}}$ mostly measure Gaussian noise. If we define $\Delta_{\mathbf{k}} = \Omega_{\mathbf{k}} - c_{\mathbf{k}}$, then we have

$\langle \Delta_{\mathbf{k}}^\dagger \Delta_{\mathbf{k}} \rangle = \frac{1}{n(1-\delta^2)}$; thus, the noise amplitude $\sqrt{\langle \Delta_{\mathbf{k}}^\dagger \Delta_{\mathbf{k}} \rangle}$ in each mode is of order $1/\sqrt{n}$ while the expectation value $c_{\mathbf{k}}$ is of order $1/\sqrt{K}$. The expected norm of the vector $(\Omega_{\mathbf{k}})_{\mathbf{k} \in \mathcal{K}}$ is given by

$$\left\langle \sum_{\mathbf{k} \in \mathcal{K}} \Omega_{\mathbf{k}}^\dagger \Omega_{\mathbf{k}} \right\rangle = \sum_{\mathbf{k} \in \mathcal{K}} |c_{\mathbf{k}}|^2 + \sum_{\mathbf{k} \in \mathcal{K}} \langle \Delta_{\mathbf{k}}^\dagger \Delta_{\mathbf{k}} \rangle = 1 + \frac{K}{n(1-\delta^2)}, \quad (15)$$

i.e. far away from unity. Furthermore, the norm itself is also subject to strong fluctuations because of the quadrature property $\langle \Delta Q_{\alpha \mathbf{k}}^4 \rangle = 3/16$. Hence the best way to obtain a normalized wavefront from the $\Omega_{\mathbf{k}}$ operators is to construct the estimate \mathbf{f} as

$$\mathbf{f}_{\mathbf{k}} = \frac{\Omega_{\mathbf{k}}}{\sqrt{\sum_{\mathbf{p}} \Omega_{\mathbf{p}}^\dagger \Omega_{\mathbf{p}}}}. \quad (16)$$

These operators trivially satisfy $\sum_{\mathbf{k}} \mathbf{f}_{\mathbf{k}}^\dagger \mathbf{f}_{\mathbf{k}} = \mathbf{1}$.

Remark: From (15) we see that, given ℓ and given α as defined by (14), beamsplitting into equal parts $r_\alpha = 1/\ell$ (giving $\delta = 0$) yields the lowest expected norm, i.e. a more accurate estimate of \mathbf{c} . Later on we will see that indeed equal splitting leads to the strongest attack.

Lemma 3 *Let $K > n$, $K \gg 1$. Let the attacker perform the challenge estimation attack with the quadrature measurements described above. Then*

$$\langle n, \mathbf{c} | f_{\mathbf{p}'}^\dagger f_{\mathbf{p}} | n, \mathbf{c} \rangle = \left[c_{\mathbf{p}'}^* c_{\mathbf{p}} \frac{n(1-\delta^2)}{K+n(1-\delta^2)} + \frac{\delta_{\mathbf{p}\mathbf{p}'}}{K+n(1-\delta^2)} \right] \left[1 - \mathcal{O}\left(\frac{1}{K}\right) \right]. \quad (17)$$

Proof: See the Appendix.

Theorem 1 *Let $K > n$, $K \gg 1$. Let the attacker perform the challenge estimation attack with the quadrature measurements. Then the occupation of the modes after SLM2 is given by*

$$\frac{\langle N_{\mathbf{k}} \rangle}{n} = \left[\frac{n(1-\delta^2)}{K+n(1-\delta^2)} |(LM\mathbf{c})_{\mathbf{k}}|^2 + \frac{1}{K+n(1-\delta^2)} \right] \left[1 - \mathcal{O}\left(\frac{1}{K}\right) \right]. \quad (18)$$

Proof: We substitute Lemma 3 into Lemma 2 and then apply Lemma 1. \square

Theorem 1 is the main result of this paper. Setting $\mathbf{k} = \mathbf{0}$ in (18) and dividing by the result for a correct response, $\langle N_{\mathbf{0}} \rangle_{\text{correct}}/n = |(LM\mathbf{c})_{\mathbf{0}}|^2$, we obtain

$$\frac{\langle N_{\mathbf{0}} \rangle_{\text{attack}}}{\langle N_{\mathbf{0}} \rangle_{\text{correct}}} \approx \frac{n(1-\delta^2) + |(LM\mathbf{c})_{\mathbf{0}}|^{-2}}{K+n(1-\delta^2)}. \quad (19)$$

For the attacker it is best to choose $\delta = 0$, i.e. equal splitting. The fraction then becomes

$$\frac{\langle N_{\mathbf{0}} \rangle_{\text{attack}}}{\langle N_{\mathbf{0}} \rangle_{\text{correct}}} \Big|_{\delta=0} \approx \frac{1 + |(LM\mathbf{c})_{\mathbf{0}}|^{-2}/n}{S+1}, \quad (20)$$

where $S = K/n$ is called the Quantum Security Parameter. Note that $|(LM\mathbf{c})_{\mathbf{0}}|^{-2} \geq 4/\pi$, where the equality holds for ideal optics at the verifier side.

It is interesting to note that the number ℓ has no impact on the security, as long as $\ell \geq 2$. Measuring two quadratures yields as much information as measuring many.

For properly chosen K and n (e.g. such that $S > 2$) the disparity in the detector's photon count allows for reliable distinction between an authentic PUF and the quadrature attack, especially when the authentication protocol contains multiple challenge-response rounds.

Remark: If we set $n = 1$ and $\mathbf{k} = \mathbf{0}$ in (18) then (18) represents approximately the attacker's success probability in one authentication round performed with a single-photon source. The result in the case of an ideal detector is $(1 + \pi/4)/(K+1)$. This is consistent with (namely lower than) the theoretical upper bound $2/(K+1)$ on the false acceptance probability as derived in [22] for the single-quantum scenario.⁶

⁶The journal version [24] contains an unfortunate mistake.

4 Discussion

We have investigated the security of the optical QR-PUF realization of Ref. [10] under the strongest class of *classical* attacks: challenge estimation attacks. We have derived results for the amount of light focused onto the detector, i.e. in the $\mathbf{k} = \mathbf{0}$ mode, by the SLM. Our theoretical result matches the experiments in Ref. [10]. Lemma 2 gives the general equation for the number of detected photons when the challenge estimation attack takes place, for any choice of measurement by the attacker.

We have investigated the special case where the measurement is constrained to arbitrary beam splitting and arbitrary quadrature measurements. To our knowledge quadratures are the most informative *practical* measurement that can be performed with current technology. We find that equal splitting and a uniform spreading of quadrature angles yields the strongest attack. Furthermore, we see that splitting beyond two beams does not improve the attack. Theorem 1 is the main result in this setting, specifying the photon count in each mode after SLM2. Eq. (20) shows the discrepancy between the case of an authentic PUF and the quadrature measurement attack: roughly a factor $S + 1$ in the photon count. This result is consistent with the theoretical bound $2/(K + 1)$ for the single-photon case.

We conclude that it is very easy to choose operational conditions such that the optical QR-PUF system is quantum-secure against quadrature-based challenge estimation, especially if the authentication protocol consists of multiple challenge-response rounds.

Acknowledgments

We thank Bas Goorden, Ad Lagendijk, and Willem Vos for discussions and support. A.P.M. acknowledges financial support from the European Research Council (grant number 279248).

A Proof of Lemma 3

We begin by introducing a constant γ as

$$\gamma = (1 - \delta^2) \frac{1}{\ell^2} \sum_{\alpha=0}^{\ell-1} \frac{e^{2i\varphi_\alpha}}{r_\alpha}. \quad (21)$$

If $\ell \geq 2$ and the splitting is uniform ($\forall_{\alpha \in \{0, \dots, \ell-1\}} : r_\alpha = 1/\ell$), then $\gamma = 0$. This is seen as follows. Substitution of $r_\alpha = 1/\ell$ into (21) gives $\gamma = \ell^{-1} \sum_{\alpha=0}^{\ell-1} e^{2i\varphi_\alpha} = \ell^{-1} \sum_{\alpha=0}^{\ell-1} \zeta^\alpha$, with $\zeta := e^{i2\pi/\ell}$ satisfying $\zeta^\ell = 1$. The sum is evaluated as $\sum_{\alpha=0}^{\ell-1} \zeta^\alpha = (1 - \zeta^\ell)/(1 - \zeta) = 0$.

Next we define

$$H_{\mathbf{k}} = \sqrt{n(1 - \delta^2)}(\Omega_{\mathbf{k}} - c_{\mathbf{k}}) = 2\sqrt{1 - \delta^2} \frac{1}{\ell} \sum_{\alpha} \frac{e^{i\varphi_\alpha}}{\sqrt{r_\alpha}} \Delta Q_{\alpha\mathbf{k}}, \quad (22)$$

with $\Omega_{\mathbf{k}}$ as specified in (13). Given the state $|n, c\rangle$, the $H_{\mathbf{k}}$ are Gaussian-distributed with $\langle H_{\mathbf{k}} \rangle = 0$. As can be seen from the properties of $\Omega_{\mathbf{k}}$ and $Q_{\alpha\mathbf{k}}$ as discussed in Section 3.3, it holds that $\langle H_{\mathbf{k}}^\dagger H_{\mathbf{k}} \rangle = 1$ and $\langle H_{\mathbf{k}}^2 \rangle = \gamma$. Since the modes are independent we have $\langle H_{\mathbf{k}'}^\dagger H_{\mathbf{k}} \rangle = \delta_{\mathbf{k}\mathbf{k}'}$. The expectation value of any odd power of $H_{\mathbf{k}}$ always yields zero. Furthermore, the quadratures, being Gaussian, satisfy $\langle \Delta Q_{\alpha\mathbf{k}}^4 \rangle = 3/16$. This yields, after some algebra, $\langle (H_{\mathbf{k}}^\dagger H_{\mathbf{k}})^2 \rangle = 2 + |\gamma|^2$. Next we define the following operators,

$$G = \frac{1}{\sqrt{2}} \sum_{\mathbf{p}} (c_{\mathbf{p}}^* H_{\mathbf{p}} + c_{\mathbf{p}} H_{\mathbf{p}}^\dagger) \quad ; \quad E = \frac{1}{\sqrt{K} \sqrt{1 + |\gamma|^2}} \left(\sum_{\mathbf{p}} H_{\mathbf{p}}^\dagger H_{\mathbf{p}} - K \right). \quad (23)$$

They have convenient properties, which directly follow from the properties of $H_{\mathbf{k}}$: $\langle G \rangle = 0$,

$\langle E \rangle = 0$, $\langle G^2 \rangle = 1 + \mathcal{O}(\gamma)$, $\langle E^2 \rangle = 1$. The product $f_{\mathbf{k}'}^\dagger f_{\mathbf{k}}$ can now be written as

$$f_{\mathbf{k}'}^\dagger f_{\mathbf{k}} = \frac{n(1-\delta^2)}{K+n(1-\delta^2)} \frac{c_{\mathbf{k}'}^* c_{\mathbf{k}} + c_{\mathbf{k}'}^* \frac{H_{\mathbf{k}}}{\sqrt{n(1-\delta^2)}} + c_{\mathbf{k}} \frac{H_{\mathbf{k}'}^\dagger}{\sqrt{n(1-\delta^2)}} + \frac{H_{\mathbf{k}'}^\dagger H_{\mathbf{k}}}{n(1-\delta^2)}}{1 + \frac{\sqrt{K}\sqrt{1+|\gamma|^2}}{K+n(1-\delta^2)} E + \frac{\sqrt{2}\sqrt{n(1-\delta^2)}}{K+n(1-\delta^2)} G}. \quad (24)$$

Since $\langle E^2 \rangle$ and $\langle G^2 \rangle$ are both of order 1, the terms (in the denominator) in which E and G appear are much smaller than 1; we are then allowed to Taylor-expand the fraction $\frac{1}{1+[\dots]E+[\dots]G}$. The small parameter in the expansion is of order $\mathcal{O}(\sqrt{K}/(K+n)) + \mathcal{O}(\sqrt{n}/(K+n)) = \mathcal{O}(1/\sqrt{K})$. (This holds for $K \gg 1$ and any $n < K$.) We stop the expansion after the 2nd order.

Using $\langle E \rangle = 0$, $\langle G \rangle = 0$ and $\langle EG \rangle = 0$, we find that the contribution from the $c_{\mathbf{k}'}^* c_{\mathbf{k}}$ term in the numerator in (24) is

$$\frac{n(1-\delta^2)c_{\mathbf{k}'}^* c_{\mathbf{k}}}{K+n(1-\delta^2)} \left[1 + \frac{1}{2} \langle E^2 \rangle \frac{K(1+|\gamma|^2)}{[K+n(1-\delta^2)]^2} + \frac{1}{2} \langle G^2 \rangle \frac{2n(1-\delta^2)}{[K+n(1-\delta^2)]^2} + \dots \right]. \quad (25)$$

Using $\langle H_{\mathbf{k}} E \rangle = 0$, $\langle H_{\mathbf{k}} G \rangle = (c_{\mathbf{k}} + \gamma c_{\mathbf{k}}^*)/\sqrt{2}$, we find for the contribution from the $c_{\mathbf{k}'}^*$ term

$$\frac{n(1-\delta^2)c_{\mathbf{k}'}^* c_{\mathbf{k}}}{K+n(1-\delta^2)} \left[-\frac{1 + \gamma c_{\mathbf{k}}^*/c_{\mathbf{k}}}{K+n(1-\delta^2)} + \dots \right]. \quad (26)$$

The $c_{\mathbf{k}}$ term yields an expression analogous to (26), but with $\gamma c_{\mathbf{k}'}^*/c_{\mathbf{k}'}^*$ instead of $\gamma c_{\mathbf{k}}^*/c_{\mathbf{k}}$. For the $H_{\mathbf{k}'}^\dagger H_{\mathbf{k}}$ term we use $\langle H_{\mathbf{k}'}^\dagger H_{\mathbf{k}} G \rangle = 0$ (odd power of H), $\langle H_{\mathbf{k}'}^\dagger H_{\mathbf{k}} E \rangle = \delta_{\mathbf{k}\mathbf{k}'} \mathcal{O}(1/\sqrt{K})$ and obtain

$$\frac{\delta_{\mathbf{k}\mathbf{k}'}}{K+n(1-\delta^2)} \left[1 - \mathcal{O}\left(\frac{1}{K}\right) \right]. \quad (27)$$

Inspecting (25) and (26), we find that the total relative correction term to the expression $n(1-\delta^2)c_{\mathbf{k}'}^* c_{\mathbf{k}}/(K+n(1-\delta^2))$ is approximately $\frac{1}{2}(K+2n)/(K+n)^2 - 2/(K+n)$, which is negative. The result (17) follows.

References

- [1] C.H. Bennett and G. Brassard. Quantum cryptography: Public key distribution and coin tossing. *IEEE International Conference on Computers, Systems and Signal Processing*, pages 175–179, 1984.
- [2] C.H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W.K. Wootters. Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Phys. Rev. Lett.*, 70:1895–1899, 1993.
- [3] J.D.R. Buchanan, R.P. Cowburn, A. Jausovec, D. Petit, P. Seem, G. Xiong, D. Atkinson, K. Fenton, D.A. Allwood, and M.T. Bryan. Forgery: ‘fingerprinting’ documents and packaging. *Nature, Brief Communications*, 436:475, 2005.
- [4] J.F. de Boer, M.C.W. van Rossum, M.P. van Albada, Th.M. Nieuwenhuizen, and A. Lagendijk. Probability distribution of multiple scattered light measured in the total transmission. *Phys. Rev. Lett.*, 73:2567, 1994.
- [5] G. DeJean and D. Kirovski. Radio frequency certificates of authenticity. In *IEEE Antenna and Propagation Symposium – URSI*, 2006.
- [6] D. Dieks. *Phys. Lett. A*, 92:271, 1982.
- [7] S. Popoff et al. Measuring the transmission matrix in optics: An approach to the study and control of light propagation in disordered media. *Phys. Rev. Lett.*, 104:100601, 2010.

- [8] B. Gassend, D.E. Clarke, M. van Dijk, and S. Devadas. Silicon physical unknown functions. In *ACM Conference on Computer and Communications Security (CCS) 2002*, pages 148–160. ACM, 2002.
- [9] J.W. Goodman. *Laser Speckle and Related Phenomena*, chapter Statistical properties of laser speckle patterns. Springer-Verlag, New York, 2nd edition, 1984.
- [10] S.A. Goorden, M. Horstmann, A.P. Mosk, B. Škorić, and P.W.H. Pinkse. Quantum-Secure Authentication with a Classical Key. <http://arxiv.org/abs/1303.0142>, 2013.
- [11] 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*, volume 4727 of *LNCS*, pages 63–80. Springer, 2007.
- [12] T. Ignatenko, G.-J. Schrijen, B. Škorić, P. Tuyls, and F.M.J. Willems. Estimating the Secrecy Rate of Physical Uncloneable Functions with the Context-Tree Weighting Method. In *Proc. IEEE International Symposium on Information Theory (ISIT) 2006*, pages 499–503, 2006.
- [13] M. Majzoobi, A. Elnably, and F. Koushanfar. FPGA time-bounded unclonable authentication. In *Information Hiding 2010*, volume 6387 of *LNCS*, pages 1–16. Springer, 2010.
- [14] D.N. Matsukevich and A. Kuzmich. Quantum state transfer between matter and light. *Science*, 306(5696):663–666, 2004.
- [15] A.P. Mosk, A. Lagendijk, G. Lerosey, and M. Fink. Controlling waves in space and time for imaging and focusing in complex media. *Nat. Photon.*, 6:283, 2012.
- [16] R. Pappu, B. Recht, J. Taylor, and N. Gershenfeld. Physical One-Way Functions. *Science*, 297:2026–2030, 2002.
- [17] A.-R. Sadeghi and D. Naccache (Eds.). *Towards Hardware-Intrinsic Security*. Springer, 2010.
- [18] P. Tuyls, G.J. Schrijen, B. Škorić, J. van Geloven, R. Verhaegh, and R. Wolters. Read-proof hardware from protective coatings. In *Cryptographic Hardware and Embedded Systems (CHES) 2006*, volume 4249 of *LNCS*, pages 369–383. Springer-Verlag, 2006.
- [19] P. Tuyls, B. Škorić, and T. Kevenaar (Eds.). *Security with Noisy Data: Private Biometrics, Secure Key Storage and Anti-Counterfeiting*. Springer, London, 2007.
- [20] P. Tuyls, B. Škorić, S. Stallinga, A.H.M. Akkermans, and W. Opehy. Information-theoretic security analysis of physical uncloneable functions. In *9th Conf. on Financial Cryptography and Data Security*, volume 3570 of *LNCS*, pages 141–155. Springer, 2005.
- [21] I.M. Vellekoop and A.P. Mosk. Focusing coherent light through opaque strongly scattering media. *Opt. Lett.*, 32:2309, 2007.
- [22] B. Škorić. Quantum Readout of Physical Unclonable Functions. <http://eprint.iacr.org/2009/369>.
- [23] B. Škorić. On the entropy of keys derived from laser speckle; statistical properties of Gabor-transformed speckle. *Journal of Optics A: Pure and Applied Optics*, 10(5):055304–055316, 2008.
- [24] B. Škorić. Quantum Readout of Physical Unclonable Functions. *International Journal of Quantum Information*, 10(1):1250001–1 – 125001–31, 2012.
- [25] B. Škorić, T. Bel, A.H.M. Blom, B.R. de Jong, H. Kretschman, and A.J.M. Nellissen. Randomized resonators as uniquely identifiable anti-counterfeiting tags. *Secure Component and System Identification Workshop*, Berlin, March 2008.
- [26] W.K. Wootters and W.H. Zurek. A single quantum cannot be cloned. *Nature*, 299:802–803, 1982.