Quantum readout of Physical Unclonable Functions:

Remote authentication without trusted readers and authenticated Quantum Key Exchange without initial shared secrets

B. Škorić

Abstract

Physical Unclonable Functions (PUFs) are physical structures that are hard to clone and have a unique challenge-response behaviour. The term PUF was coined by Pappu et al. in 2001. That work triggered a lot of interest, and since then a substantial number of papers has been written about the use of a wide variety of physical structures for different security purposes such as identification, authentication, read-proof key storage, key distribution, tamper evidence, anti-counterfeiting, software-to-hardware binding and trusted computing.

In this paper we propose a new security primitive: the **quantum-readout PUF** (QR-PUF). This is a classical PUF which is challenged using a quantum state, e.g. a single-photon state, and whose response is also a quantum state. By the no-cloning property of unknown quantum states, attackers cannot intercept challenges or responses without noticeably disturbing the readout process. Thus, a verifier who sends quantum states as challenges and receives the correct quantum states back can be certain that he is probing a specific QR-PUF without disturbances, even in the QR-PUF is far away 'in the field' and under hostile control. For PUFs whose information content is not exceedingly large, all currently known PUF-based authentication and anti-counterfeiting schemes require trusted readout devices in the field. Our quantum readout scheme has no such requirement.

Furthermore, we show how the QR-PUF authentication scheme can be interwoven with Quantum Key Exchange (QKE), leading to an authenticated QKE protocol between two parties. This protocol has the special property that it requires no a priori secret, or entangled state, shared by the two parties.

1 Introduction

1.1 Physical Unclonable Functions

The term *Physical Unclonable Function* (PUF) was coined by Pappu et al. in [18, 19]. It refers to a physical object that is hard to clone and that can be subjected to challenges, yielding different random-looking outputs. A cryptographic equivalent of such challenge-response behaviour would be a keyed hash function, where the precise structure of the object represents the key. Pappu et al. also used the term *Physical One-Way Function* (POWF). Although the use of hard-to-clone (classical) physical structures for authentication purposes dates back a long time, the work [18, 19] was the first to introduce the 'function' behaviour of such objects and to consider mathematical unclonability as well (difficulty of modelling). It was shown that an optical medium with a high density of scatterers makes an extremely good PUF: A challenge consists e.g. of the angle of incidence of a laser beam; the response is the speckle pattern resulting from multiple coherent scattering. The speckle pattern has high entropy and strongly depends on the precise locations of the scatterers, which makes the object hard to clone.

The results of Pappu et al. have sparked a lot of interest in the use of (classical) physics for different security purposes such as identification, authentication, read-proof key storage, key distribution, tamper evidence, anti-counterfeiting, software-to-hardware binding and trusted computing. By now there is a whole zoo of PUF-like systems that have appeared in the literature: optical PUFs,

delays in integrated circuits [9], dielectric properties of security coatings [21], two-dimensional fiber-optic configurations [14], radio-frequent probing of wire configurations [5] and thin-film resonators [25], laser probing of fibers in paper [3], startup values of SRAM cells [11], butterfly PUFs [15], phosphor patterns [4], phase-change memory states [16]. What all these have in common is a strong dependence of the measurement results on uncontrollable aspects of the manufacturing process. An overview of the field is given in [22].
Because of the wide variety of different physical systems and security goals involved, the terminology used in the literature can be confusing. There exist multiple definitions of a PUF, differing in their list of properties that must be satisfied. Often mentioned properties are physical unclonability, mathematical unclonability, uniqueness, tamper evidence, high response entropy, large number of different challenges and read-proofness. Descriptions like Physical One-Way Function, Physical Unknown Function, Physically Obfuscated/Obscured Key, and Physical Pseudorandom Function are sometimes used to specify which properties are most important for a certain application.
We mention explicitly that in this paper we rely only on the following properties of the pair {physical object, measurement method}:

- Physical unclonability. It is technically/financially infeasible to make a physical clone of a given QR-PUF (given full knowledge of this QR-PUF), such that it behaves exactly as the original one for the given measurement method.
- Quantum-computational unclonability. This is a new physical assumption. It is technically/financially infeasible to build a quantum computer and input/output handling which emulates a given QR-PUF (given full knowledge of this QR-PUF) with a sufficiently small delay time.
- Uniqueness. Different challenges can be applied to the QR-PUF. The number of different challenges does not have to be large. Together, the responses to these challenges (measurements) have to contain enough entropy so that all the to-be-authenticated QR-PUFs out in the field can be distinguished from each other, e.g. at least 20 bits of entropy to distinguish between a million QR-PUFs.

The entropy of optical PUFs was studied in [23, 24, 22]. We stress that the responses to all challenges, for each manufactured QR-PUF, are allowed to be public knowledge. In this paper, there is no secrecy concerning any aspect of the QR-PUFs.
Note that the PUF literature is concerned only with *classical* physics. The word 'unclonability' is an assumption about the *effort* it takes to produce a clone; it does not indicate that it is fundamentally impossible to produce one. There is no relation to the *provable* no-cloning theorem [26, 6] of unknown quantum states. Our QR-PUF is classical in itself, but the measurement is quantum, and we will make use of the no-cloning theorem to detect tampering with the measurement process.

1.2 Getting rid of the trusted remote reader

Some of the early PUF-based remote authentication protocols rely on the vastness of the PUF's entropy. At enrolment, Alice measures PUF responses for a large number of random challenges. She stores the table of challenge-response pairs (CRPs). Then the PUF is given to Bob. When Alice wants to authenticate Bob, she sends him a challenge, randomly selected from her CRP table. This is done over a public channel. Bob feeds the challenge to his PUF and measures the response. He sends the response to Alice over a public channel. If Bob's response is correct, Alice is convinced that he has access to the PUF. Alice deletes the used CRP from her list.
The security of such a protocol (assuming that Alice and Bob are honest) is based on the following physical assumptions: (I) Knowledge of eavesdropped CRPs gives negligible information about the response to a new challenge different from the eavesdropped ones; (II) It is infeasible for an attacker to characterize Bob's PUF in a short amount of time (e.g. the time needed to verify a credit card) with enough accuracy to predict the response to a new random challenge.
Although optical PUFs come a long way [23, 24] towards satisfying these assumptions, it is not clear whether high-entropy PUFs will be feasible in practice.

Even if one gives up the high-entropy property, PUFs are very useful. For instance[1], an effective authentication/anti-counterfeiting method can be designed that is based solely on the physical unclonability and uniqueness properties. Consider the following infrastructure. When a PUF gets manufactured, a PUF certification authority (PCA) enrolls it. This can be done in either of two ways. (i) A PUF identifier I and a precise characterization of the PUF (e.g. a CRP table) are entered into a publicly readable, PCA-maintained tamper proof database, or (ii) the PCA signs a digital certificate containing I and the PUF characterization, which certificate is then stored publicly (possibly attached to the PUF itself). Both options provide for a way to reliably obtain the enrolled data about a PUF. When a verifier wants to see if a certain PUF is authentic, he can check its challenge-response behaviour against the enrolled data. As long as the verifier knows for certain that responses are coming directly from a real PUF (as opposed to a mathematical emulation or a replay of data traffic), he is able to verify that a physical structure is the same as the enrolled one.
In the above scheme the verifier must be in control of the measurement device in order to prevent spoofing. This gives rise to an extra requirement for remote verification: the verifier has to trust a remotely located piece of measurement and processing hardware. One of the main results of this paper is that we achieve remote PUF authentication *without the trusted remote reader*.

1.3 Authenticated Quantum Key Exchange

Quantum Key Exchange (QKE), also known as Quantum Key Distribution (QKD), Quantum Key Agreement, and Quantum Cryptography, was first proposed in 1984 [1] (BB84). QKE is a protocol that allows Alice and Bob to establish an unconditionally secure shared secret if they have a channel at their disposal over which they can send quantum states. The security is based on the laws of quantum physics, in particular the no cloning theorem [26, 6], which ensures that eavesdropping and other attacks are detected. In its most easy to understand realization the protocol makes use of single-photon polarization states. Polarization can be measured in any direction perpendicular to light's direction of motion, i.e. there is a continuum of observables. However, the result of a measurement is a binary 'yes/no' answer.
The BB84 protocol can be summarized as follows in a nutshell. Alice generates two random bits. The first bit determines a basis: vertical/horizontal ('+') vs. diagonal ('×'). The second bit determines a direction within that basis, e.g. 0/1 for horizontal/vertical in the + basis. Alice sends a photon to Bob prepared in the state as described by the two bits. Bob randomly selects the + or × basis to measure the polarization. After a number of repetition of these steps Bob tells Alice which bases he has used. Alice discards all events where her basis and Bob's did not coincide. In the remaining cases Alice and Bob should have a shared secret bit. They publicly verify a random subset of their secret bitstring (which subset is then discarded) to see if there has been an attack. Next they perform information reconciliation (error correction to remove leftover noise) on their remaining bits and privacy amplification. The security of the protocol is derived from the fact that an eavesdropper, not knowing Alice's basis, only has a 50% probability of correctly guessing her basis and hence being able to correctly clone the photon. (If eavesdropping occurs, the probability that Alice and Bob get the same bit when they choose the same basis is only 3/4.)
A huge number of papers and books on QKE has been written since 1984. Progress has been made on all relevant aspects: single-photon sources, detectors, fiber optic cables, attacks and defense, use of entanglement [8], error-correction codes, privacy amplification and security proof methods. QKE products based on BB84 are now commercially available. Quantum observables other than polarization have been proposed for QKE, e.g. phase [13] and squeezed coherent states [10].
The classical communication channel between Alice and Bob is allowed to be public, but it has to be authenticated in order to prevent man-in-the-middle attacks. Public key crypto can provide authentication. However, the aim of QKE is to achieve *unconditional* (information-theoretic) security, i.e. not relying on the kind of computational assumptions that public key crypto needs.

[1] Another example is secure key storage, which relies only on read-proofness.

One way to achieve information-theoretic authentication is to let Alice and Bob share a short initial MAC key. Though this may look like cheating, it is not. QKE serves to indefinitely lengthen an initial shared secret. Another way to achieve authentication is to let Alice and Bob each possess one part of an entangled state [20]. This approach has a number of practical drawbacks, such as the requirement that the entangled state has to be stored for a long time.

1.4 Our contributions

In this paper we introduce the concept of a Quantum Readout PUF (QR-PUF): a classical PUF that is challenged using a quantum state, and whose response is also a quantum state. (The only practical realization we can suggest at the moment is an Optical PUF. Other options are not excluded, however.) In analogy with QKE, the no-cloning theorem ensures that eavesdropping on challenges or responses cannot go unnoticed. We rely on three physical assumptions about the QR-PUF: physical unclonability, quantum-computational unclonability and uniqueness.
We present a protocol for authenticating a QR-PUF remotely without reliance on a trusted remote reader device. The protocol can be based on the reflection properties of the QR-PUF, or on both reflection and transmission. The reflection part also serves as a distance bounding protocol. We give a security analysis, assuming that all QR-PUF properties are public, i.e. the only secrets in the protocol are the challenge states sent by the verifier. Intercept-resend attacks do not work. However, a QR-PUF can be spoofed by an attacker who has a sufficiently powerful quantum computer, combined with sufficiently fast ways of measuring qubit states as well as transferring challenge/response states into qubits and back. One of our assumptions says that such a powerful combination of quantum physical techniques is infeasible.
Then we show how the QR-PUF authentication can be intertwined with QKE. The idea is based on two main observations:

- The reflected states are used to authenticate the QR-PUF.
- The QR-PUF is completely 'transparent' with respect to transmitted states, in the sense that Alice can choose which quantum state reaches Bob after transmission through the QR-PUF, even if the transmission process is complicated.

We sketch an authenticated QKE protocol for n-dimensional challenge states. (BB84 readily follows as a special case for $n = 2$.) The security of the generated key and of the authentication are independent. The combination of QR-PUF authentication and QKE achieves three interesting security properties:

1. The initial authentication between parties is achieved *without any shared secret* such as a MAC key or entangled state.
2. The security of the initial authentication is based on *physical assumptions* about the QR-PUF, combined with trust in the enrolled data (e.g. Alice enrolled the QR-PUF herself, or obtains the data from a trusted database). Thus, we do not have unconditional security. However, we find it worth showing that physical assumptions can be used in this way.
3. Even if a MAC key is used for authentication in all QKE protocol runs other than the first run, the simultaneous QR-PUF verification allows a party to check not only if QKE is run with the same entity as the previous time, but also if that entity still has real-time access to the QR-PUF.

In this paper we only outline the theoretical concepts. Implementations issues are left for future work. The outline of this paper is as follows. In Section 2 we briefly review quantum physics notation and explain which physical assumptions we make about the challenges and responses. In Section 3 we present the QR-PUF authentication protocol, including a security analysis. Finally, we combine QR-PUF authentication with QKE in Section 4.

2 Preliminaries

2.1 Quantum physics notation

Quantum states are represented as vectors in a Hilbert space. We adopt the usual '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 only consider states satisfying $\langle\psi|\psi\rangle = 1$, so-called normalized states.
Real-valued observables are represented by Hermitian operators acting on the Hilbert space. The j'th eigenvalue of an observable X is denoted as x_j, and the corresponding eigenvector as $|x_j\rangle$. We have $X|x_j\rangle = x_j|x_j\rangle$. The scalars $\langle e_i|X|e_j\rangle$, for some basis e, are called the matrix elements of X. The eigenvectors of any Hermitian operator X form an orthonormal basis of the Hilbert space, i.e. $\langle x_a|x_b\rangle = \delta_{ab}$. The completeness of this basis (in a finite Hilbert space of dimension n) is expressed as

$$\sum_{j=1}^{n} |x_j\rangle\langle x_j| = \mathbf{1}. \tag{1}$$

We will often write $[n]$ for the set $\{1, 2, \ldots, n\}$. The operator X can be written as

$$X = \sum_{j=1}^{n} x_j |x_j\rangle\langle x_j|. \tag{2}$$

Any state ψ can be expressed in terms of an orthonormal basis,

$$|\psi\rangle = \sum_{j=1}^{n} c_j |x_j\rangle \quad ; \quad c_j = \langle x_j|\psi\rangle. \tag{3}$$

Note that $c_j \in \mathbb{C}$ and $\sum_{j=1}^{n} |c_j|^2 = 1$. Measurement of X collapses the state onto one of the eigenvectors (or eigenspaces) of X, and yields the corresponding eigenvalue as the measurement result. When a measurement of X is performed on a state $|\psi\rangle$, the probability that $|\psi\rangle$ collapses to the eigenvector $|x_j\rangle$ is given by $|\langle x_j|\psi\rangle|^2$. For more background we refer to standard textbooks on quantum mechanics.

2.2 Challenge and response quantum states

We will be working with a physical system that has 'external' degrees of freedom (such as direction of motion) as well as an 'internal' degree of freedom (e.g. spin or polarization). This is formally denoted as a tensor product of Hilbert spaces: $\mathcal{H} = \mathcal{H}_{\text{ext}} \otimes \mathcal{H}_{\text{int}}$, where $\mathcal{H}$ is the full Hilbert space. We will be dealing with three types of state:

- **Challenge.** A quantum state is moving from Alice to Bob's PUF.
- **Reflected.** The particle has interacted with Bob's PUF, and its internal state has changed. The system moves back to Alice.
- **Transmitted.** The particle has interacted with Bob's PUF, and its internal state has changed. The particle does not return to Alice, but moves on.

These three situations are represented in only *two* states of motion: moving away from Alice or towards her. Any state $|\psi\rangle \in \mathcal{H}$ can be decomposed as

$$|\psi\rangle = |\text{outgoing}\rangle \otimes |\psi_1\rangle + |\text{incoming}\rangle \otimes |\psi_2\rangle \tag{4}$$

where $|\text{outgoing}\rangle, |\text{incoming}\rangle \in \mathcal{H}_{\text{ext}}$ and $|\psi_1\rangle, |\psi_2\rangle \in \mathcal{H}_{\text{int}}$. We have $\langle\text{outgoing}|\text{incoming}\rangle = 0$. The following notation can also be used,

$$|\psi\rangle = \begin{pmatrix} |\psi_1\rangle \\ |\psi_2\rangle \end{pmatrix}. \tag{5}$$

We assume that the internal Hilbert space $\mathcal{H}_{\text{int}}$ is finite-dimensional. The number of dimensions is n.
The interaction between the quantum system and the PUF is assumed to be completely coherent: time evolution is determined by the Schrödinger equation, without any state collapse. We abstractly represent the interaction with the PUF as a unitary time evolution operator S, also known in physics as the scattering matrix or S-matrix. This operator maps 'before' states to 'after' states. Let $|\psi'\rangle \in \mathcal{H}$ be the state after the interaction, then in the notation of (5) we have

$$\begin{pmatrix} |\psi_1'\rangle \\ |\psi_2'\rangle \end{pmatrix} = S \begin{pmatrix} |\psi_1\rangle \\ |\psi_2\rangle \end{pmatrix} \quad ; \quad S = \begin{pmatrix} T & -R^\dagger \\ R & T^\dagger \end{pmatrix}. \tag{6}$$

Here the operator T (the 'transmission matrix') contains all the details of how the internal state has changed when the particle emerges at the other side of the PUF, and the operator R ('reflection matrix') does the same for reflected states. The unitary nature of the evolution ($S^\dagger S = SS^\dagger = \mathbf{1}$) implies that $T^\dagger T + R^\dagger R = \mathbf{1}$ and that T, $T^\dagger$, R and $R^\dagger$ all commute with each other[2].
Together R and T completely determine the challenge-response behaviour of the PUF. Hence, in this formalism, physical unclonability of a PUF means that it is difficult to create a PUF which behaves precisely according to some pre-specified R and T.
In most of the rest of the paper we will completely omit any reference to the external degree of freedom, for notational simplicity. Only the internal state will be written. It will always be understood that Challenge and Transmitted states are moving away from Alice, and that Reflected states are moving towards her.

2.3 Assumptions about state preparation and observables

We assume that there is a lot of freedom in doing measurements and in preparing quantum states. More precisely, we assume that Alice can prepare basically any internal state $|\psi\rangle \in \mathcal{H}_{\text{int}}$. Similarly, there is a large number of different observables that Alice and Bob can measure. (In the case of photon polarization there is even a continuum of observables, namely the polarization component in any direction.)
This freedom allows one to define a set of measurements that accurately characterize a PUF. Let X and Y be two different observables, whose eigenvectors do not coincide. Let us write $R_{ab} := \langle x_a|R|x_b\rangle = |R_{ab}|e^{i\rho_{ab}}$ for the (complex) matrix elements of R. In practice the characterization of R could consist of the following steps:

- By repeatedly applying the same challenge $|x_i\rangle$ and measuring X, an accurate estimate is obtained for the absolute values $\{|R_{ji}|^2\}_{j=1}^n$. These numbers are the probabilities of measuring $|x_j\rangle$.
- Then part of the phase information of the matrix elements can be determined by applying 'mixed' challenges, i.e. challenges that are not an eigenstate of X. For instance, repeated application of the mixed challenge $\cos\alpha|x_i\rangle + e^{-i\varphi}\sin\alpha|x_k\rangle$ and measurement of X allows one to obtain an accurate estimate for the probabilities $p_j := |R_{ji}\cos\alpha + R_{jk}e^{-i\varphi}\sin\alpha|^2$ for $j \in [n]$. Since the absolute values $|R_{ab}|$ are known, one can compute

$$\frac{p_j - |R_{ji}|^2\cos^2\alpha - |R_{jk}|^2\sin^2\alpha}{|R_{ji}|\cdot|R_{jk}|\sin\alpha\cos\alpha} = 2\cos(\varphi + \rho_{ji} - \rho_{jk}).$$

  Doing this for two (or more) different values of φ yields the phase difference $\rho_{ji} - \rho_{jk}$. Repetition of this procedure for different i, j, k yields all the phase differences *within each row of* R.

[2]We can write $T = U^{-1}\Lambda U$ and $R = U^{-1}\Gamma U$, where U is a unitary matrix and Λ and Γ are complex-valued diagonal matrices satisfying $|\Lambda_i|^2 + |\Gamma_i|^2 = 1$ for all $i \in [n]$. The basis vectors contained in U can be thought of as eigenmodes of the S-matrix, i.e. modes which are transmitted or reflected without being changed by the QR-PUF.

- The phase differences within the *columns* of R are obtained in a way analogous to the previous step. The challenges are of the form $|x_i\rangle$, but the measured observable is Y.

The above procedure, or any variant of it, gives all the complex values R_{ji} up to a global phase factor. This global phase can be chosen e.g. such that $R_{11} \in \mathbb{R}$.
Of course, the accuracy of the characterization depends on the number of repetitions N of each experiment. The relative error in the matrix elements is proportional to $1/\sqrt{N}$. Determination of T is completely analogous.
Finally we assume that a measurement of the internal state is able to discern whether or not there is a particle present at all. This statement is not trivial. Filter-based polarization measurements for instance do not see the difference between darkness and polarization perpendicular to the applied filter. Measurements using a polarization splitter and measurement of differential phase shift [13], on the other hand, do distinguish between presence and absence of a photon.

3 Remote authentication of a QR-PUF

Here we present a remote QR-PUF authentication scheme. The protocol makes use of both R and T, but it is also possible to use only R.

3.1 Attack model

We make the following assumptions about the capabilities of the attackers. Alice wants to verify if Bob has *real-time access* to a certain QR-PUF. The main attack to protect against is from Bob himself: he may be trying to convince Alice that he has access to the QR-PUF even though in reality he does not. Either he has never had access to the QR-PUF at all, or he has had access at some point in the past.
Alice and Bob can communicate over a classical channel and send quantum states over the quantum channel as described in Section 2. Bob has physical access to challenge states as well as reflected and transmitted states. He can destroy quantum states, perform measurements on them, perform unitary operations, and insert new states. However, he cannot clone a state. We assume that R and T are fully known to Bob. (Either because it is public information, or because he has had access to the QR-PUF long enough.) However, as discussed in Section 1.1, it is assumed infeasible to create a new QR-PUF whose challenge-reflection matrix is R and whose challenge-transmission matrix is T. It is also infeasible to build a quantum computer that emulates the QR-PUF with a very small time delay.
There is a source of enrollment data trusted by Alice (e.g. she enrolls QR-PUFs herself).

3.2 The QR-PUF authentication protocol

3.2.1 Enrollment phase

In the enrollment phase a QR-PUF with identifier I is accurately characterized. It is characterized by its responses to combinations of a challenge and two observables (one acting on Reflected states and one on Transmitted). We define a list $\mathcal{M}$ of such combinations,

$$\mathcal{M} = \left\{ |\psi_u\rangle, X_u, Y_u \right\}_{u=1}^{E} \tag{7}$$

Here E denotes the number of different experiments. Each such experiment is repeated N times. The list is constructed such that for given operator Y all the different challenges ψ and all the different operators X appear equally often in the list.
The measurement results can be summarized as two arrays of empirical probabilities, $\mathcal{P}^{\mathrm{R}} = \{P^{\mathrm{R}}_{uj}\}_{u\in[E],j\in[n]}$ and $\mathcal{P}^{\mathrm{T}} = \{P^{\mathrm{T}}_{uj}\}_{u\in[E],j\in[n]}$, with

$$\begin{aligned} P^{\mathrm{R}}_{uj} &= N^{-1} \cdot \#\{\text{Experiment } \mathcal{M}_u \text{ gave reflection, and eigenstate number } j \text{ of } X_u\} \\ P^{\mathrm{T}}_{uj} &= N^{-1} \cdot \#\{\text{Experiment } \mathcal{M}_u \text{ gave transmission, and eigenstate number } j \text{ of } Y_u\}. \end{aligned}$$

The triplet $(\mathcal{M}, \mathcal{P}^{\mathrm{R}}, \mathcal{P}^{\mathrm{T}})$ characterizes the QR-PUF and is assumed to be publicly known. Note that $\mathcal{M}$ may depend on I. The QR-PUF is given to Bob.

3.2.2 Authentication phase

Bob claims that he has access to the QR-PUF with identifier I. Alice fetches the enrollment data $(\mathcal{M}, \mathcal{P}^{\mathrm{R}}, \mathcal{P}^{\mathrm{T}})$ corresponding to I. She characterizes Bob's QR-PUF in the following way.

1. Alice initializes counters $\{n_a\}_{a=1}^{E}$, $\{r_{aj}\}_{a\in[E],j\in[n]}$ and $\{t_{aj}\}_{a\in[E],j\in[n]}$ to zero.

2. Alice generates a random number $u \in \{1, \ldots, E\}$. She increments the counter n_u by 1. She looks up ψ_u, X_u and Y_u in $\mathcal{M}$. She sends Y_u to Bob over the classical channel. She prepares challenge $|\psi_u\rangle$ and sends the quantum state to Bob.

3. Alice performs a measurement of the observable X_u on the reflected quantum state. The result of the measurement is either an eigenvalue of X_u or '$\perp$' (denoting the absence of a returning state). If a state returns too late[3], the protocol is aborted. If she observes the j'th eigenvalue of X_u, she increases the counter r_{uj} by 1.

4. Bob performs a measurement of the observable Y_u on the transmitted quantum state. The result is either an eigenvalue of Y_u or $\perp$. Bob sends the result (the ordinal number $j \in [n]$ of the eigenvalue, or '$\perp$') to Alice over the classical channel. If Alice receives the number j, she increases the counter t_{uj} by 1.

5. Alice and Bob repeat steps 2 to 4 a number of times equal to $E \cdot N'$. If the number of $\perp$ events is significantly larger than expected, they abort the protocol before all the repetitions are finished.

6. Alice computes her own empirical probabilities as

$$\tilde{P}^{\mathrm{R}}_{uj} = r_{uj} / \textstyle\sum_{j'}(r_{uj'} + t_{uj'}) \qquad ; \qquad \tilde{P}^{\mathrm{T}}_{uj} = t_{uj} / \textstyle\sum_{j'}(r_{uj'} + t_{uj'}). \tag{8}$$

   She verifies if the distance between $\mathcal{P}^{\mathrm{R}}$ and $\tilde{\mathcal{P}}^{\mathrm{R}}$ and between $\mathcal{P}^{\mathrm{T}}$ and $\tilde{\mathcal{P}}^{\mathrm{T}}$ are both sufficiently small, according to some distance metric. She also checks if the ratio $n_u^{-1} \sum_j (r_{uj}+t_{uj})$ (in the case of photons this is 1 minus the relative photon loss) is consistent with the experimental conditions (e.g. length of the optical fiber). If everything checks out, she is convinced that Bob has real-time access to the QR-PUF with identifier I.

3.3 Security of the QR-PUF authentication

We consider the case that attacker Bob does not have access to the QR-PUF, but does know R, T and $(\mathcal{M}, \mathcal{P}^{\mathrm{R}}, \mathcal{P}^{\mathrm{T}})$. Without giving rigorous proofs, we argue that intercept-recend attacks are thwarted by the no-cloning theorem. On the other hand, our authentication protocol is vulnerable to attacks that combine quantum teleportation and quantum computing. It is difficult to specify precisely how powerful a quantum computer has to be for a successful attack.

3.3.1 Intercept-resend atacks

We will use the following notation. Let $\mathcal{C}$ denote the set of challenges that appear in $\mathcal{M}$. Let t_ψ, for $\psi \in \mathcal{C}$, be the transmission probability for challenge ψ, i.e. $t_\psi = \sum_{a=1}^{n} |\langle e_a|\psi\rangle|^2 = \langle\psi|T^\dagger T|\psi\rangle$. The t_ψ are public knowledge. Let $Z \to z_a$ denote the event that measurement of observable Z yields eigenvalue z_a.
Bob chooses an observable Z and performs a measurement of Z on the incoming state $|\psi\rangle$, getting result z_a. (The probability of this event is $|\langle z_a|\psi\rangle|^2$). He uses the outcome z_a to guess what ψ was

[3]This is a form of distance bounding. If Alice has a rough idea where Bob should be located (e.g. which continent), she knows what round-trip time to expect.

before he destroyed it. Since $\mathcal{M}$ is constructed in such a way that Bob's knowledge of Y gives him no information about ψ, he has no other clues. He computes the conditional probabilities $g_{\chi|Za}$ that Alice sent challenge χ, given his measurement result $Z \to z_a$,

$$g_{\chi|Za} = \frac{\text{Prob}[\psi = \chi, Z \to z_a]}{\text{Prob}[Z \to z_a]} = \frac{|\langle z_a|\chi\rangle|^2}{\sum_{\chi' \in \mathcal{C}} |\langle z_a|\chi'\rangle|^2}. \tag{9}$$

The numbers $\{g_{\chi|Za}\}_{\chi\in\mathcal{C}}$ represent all his knowledge about ψ. Due to his incomplete knowledge, he is unable to correctly reconstruct ψ with high enough probability and therefore also unable to prepare a correct response state $R|\psi\rangle$ with high enough probability; furthermore he will not be able to give Alice the correct statistics $|\langle y|T|\psi\rangle|^2$. These statements are substantiated in Appendix A.

3.3.2 Attack by quantum computer

It is possible to break the QR-PUF authentication scheme if one has [i] a sufficiently powerful quantum computer (QC); [ii] a sufficiently fast way of transferring challenge states into the QC's qubits and transferring the computation result back to a response state; [iii] a sufficiently fast way of measuring the state of the QC.

We give a high-level description of the attack. It has three steps. First the challenge state $|\psi_u\rangle$ is transferred to the working memory of the QC (qubits or higher order digits). This can be done without measuring ψ_u, namely by quantum teleportation [2], in particular a form of teleportation that transfers states from one type of physical system to another [7, 17]. There are two sets of quantum digits: part 1 for the 'outgoing' part of Hilbert space and part 2 for the 'incoming' part (see Section 2). The challenge is transferred to part 1. Part 2 is initialized into the zero state.

Then the QC does a computation that has the effect of applying the S-matrix (6) to $\binom{|\psi_u\rangle}{0}$. This is possible in theory since R and T are known publicly, and applying S is a unitary operation. The computation results in an entangled state between parts 1 and 2 of the quantum memory: a superposition of $T\psi_u \otimes 0$ and $0 \otimes R\psi_u$.

Finally a measurement is performed on part 1 of the memory, using an equivalent of the operator Y_u. (We assume that such an operator can be constructed, since the original Hilbert space and the Hilbert space of part 1 of the memory are equivalent.) If the measurement results in eigenstate $j \in [n]$ of Y_u, then j is sent to Alice over the classical channel. If the zero state is detected, then part 2 of the memory contains $R\psi_u$; this part is teleported to a response state and sent to Alice over the quantum channel.

Our assumption denoted as 'quantum-computational unclonability' states that it is either technically/financially too difficult to pull off the above given steps or, if all the hardware works, the attack is too slow.

It is not yet clear to us how to make quantitative statements about the difficulty of launching an effective 'quantum' attack. For instance, the challenge ψ could comprise a random choice of photon wavelength λ, with the S-matrix depending on λ. How is the information about λ transferred to the quantum memory? How does the attacker know which S-matrix to apply in the quantum computation step? Does he have to construct a big unitary operation that acts on all wavelengths simultaneously? That would certainly strain the QC hardware.

3.4 Remarks

The number N' will in general be smaller than the number of repetitions N at enrollment, because of time constraints. Thus, the characterization of the PUF during the authentication phase will be less accurate than at enrollment. This is not a problem as long as enough accuracy remains.

The Transmission part of the response may be omitted from the protocol if R in itself already satisfies the physical assumptions.

The authentication scheme does not tell Alice if Bob is located right next to the enrolled QR-PUF; she merely learns whether Bob can communicate with it real-time.

4 Combining QR-PUF authentication with QKE

In this section we take the version of the authentication protocol that is based solely on R and combine it with a Quantum Key Exchange protocol that is performed on the Transmitted states. The intuition is as follows. Since Alice knows T, she can prepare $|\psi\rangle = T^{-1}|\varphi\rangle$, for arbitrary φ, such that Bob receives φ if transmission occurs. In this sense the QR-PUF is 'transparent' for the purpose of sending specific states to Bob. In particular this means that Alice and Bob can run a Quantum Key Exchange protocol *through the QR-PUF*. Some of Alice's challenges (at unpredictable times) will will be reflected back to Alice. She uses these to characterize the QR-PUF. In this way Alice is simultaneously authenticating the QR-PUF and generating a shared secret with Bob. The security of the QKE and of the authentication are independent of each other. Therefore this section does not contain a separate security analysis.

4.1 Attack model

Bob claims that he possesses a QR-PUF with identifier I. Alice's goal is to authenticate QR-PUF I and to generate a shared secret key with the party possessing that QR-PUF. Eve's goal is to learn the generated key.

We make the following assumptions. Alice is honest. Bob may be acting in one of the following ways: (i) He is honest. (ii) He has access to the QR-PUF but does not hold it personally. He is in collusion with the party holding the QR-PUF. (iii) He has access to the QR-PUF but does not hold it personally. The party holding the QR-PUF is not cooperating with him. (iv) He does not have access to the QR-PUF.

In cases (i) and (ii) the protocol should result in authentication and a shared key, even though in case (ii) the secret key is shared between Alice and the PUF holder, while Bob may not even know the key. Case (iii) should result in authentication without a shared key. Case (iv) should not even result in authentication.

Eve has physical access to Challenge and Reflected states, but *not* to Transmitted states. (Only the QR-PUF holder has access to those.) She can destroy quantum states, perform measurements on them, perform unitary operations, and insert new states. However, she cannot clone a state.

We assume that the R and T matrices are both public information. Creation or real-time emulation of a QR-PUF whose challenge-response matrix is R is assumed infeasible. We make no such assumption about the T matrix.

There is a source of enrollment data trusted by Alice and Bob.

4.2 The authenticated Quantum Key Exchange protocol

4.2.1 Enrollment phase

A QR-PUF is labeled with identifier I. Two observables B^0, B^1 are selected for future use on Transmitted states. Their eigenstates satisfy $|\langle b_i^0|b_j^1\rangle|^2 = 1/n$ for all[4] $i, j \in [n]$. The R and T matrix are accurately measured. A list $\mathcal{L}$ of challenge-operators pairs is constructed,

$$\mathcal{L} = \{\psi_u, X_u\}_{u=1}^{E}, \tag{10}$$

whose purpose is to characterize the R matrix. As in Section 3.2, each experiment is repeated N times and an array $\mathcal{P}^{\mathrm{R}} = \{P_{uj}^{\mathrm{R}}\}_{u\in[E],j\in[n]}$ of empirical probabilities is constructed,

$$P_{uj}^{\mathrm{R}} := \rho_u^{-1} \cdot \#\{\text{Experiment } \mathcal{L}_u \text{ gave reflection, and eigenstate } j \text{ of } X_u\}.$$

[4]Such inner products can be achieved for instance by having

$$|b_k^1\rangle = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}(e^{-i2\pi/n})^{kj}|b_j^0\rangle \quad ; \quad |b_a^0\rangle = \frac{1}{\sqrt{n}}\sum_{k=1}^{n}(e^{i2\pi/n})^{ka}|b_k^1\rangle.$$

Here ρ_u stands for the number of times reflection occurs. The list $\mathcal{L}$ is tuned such that for every ψ in $\mathcal{L}$ the state $T|\psi\rangle$ is an eigenvector of either B^0 or B^1. All eigenvectors occur equally often. The T, $\mathcal{L}$, $\mathcal{P}^{\mathrm{R}}$, B^0 and B^1 are assumed to be public knowledge. Note that $\mathcal{L}$ and the choice of B^0, B^1 may depend on I. The QR-PUF is given to Bob.

4.2.2 Authentication and key agreement phase

Alice and Bob both fetch the enrollment data for PUF I. Then they perform the following steps.

1. Alice initializes counters t, $\{n_a\}_{a=1}^{E}$, $\{r_{aj}\}_{a\in[E],j\in[n]}$ to zero. She initializes a set $\mathcal{V}$ to $\emptyset$. Bob initializes a counter τ to zero.
2. The following steps are repeated M times.
   (a) Alice increases t by 1. She draws a random number $u \in [E]$ and selects experiment $\mathcal{L}_u$. She increases the counter n_u by 1. She stores $\beta_t \in \{0,1\}$ and $J_t \in [n]$, where β_t and J_t are defined such that $T\psi_u$ is the J'th eigenvalue of B^β. She prepares the state ψ_u and sends it to Bob. She performs a measurement of X_u on the reflected state. If she measures $\perp$, she adds t to the set $\mathcal{V}$. If she measures the c'th eigenvalue of X_u, then she increases the counter r_{uc} by 1.
   (b) Bob increases τ by 1. He draws a random bit $b_\tau \in \{0,1\}$. He performs a measurement of B^{b_τ} on the transmitted state. He stores the outcome as $j_\tau \in \{[n], \perp\}$.
3. Alice verifies if her empirical probabilities $r_{uj}/\sum_{j'} r_{uj'}$ are sufficiently close to the enrolled P^{R}_{uj} and if the ratio $n_u^{-1}\sum_j r_{uj}$ is consistent with ρ_u/N and the experimental conditions. If not, the protocol is aborted.
4. Bob creates a set $\mathcal{W} = \{\tau|\ j_\tau \in [n]\}$, i.e. a list of times when he actually did measure a quantum state. He sends $\vec{b}$ and $\mathcal{W}$ to Alice over the classical channel.
5. Alice checks if the difference between $\mathcal{V}$ and $\mathcal{W}$ is sufficiently small. If not, then the protocol is aborted.
6. Alice computes the overlap list $\mathcal{T} = \{t \in \mathcal{V}\cap\mathcal{W}|\ b_t = \beta_t\}$ of events where Alice and Bob both observed Transmission *and* Bob's operator $B^{0/1}$ matched with the eigenstate that Alice sent. Alice generates a random subset $\mathcal{S} \subset \mathcal{T}$ and sends $\mathcal{S}$ to Bob.
7. Bob returns $\vec{j}_{\mathcal{S}}$ to Alice. (The subscript $\mathcal{S}$ denotes selection of j_τ with $\tau \in \mathcal{S}$.)
8. Alice checks if $\vec{j}_{\mathcal{S}}$ is sufficiently close to $\vec{J}_{\mathcal{S}}$. If not, the protocol is aborted.
9. Alice sends $\mathcal{T}\setminus\mathcal{S}$ to Bob.
10. Alice and Bob now have almost identical secret n-ary strings. Alice has $\vec{J}_{\mathcal{T}\setminus\mathcal{S}}$, Bob has $\vec{j}_{\mathcal{T}\setminus\mathcal{S}}$. They perform information reconciliation and privacy amplification to agree on a noise-free near-uniform secret string.

4.3 Remarks

- Alice does not know if she has a shared secret with Bob, but she does know that she has a shared secret with a party that has real-time access to the Transmitted states leaving the QR-PUF.
- We have deliberately been vague about the choice of information reconciliation and privacy amplification scheme. These are independent of the method for achieving the shared noisy secret.

- We have also been vague about the meaning of the word 'sufficient' in steps 5 and 8. We do not wish to dwell on these issues as they have been adequately treated in the existing literature on QKE.
- The protocol presented in Section 4.2.2 may be modified in many ways without changing its essential properties. Alice may reveal $\mathcal{T}$ to Bob at an earlier stage. She may send $\vec{J}_{\mathcal{S}}$ to Bob. The moment of checking the empirical reflection probabilities may be shifted.
- If Alice and Bob want *mutual* authentication, the protocol can be run twice: first with Alice authenticating Bob's QR-PUF, then the other way round. The two protocols may also be run intertwined, i.e. alternating their steps 2a,2b before proceeding to step 3.

5 Summary and future work

We have introduced a new security primitive, the Quantum Readout PUF (QR-PUF), which is a classical PUF which can be read out using quantum states. We have shown how QR-PUFs can be used to achieve remote authentication without a trusted remote reader device. The security is based on two well known physical assumptions, physical unclonability and uniqueness, and one new physical assumption, quantum-computational unclonability. The no-cloning theorem guarantees that intercept-resend attacks will be detected. Our authentication scheme is vulnerable to a three-step attack employing quantum teleportation and a quantum computer. For this reason we need the assumption of quantum-computational unclonability, which states that this kind of attack, while possible in theory, is infeasible in practice because of technical or financial issues. What makes it especially difficult is the fact that our protocol doubles as a distance bounding scheme; all three steps of the attack have to be extremely fast.
For QR-PUFs whose responses have a Transmission part and a Reflection part, we have sketched how QR-PUF authentication can be intertwined with Quantum Key Exchange. This combination achieves authenticated QKE without the need for an initial shared secret (short MAC key or an entangled state).
We have sketched our protocols in a fairly theoretical way, almost without discussing implementation issues, leaving open a lot of questions. The biggest question is how to construct a QR-PUF in the first place. The most practical option seems to be a partially transmissive optical PUF that is challengeable by single photon states through an optical fiber. The main difficulty is to make sure that the uniqueness and physical unclonability properties hold, in spite of the limited number of challenges that can be passed through a fiber. Fibers carry a limited number of transversal modes, while such modes are the main way of challenging an ordinary speckle-producing optical PUF [18, 19]. We are aided by the fact that multiple wavelengths are available.
Another question is how to quantify the difficulty of the 'quantum' attack (Section 3.3.2), which is the most serious threat to QR-PUFs. Here too the availability of different wavelengths seems to help us.
Finally there are various other issues that we glossed over, such as state preparation, existence of the right observables, detector efficiency, details of error correction and privacy amplification, and truly rigorous security proofs. Also we did not specify any numbers or set sizes in the authenticated QKE protocol. This was done on purpose, as the main aim of this paper is to present the basic ideas. Implementation details and rigorous security proofs are left for future work.
Our protocols can be extended in a number of obvious ways. For instance, EPR pairs can be used, as well as anti-eavesdropping countermeasures like 'decoy states' [12]. The QR-PUF can be used for Quantum Oblivious Transfer. Another option is transmitting states through more than one QR-PUF. It would also be interesting to see if one can construct a 'quantum PUF', i.e. a PUF that has actual quantum behaviour, resulting in nontrivial (perhaps nonlinear) interaction between the challenge state and the PUF.

Acknowledgements

We kindly thank Pim Tuyls, Berry Schoenmakers, Allard Mosk and Andrea Fiore for useful discussions.

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, July 2005.

[4] C.N. Chong, D. Jiang, J. Zhang, and L. Guo. Anti-counterfeiting with a random pattern. In *SECURWARE*, pages 146–153. IEEE, 2008.

[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] L.M. Duan, M. Lukin, J.I. Cirac, and P. Zoller. Long-distance quantum communication with atomic ensembles and linear optics. *Nature*, 414:413–418, 2001.

[8] A.K. Ekert. Quantum cryptography based on Bells theorem. *Phys. Rev. Lett.*, 67:661 – 663, 1991.

[9] B. Gassend, D.E. Clarke, M. van Dijk, and S. Devadas. Silicon physical unknown functions. In V. Atluri, editor, *ACM Conference on Computer and Communications Security — CCS 2002*, pages 148–160. ACM, November 2002.

[10] D. Gottesman and J. Preskill. Secure quantum key exchange using squeezed states, 2000. arXiv:quant-ph/0008046v2.

[11] J. Guajardo, S.S. Kumar, G.J. Schrijen, and P. Tuyls. FPGA intrinsic PUFs and their use for IP protection. In P. Paillier and I. Verbauwhede, editors, *CHES*, volume 4727 of *LNCS*, pages 63–80. Springer, 2007.

[12] W.-Y. Hwang. Quantum key distribution with high loss: Toward global secure communication. *Phys.Rev.Lett.*, 91:057901, 2003.

[13] K. Inoue, E. Waks, and Y. Yamamoto. Differential phase shift quantum key distribution. *Phys. Rev. Lett.*, 89:037902, 2002.

[14] D. Kirovski. Toward an automated verification of certificates of authenticity. In J.S. Breese, J. Feigenbaum, and M.I. Seltzer, editors, *ACM Conference on Electronic Commerce*, pages 160–169. ACM, 2004.

[15] S.S. Kumar, J. Guajardo, R. Maes, G.J. Schrijen, and P. Tuyls. The Butterfly PUF: Protecting IP on every FPGA. In M. Tehranipoor and J. Plusquellic, editors, *HOST*, pages 67–70. IEEE Computer Society, 2008.

[16] K. Kursawe, A.-R. Sadeghi, D. Schellekens, B. Škorić, and P. Tuyls. Reconfigurable physical unclonable functions. In *HOST*, 2009.

[17] D.N. Matsukevich and A. Kuzmich. Quantum state transfer between matter and light. *Science*, 306(5696):663–666, 2004.

[18] R. Pappu. *Physical One-Way Functions*. PhD thesis, MIT, 2001.

[19] R. Pappu, B. Recht, J. Taylor, and N. Gershenfeld. Physical One-Way Functions. *Science*, 297:2026–2030, Sept. 2002.

[20] B.S. Shi, J. Li, J.M. Liu, X.F. Fan, and G.C. Guo. Quantum key distribution and quantum authentication based on entangled state. *Phys.Lett.A*, 281:83–87, 2001.

[21] P. Tuyls, G.J. Schrijen, B. Škorić, J. van Geloven, R. Verhaegh, and R. Wolters. Read-proof hardware from protective coatings. In L. Goubin and M. Matsui, editors, *Cryptographic Hardware and Embedded Systems — CHES 2006*, volume 4249 of *LNCS*, pages 369–383. Springer-Verlag, 2006.

[22] P. Tuyls, B. Škorić, and T. Kevenaar. *Security with Noisy Data: Private Biometrics, Secure Key Storage and Anti-Counterfeiting*. Springer, London, 2007.

[23] P. Tuyls, B. Škorić, S. Stallinga, A.H.M. Akkermans, and W. Ophey. Information-theoretic security analysis of physical uncloneable functions. In A.S. Patrick and M. Yung, editors, *9th Conf. on Financial Cryptography and Data Security*, volume 3570 of *LNCS*, pages 141–155. Springer, 2005.

[24] 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.

[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.

A Analysis of probabilities in the authentication protocol

Let Bob have a (probabilistic) strategy for deciding which response to send back to Alice as a function of his operator Z and the measurement result z_a. We will be very generous, allowing Bob to choose arbitrary Hermitian Z, and to send arbitrary quantum states, i.e. not necessarily of the form $R|\psi\rangle$ for some $\psi \in \mathcal{C}$. We write

$$P_Z := \text{Prob[Bob selects observable } Z], \tag{11}$$

$$\begin{aligned} Q_{Za}(\varphi) &:= \text{Prob[Bob sends } |\varphi\rangle \text{ and } \perp \mid Z, z_a] \\ q_{Za}(j) &:= \text{Prob[Bob sends no state and } j \mid Z, z_a]. \end{aligned} \tag{12}$$

Here the probability is with respect to Bob's probabilistic strategy.

Total transmission probability

First we show that Bob is able to successfully fake the total transmission probability for any challenge $\psi \in \mathcal{C}$. In Section 3.3 we saw that the correct value is $t_\psi = \langle\psi|T^\dagger T|\psi\rangle$. His strategy

(11,12) gives rise to a transmission rate t'_ψ,

$$\begin{aligned} t'_\psi &= \sum_{j=1}^{n} \sum_{Z} P_Z \sum_{a=1}^{n} |\langle z_a|\psi\rangle|^2 q_{Za}(j) \\ &= \langle\psi| \left(\sum_{Z} P_Z \sum_{a=1}^{n} [\sum_{j=1}^{n} q_{Za}(j)] |z_a\rangle\langle z_a| \right) |\psi\rangle. \end{aligned} \tag{13}$$

This must equal $\langle\psi|T^\dagger T|\psi\rangle$ for all ψ, so Bob's task is to finetune his strategy such that the operator in $(\cdot)$ brackets is precisely $T^\dagger T$. This is achieved as follows. Bob sets $Z = T^\dagger T$. Note that this is a Hermitian operator with positive eigenvalues $\{\lambda_u\}_{u=1}^n$. Furthermore Bob sets $\sum_j q_{Za}(j) = \lambda_a$, which is possible since λ_a is real-valued and positive. Substitution into (13) gives $\sum_a \lambda_a |\lambda_a\rangle\langle\lambda_a| = T^\dagger T$ for the operator, and $t'_\psi = t_\psi$.

One specific transmission probability

Let us now consider how Bob can simulate the correct probability for the event $Y \to y_j$ for a *single* fixed j, for all challenges. The correct probability is given by

$$p_j := |\langle y_j|T|\psi\rangle|^2 = \langle\psi| \left(T^\dagger |y_j\rangle\langle y_j|T \right) |\psi\rangle, \tag{14}$$

while Bob's strategy (11,12) gives

$$\begin{aligned} p'_j &:= \sum_{Z} P_Z \sum_{a=1}^{n} |\langle z_a|\psi\rangle|^2 q_{Za}(j) \\ &= \langle\psi| \left(\sum_{Z} P_Z \sum_{a=1}^{n} q_{Za}(j) |z_a\rangle\langle z_a| \right) |\psi\rangle. \end{aligned} \tag{15}$$

Since Bob has to make sure that $p'_j = p_j$ for all ψ, he has to finetune his strategy such that the operators inside the $(\cdot)$ brackets in (14) and (15) are equal. Bob selects an observable Z^* such that its b'th eigenvector precisely points in the direction of $T^\dagger|y_j\rangle$, i.e. $|z_b^*\rangle = \kappa_j T^\dagger |y_j\rangle$, where κ_j is a normalization constant. Furthermore he sets $P_Z = \delta_{Z,Z^*}$ and $q_{Z^*a}(j) = |\kappa_j|^{-2}\delta_{ab}$. Substitution into (15) gives $T^\dagger|y_j\rangle(T^\dagger|y_j\rangle)^\dagger = T^\dagger|y_j\rangle\langle y_j|T$ for the operator, and hence $p'_j = p_j$.

All transmission probabilities

Above we have seen that Bob has to choose his observable Z^* according to the event $(Y \to y_j)$ that he wants to simulate correctly, i.e. his choice for Z^* depends on j. It is clear that he cannot do this for all $j \in [n]$ simultaneously.

One specific reflection probability

We consider the case that Bob wants to simulate one specific reflection probability, namely of the event $X \to x_j$. The correct probability of this event is

$$w_j := |\langle x_j|R|\psi\rangle|^2 = \langle\psi| \left(R^\dagger |x_j\rangle\langle x_j|R \right) |\psi\rangle, \tag{16}$$

while Bob's strategy (11,12) gives

$$\begin{aligned} w'_j &= \sum_{Z} P_Z \sum_{a=1}^{n} |\langle z_a|\psi\rangle|^2 \sum_{\varphi} Q_{Za}(\varphi) |\langle x_j|\varphi\rangle|^2 \\ &= \langle\psi| \left(\sum_{Z} P_Z \sum_{a=1}^{n} |z_a\rangle\langle z_a| \sum_{\varphi} Q_{Za}(\varphi) |\langle x_j|\varphi\rangle|^2 \right) |\psi\rangle. \end{aligned} \tag{17}$$

Again, Bob wants to simulate the correct probability for all ψ, hence the operators between () brackets in (16) and (17) have to be identical. Bob has to set $P_Z = \delta_{ZZ^*}$ with Z^* such that for

some $b \in [n]$ it holds that $|z_b^*\rangle = \nu_j R^\dagger |x_j\rangle$, where ν_j is a normalization constant. Furthermore he has to set $Q_{Z^* a}(\varphi) = \delta_{ab} F(\varphi) |\nu_j|^{-2}$ where F has to satisfy $\sum_\varphi F(\varphi) |\langle x_j | \varphi \rangle|^2 = 1$.

All reflection probabilities

We showed above that Bob has to choose his Z^* as a function of j if he wants to successfully manipulate Alice's x_j measurements. Bob cannot do this for all $j \in [n]$ simultaneously.