Robust Quantum Public-Key Encryption with Applications to Quantum Key Distribution

Giulio Malavolta[1,2] and Michael Walter[3]

[1]Bocconi University, Milan, Italy
[2]Max Planck Institute for Security and Privacy, Bochum, Germany
[3]Ruhr-Universität Bochum, Bochum, Germany

Abstract

Quantum key distribution (QKD) allows Alice and Bob to agree on a shared secret key, while communicating over a public (untrusted) quantum channel. Compared to classical key exchange, it has two main advantages: (i) The key is *unconditionally* hidden to the eyes of any attacker, and (ii) its security assumes only the existence of authenticated classical channels which, in practice, can be realized using Minicrypt assumptions, such as the existence of digital signatures. On the flip side, QKD protocols typically require multiple rounds of interactions, whereas classical key exchange can be realized with the minimal amount of two messages using public-key encryption. A long-standing open question is whether QKD requires more rounds of interaction than classical key exchange.

In this work, we propose a two-message QKD protocol that satisfies *everlasting* security, assuming only the existence of quantum-secure one-way functions. That is, the shared key is unconditionally hidden, provided computational assumptions hold during the protocol execution. Our result follows from a new construction of quantum public-key encryption (QPKE) whose security, much like its classical counterpart, only relies on authenticated *classical* channels.

1 Introduction

Quantum key distribution (QKD) [5] enables Alice and Bob to exchange a secret key over a public (untrusted) quantum channel. Compared to classical key exchange, it offers two main advantages: (i) It hides the key *unconditionally* or *information-theoretically* to the eyes of any (possibly unbounded and quantum) attacker, and (ii) it relies only on the existence of authenticated classical channels, which in practice can be instantiated using Minicrypt [13] computational assumptions. The former is plainly impossible to achieve without quantum information, and we also have strong evidence [14] that classical key exchange requires more structured (Cryptomania) computational assumptions.

Over the past decades, QKD has inspired a staggering amount of research, ranging from profound theoretical works [20, 18, 19, 23] all the way to large-scale experiments [15, 17, 26]. As such, QKD is one of the most studied topics in the theory of quantum information. Despite the vast literature on the topic, QKD protocols still do not outperform classical key exchange in all aspects: Whereas classical key exchange can be realized using two messages [6], which is optimal, to the best of our knowledge all known QKD protocols require more than two rounds of interaction.

This raises the question of whether more rounds of interaction are really necessary for QKD. Apart from its theoretical importance, there are also practical reasons for addressing this question. Indeed, two-message protocols are particularly desirable for practical scenarios where parties may drop offline during the protocol execution, may want to send their message at a later point in time, or do not want to keep a state across rounds. Arguably, this minimal interaction pattern is the property that makes traditional cryptographic primitives, such as public-key encryption, so useful. However, despite this strong motivation and almost fifty years of intense research, the optimal round complexity of QKD is still an open problem.

1.1 Our Results

In this work we show that two messages are sufficient to build a QKD protocol with everlasting security. Specifically, we prove the following statement:

If quantum-secure one-way functions exist, then there exists two-message everlasting QKD.

Since two messages are clearly necessary for QKD, our protocol achieves the optimal round complexity. Furthermore, only the first message (from Alice to Bob) is quantum, whereas Bob's response is entirely classical. The protocol satisfies the strong notion of *everlasting security*: As long as the attacker runs in quantum polynomial time during the execution of the protocol, the shared key is hidden in an *information-theoretic* sense.

We view our approach as a big departure from the traditional design of QKD protocols [5, 9], and it is inspired instead by recent works on cryptography with certified deletion [4] and quantum public-key encryption [3]. Both our protocol and its analysis are entirely elementary and they are simple enough to be fully described in an undergraduate class. For comparison, it took more than ten years for researchers to establish a formal proof of the first QKD protocol [5].

Our main technical ingredient is a new framework for building quantum public-key encryption (QPKE), a cryptographic primitive that allows Alice to sample a public key consisting of a quantum state ρ and a classical string pk. The scheme is *robust*, in the sense that security is guaranteed to hold even if the distinguisher is given pk, and it is allowed to tamper arbitrarily with the quantum state ρ. We present two instances of QPKE:

- (Everlasting Security) In our first scheme, the message m remains *information-theoretically* hidden, provided that the distinguisher was computationally bounded during the execution of the protocol. This property holds if the distinguisher is given a *single copy* of the public key, which is sufficient to build QKD.

- (Computational Security) In our second scheme, the message m is computationally hidden, i.e., we only require security against a computationally bounded distinguisher. While this is a weaker security notion, the advantage of the scheme is that security holds even if the distinguisher is given *arbitrarily many* copies of the public key.

The fact that the first construction is only secure in a model where we give access to a single copy of the public key is not a coincidence. It is shown [3] that given enough (but polynomially-many) copies of the public key, an unbounded adversary can launch a key-recovery attack. This means that there does not exist any QPKE with everlasting security against a distinguisher that sees arbitrarily-many copies of the public key, and it justifies the need for a weaker security notion (computational security).

Everlasting Security. We point out a subtle difference between the attacker model that we consider in this work, compared to the standard attacker model for QKD. The latter, considered for instance in [24, 25], models the attacker as a computationally unbounded quantum channel, that is however not allowed to tamper with the information sent over the *classical channels*. That is, it only assumes the existence of authenticated classical channels, but otherwise does not impose any restriction on the runtime of the distinguisher. On the other hand, in this work we consider – in addition to the existence of authenticated classical channels – an attacker that runs in *quantum polynomial time* during the protocol execution, but it is allowed to be unbounded once the protocol terminates. That is, we prove *everlasting security* in the sense of [21].

While technically different, we argue that for most practical scenarios the two models are in fact equivalent. The assumption of an authenticated classical channel is most often justified by having each party sign their own messages with a digital signature, which would require computational assumptions to hold (at the very least) during the execution of the protocol. In this sense, the mere existence of authenticated classical channels already restricts the attacker to run in quantum polynomial time during the protocol execution (as otherwise it could just break the security of the digital signature).

Finally, it is not hard to show that everlasting security is the *best possible* security notion for QPKE, i.e., there exists a generic attack against any QPKE scheme, if the distinguisher is allowed to run in unbounded time during the execution of the protocol. The attack works even in the presence of authenticated classical channels and succeeds with certainty. For completeness, we report the proof of this fact in Appendix A.

1.2 Concurrent and Related Work

A concurrent work [16] obtains similar results on robust QPKE, where security holds only against a distinguisher that is allowed to tamper arbitrarily with the quantum portion of the public key. Compared to our work, they only consider the setting of *computational* security, whereas we view the scheme with *everlasting* security as the main contribution of our work, which is the one that enables our two-message QKD protocol. Even focusing on the computational settings, our schemes share many similarities but, interestingly, they are not identical. At a technical level, their approach is based on one-time signatures for Wiesner states, whereas our approach can (in retrospect) be thought of as one-time signing the $|+\rangle$ state. On the other hand, [16] presents a scheme where the public key is a pure state and furthermore their schemes achieve the stronger notion of CCA-security, which we do not consider in this work.

Finally, both the present work and [16] can be seen as a follow-up to [3], that formally introduced the notion of QPKE. However, the security definition presented in [3] is in a much weaker adversarial model, where the public keys are distributed via an authenticated *quantum* channel. In contrast, both our work and [16] require only authenticated *classical* channels, which is the same assumption as in traditional PKE.

1.3 Open Problems

Our work leaves open a series of questions, that we hope will inspire further research in this area. For starters, our protocols are described in the presence of perfect (noiseless) quantum channels. Any practical protocol would need to withstand the presence of noise. While theoretically one could simply encode all states using a quantum error correcting code, this may lead to poor concrete

efficiency. We leave open the question of investigating variants our non-interactive QKD protocol that are efficient in the presence of noise.

A compelling aspect of standard QKD protocols such as [5] is that all quantum states consists of tensor products of single qubits, whereas our protocols require coherent superpositions of many-qubit states. The former property is desirable, since it allows the experimental realization of the protocol on present-day quantum hardware. We view the problem of constructing a non-interactive QKD protocol in this qubit-by-qubit model as fascinating research direction, and believe that it might require substantially different techniques from the present approach.

1.4 Overview of the Solution

Our main technical contribution is a new recipe to construct QPKE, whose definition we recall next. The syntax of QPKE consists of three algorithms: key generation, encryption, and decryption. The key generation algorithm produces a classical key pair $(\mathsf{sk}, \mathsf{pk})$, along with a quantum state ρ. The pair (pk, ρ) makes up the public key. Given this public key, anyone can compute a classical ciphertext ct encrypting a given message m, which can only be decrypted by the owner of the secret key. In terms of security, we require that the message m should be hidden, even if an efficient attacker is allowed to tamper arbitrarily with the quantum state associated with the public key.

To understand the challenge, let us recall that the security of existing quantum PKEs [3] only apply if the (quantum) public key is honestly delivered to the encrypter. In other words, such schemes implicitly assume the presence of authenticated quantum channels. This is in contrast with the standard model for QKD, which requires only authenticated *classical* channels. As discussed before, this model is justified by the fact that it is easy to implement in practice (e.g., by signing all messages), whereas authenticating quantum states is a notoriously difficult problem, and there is evidence that it is in fact not possible in full generality [2].

Fortunately, full authentication of quantum channels is not necessary for constructing quantum PKE: For instance, it is acceptable if some malformed state passes the authentication check, so long as the encryption algorithm results in an "undecryptable" cipher. In this work, we show how to achieve this, assuming the existence of (one-time) digital signatures. In more details, a public key in our scheme consists of a uniform superposition of two valid message-signature pairs

$$\frac{|0,\sigma_0\rangle + |1,\sigma_1\rangle}{\sqrt{2}} \tag{1.1}$$

where σ_b is a valid signature on b, under some verification key vk. Given such state, the encrypter authenticates the states by applying the projection

$$\Pi = \sum_{\sigma\in\Sigma_0} |0,\sigma\rangle\langle 0,\sigma| + \sum_{\sigma\in\Sigma_1} |1,\sigma\rangle\langle 1,\sigma| \,.$$

Where Σ_b is the set of all valid signatures on b. Note that Π can be implemented efficiently given vk by running the verification algorithm coherently, and measuring the bit that denotes acceptance/rejection. If this test passes, the encrypter encodes its message $m \in \{0,1\}$ by applying a conditional phase flip to the resulting state. In an honest run of the protocol, this results in the state

$$\frac{|0,\sigma_0\rangle + (-1)^m |1,\sigma_1\rangle}{\sqrt{2}}$$

which is efficiently decodable by decrypter, who knows both σ_0 and σ_1, by measuring the state in the corresponding basis. Our analysis boils down to showing that for any state that passes one of the following two events must have occurred:

- The state is identical to the one in Eq. (1.1). In which case, the attacker must have done nothing, and therefore the state was honestly delivered to the encrypter. One can then show that the message m is hidden by appealing to the *distinguishing implies swapping* principle [1, 12].
- The state was measured in the computational basis. Note that in this case, the state is in the image of the projector Π, and therefore it will pass the authentication check. However, a phase flip on any basis state only adds a global phase, and therefore has no effect on the ciphertext. Therefore, the message is information-theoretically hidden.

Once we built QPKE, it appears to be an easy exercise to construct a two-message QKD protocol: Alice can sample a public key and send it to Bob, who replies with an encryption of a randomly sampled key $k \in \{0,1\}^\lambda$. However, there is a subtle aspect in the analysis of this protocol, where one needs to ensure that an attacker cannot cause Alice and Bob to agree on *different keys*. Note that this does not follow immediately from the security of the QPKE, which only protects privacy of the plaintext. We once again rely on one-time signatures to (provably) prevent this class of attacks.

1.5 Organization of this Paper

In Section 2, we discuss preliminaries in quantum information and cryptography and we prove some useful technical statements. In Section 3, we define the notion of QPKE, and we present two constructions (with different tradeoffs) whose security can be reduced to the one-wayness of any post-quantum one-way function. In Section 4 we present the formal description and the analysis of our two-message QKD protocol.

2 Preliminaries

Throughout this work, we denote the security parameter by λ. We denote by 1^λ the all-ones string of length λ. We say that a function negl is *negligible* in the security parameter λ if $\mathsf{negl}(\lambda) = \lambda^{-\omega(1)}$. For a finite set S, we write $x \leftarrow S$ to denote that x is sampled uniformly at random from S. We write Tr for the trace of a matrix or operator.

2.1 Quantum Information

In this section, we provide some preliminary background on quantum information. For a more in-depth introduction, we refer the reader to [22]. A *register* x consisting of n qubits is given by a Hilbert space $(\mathbb{C}^2)^{\otimes n}$ with name or label x. Given two registers x and y, we write $\mathsf{x} \otimes \mathsf{y}$ for the composite register, with Hilbert space the tensor product of the individual registers' Hilbert spaces. A *pure quantum state* on register x is a unit vector $|\Psi\rangle_{\mathsf{x}} \in (\mathbb{C}^2)^{\otimes n}$. A *mixed quantum state* on register x is represented by a density operator ρ_{x} on $(\mathbb{C}^2)^{\otimes n}$, which is a positive semi-definite Hermitian matrix with trace 1. Any pure state $|\Psi\rangle_{\mathsf{x}}$ can also be regarded as a mixed state $\rho_{\mathsf{x}} = |\Psi\rangle_{\mathsf{x}} \langle\Psi|_{\mathsf{x}}$, but there

are mixed states that are not pure. In the above, we use subscripts to denote registers, but we often omit these when clear from context. We adopt the convention that

$$\{|0\rangle, |1\rangle\} \text{ and } \left\{\frac{|0\rangle + |1\rangle}{\sqrt{2}}, \frac{|0\rangle - |1\rangle}{\sqrt{2}}\right\}$$

denote the *computational* and the *Hadamard basis* states, respectively.

A *quantum channel* F is a completely-positive trace-preserving (CPTP) map from a register x to a register y. That is, on input any density matrix ρ_{x}, the operation F produces $F(\rho_{\mathsf{x}}) = \tau_{\mathsf{y}}$, another state on register y, and the same is true when we apply F to the x register of a quantum state ρ_{xz}. For any unitary operator U, meaning $U^\dagger U = UU^\dagger = \mathsf{Id}$, one obtains a quantum channel that maps input states ρ to output states $\tau := U\rho U^\dagger$. The Pauli operators $\mathsf{X}, \mathsf{Y}, \mathsf{Z}$ are 2×2 matrices that are unitary and Hermitian. More specifically:

$$\mathsf{X} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \mathsf{Y} = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \mathsf{Z} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

A *projector* Π is a Hermitian operator such that $\Pi^2 = \Pi$. A *projective measurement* is given by a collection of projectors $\{\Pi_j\}_j$ such that $\sum_j \Pi_j = \mathsf{Id}$. Given a state ρ, the measurement yields outcome j with probability $p_j = \mathsf{Tr}(\Pi_j \rho)$, upon which the state changes to $\Pi_j \rho \Pi_j / p_j$ (this can be modeled by a quantum channel, but we will not need this). For any two registers x and y, the partial trace Tr_{y} is the unique channel from $\mathsf{x} \otimes \mathsf{y}$ to x such that $\mathsf{Tr}_{\mathsf{y}}(\rho_{\mathsf{x}} \otimes \tau_{\mathsf{y}}) = \mathsf{Tr}_{\mathsf{y}}(\tau_{\mathsf{y}})\rho_{\mathsf{x}}$ for all ρ_{x} and τ_{y}.

The *trace distance* between two states ρ and τ, denoted by $\mathsf{Td}(\rho, \tau)$ is defined as

$$\mathsf{Td}(\rho, \tau) = \frac{1}{2}\|\rho - \tau\|_1 = \frac{1}{2}\mathsf{Tr}\left(\sqrt{(\rho - \tau)^\dagger(\rho - \tau)}\right).$$

The operational meaning of the trace distance is that $\frac{1}{2}(1 + \mathsf{Td}(\rho, \tau))$ is the maximal probability that two states ρ and τ can be distinguishdd by any (possibly unbounded) quantum channel or algorithm. If $\tau = |\Phi\rangle\langle\Phi|$ is a pure state, we have the following version of the Fuchs-van de Graaf inequalities:

$$1 - \langle\Phi|\rho|\Phi\rangle \leq \mathsf{Td}(\rho, \tau) \leq \sqrt{1 - \langle\Phi|\rho|\Phi\rangle}. \tag{2.1}$$

Quantum Algorithms. A non-uniform *quantum polynomial-time (QPT) machine* $\{\mathcal{A}_\lambda\}_{\lambda\in\mathbb{N}}$ is a family of polynomial-size quantum machines $\mathcal{A}_\lambda$, where each is initialized with a polynomial-size advice state $|\alpha_\lambda\rangle$. Each $\mathcal{A}_\lambda$ can be described by a CPTP map. A quantum interactive machine is simply a sequence of quantum channels, with designated input, output, and work registers. We say that two probability distributions $\mathcal{X}$ and $\mathcal{Y}$ are *computationally indistinguishable* if there exists a negligible function negl such that for all QPT algorithms $\mathcal{A}_\lambda$ it holds that

$$\left|\Pr\left[1 \leftarrow \mathcal{A}_\lambda(x) : x \leftarrow \mathcal{X}\right] - \Pr\left[1 \leftarrow \mathcal{A}_\lambda(y) : y \leftarrow \mathcal{Y}\right]\right| = \mathsf{negl}(\lambda).$$

We say that they are *statistically indistinguishable* if the same holds for all (possibly unbounded) algorithms.

Distinguishing Implies Swapping. We recall the formal statement of the equivalence between distinguishing states and swapping on the conjugate basis. This was proven in [1] and below we show a rephrased version borrowed from [12]. We actually only state one direction of the implication (the converse is also shown to be true in [1]), since it is the one needed for our purposes.

Theorem 2.1 (Distinguishing Implies Swapping [1])**.** Let $|\Psi\rangle$ and $|\Phi\rangle$ be orthogonal n-qubit states, and suppose that a QPT distinguisher $\mathcal{A}_\lambda$ distinguishes $|\Psi\rangle$ and $|\Phi\rangle$ with advantage δ without using any ancilla qubits. Then, there exists a polynomial-time computable unitary U over n-qubit states such that

$$\frac{|\langle y|\, U\, |x\rangle + \langle x|\, U\, |y\rangle|}{2} = \delta \text{ where } |x\rangle = \frac{|\Psi\rangle + |\Phi\rangle}{\sqrt{2}} \text{ and } |y\rangle = \frac{|\Psi\rangle - |\Phi\rangle}{\sqrt{2}}.$$

Moreover, if $\mathcal{A}_\lambda$ does not act on some qubits, then U also does not act on those qubits.

2.2 Information Theory

Recall the definition of the min-entropy of a random variable X as

$$H_\infty(X) = -\log\left(\mathsf{max}_x \Pr[X = x]\right).$$

We recall the definition of average conditional min-entropy in the following.

Definition 2.2 (Average Conditional Min-Entropy)**.** Let X be a random-variable supported on a finite set $\mathcal{X}$ and let Z be a (possibly correlated) random variable supported on a finite set $\mathcal{Z}$. The average-conditional min-entropy $\tilde{H}_\infty(X|Z)$ is defined as

$$\tilde{H}_\infty(X|Z) = -\log\left(\mathbb{E}_z\left[\mathsf{max}_{x\in\mathcal{X}} \Pr[X = x|Z = z]\right]\right).$$

It is shown in [8, 7] that the average conditional min-entropy satisfies a *chain rule*, that is

$$\tilde{H}_\infty(X|Z) \geq H_\infty(X) - H_0(Z) \tag{2.2}$$

where $H_0(Z)$ denotes the logarithm of the size of the support of Z. Next, we recall the definition of a seeded randomness extractor.

Definition 2.3 (Extractor)**.** A function $\mathsf{Ext} : \{0,1\}^d \times \mathcal{X} \to \{0,1\}^\ell$ is called a seeded strong average-case (k,ε)-extractor, if it holds for all random variables X with support $\mathcal{X}$ and Z defined on some finite support that if $\tilde{H}_\infty(X|Z) \geq k$, then it holds that the statistical distance of the following distributions is a most ε

$$(\mathsf{seed}, \mathsf{Ext}(\mathsf{seed}, X), Z) \approx_\varepsilon (\mathsf{seed}, U, Z)$$

where $\mathsf{seed} \leftarrow \{0,1\}^d$ and $U \leftarrow \{0,1\}^\ell$.

Recall that a hash function $\mathsf{Hash} : \mathcal{X} \to \mathcal{Y}$ is a universal hash if for all $x \neq x' \in \mathcal{X}$ it holds that

$$\Pr[\mathsf{Hash}(x) = \mathsf{Hash}(x')] \leq \frac{1}{|\mathcal{Y}|}$$

where the probability is taken over the choice of the hash function. It is shown [8, 7] that any universal hash function is an average-case randomness extractor.

Lemma 2.4 (Leftover Hash Lemma)**.** Let X be a random-variable supported on a finite set $\mathcal{X}$ and let Z be a (possibly correlated) random variable supported on a finite set $\mathcal{Z}$ such that $\tilde{H}_\infty(X|Z) \geq k$. Let $\mathsf{Hash} : \mathcal{X} \to \{0,1\}^\ell$, where $\ell \leq k - 2\log\left(\frac{1}{\varepsilon}\right)$, be a family of universal hash functions. Then Hash is a seeded strong average-case (k,ε)-extractor.

2.3 Pseudorandom Functions

We recall the notion of a pseudorandom function [11]. A pseudorandom function (PRF) is a keyed function

$$\mathsf{PRF} : \{0,1\}^\lambda \times \{0,1\}^\lambda \to \{0,1\}^\lambda$$

that is computationally indistinguishable from a truly random function. More precisely, we require that there exists a negligible function negl such that for all QPT $\{\mathcal{A}_\lambda\}$, for all $\lambda \in \mathbb{N}$, it holds that the following distributions are computationally indistinguishable

$$\mathcal{A}_\lambda(1^\lambda)^{\mathsf{PRF}(k,\cdot)} \approx \mathcal{A}_\lambda(1^\lambda)^{f(\cdot)}$$

where $k \leftarrow \{0,1\}^\lambda$ and f is a uniformly-sampled function. It is well-known that quantum-secure PRFs can be built from any one-way function [11].

2.4 One-Time Signatures

We recall the notion of a one-time signature scheme [10].

Definition 2.5 (One-Time Signature)**.** A *one-time signature (OTS)* scheme is defined as a tuple of algorithms $(\mathsf{SGen}, \mathsf{Sign}, \mathsf{Ver})$ such that:

- $(\mathsf{vk}, \mathsf{sk}) \leftarrow \mathsf{SGen}(1^\lambda)$: A polynomial-time algorithm which, on input the security parameter 1^λ, outputs two bit strings vk and sk.
- $\sigma \leftarrow \mathsf{Sign}(\mathsf{sk}, m)$: A polynomial-time algorithm which, on input the signing key sk and a message m, outputs signature σ.
- $\{0,1\} \leftarrow \mathsf{Ver}(\mathsf{vk}, m, \sigma)$: A polynomial-time algorithm which, on input the verification key vk, a message m, and a signature σ, returns a bit denoting accept or reject.

The OTS scheme is *correct* if for all $\lambda \in \mathbb{N}$ and all messages m it holds that

$$\Pr\left[1 = \mathsf{Ver}(\mathsf{vk}, m, \mathsf{Sign}(\mathsf{sk}, m)) : (\mathsf{vk}, \mathsf{sk}) \leftarrow \mathsf{SGen}(1^\lambda)\right] = 1.$$

Next we define the notion of strong existential unforgeability for OTS, which states that one should not be able to produce a different signature (even if on the same message) than the one provided by the signing oracle. It is well-known that strongly unforgeable signatures can be constructed from any one-way function (OWF) [10]. For convenience we define a slightly weaker notion, where the message to be signed is fixed in advance – this notion is clearly implied by the standard one, where the attacker can query the signing oracle adaptively.

Definition 2.6 (Strong Existential Unforgeability)**.** We say that an OTS scheme $(\mathsf{SGen}, \mathsf{Sign}, \mathsf{Ver})$ satisfies *(quantum-secure) strong existential unforgeability* if there exists a negligible function negl such that for all QPT $\{\mathcal{A}_\lambda\}$, for all $\lambda \in \mathbb{N}$, and for all messages m, it holds that

$$\Pr\left[1 = \mathsf{Ver}(\mathsf{vk}, m^*, \sigma^*) \text{ and } (m^*, \sigma^*) \neq (m, \sigma) : \begin{array}{l}(\mathsf{vk}, \mathsf{sk}) \leftarrow \mathsf{SGen}(1^\lambda);\\ \sigma \leftarrow \mathsf{Sign}(\mathsf{sk}, m);\\ (m^*, \sigma^*) \leftarrow \mathcal{A}_\lambda(\mathsf{vk}, m, \sigma)\end{array}\right] = \mathsf{negl}(\lambda).$$

Indistinguishability of Signature States. We provide a formal statement and a proof of the indistinguishability of our signature states and the corresponding classical mixture. This proof is inspired by, and closely follows, the work of [12].

Lemma 2.7. Let $(\mathsf{SGen}, \mathsf{Sign}, \mathsf{Ver})$ be an OTS scheme that satisfies strong existential unforgeability. Then the following distribution ensambles are computationally indistinguishable

$$\left\{ \frac{|0,\sigma_0\rangle + |1,\sigma_1\rangle}{\sqrt{2}} \frac{\langle 0,\sigma_0| + \langle 1,\sigma_1|}{\sqrt{2}}, (\mathsf{vk}_0, \mathsf{vk}_1) \right\} \approx \left\{ \frac{|0,\sigma_0\rangle\langle 0,\sigma_0| + |1,\sigma_1\rangle\langle 1,\sigma_1|}{2}, (\mathsf{vk}_0, \mathsf{vk}_1) \right\}$$

where $(\mathsf{vk}_b, \mathsf{sk}_b) \leftarrow \mathsf{SGen}(1^\lambda)$ and $\sigma_b \leftarrow \mathsf{Sign}(\mathsf{sk}_b, b)$, for $b \in \{0,1\}$.

Proof. By convexity,[1] it suffices to show that no QPT adversary acting on registers v and x can distinguish between the states $|\Psi_0\rangle$ and $|\Psi_1\rangle$ with non-negligible probability, where

$$|\Psi_b\rangle = \sum_{(\mathsf{vk},\mathsf{sk})} \sqrt{D(\mathsf{vk},\mathsf{sk})}\, |\mathsf{vk},\mathsf{sk}\rangle_{\mathsf{s}} |\mathsf{vk}\rangle_{\mathsf{v}} \otimes \frac{|0,\sigma_0\rangle_{\mathsf{x}} + (-1)^b |1,\sigma_1\rangle_{\mathsf{x}}}{\sqrt{2}}$$

and we adopt the following convention for the notation:

$$\mathsf{sk} = \{\mathsf{sk}_0, \mathsf{sk}_1\} \; ; \; \mathsf{vk} = \{\mathsf{vk}_0, \mathsf{vk}_1\} \; ; \; \sigma_b = \mathsf{Sign}(\mathsf{sk}_b, b) \text{ and } D(\mathsf{vk},\mathsf{sk}) = \Pr[(\mathsf{vk},\mathsf{sk}) = \mathsf{SGen}(1^\lambda)].$$

Assume towards contradiction that there exists a QPT distinguisher acting on registers v, x, as well as an auxiliary register $|\alpha_\lambda\rangle_{\mathsf{a}}$ that succeeds with probability δ. Then, by Theorem 2.1 there exists a polynomial-time computable unitary U such that

$$\frac{1}{2} \left| \begin{array}{l} \langle \Psi_1'|_{\mathsf{s},\mathsf{v},\mathsf{x}} \langle \alpha_\lambda|_{\mathsf{a}} (U_{\mathsf{v},\mathsf{x},\mathsf{a}} \otimes \mathsf{Id}_{\mathsf{s}}) |\Psi_0'\rangle_{\mathsf{s},\mathsf{v},\mathsf{x}} |\alpha_\lambda\rangle_{\mathsf{a}} \\ + \langle \Psi_0'|_{\mathsf{s},\mathsf{v},\mathsf{x}} \langle \alpha_\lambda|_{\mathsf{a}} (U_{\mathsf{v},\mathsf{x},\mathsf{a}} \otimes \mathsf{Id}_{\mathsf{s}}) |\Psi_1'\rangle_{\mathsf{s},\mathsf{v},\mathsf{x}} |\alpha_\lambda\rangle_{\mathsf{a}} \end{array} \right| = \delta$$

where

$$|\Psi_b'\rangle = \frac{|\Psi_0\rangle + (-1)^b |\Psi_1\rangle}{\sqrt{2}} = \sum_{(\mathsf{vk},\mathsf{sk})} \sqrt{D(\mathsf{vk},\mathsf{sk})}\, |\mathsf{vk},\mathsf{sk}\rangle_{\mathsf{s}} |\mathsf{vk}\rangle_{\mathsf{v}} \otimes |b,\sigma_b\rangle_{\mathsf{x}}.$$

Consequently, it must be the case that either

- $\langle \Psi_1'|_{\mathsf{s},\mathsf{v},\mathsf{x}} \langle \alpha_\lambda|_{\mathsf{a}} (U_{\mathsf{v},\mathsf{x},\mathsf{a}} \otimes \mathsf{Id}_{\mathsf{s}}) |\Psi_0'\rangle_{\mathsf{s},\mathsf{v},\mathsf{x}} |\alpha_\lambda\rangle_{\mathsf{a}} \geq \delta$, or
- $\langle \Psi_0'|_{\mathsf{s},\mathsf{v},\mathsf{x}} \langle \alpha_\lambda|_{\mathsf{a}} (U_{\mathsf{v},\mathsf{x},\mathsf{a}} \otimes \mathsf{Id}_{\mathsf{s}}) |\Psi_1'\rangle_{\mathsf{s},\mathsf{v},\mathsf{x}} |\alpha_\lambda\rangle_{\mathsf{a}} \geq \delta$.

Without loss of generality we assume that the former holds, but the argument works symmetrically also for the latter case. We will show that this leads to a contradiction with a reduction against the one-time unforgeability of OTS. In fact we will consider an even weaker definition where the adversary receives *no signature*.

On input a verification key vk_1 and an advice $|\alpha_\lambda\rangle_{\mathsf{a}}$ the reduction samples a uniform $(\mathsf{vk}_0, \mathsf{sk}_0) \leftarrow \mathsf{SGen}(1^\lambda)$, and sets $\mathsf{vk} = \{\mathsf{vk}_0, \mathsf{vk}_1\}$. Then it computes $\sigma_0 \leftarrow \mathsf{Sign}(\mathsf{sk}_0, 0)$ and

$$U_{\mathsf{v},\mathsf{x},\mathsf{a}} |\mathsf{vk}\rangle_{\mathsf{v}} |0,\sigma_0\rangle_{\mathsf{x}} |\alpha_\lambda\rangle_{\mathsf{a}}$$

[1] Note that the mixed state $\frac{|0,\sigma_0\rangle\langle 0,\sigma_0| + |1,\sigma_1\rangle\langle 1,\sigma_1|}{2}$ is a convex combination of $\frac{|0,\sigma_0\rangle + (-1)^b |1,\sigma_1\rangle}{\sqrt{2}} \frac{\langle 0,\sigma_0| + (-1)^b \langle 1,\sigma_1|}{\sqrt{2}}$ for $b \in \{0,1\}$.

and returns the result of a measurement of the x register in the computational basis.

We now analyze the success probability of the reduction in producing a valid signature, for a fixed key pair $(\mathsf{vk}_1, \mathsf{sk}_1)$. Let us denote by Σ_1 the set of all valid signatures on 1 under vk_1, and by $\sigma_1 = \mathsf{Sign}(\mathsf{sk}_1, 1)$. Then we have that the success probability of the reduction equals

$$\begin{aligned}
&\sum_{\sigma_1' \in \Sigma_1} |\Sigma_1| \cdot \left\| \langle 1, \sigma_1' |_{\mathsf{x}} U_{\mathsf{v},\mathsf{x},\mathsf{a}} |\mathsf{vk}\rangle_{\mathsf{v}} |0, \sigma_0\rangle_{\mathsf{x}} |\alpha_\lambda\rangle_{\mathsf{a}} \right\|^2 \\
&\qquad \geq \| \langle 1, \sigma_1 |_{\mathsf{x}} U_{\mathsf{v},\mathsf{x},\mathsf{a}} |\mathsf{vk}\rangle_{\mathsf{v}} |0, \sigma_0\rangle_{\mathsf{x}} |\alpha_\lambda\rangle_{\mathsf{a}} \|^2 \\
&\qquad \geq |\langle \mathsf{vk}|_{\mathsf{v}} \langle 1, \sigma_1 |_{\mathsf{x}} \langle \alpha_\lambda |_{\mathsf{a}} U_{\mathsf{v},\mathsf{x},\mathsf{a}} |\mathsf{vk}\rangle_{\mathsf{v}} |0, \sigma_0\rangle_{\mathsf{x}} |\alpha_\lambda\rangle_{\mathsf{a}}|^2 \\
&\qquad = |\langle \mathsf{vk}, \mathsf{sk}|_{\mathsf{s}} \langle \mathsf{vk}|_{\mathsf{v}} \langle 1, \sigma_1 |_{\mathsf{x}} \langle \alpha_\lambda |_{\mathsf{a}} (U_{\mathsf{v},\mathsf{x},\mathsf{a}} |\mathsf{vk}\rangle_{\mathsf{v}} \otimes \mathsf{Id}_{\mathsf{s}}) |\mathsf{vk}, \mathsf{sk}\rangle_{\mathsf{s}} |0, \sigma_0\rangle_{\mathsf{x}} |\alpha_\lambda\rangle_{\mathsf{a}}|^2
\end{aligned}$$

where the second inequality follows from the fact that inserting $\langle \mathsf{vk}|_{\mathsf{v}}$ and $\langle \alpha_\lambda|_{\mathsf{a}}$ can only decrease the norm. Now, over the random choice of $(\mathsf{vk}, \mathsf{sk})$ the success probability of the reduction can be lower bounded by

$$\begin{aligned}
&\mathbb{E}_{(\mathsf{vk},\mathsf{sk})} \left[|\langle \mathsf{vk}, \mathsf{sk}|_{\mathsf{s}} \langle \mathsf{vk}|_{\mathsf{v}} \langle 1, \sigma_1 |_{\mathsf{x}} \langle \alpha_\lambda |_{\mathsf{a}} (U_{\mathsf{v},\mathsf{x},\mathsf{a}} |\mathsf{vk}\rangle_{\mathsf{v}} \otimes \mathsf{Id}_{\mathsf{s}}) |\mathsf{vk}, \mathsf{sk}\rangle_{\mathsf{s}} |0, \sigma_0\rangle_{\mathsf{x}} |\alpha_\lambda\rangle_{\mathsf{a}}|^2 \right] \\
&\geq \left| \mathbb{E}_{(\mathsf{vk},\mathsf{sk})} \left[\langle \mathsf{vk}, \mathsf{sk}|_{\mathsf{s}} \langle \mathsf{vk}|_{\mathsf{v}} \langle 1, \sigma_1 |_{\mathsf{x}} \langle \alpha_\lambda |_{\mathsf{a}} (U_{\mathsf{v},\mathsf{x},\mathsf{a}} |\mathsf{vk}\rangle_{\mathsf{v}} \otimes \mathsf{Id}_{\mathsf{s}}) |\mathsf{vk}, \mathsf{sk}\rangle_{\mathsf{s}} |0, \sigma_0\rangle_{\mathsf{x}} |\alpha_\lambda\rangle_{\mathsf{a}} \right] \right|^2 \\
&\geq \delta
\end{aligned}$$

where the first inequality follows from Jensen's inequality. This contradicts the unforgeability of OTS and concludes our proof. □

3 Quantum Public-Key Encryption

In the following we define and construct the central cryptographic primitive of this work, which we refer to as *quantum-public-key encryption*.

3.1 Definitions

The syntax for this primitive is taken almost in verbatim from [3], although in this work we consider a stronger notion of security. For notational convenience, we define the primitive for encrypting one-bit messages, but it is easy to generalize the notion and the corresponding construction to multiple bits, via the standard bit-by-bit encryption. Security of the multi-bit construction follows by a standard hybrid argument.

Definition 3.1 (QPKE)**.** A *quantum-public-key encryption (PKE)* scheme is defined as a tuple of algorithms $(\mathsf{SKGen}, \mathsf{PKGen}, \mathsf{Enc}, \mathsf{Dec})$ such that:

- $\mathsf{sk} \leftarrow \mathsf{SKGen}(1^\lambda)$: A PPT algorithm which, on input the security parameter 1^λ outputs a secret bit string sk.
- $(\rho, \mathsf{pk}) \leftarrow \mathsf{PKGen}(\mathsf{sk})$: A QPT algorithm which, on input the secret key sk, outputs a (possibly mixed) quantum state ρ and a bit strings pk.
- $\mathsf{ct} \leftarrow \mathsf{Enc}(\rho, \mathsf{pk}, m)$: A QPT algorithm which, on input the public key (ρ, pk) and a message $m \in \{0,1\}$, outputs a ciphertext ct.

- $m \leftarrow \mathsf{Dec}(\mathsf{sk}, \mathsf{ct})$: A QPT algorithm which, on input the secret key sk and the ciphertext ct, outputs a message $m \in \{0,1\}$.

A QPKE scheme $(\mathsf{SKGen}, \mathsf{PKGen}, \mathsf{Enc}, \mathsf{Dec})$ satisfies *correctness* if for all $\lambda \in \mathbb{N}$ and all $m \in \{0,1\}$ it holds that:

$$\Pr\left[m = \mathsf{Dec}(\mathsf{sk}, \mathsf{ct}) : \mathsf{sk} \leftarrow \mathsf{SKGen}(1^\lambda); (\rho, \mathsf{pk}) \leftarrow \mathsf{PKGen}(\mathsf{sk}); \mathsf{ct} \leftarrow \mathsf{Enc}(\rho, \mathsf{pk}, m)\right] = 1.$$

Everlasting Security. Next, we define the security notion of *everlasting security* for QPKE. Informally, we require that the message is unconditionally hidden from the eyes of an attacker, even if a QPT attacker is allowed to tamper with the public key arbitrarily. However, the attacker is supplied a *single copy* of the public key.

Definition 3.2 (Everlasting Security)**.** For a family of QPT algorithms $\{\mathcal{A}_\lambda\}_{\lambda\in\mathbb{N}}$, we define the experiment $\mathsf{Exp}^{\mathcal{A}_\lambda}(1^\lambda, m)$ as follows:

1. Sample $\mathsf{sk} \leftarrow \mathsf{SKGen}(1^\lambda)$ and $(\rho, \mathsf{pk}) \leftarrow \mathsf{PKGen}(\mathsf{sk})$ and send the corresponding public key (ρ, pk) to $\mathcal{A}_\lambda$.
2. $\mathcal{A}_\lambda$ returns two quantum registers. The first register is parsed as the modified public-key register, whereas the second register is arbitrary and will be referred to as the adversary's internal register.
3. Compute ct by applying the map defined by $\mathsf{Enc}(\cdot, \mathsf{pk}, m)$ to the public-key register returned by the adversary in the previous round.
4. The output of the experiment is defined to be the joint state of ct and the internal register of the adversary.

Then we say that a QPKE scheme $(\mathsf{SKGen}, \mathsf{PKGen}, \mathsf{Enc}, \mathsf{Dec})$ satisfies *everlasting security* if there exists a negligible function negl such that for all $\lambda \in \mathbb{N}$ and all QPT $\mathcal{A}_\lambda$ it holds that

$$\mathsf{Td}\left(\mathsf{Exp}^{\mathcal{A}_\lambda}(1^\lambda, 0), \mathsf{Exp}^{\mathcal{A}_\lambda}(1^\lambda, 1)\right) = \mathsf{negl}(\lambda).$$

Let us comment on the definition as stated above. First, we remark that the definition can be easily extended to the case of multi-bit messages, provided that the syntax of the encryption scheme is extended accordingly. We also mention that an alternative definition might also allow the adversary to do some arbitrary post-processing on the output of the experiment. However, our definition is equivalent (and arguably simpler) by the monotonicity of the trace distance.

An important point of our definition (which distinguishes it from prior works) is that the attacker is allowed to modify the quantum states arbitrarily, although it cannot tamper with the classical information (such as pk or ct). This models the presence of *authenticated classical channels*, which are assumed to deliver the classical information faithfully. Note that the same assumption is also present (although somewhat more implicitly) for the standard notion of *classical* PKE, where the encryption algorithm in the CPA/CCA-security experiment is always provided as input the correct public key sampled by the challenger. This restriction is of course necessary, since if the attacker is allowed to choose the public key arbitrarily, then the definition would be impossible to achieve.

Finally, we mention that a stronger definition would allow the adversary to see a polynomial number of copies of the public key, instead of a single one. Unfortunately the work of [3] shows a key

recovery attack against any QPKE, if the attacker is given sufficiently many copies of the public key and it is allowed to run in unbounded time. This immediately rules out any QPKE with everlasting security in the presence of polynomial copies of the public key, since an unbounded distinguisher can simply run such key recovery algorithm, and decrypt the challenge ciphertext using the honest decryption algorithm. To overcome this limitation, we define the notion of computational security.

Computational Security. We define the weaker notion of computational security for QPKE, where the message is only required to be kept hidden against computationally bounded adversary. The upshot is that this can hold even if the adversary is given access to multiple copies of the public key. We present a formal definition below.

Definition 3.3 (Computational Security)**.** For a family of QPT algorithms $\{\mathcal{A}_\lambda\}_{\lambda\in\mathbb{N}}$, we define the experiment $\mathsf{Exp}^{\mathcal{A}_\lambda}(1^\lambda, m, n)$ as follows:

1. Sample $\mathsf{sk} \leftarrow \mathsf{SKGen}(1^\lambda)$ and $\{(\rho_i, \mathsf{pk}_i) \leftarrow \mathsf{PKGen}(\mathsf{sk})\}_{i=1,\dots,n}$ and send the corresponding public keys (ρ_i, pk_i) to $\mathcal{A}_\lambda$.
2. $\mathcal{A}_\lambda$ returns two quantum registers. The first register is parsed as the modified public-key register, whereas the second register is arbitrary and will be referred to as the adversary's internal register.
3. Compute ct by applying the map defined by $\mathsf{Enc}(\cdot, \mathsf{pk}_1, m)$ to the public-key register returned by the adversary in the previous round.
4. The output of the experiment is defined to be the joint state of ct and the internal register of the adversary.

Then we say that a QPKE scheme $(\mathsf{SKGen}, \mathsf{PKGen}, \mathsf{Enc}, \mathsf{Dec})$ satisfies *computational security* if there exists a negligible function negl such that for all $\lambda \in \mathbb{N}$, all polynomials $n = n(\lambda)$, and all QPT $\mathcal{A}_\lambda$ it holds that the distributions

$$\mathsf{Exp}^{\mathcal{A}_\lambda}(1^\lambda, 0, n) \approx \mathsf{Exp}^{\mathcal{A}_\lambda}(1^\lambda, 1, n)$$

are computationally indistinguishable.

3.2 Everlasting Secure QPKE

We describe our scheme below. As the only computational ingredient, we assume the existence of a quantum-secure strongly existentially unforgeable one-time signature scheme $(\mathsf{SGen}, \mathsf{Sign}, \mathsf{Ver})$, see Section 2.4. As discussed, this can be constructed from any quantum-secure one-way function.

- $\mathsf{SKGen}(1^\lambda)$:
  - Sample two key pairs $(\mathsf{sk}_0, \mathsf{vk}_0) \leftarrow \mathsf{SGen}(1^\lambda)$ and $(\mathsf{sk}_1, \mathsf{vk}_1) \leftarrow \mathsf{SGen}(1^\lambda)$.
  - Compute $\sigma_0 \leftarrow \mathsf{Sign}(\mathsf{sk}_0, 0)$ and $\sigma_1 \leftarrow \mathsf{Sign}(\mathsf{sk}_1, 1)$.
  - Sample a bit $d_0 \leftarrow \{0,1\}$.
  - Return $\mathsf{sk} = (\mathsf{vk}_0, \mathsf{vk}_1, \sigma_0, \sigma_1, d_0)$.
- $\mathsf{PKGen}(\mathsf{sk})$:

– Define the state

$$|\Psi\rangle = \frac{|0,\sigma_0\rangle + (-1)^{d_0}\,|1,\sigma_1\rangle}{\sqrt{2}}.$$

This state is efficiently computable by preparing an EPR pair and CNOT-ing the bits of the signatures into an auxiliary register, controlled on the value of the first qubit. The relative phase can be then added by a controlling the application of Z with d_0.

– Set the quantum part of the public key ρ to be the state $|\Psi\rangle$ and set the classical part of the public key and the classical secret key to $\mathsf{pk} = (\mathsf{vk}_0, \mathsf{vk}_1)$.

- $\mathsf{Enc}(\rho, \mathsf{pk}, m)$:

  – Project ρ onto the subspace of valid signatures of 0 and 1, under vk_0 and vk_1, respectively. More precisely, denote by Σ_0 and Σ_1 the set of accepting signatures on 0 and 1, under vk_0 and vk_1, respectively, and consider the projector

  $$\Pi = \sum_{\sigma\in\Sigma_0} |0,\sigma\rangle\langle 0,\sigma| + \sum_{\sigma\in\Sigma_1} |1,\sigma\rangle\langle 1,\sigma|\,.$$

  Apply the projective measurement $\{\Pi, \mathsf{Id} - \Pi\}$, and abort the execution (return $\perp$) if the measurement returns the second outcome. Note that this measurement can be implemented efficiently by running the verification algorithm coherently, CNOT-ing the output qubit onto a separate register, and measuring this register.

  – Measure the residual state in the Hadamard basis, to obtain a bit string (d_1, d_2), where we denote by $d_1 \in \{0,1\}$ the first bit of the measurement outcome and by d_2 the rest.

  – Return the following as the classical ciphertext:

  $$\mathsf{ct} = (m \oplus d_1, d_2). \tag{3.1}$$

- $\mathsf{Dec}(\mathsf{sk}, \mathsf{ct})$:

  – Parse $\mathsf{ct} = (\mathsf{ct}_1, \mathsf{ct}_2)$, where $\mathsf{ct}_1 \in \{0,1\}$ is one bit, and return

  $$m = d_0 \oplus \mathsf{ct}_1 \oplus \mathsf{ct}_2 \cdot (\sigma_0 \oplus \sigma_1). \tag{3.2}$$

Analysis. We claim that the scheme satisfies correctness. First, observe that the state ρ taken as input by the encryption algorithm is in the image of the projector Π as defined above. Consequently, applying the projective measurement $\{\Pi, \mathsf{Id} - \Pi\}$ returns the outcome associated with Π with certainty and does not change the state. Applying the Hadamard transformation to the state $|\Psi\rangle$ gives

$$\mathsf{H}\,|\Psi\rangle \propto \sum_{d_1,d_2} (-1)^{(d_1,d_2)\cdot(0,\sigma_0)}\,|d_1,d_2\rangle + (-1)^{d_0\oplus(d_1,d_2)\cdot(1,\sigma_1)}\,|d_1,d_2\rangle = \sum_{d_1,d_2\,:\,(d_1,d_2)\cdot(1,\sigma_0\oplus\sigma_1)=d_0} |d_1,d_2\rangle\,,$$

omitting overall normalization factors. Therefore, a measurement returns a uniformly random bit string (d_1, d_2) satisfying

$$d_1 \oplus d_2 \cdot (\sigma_0 \oplus \sigma_1) = d_0.$$

Substituting Eq. (3.1) in Eq. (3.2) and using this relation, we obtain

$$d_0 \oplus \mathsf{ct}_1 \oplus \mathsf{ct}_2 \cdot (\sigma_0 \oplus \sigma_1) = d_0 \oplus (m \oplus d_1) \oplus d_2 \cdot (\sigma_0 \oplus \sigma_1) = m,$$

as desired. Next, we show that the scheme satisfies everlasting security.

Theorem 3.4 (Everlasting security)**.** If quantum-secure one-way functions exist, then the QPKE $(\mathsf{SKGen}, \mathsf{PKGen}, \mathsf{Enc}, \mathsf{Dec})$ satisfies everlasting security.

Proof. We proceed by defining a series of hybrid experiments that we show to be indistinguishable from the eyes of any (possibly unbounded) algorithm. For convenience, we define

$$\mathsf{Adv}(i) = \mathsf{Td}\left(\mathsf{Hyb}_i^{\mathcal{A}_\lambda}(1^\lambda, 0), \mathsf{Hyb}_i^{\mathcal{A}_\lambda}(1^\lambda, 1)\right).$$

- $\mathsf{Hyb}_0^{\mathcal{A}_\lambda}(1^\lambda, b)$: This is the original experiment $\mathsf{Exp}^{\mathcal{A}_\lambda}(1^\lambda, b)$, as defined in Definition 3.2.
- $\mathsf{Hyb}_1^{\mathcal{A}_\lambda}(1^\lambda, b)$: In this experiment, we modify the PKGen algorithm to measure the state $|\Psi\rangle$ in the computational basis, before outputting ρ.

Since the result sk of the SKGen algorithm is not used in the experiment, we only need to argue that the reduced states of (pk, ρ) are unchanged by this modification. This is indeed the case, since adding a random phase is equivalent to measuring in the computational basis by a standard Pauli Z-twirl argument. Thus the two experiments are identical from the perspective of the adversary and therefore $\mathsf{Adv}(0) = \mathsf{Adv}(1)$.

- $\mathsf{Hyb}_2^{\mathcal{A}_\lambda}(1^\lambda, b)$: In this experiment, we further modify the PKGen algorithm to sample the state ρ as follows. Flip a random coin $c \leftarrow \{0, 1\}$. If $c = 0$ then return $|0, \sigma_0\rangle\langle 0, \sigma_0|$, and otherwise return $|1, \sigma_1\rangle\langle 1, \sigma_1|$.

Observe that the state ρ returned by the modified PKGen algorithm is the classical mixture

$$\rho = \frac{|0, \sigma_0\rangle\langle 0, \sigma_0| + |1, \sigma_1\rangle\langle 1, \sigma_1|}{2}$$

which is identical to the state returned in the previous hybrid. Therefore $\mathsf{Adv}(1) = \mathsf{Adv}(2)$. Next, let us denote by ρ^* the reduced density matrix of the modified public-key register returned by the adversary in step 2 of the experiment. We will assume that the state ρ^* returned by the adversary is such that the encryption algorithm accepts (i.e., does not abort) with non-negligible probability (for otherwise $\mathsf{Adv}(2) = \mathsf{negl}(\lambda)$ and we are done). We can then establish that the state ρ^*, conditioned on the verification succeeding, must be negligibly close in trace distance from the state produced by the PKGen algorithm. More precisely:

Claim 3.5. There exists a negligible function negl such that, for any $c \in \{0, 1\}$ and for ρ^* the state returned by the adversary conditional on the result of the coin toss in the Gen algorithm being equal to c,

$$\mathsf{Td}\left(\frac{\Pi\rho^*\Pi}{\mathsf{Tr}\left(\Pi\rho^*\right)}, |c, \sigma_c\rangle\langle c, \sigma_c|\right) = \mathsf{negl}(\lambda).$$

Proof of Claim 3.5. The proof follows by a reduction to the unforgeability of the OTS. Indeed, assume for sake of contradiction that the post-measurement state is non-negligibly far from $|c, \sigma_c\rangle$ in trace distance. By Eq. (2.1), the latter is equivalent to saying that if we measure in the computational basis then the probability of obtaining outcome (c, σ_c) is non-negligibly smaller than 1. Since the post-measurement state is supported on range of Π, it follows that if we measure in the computational basis then we must with non-negligible probability obtain an outcome (z, σ) such that

$$(z, \sigma) \neq (c, \sigma_c) \quad \text{and} \quad \mathsf{Ver}(\mathsf{vk}_z, z, \sigma) = 1,$$

that is, $(z, \sigma) \neq (c, \sigma_c)$ is a valid message-signature pair. As the adversary along with the projective measurement $\{\Pi, I - \Pi\}$ and the standard basis measurement run in quantum polynomial time, and the projective measurement returns Π with non-negligible probability, this contradicts the strong existential unforgeability of the OTS scheme. □

We conclude by establishing that the message m is statistically hidden in the last experiment.

Claim 3.6. There exists a negligible function negl such that $\mathsf{Adv}(2) = \mathsf{negl}(\lambda)$.

Proof of Claim 3.6. By Claim 3.5, the post-measurement state is negligibly close to the state $|c, \sigma_c\rangle$. As the latter is a pure state and extensions of pure states are always in tensor product, by Uhlmann's theorem it follows that the post-measurement register is negligibly close to being in tensor product with the internal register of the adversary. Thus it suffices to show that the output distribution of the Enc algorithm does not depend on the message m when given as input $|c, \sigma_c\rangle$. To see this, observe that the rotated state in the Hadamard basis is up to overall normalization given by

$$\mathsf{H} |c, \sigma_c\rangle \propto \sum_{d'} (-1)^{d' \cdot (c, \sigma_c)} \left| d' \right\rangle,$$

and therefore a measurement returns a uniformly random bit string $d' = (d_1, d_2)$. In particular, $\mathsf{ct} = (m \oplus d_1, d_2)$ has the same distribution for $m \in \{0, 1\}$. As explained above, it is also negligibly close to being independent from the internal register of the adversary, and thus the claim follows. □

By Claim 3.6 we have that

$$\mathsf{Adv}(0) = \mathsf{Adv}(1) = \mathsf{Adv}(2) = \mathsf{negl}(\lambda),$$

and this concludes the proof of Theorem 3.4. □

3.3 Computationally Secure QPKE

Our scheme assumes the existence of a quantum-secure strongly existentially unforgeable one-time signature scheme $(\mathsf{SGen}, \mathsf{Sign}, \mathsf{Ver})$ and a quantum-secure pseudorandom function PRF. Both such building blocks can be constructed assuming any quantum-secure one-way function.

- $\mathsf{SKGen}(1^\lambda)$:
  - Sample a key $k \leftarrow \{0, 1\}^\lambda$ and set $\mathsf{sk} = k$.
- $\mathsf{PKGen}(\mathsf{sk})$:

– Sample two uniform $(r_0, r_1) \leftarrow \{0,1\}^\lambda$ and compute

$$(\mathsf{sk}_0, \mathsf{vk}_0) \leftarrow \mathsf{SGen}(1^\lambda; \mathsf{PRF}(k, r_0)) \quad \text{and} \quad (\mathsf{sk}_1, \mathsf{vk}_1) \leftarrow \mathsf{SGen}(1^\lambda; \mathsf{PRF}(k, r_1)).$$

– Compute $\sigma_0 \leftarrow \mathsf{Sign}(\mathsf{sk}_0, 0)$ and $\sigma_1 \leftarrow \mathsf{Sign}(\mathsf{sk}_1, 1)$.

– Define the state

$$|\Psi\rangle = \frac{|0, \sigma_0\rangle + |1, \sigma_1\rangle}{\sqrt{2}}.$$

This state is efficiently computable by preparing an EPR pair and CNOT-ing the bits of the signatures into an auxiliary register, controlled on the value of the first qubit.

– Set the quantum part of the public key ρ to be the state $|\Psi\rangle$ and set the classical part of the public key and the classical secret key to $\mathsf{pk} = (\mathsf{vk}_0, \mathsf{vk}_1, r_0, r_1)$.

- $\mathsf{Enc}(\rho, \mathsf{pk}, m)$:

  – Project ρ onto the subspace of valid signatures of 0 and 1, under vk_0 and vk_1, respectively. More precisely, denote by Σ_0 and Σ_1 the set of accepting signatures on 0 and 1, under vk_0 and vk_1, respectively, and consider the projector

$$\Pi = \sum_{\sigma \in \Sigma_0} |0, \sigma\rangle\langle 0, \sigma| + \sum_{\sigma \in \Sigma_1} |1, \sigma\rangle\langle 1, \sigma| .$$

  Apply the projective measurement $\{\Pi, \mathsf{Id} - \Pi\}$, and abort the execution (return $\perp$) if the measurement returns the second outcome. Note that this measurement can be implemented efficiently by running the verification algorithm coherently, CNOT-ing the output qubit onto a separate register, and measuring this register.

  – Apply the Z^m operator to the first qubit of ρ, classically controlled on the message m.

  – Set ct to be the residual state, along with (r_0, r_1).

- $\mathsf{Dec}(\mathsf{sk}, \mathsf{ct})$:

  – Use the secret key k to recompute

$$(\mathsf{sk}_0, \mathsf{vk}_0) \leftarrow \mathsf{SGen}(1^\lambda; \mathsf{PRF}(k, r_0)) \quad \text{and} \quad (\mathsf{sk}_1, \mathsf{vk}_1) \leftarrow \mathsf{SGen}(1^\lambda; \mathsf{PRF}(k, r_1)).$$

  along with $\sigma_0 \leftarrow \mathsf{Sign}(\mathsf{sk}_0, 0)$ and $\sigma_1 \leftarrow \mathsf{Sign}(\mathsf{sk}_1, 1)$.

  – Measure the quantum state of ct in the

$$\left\{ \frac{|0, \sigma_0\rangle + |1, \sigma_1\rangle}{\sqrt{2}}, \frac{|0, \sigma_0\rangle - |1, \sigma_1\rangle}{\sqrt{2}} \right\}$$

  basis. And return the corresponding outcome.

Analysis. To see why the scheme satisfies correctness, first observe that the state $|\Psi\rangle$ as defined in the PKGen algorithm lies in the image of the projector Π. Therefore the projective measurement $\{\Pi, \mathsf{Id} - \Pi\}$ acts as the identity on $|\Psi\rangle$. Then, the state output by the encryption algorithm corresponds to

$$(\mathsf{Z}^m \otimes \mathsf{Id})\frac{|0,\sigma_0\rangle + |1,\sigma_1\rangle}{\sqrt{2}} = \frac{|0,\sigma_0\rangle + (-1)^m\,|1,\sigma_1\rangle}{\sqrt{2}}.$$

Therefore, the output of the decryption algorithm equals m with certainty. Next, we show that the scheme is computationally secure.

Theorem 3.7 (Computational security)**.** If quantum-secure one-way functions exist, then the QPKE $(\mathsf{SKGen}, \mathsf{PKGen}, \mathsf{Enc}, \mathsf{Dec})$ satisfies computational security.

Proof. We proceed by defining a series of hybrid experiments that we show to be indistinguishable from the eyes of any QPT algorithm.

- $\mathsf{Hyb}_0^{\mathcal{A}_\lambda}(1^\lambda, b)$: This is the original experiment $\mathsf{Exp}^{\mathcal{A}_\lambda}(1^\lambda, b, n)$, as defined in Definition 3.3.
- $\mathsf{Hyb}_1^{\mathcal{A}_\lambda}(1^\lambda, b)$: In this experiment, we simulate the output of the PRF by lazy sampling, i.e., every time that the PKGen algorithm is invoked, the experiment sample a uniform tuple $(r_0, r_1, \tilde{r}_0, \tilde{r}_1)$ and uses the latter pair as the randomness for the OTS scheme. To keep things consistent, the experiment maintains a list of all such tuples.

Note that the only difference between these two hybrids is in the way the random coins of the OTS are sampled. Therefore, by the pseudorandomness of PRF, the two hybrids are computationally indistinguishable. Note that, in the second hybrid, different copies of the public key are now independent from each other.

- $\mathsf{Hyb}_2^{\mathcal{A}_\lambda}(1^\lambda, b)$: In this experiment, we further modify the *first invocation* of the PKGen algorithm to sample the state ρ as follows. Flip a random coin $c \leftarrow \{0,1\}$. If $c = 0$ then return $|0,\sigma_0\rangle\langle 0,\sigma_0|$, and otherwise return $|1,\sigma_1\rangle\langle 1,\sigma_1|$.

Observe that the state ρ returned by the modified PKGen algorithm is the classical mixture

$$\rho = \frac{|0,\sigma_0\rangle\langle 0,\sigma_0| + |1,\sigma_1\rangle\langle 1,\sigma_1|}{2}.$$

By a direct application of Lemma 2.7, we can conclude that the two hybrids are computationally indistinguishable. At this point, we can appeal to Claim 3.5 (in the proof of Theorem 3.4) to establish that the state ρ^*, as returned by the adversary in the security experiment, must be within negligible trace distance from a basis state. The proof is concluded by noting that the distributions of an encryption of 0 and an encryption of 1 are identical, up to a global phase, if the algorithm is called on input any basis state. □

4 Two-Message Quantum Key Distribution

In the following we outline how to use a QPKE scheme $(\mathsf{SKGen}, \mathsf{PKGen}, \mathsf{Enc}, \mathsf{Dec})$ to construct a QKD protocol with a minimal number of two rounds of interaction, as announced in the introduction.

4.1 Definitions

We give a formal definition of quantum key distribution in the everlasting settings, i.e., where an attacker is required to be computationally bounded only during the execution of the protocol. For convenience, we adopt a syntax specific for two-message protocols, but the definitions can be extended to the more general interactive settings canonically.

Definition 4.1 (Two-Message QKD)**.** A *quantum key distribution (QKD)* scheme is defined as a tuple of algorithms $(\mathsf{QKDFirst}, \mathsf{QKDSecond}, \mathsf{QKDDecode})$ such that:

- $(\mathsf{msg}, \mu, \mathsf{st}) \leftarrow \mathsf{QKDFirst}(1^\lambda)$: A QPT algorithm which, on input the security parameter 1^λ outputs a message, consisting of a classical component msg and a (possibly mixed) quantum state μ, and an internal state st.
- $\{(\mathsf{resp}, \eta, k), \bot\} \leftarrow \mathsf{QKDSecond}(\mathsf{msg}, \mu)$: A QPT algorithm which, on input the first message (msg, μ), outputs a response, consisting of a classical component resp and a (possibly mixed) quantum state η, along with a key $k \in \{0,1\}^\lambda$, or a distinguished symbol $\bot$, denoting rejection.
- $\{k, \bot\} \leftarrow \mathsf{QKDDecode}(\mathsf{st}, \mathsf{resp}, \eta)$: A QPT algorithm which, on input the internal state st, and the response (resp, η), returns a key $k \in \{0,1\}^\lambda$ or a distinguished symbol $\bot$, denoting rejection.

We say that a QKD scheme $(\mathsf{QKDFirst}, \mathsf{QKDSecond}, \mathsf{QKDDecode})$ satisfies *correctness* if for all $\lambda \in \mathbb{N}$ it holds that:

$$\Pr\left[\bot = \mathsf{QKDSecond}(\mathsf{msg}, \mu) : (\mathsf{msg}, \mu, \mathsf{st}) \leftarrow \mathsf{QKDFirst}(1^\lambda)\right] = 0$$

and

$$\Pr\left[k = \mathsf{QKDDecode}(\mathsf{st}, \mathsf{resp}, \eta) : \begin{array}{l}(\mathsf{msg}, \mu, \mathsf{st}) \leftarrow \mathsf{QKDFirst}(1^\lambda); \\ (\mathsf{resp}, \eta, k) \leftarrow \mathsf{QKDSecond}(\mathsf{msg}, \mu)\end{array}\right] = 1.$$

Everlasting Security. Next, we define the security notion of *everlasting security* for QKD, which consists of two properties. Privacy requires that the key k should be hidden unconditionally in the presence of an adversary that is computationally bounded during the execution of the protocol. In addition, as standard for QKD, we assume the existence of an authenticated classical channel, which is modeled by not allowing the adversary to tamper with the classical messages. On the other hand, verifiability requires that no computationally bounded adversary should be able to cause Alice and Bob to disagree on the key, without any of them noticing.

Definition 4.2 (Everlasting Security)**.** For a family of QPT algorithms $\{\mathcal{A}_\lambda\}_{\lambda\in\mathbb{N}}$, we define the experiment $\mathsf{QKDSec}^{\mathcal{A}_\lambda}(1^\lambda)$ as follows:

1. Sample $(\mathsf{msg}, \mu, \mathsf{st}) \leftarrow \mathsf{QKDFirst}(1^\lambda)$ and send the corresponding first message (msg, μ) to $\mathcal{A}_\lambda$.
2. $\mathcal{A}_\lambda$ returns two quantum registers. The first register is parsed as the modified first message, whereas the second register is arbitrary and will be referred to as the adversary's internal register.
3. Compute $\{(\mathsf{resp}, \eta, k_0), \bot\}$ by applying the map defined by $\mathsf{QKDSecond}(\mathsf{msg}, \cdot)$ to the modified first message register returned by the adversary in the previous round.

(a) If the above message is $\perp$, set $(k_0, k_1) = (\perp, \perp)$ and conclude the experiment.

(b) Otherwise, return (resp, η) to the adversary, along with its internal state.

4. $\mathcal{A}_\lambda$ returns once again two quantum registers. The first register is parsed as the modified response, whereas the second register is the adversary's internal register.

5. Compute $\{k_1, \perp\}$ by applying the map defined by $\mathsf{QKDDecode}(\mathsf{st}, \mathsf{resp}, \cdot)$ to the modified response register returned by the adversary in the previous round. If the result is $\perp$, then set $k_1 = \perp$.

6. The output of the experiment is defined to be the internal register of the adversary.

Then we say that a QKD scheme $(\mathsf{QKDFirst}, \mathsf{QKDSecond}, \mathsf{QKDDecode})$ satisfies *everlasting security* if the following properties hold.

- (Privacy) There exists a negligible function negl such that for all $\lambda \in \mathbb{N}$ and all QPT $\mathcal{A}_\lambda$ it holds that

$$\mathsf{Td}\left(\left\{\mathsf{QKDSec}^{\mathcal{A}_\lambda}(1^\lambda), k_0, k_1\right\}, \left\{\mathsf{QKDSec}^{\mathcal{A}_\lambda}(1^\lambda), \tilde{k}_0, \tilde{k}_1\right\}\right) = \mathsf{negl}(\lambda)$$

  where the variables k_0 and k_1 are defined in the experiment, whereas $\tilde{k}_b$, for $b \in \{0,1\}$, is defined as

$$\begin{cases} \tilde{k}_b = \perp & \text{if } k_b = \perp \\ \tilde{k}_b \leftarrow \{0,1\}^\lambda & \text{otherwise.} \end{cases}$$

- (Verifiability) There exists a negligible function negl such that for all $\lambda \in \mathbb{N}$ and all QPT $\mathcal{A}_\lambda$ it holds that

$$\Pr[k_0 \neq k_1 \text{ and } k_1 \neq \perp] = \mathsf{negl}(\lambda)$$

  where k_0 and k_1 are defined in the experiment.

Note that the above definition of verifiability is tight for two message protocols: An adversary can easily cause a disagreement between the keys by doing nothing on the first round, and blocking the second message. In this case k_0 would be a valid key (by correctness), whereas k_1 would be set to $\perp$, since the second message was never delivered.

4.2 Two-Message QKD from QPKE

We are now ready to describe our QKD protocol. Our ingredients are a QPKE scheme $(\mathsf{SKGen}, \mathsf{PKGen}, \mathsf{Enc}, \mathsf{Dec})$ and a OTS scheme $(\mathsf{SGen}, \mathsf{Sign}, \mathsf{Ver})$, which can be both constructed from one-way functions. Additionally, we will use a universal hash function family

$$\mathsf{Hash} : \{0,1\}^{4\lambda} \to \{0,1\}^\lambda$$

which exist unconditionally. For convenience, we denote by $s(\lambda)$ the size of a signature for a message of size λ. We present the protocol below.

- $\mathsf{QKDFirst}(1^\lambda)$:

– For all $i = 1, \ldots, 4\lambda + s(4\lambda)$ sample a QPK key pair

$$\mathsf{sk}_i \leftarrow \mathsf{SKGen}(1^\lambda) \quad \text{and} \quad (\rho_i, \mathsf{pk}_i) \leftarrow \mathsf{PKGen}(\mathsf{sk}_i).$$

– Set the first message to $\mathsf{msg} = (\mathsf{pk}_1, \ldots, \mathsf{pk}_{4\lambda+s(4\lambda)})$ and $\mu = \rho_1 \otimes \cdots \otimes \rho_{4\lambda+s(4\lambda)}$.

- $\mathsf{QKDSecond}(\mathsf{msg}, \mu)$:
  – Sample a OTS key pair $(\mathsf{vk}, \mathsf{zk}) \leftarrow \mathsf{SGen}(1^\lambda)$, a key $k \leftarrow \{0,1\}^{4\lambda}$, and a universal hash function Hash.
  – Compute $\sigma \leftarrow \mathsf{Sign}(\mathsf{zk}, k)$.
  – For all $i = 1, \ldots, 4\lambda + s(4\lambda)$ compute

$$\{\mathsf{ct}_i \leftarrow \mathsf{Enc}(\rho_i, \mathsf{pk}_i, k_i)\}_{i \le 4\lambda} \quad \text{and} \quad \{\mathsf{ct}_i \leftarrow \mathsf{Enc}(\rho_i, \mathsf{pk}_i, \sigma_i)\}_{i > 4\lambda}$$

  where $k = (k_1, \ldots, k_{4\lambda})$ and $\sigma = (\sigma_1, \ldots, \sigma_{s(4\lambda)})$.
  – If any of the encryption procedures fails, return $\perp$.
  – Else, set the response as $\mathsf{resp} = (\mathsf{Hash}, \mathsf{vk}, \mathsf{ct}_1, \ldots, \mathsf{ct}_{4\lambda+s(4\lambda)})$, no quantum state is present.
  – Set the key to $K = \mathsf{Hash}(k)$.

- $\mathsf{QKDDecode}(\mathsf{st}, \mathsf{resp})$:
  – For all $i = 1, \ldots, 4\lambda + s(4\lambda)$ compute

$$\{k_i \leftarrow \mathsf{Dec}(\mathsf{sk}_i, \mathsf{ct}_i)\}_{i \le 4\lambda} \quad \text{and} \quad \{\sigma_i \leftarrow \mathsf{Dec}(\mathsf{sk}_{\lambda+i}, \mathsf{ct}_{\lambda+i})\}_{i > 4\lambda}.$$

  – If $\mathsf{Ver}(\mathsf{vk}, k, \sigma) \neq 1$ return $\perp$.
  – Else, return $K = \mathsf{Hash}(k)$.

Correctness follows immediately from the correctness of the underlying building blocks. Next we prove that the scheme is private and verifiable.

Theorem 4.3. If quantum-secure one-way functions exist, then the QKD ($\mathsf{QKDFirst}, \mathsf{QKDSecond}$, $\mathsf{QKDDecode}$) satisfies everlasting security (privacy and verifiability).

Proof. We first show that the scheme satisfies privacy. Let us change the syntax of the experiment to explicitly include the view of the adversary the flag $\mathsf{abort} \in \{0,1\}$ that denotes whether the experiment aborted or not in step 5. Then, it suffices to show that the distribution of the keys K_0 (as defined step 3 of the experiment) and K_1 (as defined in step 5 of the experiment) are statistically close to uniform, conditioned on the variables

$$\left\{\mathsf{QKDSec}^{\mathcal{A}_\lambda}(1^\lambda), \mathsf{abort}\right\}.$$

As a first step, we claim that the min-entropy of k is $H_\infty(k) \ge 3\lambda$ in the above view. To show this, we will first consider a modified distribution, where the distinguisher *is not provided* the variable abort. We then define $\mathsf{Hyb}_0^{\mathcal{A}_\lambda}(1^\lambda)$ to be the output of the original experiment $\mathsf{QKDSec}^{\mathcal{A}_\lambda}(1^\lambda)$ as defined in Definition 4.2. Then for $i = 1, \ldots, 4\lambda + s(4\lambda)$ we define the hybrid $\mathsf{Hyb}_i^{\mathcal{A}_\lambda}(1^\lambda)$ as follows.

- $\mathsf{Hyb}_i^{\mathcal{A}_\lambda}(1^\lambda)$: This is defined as the previous hybrid, except that the i-th ciphertext ct_i is computed as
$$\mathsf{ct}_i \leftarrow \mathsf{Enc}(\rho_i, \mathsf{pk}_i, 0).$$

The statistical indistinguishability of neighbouring outputs follows immediately from the everlasting security of the QPKE scheme, i.e.,

$$\mathsf{Td}\left(\mathsf{Hyb}_{i-1}^{\mathcal{A}_\lambda}(1^\lambda), \mathsf{Hyb}_i^{\mathcal{A}_\lambda}(1^\lambda)\right) = \mathsf{negl}(\lambda) \text{ for all } i = 1, \ldots, 4\lambda + s(4\lambda).$$

Note that in the last hybrid $\mathsf{Hyb}_{4\lambda+s(4\lambda)}^{\mathcal{A}_\lambda}(1^\lambda)$ the view of the adversary is formally independent from the key k and therefore k has exactly 4λ bits of entropy. An application of Lemma 2.4 already shows that K_0 is statistically close to uniform, since its distribution is independent of the event abort.

This is however not the case for K_1, since whether or not $K_1 = \perp$ depends on the event abort. Applying the same argument backwards, we lose only a negligible summand in the entropy of k, and so we can conclude that $H_\infty(k) \geq 4\lambda - 1$ in the original experiment. By Eq. (2.2) (chain rule for average-case min entropy), we have that, conditioned on the event abort, it holds that

$$\tilde{H}_\infty(k|\mathsf{abort}) \geq H_\infty(k) - 2 > 3\lambda$$

since $\mathsf{abort} \in \{0, 1\}$. By Lemma 2.4, the statistical distance of K_1 from uniform is bounded from above by $\varepsilon = 2^{-\lambda}$ since

$$\tilde{H}_\infty(k|\mathsf{abort}) - 2\log\left(\frac{1}{\varepsilon}\right) \geq 3\lambda - 2\lambda = \lambda = \ell.$$

This shows that both K_0 and K_1 are statistically close to uniform, and concludes the proof of everlasting security.

As for verifiability, let us assume towards contradiction that there exists a QPT adversary that causes a key mismatch $k_0 \neq k_1$ while not causing the decoding algorithm to reject, i.e., $k_1 \neq \perp$. Then it must be the case that the adversary is able to produce a valid message-signature pair (k_1, σ^*) under vk, for a message $k_1 \neq k_0$. Since the adversary runs in quantum polynomial time, this contradicts the unforgeability of the OTS scheme and concludes our proof. □

Acknowledgments

The authors would like to thank Khashayar Barooti for many discussion on quantum public key encryption and Takashi Yamakawa for suggesting a proof of Theorem A.1. G.M. was partially funded by the European Research Council through an ERC Starting Grant (Grant agreement No. 101077455, ObfusQation), by the German Federal Ministry of Education and Research (BMBF) in the course of the 6GEM research hub under grant number 16KISK038, and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2092 CASA – 390781972. M.W. acknowledges support by the the European Union (ERC, SYMOPTIC, 101040907), by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2092 CASA - 390781972, by the BMBF through project QuBRA, and by the Dutch Research Council (NWO grant OCENW.KLEIN.267). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

References

[1] Aaronson, S., Atia, Y., Susskind, L.: On the hardness of detecting macroscopic superpositions. Electron. Colloquium Comput. Complex. **TR20-146** (2020), https://eccc.weizmann.ac.il/report/2020/146

[2] Alagic, G., Gagliardoni, T., Majenz, C.: Can you sign a quantum state? Quantum **5**, 603 (Dec 2021). https://doi.org/10.22331/q-2021-12-16-603, https://doi.org/10.22331/q-2021-12-16-603

[3] Barooti, K., Grilo, A.B., Huguenin-Dumittan, L., Malavolta, G., Sattath, O., Vu, Q., Walter, M.: Public-key encryption with quantum keys. CoRR **abs/2306.07698** (2023). https://doi.org/10.48550/arXiv.2306.07698, https://doi.org/10.48550/arXiv.2306.07698

[4] Bartusek, J., Khurana, D.: Cryptography with certified deletion. CoRR **abs/2207.01754** (2022). https://doi.org/10.48550/arXiv.2207.01754, https://doi.org/10.48550/arXiv.2207.01754

[5] Bennett, C.H., Brassard, G.: Quantum cryptography: Public key distribution and coin tossing. In: Proceedings of IEEE International Conference on Computers, Systems, and Signal Processing. p. 175. India (1984)

[6] Diffie, W., Hellman, M.E.: New directions in cryptography. IEEE Trans. Inf. Theory **22**(6), 644–654 (1976). https://doi.org/10.1109/TIT.1976.1055638

[7] Dodis, Y., Ostrovsky, R., Reyzin, L., Smith, A.D.: Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. SIAM J. Comput. **38**(1), 97–139 (2008). https://doi.org/10.1137/060651380, https://doi.org/10.1137/060651380

[8] Dodis, Y., Reyzin, L., Smith, A.D.: Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. In: Cachin, C., Camenisch, J. (eds.) Advances in Cryptology - EUROCRYPT 2004, International Conference on the Theory and Applications of Cryptographic Techniques, Interlaken, Switzerland, May 2-6, 2004, Proceedings. Lecture Notes in Computer Science, vol. 3027, pp. 523–540. Springer (2004). https://doi.org/10.1007/978-3-540-24676-3_31, https://doi.org/10.1007/978-3-540-24676-3_31

[9] Ekert, A.K.: Quantum cryptography based on Bell's theorem. Physical review letters **67**(6), 661 (1991)

[10] Goldreich, O.: The Foundations of Cryptography - Volume 2: Basic Applications. Cambridge University Press (2004). https://doi.org/10.1017/CBO9780511721656

[11] Goldreich, O., Goldwasser, S., Micali, S.: How to construct random functions. Journal of the ACM (JACM) **33**(4), 792–807 (1986)

[12] Hhan, M., Morimae, T., Yamakawa, T.: From the hardness of detecting superpositions to cryptography: Quantum public key encryption and commitments. IACR Cryptol. ePrint Arch. p. 1375 (2022), https://eprint.iacr.org/2022/1375

[13] Impagliazzo, R.: A personal view of average-case complexity. In: Proceedings of Structure in Complexity Theory. Tenth Annual IEEE Conference. pp. 134–147. IEEE (1995)

[14] Impagliazzo, R., Rudich, S.: Limits on the provable consequences of one-way permutations. In: Johnson, D.S. (ed.) Proceedings of the 21st Annual ACM Symposium on Theory of Computing, May 14-17, 1989, Seattle, Washington, USA. pp. 44–61. ACM (1989). https://doi.org/10.1145/73007.73012, https://doi.org/10.1145/73007.73012

[15] Jouguet, P., Kunz-Jacques, S., Leverrier, A., Grangier, P., Diamanti, E.: Experimental demonstration of long-distance continuous-variable quantum key distribution. Nature Photonics **7**(5), 378–381 (2013)

[16] Kitagawa, F., Morimae, T., Nishimaki, R., Yamakawa, T.: Quantum public-key encryption with tamper-resilient public keys from one-way functions. CoRR **abs/2304.01800** (2023). https://doi.org/10.48550/arXiv.2304.01800, https://doi.org/10.48550/arXiv.2304.01800

[17] Korzh, B., Lim, C.C.W., Houlmann, R., Gisin, N., Li, M.J., Nolan, D., Sanguinetti, B., Thew, R., Zbinden, H.: Provably secure and practical quantum key distribution over 307 km of optical fibre. Nature Photonics **9**(3), 163–168 (2015)

[18] Lo, H.K., Chau, H.F.: Unconditional security of quantum key distribution over arbitrarily long distances. Science **283**(5410), 2050–2056 (1999)

[19] Mayers, D.: Unconditional security in quantum cryptography. Journal of the ACM (JACM) **48**(3), 351–406 (2001)

[20] Mayers, D., Yao, A.C.: Quantum cryptography with imperfect apparatus. In: 39th Annual Symposium on Foundations of Computer Science, FOCS '98, November 8-11, 1998, Palo Alto, California, USA. pp. 503–509. IEEE Computer Society (1998). https://doi.org/10.1109/SFCS.1998.743501

[21] Müller-Quade, J., Unruh, D.: Long-term security and universal composability. J. Cryptol. **23**(4), 594–671 (2010). https://doi.org/10.1007/s00145-010-9068-8

[22] Nielsen, M.A., Chuang, I.L.: Quantum Computation and Quantum Information (10th Anniversary edition). Cambridge University Press (2016)

[23] Renner, R.: Security of quantum key distribution. International Journal of Quantum Information **6**(01), 1–127 (2008)

[24] Shor, P.W., Preskill, J.: Simple proof of security of the BB84 quantum key distribution protocol. Physical review letters **85**(2), 441 (2000)

[25] Tomamichel, M., Leverrier, A.: A largely self-contained and complete security proof for quantum key distribution. Quantum **1**, 14 (2017)

[26] Yin, J., Cao, Y., Li, Y.H., Liao, S.K., Zhang, L., Ren, J.G., Cai, W.Q., Liu, W.Y., Li, B., Dai, H., et al.: Satellite-based entanglement distribution over 1200 kilometers. Science **356**(6343), 1140–1144 (2017)

A Impossibility of Unconditionally Secure QPKE

In the following we describe a simple argument that rules out the existence of *unconditionally* secure QPKE, even if the adversary is given access to a single copy of the public key. In more details, we show an adversary that can break the security of any QPKE if it is allowed to be unbounded also *during* the protocol execution. The following proof was suggested by Takashi Yamakawa, who should be credited for the argument.

Theorem A.1 (Unconditional Security)**.** There does not exists an unconditionally secure QPKE.

Proof. We provide a description of our generic attacker that, running in unbounded time, wins the experiment defined in Definition 3.2 with certainty. The attacker proceeds as follows.

- On input a state ρ and a bitstring pk, enter the following loop:
  - Sample a secret key $\mathsf{sk}^* \leftarrow \mathsf{SKGen}(1^\lambda)$ uniformly.
  - Run $(\rho^*, \mathsf{pk}^*) \leftarrow \mathsf{PKGen}(\mathsf{sk})$.
  - If $\mathsf{pk}^* = \mathsf{pk}$ exit the loop and return $(\rho^*, \mathsf{pk}^*, \mathsf{sk}^*)$.
  - Else, start over.
- Let $(\rho^*, \mathsf{pk}^*, \mathsf{sk}^*)$ be the tuple output by the above loop. Return ρ^* to the challenger.
- Upon receiving ct, use sk^* to decrypt the message.

To show that the attack always succeeds, it suffices to observe that the internal loop eventually returns a tuple $(\rho^*, \mathsf{pk}^*, \mathsf{sk}^*)$ such that

$$\mathsf{pk}^* = \mathsf{pk} \quad \text{and} \quad (\rho^*, \underbrace{\mathsf{pk}^*}_{=\mathsf{pk}}) = \mathsf{PKGen}(\mathsf{sk}^*)$$

In particular, this means that the algorithm Enc run by the challenger is run on a valid pair (ρ^*, pk). Therefore, by the correctness of QPKE, the secret key sk^* recovers the correct message with certainty. □

We point out that the same attack works also against protocol with imperfect correctness. For a protocol where decryption succeeds with probability $1 - \varepsilon$, the same attack also succeeds with probability $1 - \varepsilon$: Observe that sampling the triple $(\rho, \mathsf{pk}, \mathsf{sk})$ honestly, is identical to sampling a random pk, and then a pair (ρ^*, sk^*) that is consistent with pk (by rejection sampling). This implies that the correctness guarantee also applies to $(\rho^*, \mathsf{pk}, \mathsf{sk}^*)$ as sampled by the attacker. Since $1 - \varepsilon$ must be non-negligible, we have an attacker that succeeds with non-negligible probability.

Furthermore, we also mention that a similar attack can be applied against general quantum key agreement protocols. The attack works identically (except for obvious syntactical modifications) even if Bob's message is quantum, and results in the adversary and Bob sharing a key, without Bob noticing any difference, by the correctness of the protocol. However, we also point out that the same attack does not allow the attacker to fool Alice, nor it rules out the fact that Alice could notice that something went wrong.