New security notions and feasibility results for authentication of quantum data

Sumegha Garg
Princeton

Henry Yuen
MIT

Mark Zhandry
Princeton

Abstract

We give a new class of security definitions for authentication in the quantum setting. Our definitions capture and strengthen several existing definitions, including superposition attacks on *classical* authentication, as well as full authentication of quantum data. We argue that our definitions resolve some of the shortcomings of existing definitions.

We then give several feasibility results for our strong definitions. As a consequence, we obtain several interesting results, including: (1) the classical Carter-Wegman authentication scheme with 3-universal hashing is secure against superposition attacks, as well as adversaries with quantum side information; (2) quantum authentication where the entire key can be re-used if verification is successful; (3) conceptually simple constructions of quantum authentication; and (4) a conceptually simple QKD protocol.

1 Introduction

Authenticating messages is one of the fundamental operations in classical cryptography. A sender Alice and receiver Bob share a secret key k, and Alice wishes to send a message m over an insecure channel to Bob, ensuring that the message was not tampered with in transit. Alice will affix a "signature" σ to m using the key k and send the message/signature pair (m, σ) to Bob. Bob receives some potentially altered pair (m', σ'), and will then verify that σ' is a valid signature on m'. If verification passes, Bob accepts m', and if verification fails, Bob ignores the message and discards it. The guarantee is that, even if the adversary has arbitrarily tamper with the communication channel, as long as the adversary does not know the secret key k, either Bob rejects, or the message he receives is m. Intuitively, this means the adversary cannot do anything but forward the message as is or send a junk message that is always rejected. We generally require that security holds for *any* m, reflecting the possibility that the adversary may be able to affect the message being sent. Such a (symmetric key) authentication protocol is usually referred to as a Message Authentication Code (MAC). As long as k is only used to authenticate a single message, information-theoretic security can be achieved: no computationally unbounded adversary can modify the message. Put another way, information-theoretic classical one-time MACs exist [WC81].

Just as authentication is fundamental to classical cryptography, it will continue to be an important tool in the coming age of quantum computers. In this work, we investigate authentication in the quantum setting. Namely, we explore both quantum attacks on *classical* protocols, as well as full-fledged quantum protocols for authenticating quantum data.

Quantum Attacks on Classical Protocols. A recent series of works [BDF+11, DFNS13, BZ13a, BZ13b, Zha12, KLLNP16] have studied quantum superposition attacks on classical cryptosystems. In the case of message authentication codes, an adversary in such an attack is able to trick the sender into signing a superposition of messages. That is, the sender computes the map $|m\rangle \mapsto |m, \sigma_m\rangle$ in superposition, where σ_m is the signature on m. The adversary chooses some message superposition $\sum_m \alpha_m |m\rangle$, and the sender then applies the map, giving the adversary $\sum_m \alpha_m |m, \sigma_m\rangle$.

At this point, it is unclear what the security definition should actually be. Clearly, the adversary can tamper with the signed state: he can, for example, measure the entire state in the standard basis, obtaining the pair (m, σ_m) with probability $|\alpha_m|^2$. Then m, σ_m will pass verification, but will be different from the signed state the adversary received. If the adversary can change the message state, what sort of guarantees can we hope for?

Boneh and Zhandry [BZ13a] give the first definition of security for classical authentication against superposition attacks. They argue that, at a minimum, the adversary given a single signed superposition should only be able to produce a single signed message; he should not be able to produce both valid signed messages m, σ_m and $m', \sigma_{m'}$ for $m \neq m'$. In the classical setting, this requirement is equivalent to the traditional MAC security definition: an adversary who intercepts the signed message (m, σ), and is able to maul the message into (m', σ'), can also produce two signed messages: namely the original senders message (m, σ) and the mauled message (m', σ').

However, the Boneh-Zhandry definition has some unsatisfying properties. For example, consider the case where the sender only signs messages that start with the email address of some intended recipient, say, `bob@gmail.com`. Suppose the adversary tricks the sender into a signing a superposition of messages that all begin with `bob@gmail.com`, but then manipulates the signed superposition into a different superposition that includes valid signed messages that *do not* start with `bob@gmail.com`. Clearly, this is an undesirable outcome. Unfortunately, the Boneh-Zhandry definition does not rule out such attacks — it only rules out the possibility of an adversary producing $q + 1$ valid signed messages when given q signed superpositions. The situation illustrated here, however, is that the adversary is given *one* signed superposition, and now wants to produce *one* valid signed message that was not part of the original superposition.

Along similar lines, suppose an adversary tricks the sender into signing a uniform superposition on messages, and then produces a classical signed message (m, σ). From the sender's perspective, each message has weight $\frac{1}{|\mathcal{M}|}$, where $\mathcal{M}$ is the message space. The sender cannot prevent the adversary from measuring the message state to produce (m, σ) for a random m. However, it would be reasonable to expect that the adversary cannot bias the output of this measurement to obtain, say, (m^*, σ_{m^*}) with probability much higher than $\frac{1}{|\mathcal{M}|}$. Again, Boneh and Zhandry's definition does not preclude such a biasing, since the adversary only ever obtains a single signed message. Thus, the Boneh-Zhandry definition does not capture natural non-malleability properties one would hope for from an authentication scheme.

Boneh and Zhandry's definitions suffers from these weaknesses because it only considers what types of outputs the adversary can produce, ignoring the relationships between the output and the original signed state. In the classical setting, the two approaches are actually equivalent, but in the quantum setting this is not the case.

Quantum Authentication of Quantum Data. Barnum et al. [BCG+02] investigate the possibility of authenticating quantum data using a quantum protocol. They present a definition of non-interacting quantum authentication where, conditioned on the protocol succeeding, the sender has effectively teleported a quantum state to the receiver (provided that the probability of success is not too small). They then give a scheme which attains this definition. Interestingly, they show that quantum state authentication also implies quantum state *encryption*. Roughly, they argue that authentication in one basis (say, the computational or Fourier basis) implies encryption in the complementary basis. Their definition corresponds to authentication in all bases, which gives encryption in all bases.

However, their general definition of quantum authentication has some shortcomings: first, it does not explicitly handle the case of when the adversary has some quantum side information about the message. Second, the security definition averages over the secret key shared between the sender and receiver. Suppose Alice sends Bob the authenticated state $\mathsf{Auth}_k(\rho)$ using key k. Bob receives a (possibly tampered) state σ_k, and proceeds to verify the authentication. Let τ_k denote the Bob's state *conditioned* on successful verification. Roughly speaking, the definition given by Barnum, et al. state that the *average* state $\mathbb{E}_k \tau_k$ is close to the original state ρ. However, this does not immediately imply that τ_k is close to the original state ρ *with high probability*, which is a much more useful condition. When there is no quantum side information, their definition does in fact imply a "with high probability" statement, but this implication no longer seems to hold when the adversary can manipulate the side information.

The work of Hayden, Leung, and Mayers [HLM11] later showed that the protocol given by [BCG+02] actually has *universal composable security*, which implies that it remains secure in the presence of side information. However, no general definition for authentication with quantum side information was given.

Furthermore, [HLM11] show that the secret key used in the Barnum, et al. protocol can be partially *re-used* in further applications without compromising their security. When authenticating classical information, the key can even be re-used in its entirety [DPS05]; as long as verification never fails, an unbounded number of messages can be authenticated. This is quite surprising, since in the classical setting such re-usability cannot be obtained without computational assumptions.

Again, unfortunately, the key re-usability property does not follow from the general security definition alone, but follows from an analysis of the particular [BCG+02] protocol. Moreover, it has been an open question of whether there is a quantum authentication scheme to allow for full re-usability of the key upon successful verification.

Other works have presented other schemes for quantum state authentication as well. In [ABOE10], two separate quantum authentication schemes were given (the Clifford authentication scheme, and the polynomial authentication scheme), for use in delegated quantum computation. These schemes were shown to satisfy the Barnum, et al. security definition, which does not consider side information and averages over the key. The work [BGS13] presented the so-called trap authentication scheme, but its security was proven in the context of the much more complicated functionality of *quantum one-time programs*. Recently, Broadbent and Wainewright [BW16] showed that the Clifford scheme and the trap scheme are have Universally Composable (UC) security, which implies that they are secure in the presence of quantum side information. However, their

security definition (which also appears in [DNS12]) still appears to average over the key.

1.1 This Work

In this work, we address the above limitations by giving new security notions for authentication in the quantum setting. More generally, we present an abstract framework of security for both classical and quantum authentication schemes that not only captures existing security definitions (such as the Boneh-Zhandry definition for classical protocols or the Barnum, et al. definition of quantum state authentication), but also is more powerful in that it strongly *characterizes* the (effective) behavior of an adversary. In particular, the adversary may have access to quantum side information with the message state that is being authenticated. The characterization of the adversary's admissable actions is what allows us to easily deduce many desirable security properties (such as unforgeability, key reuse, and more). Furthermore, we will show that various natural authentication protocols satisfy our security definitions.

Our abstract security framework is inspired by the simulation paradigm in classical cryptography. In our framework, one first defines a class $\mathscr{A}$ of *ideal adversaries*. Intuitively, ideal adversaries are those that cannot be avoided in a real execution of an authentication protocol, such as those that discard messages, or ones that carry out actions explicitly allowed by the protocol. For example, in the case of classical protocols, one can define the class of ideal adversaries to be ones that "behave classically" on the message state – that is, they're restricted to measurements in the computational basis. In the case of quantum authentication, an ideal adversary can *only* act on the side information, but otherwise acts as the identity on the authenticated message.

An authentication protocol P satisfies our security definition with respect to the class $\mathscr{A}$ if for any adversary (not necessarily ideal), its behavior in the protocol P can be approximately simulated by an ideal adversary in $\mathscr{A}$. We take the most general notion of simulation possible: the joint state of the secret key, the message state after the receiver's verification procedure (after an arbitrary adversary's action), and the quantum side information held by the adversary must be (up to some error) indistinguishable from the joint state arising from the actions of *some* ideal adversary from the class $\mathscr{A}$.

We now discuss how security for both classical authentication schemes and fully quantum authentication protocols can be defined in this framework.

A new security definition for classical authentication. The Boneh-Zhandry definition focuses on what classical signed messages an adversary can produce, treating the superposition access to the sender as a tool to mount stronger attacks. Here, we instead think of a classical protocol giving rise to a weak form of authentication of quantum messages, where a superposition is authenticated by classically signing each message in the superposition. That is, a state $\sum_m \alpha_m |m\rangle$ is authenticated as the state $\sum_m \alpha_m |m, \sigma_m\rangle$. The state is similarly verified in superposition by running the classical verification algorithm in superposition, and measuring the result of the computation.

More generally, we think of the protocol acting on messages states that may be entangled with an adversary. For example, the sender could sign the $\mathcal{M}$ part of the state $\sum_m \alpha_m |m\rangle^{\mathcal{M}} \otimes |\varphi_m\rangle^{\mathcal{Z}}$, where the adversary has control of the quantum side information $|\varphi_m\rangle^{\mathcal{Z}}$ states. The signed state then would become $\sum_m \alpha_m |m, \sigma_m\rangle^{\mathcal{MT}} \otimes |\varphi_m\rangle^{\mathcal{Z}}$. Signing mixed states can also be expressed in this way, simply by purifying the mixture. By thinking of the protocol in this way, we are able to give

security definitions that actually consider the relationship between the sender's signed state and the final state the adversary produces.

Clearly, such a classical scheme cannot fully protect the quantum state. An adversary could, for example measure m, σ_m, or any subset of bits of the state, and keep the result of such a measurement in his own private space. Also, the adversary can choose to replace the signed state with junk if the outcome of some measurement is 1, and forward the signed state if the outcome of a measurement is 0. None of these actions would be detected by the classical verification procedure.

Our security definition for classical protocols says that, roughly, an arbitrary adversary can be simulated by an ideal adversary that can only do the following: perform some measurement in the computational basis (perhaps perturbing his own private qubits based on the result of the measurement), and then perhaps conditionally replacing the state with junk. We also extend the definition to handle side information the adversary may have about the message state; for example, the adversary may possess the purification of the message state. Thus, our definition is essentially the best one could hope for, since is disallows the adversary from doing anything other than operations that are trivially possible on *any* classical protocol.

Our definition readily implies the Boneh-Zhandry security definition for one-time MACs, and does not suffer from the weakness of their definition[1]. Finally, we show that the classical Carter-Wegman MAC that uses three-universal hashing is sufficient for achieving this strong security definition.

Definitions for Quantum Authentication. We next turn to quantum protocols for authenticating quantum messages. For general quantum protocols, the adversary can always do the following. He can always act non-trivially on his own private workspace – the verification procedure can never detect this. Otherwise, he can forward the authenticated state as is, without recording any information about the state, or he can send junk to the receiver. Our strongest definition of security – which we call *total authentication* – says that this is essentially all an adversary can do in a secure quantum authentication protocol. In other words, a real adversary in a total authentication protocol can be approximated by an ideal adversary that behaves trivially on the authenticated state.

Our definition strengthens Barnum et al.'s definition, and due to the fact that we consider side information about the plaintext state, we obtain security guarantees that are similar to the universally composable variant of their definition [HLM11, DNS12, BW16]. However, our definition is actually strictly stronger, due to the fact that we consider the receiver's view to include the authentication key as well as whatever information the adversary may learn about the key. The ideal adversary must approximate the real adversary, even considering the entire key. In contrast, existing definitions trace out the key — either partially or entirely — and therefore do not directly consider *arbitrary* information the adversary may learn about the key. Our security definition of total authentication thus rules out the possibility of the adversary learning anything about the key (because the ideal adversary does not interact with the authenticated state at all).

This fact has interesting consequences. For example, our definition immediately implies that,

[1]One limitation of our definition is that we consider the signature registers as being initialized by the signer. Boneh and Zhandry, in contrast, allow the registers to be initialized by the adversary, with the signature being XORed into the registers

upon successful verification by the receiver, the key can actually be completely recycled to authenticate a new message. This is because, upon successful verification, the key is completely hidden from the adversary and can therefore be used again in the same protocol. We note that key recycling from quantum authentication was studied before by [HLM11], but they were only able to demonstrate that *part* of the key in the Barnum, et al. protocol is reusable. Furthermore, no prior definition for authentication of quantum data directly implies key re-usability, and no prior protocol for quantum messages gets full key re-usability upon successful verification.

Our definition also gives a conceptually simple quantum key distribution (QKD) protocol[2]. Alice prepares a maximally entangled state, chooses a random key k, and authenticates half the state with the key. She then sends the authenticated half to Bob, keeping the unauthenticated half to herself. When Bob receives the state, he sends a a "received" message back to Alice, who then sends the key k to Bob. Bob verifies the state using the key. Even though the adversary eventually sees the authentication key k, he does not know the key when he intercepts the quantum state, and must therefore interact with the state without the key. If Bob's verification passes, it implies, roughly, that the adversary could not have tampered with the state (by the security of total authentication); in particular, the adversary could not have learned any information about the maximally entangled state. Therefore, Alice and Bob measure their halves of the maximally entangled state and obtain a shared key that is unknown to the eavesdropper. If Bob's verification rejects, the two try again. Though this is not a practical QKD scheme (because any tampering by the adversary would cause Alice and Bob to abort), it is conceptually very simple and illustrates the power of our definitions.

Next, we exhibit a protocol meeting our strong security notion. We present an authentication scheme based on *unitary designs*, which are efficiently sampleable distributions over unitary matrices that behave much like the uniform distribution over unitaries when only considering low degree moments. The protocol is simple: to authenticate a quantum state ρ, first the state ρ is padded with some number zero qubits, so that the state looks like $\rho \otimes |0\rangle\langle 0|^{\otimes s}$. Then, using the secret key k the sender selects a random unitary U_k from an appropriate unitary design. The state $U_k \rho \otimes |0\rangle\langle 0|^{\otimes s} U_k^\dagger$ is then sent across the quantum channel. To verify, the receiver applies the inverse unitary $U_k^\dagger$ and checks that the last s qubits are all 0. Recall that in the classical setting, padding a message before applying a non-malleable encryption gives authenticated encryption. Thus, our construction of authentication from unitary designs generalizes this idea to the quantum setting.

This scheme is very similar to the *non-malleable quantum encryption* scheme based on unitary 2-designs that was proposed by Ambainis, Bouda, and Winter [ABW09]. However, their scheme does not provide any authentication, and does not consider quantum side information.

Finally, we also give a definition of *total authentication with key leakage*. This is a notion of security where the real adversary can be simulated by an ideal trivial adversary that only acts on its own private workspace, *but in a manner that may depend on the key*. This is slightly weaker notion of security than total authentication, but it still implies simple QKD and some amount of key reuse. We note that the work of [HLM11] essentially show that the Barnum et al. protocol satisfies total authentication with (minor) key leakage.

We give a simple authentication scheme that achieves this: first, one classically authenticates,

[2]The observation that quantum authentication implies a form of QKD is due to Charlie Bennett.

performs the quantum Fourier transform, and classically authenticates again using a fresh key. We call this the "Auth-QFT-Auth" protocol, and show that it achieves total authentication where the key used in the second authentication may leak. In exchange we obtain secrecy for the quantum message as well as the key from the first authentication. This illustrates the surprising versatility of classical authentication schemes: combined with one quantum step (the Fourier transform), it can give full quantum authentication. This also gives a conceptually simple alternative to the protocol of [BCG+02].

Outline. In the next section we cover some preliminaries and notation. In Sections 3 and 4 we formally present the fundamental security definitions used in our paper. In Sections 5 and 6 we prove that our definitions satisfy the properties expected of authentication schemes. In Section 7, we analyze the security of the Carter-Wegman MAC with 3-universal hashing within our security framework. In Section 8 we present and analyze the Auth-QFT-Auth scheme. In Section 9 we present and analyze the unitary design scheme.

2 Preliminaries

Quantum information. We assume basic familiarity with quantum computing concepts, such as states, measurements, and unitary operations. We will use caligraphic letters to denote Hilbert spaces, such as $\mathcal{H}$, $\mathcal{M}$, $\mathcal{T}$, $\mathcal{K}$, and so on. We write $\mathrm{S}(\mathcal{H})$ to denote the set of unit vectors in $\mathcal{H}$. For two Hilbert spaces $\mathcal{H}$ and $\mathcal{M}$, we write $\mathrm{L}(\mathcal{H}, \mathcal{M})$ to denote the set of matrices that map $\mathcal{H}$ to $\mathcal{M}$. We abbreviate $\mathrm{L}(\mathcal{H}, \mathcal{H})$ as simply $\mathrm{L}(\mathcal{H})$. The following are important subsets of $L(\mathcal{H})$ that we'll use throughout this paper.

- $\mathrm{D}(\mathcal{H})$ denotes the set of *density matrices* on $\mathcal{H}$; that is, positive semidefinite operators on $\mathcal{H}$ with unit trace.
- $\mathrm{D}_{\leq}(\mathcal{H})$ denotes the set of *subnormalized* density matrices on $\mathcal{H}$; that is, positive semidefinite operators on $\mathcal{H}$ with trace at most one.
- $\mathrm{U}(\mathcal{H})$ denotes the set of unitary matrices acting on $\mathcal{H}$. For an integer N, we will also write $\mathrm{U}(N)$ to denote the set of all $N \times N$ complex unitary matrices.

Another important class of operators are *isometries*: these are like unitaries, except that can append ancilla qubits. We say that a map $V \in \mathrm{L}(\mathcal{H}, \mathcal{M})$ is an isometry if for all vectors $|\psi\rangle \in \mathcal{H}$, $\|V|\psi\rangle\| = \||\psi\rangle\|$. Note that this requires $\dim(\mathcal{M}) \geq \dim(\mathcal{H})$. We will let $\mathrm{J}(\mathcal{H}, \mathcal{M})$ denote the set of isometries in $\mathrm{L}(\mathcal{H}, \mathcal{M})$.

We use $\mathbb{I}$ to denote the identity matrix. For a Hilbert space $\mathcal{H}$, we let $|\mathcal{H}|$ denote the dimension of $\mathcal{H}$.

We will typically decorate states and unitaries with subscripts to denote which spaces they act on. For example, let $\mathcal{Y}$ and $\mathcal{Z}$ be two Hilbert spaces. Let $U \in \mathrm{U}(\mathcal{Y})$ and let $V \in \mathrm{U}(\mathcal{Y} \otimes \mathcal{Z})$. Then when we write the product $U^{\mathcal{Y}} V^{\mathcal{Y}\mathcal{Z}}$ we mean the $(U^{\mathcal{Y}} \otimes \mathbb{I}^{\mathcal{Z}}) V^{\mathcal{Y}\mathcal{Z}}$; we will often omit mention of the identity unitary when it is clear from context.

Another convention is the implicit partial trace. For example, let $\rho^{\mathcal{K}\mathcal{M}} \in \mathrm{D}(\mathcal{K} \otimes \mathcal{M})$. Then $\rho^{\mathcal{M}} = \mathrm{Tr}_{\mathcal{K}}(\rho^{\mathcal{K}\mathcal{M}})$. Additionally, given a pure state $|\rho\rangle$, we will let ρ denote the rank one density matrix $|\rho\rangle\langle\rho|$.

Superoperators. In this paper we will consider *superoperators*, which are linear maps that act on a vector space of linear maps. For Hilbert spaces $\mathcal{H}$ and $\mathcal{M}$, let $\mathrm{T}(\mathcal{H}, \mathcal{M})$ denote the set of all linear maps that take elements of $\mathrm{L}(\mathcal{H})$ to $\mathrm{L}(\mathcal{M})$. While superoperators can be very general, we will focus on superoperators $\mathcal{O} \in \mathrm{T}(\mathcal{H}, \mathcal{M})$ that are *completely positive* and *trace non-increasing*, which have the following characterization: there exists an alphabet Σ and set of matrices (not necessarily Hermitian) $\{A_a\}_{a\in\Sigma} \subset \mathrm{L}(\mathcal{H}, \mathcal{M})$ such that

1. $\mathcal{O}(X) = \sum_{a\in\Sigma} A_a X A_a^\dagger$ for all $X \in \mathrm{L}(\mathcal{H})$, and
2. $\sum_{a\in\Sigma} A_a^\dagger A_a \preceq \mathbb{I}^{\mathcal{H}}$.

For the rest of this paper, when we speak of superoperators, we will always mean completely positive, trace non-increasing superoperators. Although the definition of superoperators is rather abstract, they capture general quantum operations on arbitrary quantum states, including post-selection, as demonstrated by Stinespring's dilation theorem:

Theorem 1 (Stinespring's dilation theorem)**.** *A map $\mathcal{O} \in \mathrm{T}(\mathcal{H}, \mathcal{M})$ is a completely positive, trace non-increasing superoperator if and only if there exists auxiliary Hilbert spaces $\mathcal{Z}, \mathcal{Z}'$, an isometry $V \in \mathrm{J}(\mathcal{H} \otimes \mathcal{Z}, \mathcal{M} \otimes \mathcal{Z}')$, and a projector Π acting on $\mathcal{M} \otimes \mathcal{Z}'$ such that for all density matrices $\rho \in \mathrm{D}(\mathcal{H})$, we have*

$$\mathcal{O}(\rho) = \mathrm{Tr}_{\mathcal{Z}'}(\Pi V \rho V^\dagger \Pi).$$

Matrix norms and distance measures. We will make use of several matrix norms and distance measures in this paper.

Given a (not necessarily unit) vector $|\psi\rangle \in \mathcal{H}$, we use $\||\psi\rangle\|_2$ to denote the Euclidean norm of $|\psi\rangle$.

The most matrix norm important is the *trace norm* of a linear operator X, defined to be $\|X\|_1 = \mathrm{Tr}(\sqrt{X^\dagger X})$. Correspondingly, the *trace distance* between density matrices ρ, σ is defined to be $\|\rho - \sigma\|_1$. The operational significance of the trace distance is that $\|\rho - \sigma\|_1$ denotes the maximum bias with which one can distinguish between ρ and σ using any quantum operation.

The next norm we will make use of is the *Frobenius norm* of a linear operator X, which is defined to be $\|X\|_2 = \sqrt{\mathrm{Tr}(X^\dagger X)}$. A useful property of the Frobenius norm is that $\|X\|_2 = \sqrt{\sum_{ij} |X_{ij}|^2}$, where the sum is over all the matrix entries of X (with respect to any basis).

The *operator norm* (also known as the *spectral norm*) of an operator $X \in \mathrm{L}(\mathcal{H})$ is to defined to be $\|X\|_\infty = \sup_{|v\rangle\in\mathrm{S}(\mathcal{H})} \|X|v\rangle\|_2$, where the supremum is over all unit vectors in $\mathcal{H}$.

Fact 2. *Let $|\psi\rangle, |\theta\rangle \in \mathrm{S}(\mathcal{H})$. Then*

$$\|\psi - \theta\|_1 \leq 2\||\psi\rangle - |\theta\rangle\|_2$$

where recall that $\psi = |\psi\rangle\langle\psi|$ and $\theta = |\theta\rangle\langle\theta|$.

3 Definitions

Spaces. We let $\mathcal{K}$ denote the **key space**, $\mathcal{M}$ denote the **message space**, and $\mathcal{Y}$ denote the **authenticated space**.

Authentication scheme. An δ-authentication scheme is a pair of keyed superoperators $\mathsf{Auth}, \mathsf{Ver}$ where

- Auth_k for $k \in \mathcal{K}$ is a superoperator mapping $\mathrm{D}(\mathcal{M})$ to $\mathrm{D}(\mathcal{Y})$.
- Ver_k for $k \in \mathcal{K}$ is a superoperator mapping $\mathrm{D}(\mathcal{Y})$ to $\mathrm{D}(\mathcal{M})$.

satisfying the (approximate) correctness requirements that for any (potentially mixed) quantum state $\rho \in \mathrm{D}(\mathcal{M})$,

$$\left\| \left(\mathop{\mathbb{E}}_k |k\rangle\langle k| \otimes \mathsf{Ver}_k(\mathsf{Auth}_k(\rho)) \right) - \mathop{\mathbb{E}}_k |k\rangle\langle k| \otimes \rho \right\|_1 < \delta \tag{1}$$

where $\|\cdot\|_1$ denotes the trace norm.

This definition of authentication scheme is more general than we need in this paper. Throughout this work, we shall exclusively work with exact authentication schemes; that is, authentication schemes where $\mathsf{Ver}_k(\mathsf{Auth}_k(\rho)) = \rho$ for all k. Furthermore, we will assume that Auth_k behaves as an isometry taking $\mathcal{M}$ to $\mathcal{Y}$ (i.e. it isn't probabilistic).

We will treat Ver_k as a filter that only accepts states that were properly authenticated. More formally, we view $\mathsf{Ver}_k(\tau)$ as first projecting the input state τ onto the subspace of $\mathcal{Y}$ that is the image of $\mathcal{M}$ under Auth_k. Then, it applies the inverse isometry Auth_k on this projection ("undoes the authentication"). Thus $\mathsf{Ver}_k(\cdot)$ is not a trace-preserving quantum operation.

Note that normally, the verification or decoding procedure of an authentication scheme (e.g., as defined in [BCG+02]) is a trace preserving operation that additionally generates an additional bit b indicating whether verification accepted or rejected. Then the correctness requirement above would additionally require that $b = 1$ with probability (negligibly close to) 1, for inputs obtained by running Auth_k on some state. However, we note that this is equivalent to the formulation above. Indeed, starting from a verification operator Ver'_k that additionally outputs b, we obtain an operator Ver_k that projects onto $b = 1$, and then discards b. If $b = 0$ with non-negligible probability, then the trace of the result would be smaller than that of ρ. Hence, the result could not be close in trace distance. Therefore, a small trace distance implies that $b = 1$ with overwhelming probability. This view of the verification procedure Ver_k as a filter will be more useful in our paper.

We will also use Auth and Ver to denote the operators

$$\mathsf{Auth}(\cdot) = \sum_k |k\rangle\langle k| \otimes \mathsf{Auth}_k(\cdot) \qquad \mathsf{Ver}(\cdot) = \sum_k |k\rangle\langle k| \otimes \mathsf{Ver}_k(\cdot).$$

Classical Authentication. In a classical authentication protocol, the authentication operator Auth_k is specified by a classical function $\mathsf{Auth}_k : \mathcal{M} \mapsto \mathcal{Y}$ acting on the computational basis, run in superposition on the input state. The verification operator behaves the same as described above: Ver_k projects onto the subspace of $\mathcal{Y}$ spanned by classical strings $\mathsf{Auth}_k(m)$ for all $m \in \mathcal{M}$, and then applies the inverse map Auth_k^{-1}.

Oftentimes we will want to project onto the space of valid authenticated messages, without undo-ing the authentication. We use the operator Check to denote this:

$$\mathsf{Check}_k = \sum_m |\mathsf{Auth}_k(m)\rangle\langle\mathsf{Auth}_k(m)|$$

We will also let $\mathsf{Check}(\cdot) = \sum_k |k\rangle\langle k| \otimes \mathsf{Check}_k(\cdot)$.

Message authentication codes. A message authentication code (or MAC) is special type of classical authentication scheme $(\mathsf{Auth}, \mathsf{Ver})$ where for a message m, $\mathsf{Auth}_k(m) = (m, \sigma(k, m))$, where we call $\sigma(k, m)$ the *message tag*. We treat Ver_k as an operator that projects out messages that do not have valid tags, and for messages with valid tags, Ver_k will strip the tags away:

$$\mathsf{Ver}_k = \sum_m |m\rangle\langle m, \sigma(k, m)|.$$

In the case of a MAC, the check operator looks like:

$$\mathsf{Check}_k = \sum_m |m, \sigma(k, m)\rangle\langle m, \sigma(k, m)|.$$

Adversaries. The way we model adversaries is the most general – and the most conservative – way possible: the adversary prepares the initial message state $|\rho\rangle^{\mathcal{MZ}}$, where we can assume that the adversary possesses the purification of $\rho^{\mathcal{M}}$. After the state is authenticated with a secret key k, the adversary gets to attack the $\mathcal{YZ}$ spaces with an arbitrary completely positive trace non-increasing superoperator $\mathcal{O}$. After this attack, the state is un-authenticated with the same key k.

We don't require the superoperator $\mathcal{O}$ to be trace preserving; this is to allow adversaries to *discard* certain measurement outcomes (or, alternatively, *post-select* on measurement outcomes, without renormalizing). While this may seem to give the adversary far too much power, in our security definitions we take into account the probability of the event that the adversary post-selects on. If this probability is too small, the security guarantees are meaningless, which is necessary. Allowing for superoperators to be trace non-preserving will help make our definitions clean to state.

4 Security Framework for Quantum Authentication

We now give a framework of security definition for authentication protocols in the quantum setting, involving adversaries that may possess side information that is entangled with the messages. Our security definitions generalizes some of the known classical and quantum authentication definitions.

We present our security definitions using the real/ideal paradigm. Let $(\mathsf{Auth}, \mathsf{Ver})$ be an authentication protocol, with key space $\mathcal{K}$, message space $\mathcal{M}$, and authenticated space $\mathcal{Y}$. Let $\mathcal{Z}$ denote the space of auxiliary side information.

Definition 3. *Let* $(\mathsf{Auth}, \mathsf{Ver})$ *be an authentication scheme. Let* $\mathscr{A} \subseteq \mathrm{T}(\mathcal{YZ}, \mathcal{YZ})$ *denote a set of ideal adversaries. The scheme* $(\mathsf{Auth}, \mathsf{Ver})$ *is* ε-reduces to $\mathscr{A}$-adversaries *iff the following holds: for all initial message states* $|\rho\rangle^{\mathcal{MZ}}$*, for all adversaries* $\mathcal{O} \in \mathrm{T}(\mathcal{YZ}, \mathcal{YZ})$*, there exists an ideal adversary* $\mathcal{I} \in \mathscr{A}$ *such that the following (not necessarily normalized) states are* ε*-close in trace distance:*

- *(Real experiment)* $\mathbb{E}_k |k\rangle\langle k| \otimes \left[(\mathsf{Ver}_k \otimes \mathbb{I}^{\mathcal{Z}}) \circ \mathcal{O} \circ (\mathsf{Auth}_k \otimes \mathbb{I}^{\mathcal{Z}})\right](\rho^{\mathcal{MZ}})$
- *(Ideal experiment)* $\mathbb{E}_k |k\rangle\langle k| \otimes \left[(\mathsf{Ver}_k \otimes \mathbb{I}^{\mathcal{Z}}) \circ \mathcal{I} \circ (\mathsf{Auth}_k \otimes \mathbb{I}^{\mathcal{Z}})\right](\rho^{\mathcal{MZ}})$

Intuitively, our security definition states that for an authentication scheme $(\mathsf{Auth}, \mathsf{Ver})$ that is $\mathscr{A}$-secure, for all initial message states $\rho^{\mathcal{MZ}}$, an *arbitrary* adversary that acts on an authenticated state $\mathsf{Auth}_k(\rho^{\mathcal{MZ}})$ is *reduced* to an "ideal adversary" in $\mathscr{A}$; behaving differently will cause the verification procedure to abort. In other words, "all the adversary can do" is behave like some adversary in the class $\mathscr{A}$.

A comment about normalization. It is important that the states of the real experiment and ideal experiment are not requiried to have unit trace. This is because their trace corresponds to the probability that the verification procedure accepts. If the probability of acceptance is smaller than ε, then the security guarantee is vacuous. Intuitively, this corresponds to situations such as the adversary successfully guessing the secret key k, so we cannot expect any security guarantee in that setting. However, if the probability of acceptance is significantly larger than ε, then we can condition on acceptance, and still obtain a meaningful security guarantee: the distance between the (renormalized) real experiment and ideal experiments is small.

We now specialize the above definition to some important classes of ideal adversaries that we will consider in this paper. Note that for two classes of ideal adversaries $\mathscr{A}$ and $\mathscr{A}'$, if $\mathscr{A} \subset \mathscr{A}'$, then an authentication scheme reducing to $\mathscr{A}$-adversaries implies reducing to $\mathscr{A}'$-adversaries. Hence reducing to $\mathscr{A}$-adversaries is a stronger security guarantee.

Comment on UC Security: Our various security definitions are equivalent to UC security for quantum authentication with different restrictions on the simulator.

4.1 Basis-dependent authentication

We first define a notion of security of authentication schemes that reduce to a *basis-respecting* adversary.

Definition 4 (Basis-respecting adversaries)**.** *Let $\mathcal{B} = \{|\psi\rangle\}$ denote an orthonormal basis for $\mathcal{Y}$. Then an adversary $\mathcal{I} \in \mathrm{T}(\mathcal{YZ}, \mathcal{YZ})$ is* $\mathcal{B}$-respecting *iff it can be written as*

$$\mathcal{I}(\sigma) = \mathrm{Tr}_{\mathcal{Z}'}(\Pi V \sigma V^\dagger \Pi)$$

for all $\sigma \in \mathrm{D}(\mathcal{YZ})$, where Π is a projector acting on $\mathcal{ZZ}'$, and $V \in \mathrm{J}(\mathcal{YZ}, \mathcal{YZZ}')$ is an isometry that can be written as

$$V = \sum_{\psi \in \mathcal{B}} |\psi\rangle\langle\psi|^{\mathcal{Y}} \otimes V_\psi$$

where for each ψ, $V_\psi \in \mathrm{J}(\mathcal{Z}, \mathcal{ZZ}')$ is some isometry.

Without the second condition on V, by Stinespring's Dilation Theorem every superoperator can be written as $\mathcal{I}(\sigma) = \mathrm{Tr}_{\mathcal{Z}'}(\Pi V \sigma V^\dagger \Pi)$ for some choice of isometry V and projector Π. However, the second condition forces V to respect the basis $\mathcal{B}$. Intuitively, a basis-respecting adversary

first measures the $\mathcal{Y}$ register in the $\mathcal{B}$ basis, and based on the measurement outcome, performs some further isometry on the side information in $\mathcal{Z}$. When $\mathcal{B}$ is simply the computational basis, then the adversary treats the $\mathcal{Y}$ register as classical.

Definition 5 (Security relative to a basis)**.** *Let $\mathcal{B}$ be a basis for $\mathcal{Y}$. An authentication scheme* $(\mathsf{Auth}, \mathsf{Ver})$ ε-authenticates relative to basis $\mathcal{B}$ *iff it ε-reduces to the class of $\mathcal{B}$-respecting adversaries.*

Intuitively, our new definition captures the "best possible" security definition for *classical* authentication protocols. With a classical protocol, the adversary can perform arbitrary measurements on the authenticated space without detection by the verification algorithm. Because measurements are now undetectable, the adversary can also perform σ-dependent operations to the auxiliary registers, where σ is the classical authenticated message observed in the authenticated registers. For example, he can copy σ into the auxiliary space. He can also now choose to abort or not depending on σ. However, he should not be able to turn σ into $\sigma' \neq \sigma$.

In Section 5, we establish two properties that follow from our basis-dependent security definition. First, we show that from the point of view of an adversary, the state which was authenticated in superposition is indistinguishable from having been measured in the basis $\mathcal{B}$. Showing this uses our definition crucially: we reduce all potential distinguishers into adversaries that must behave in a basis-respecting manner, but then such an adversary cannot tell whether the state was measured or not.

Next, we show that our definition implies unforgeability: the adversary cannot produce two valid signed messages with non-negligible probability, when given access to only one superposition. Thus, our definition subsumes the Boneh-Zhandry security definition for one-time MACs.

In Section 7 we show that the classical Carter-Wegman MAC where the message m is appended with $h(m)$, where $h(\cdot)$ is drawn from a three-wise independent hash family, is a scheme that authenticates relative to the computational basis.

Theorem 6. *The Carter-Wegman MAC with three-universal hashing is $O(\sqrt{|\mathcal{M}|/|\mathcal{T}|})$-authenticating relative to the computational basis, where $\mathcal{T}$ is the range of the hash family.*

4.2 Total authentication

Here, we will define the strongest possible notion of secure quantum authentication.

Definition 7 (Oblivious adversary)**.** *An adversary $\mathcal{I} \in \mathrm{T}(\mathcal{YZ}, \mathcal{YZ})$ is* oblivious *iff there exists a superoperator $\mathcal{O} \in \mathrm{T}(\mathcal{Z}, \mathcal{Z})$ such that*

$$\mathcal{I}(\sigma) = (\mathbb{I}^{\mathcal{Y}} \otimes \mathcal{O})(\sigma)$$

for all $\sigma \in \mathrm{D}(\mathcal{YZ})$.

In other words, an oblivious adversary does not act at all on the authenticated message, and only acts on the auxiliary side information that it possesses about the state.

Definition 8 (Total authentication)**.** *An authentication scheme* $(\mathsf{Auth}, \mathsf{Ver})$ ε-totally authenticates *iff it ε-reduces to the class of oblivious adversaries.*

This is a generalization of the Barnum et al. definition to handle arbitrary auxiliary information about the input state. This is the strongest possible notion of security: for any authentication scheme, an adversary can always mount the following trivial attacks. First, he can arbitrarily modify the unauthenticated auxiliary state. Note that he cannot necessarily modify the contents of the auxiliary state based on the authenticated state, since this amounts to some measurement on the authenticated state, which verification may detect. Second, he can choose to either forward the authenticated state as is, or abort and forward nothing (equivalently, forward a junk state that is guaranteed to reject upon verification). Moreover, he can choose whether to abort or forward based on the contents of the auxiliary registers, and can even abort/forward in superposition. However, in an authentication scheme that totally authenticates the adversary can *only* behave in such trivial ways. Such trivial attacks are incorporated in our set of idea adversaries as $\mathcal{I}(\sigma)$ is a superoperator.

In Section 6 we establish a few properties of this definition. The first is that a totally authenticating scheme yields encryption of the quantum state. Barnum, et al. showed that quantum state authentication implies quantum state encryption [BCG$^+$02]. However, they did not take into account quantum side information. We show that our definition very easily implies encryption even when the adversary may be entangled with the message state.

Then, we show how our notion of total authentication gives rise to a conceptually simple version of quantum key distribution (QKD). [HLM11] have already observed that the universal composability of the Barnum et al. protocol implies that it can be used to perform QKD as well. Thus while our application of quantum authentication to QKD is not novel, we use this as another opportunity to showcase the strength of our definition. We also show how our definition easily implies full key reuse.

In Section 9 we present a scheme achieves total authentication, and hence is the strongest possible authentication scheme in the quantum setting. To our knowledge, this is the first authentication scheme that achieves this level of security. As described in the introduction, this is based on applying an (approximate) unitary design on the input state padded with some number s of $|0\rangle$ qubits.

Theorem 9. *The unitary design scheme is $2^{-s/2}$-totally authenticating, where s is the number of extra $|0\rangle$ qubits.*

As a consequence, this yields an authentication scheme where the key can be recycled fully, conditioned on successful verification by the receiver. In contrast, the protocol of Barnum et al. is not known to possess this property; [HLM11] showed that most of the key can be securely recycled.

4.3 Total authentication with key leakage

Finally, we introduce a slight weakening of the definition of total authentication above: we consider schemes that achieve total authentication of quantum data, but incur some *key leakage*. We model this in the following way: let $K = |\mathcal{K}|$ (the size of the keyspace), and let $K' \leq K$. Define a *key leakage function* $\ell : \mathcal{K} \mapsto \{0,1\}^{\log K'}$. If K' is strictly less than K, then $\ell(k)$ must necessarily lose information about the key $k \in \mathcal{K}$, but it will leak some information about it.

In a total authentication scheme with key leakage, an arbitrary adversary is reduced to an oblivious adversary (i.e., is forced to only act on the side information), but the manner in which it acts on the side information *may depend on* $\ell(k)$.

Definition 10 (Authentication with key leakage)**.** *Let* $(\mathsf{Auth}, \mathsf{Ver})$ *be an authentication scheme. Let* $K' \leq |\mathcal{K}|$ *and let* $\ell : \mathcal{K} \to \{0,1\}^{\log K'}$ *be a key leakage function. Let* $\mathscr{A} \subseteq \mathrm{T}(\mathcal{YZ}, \mathcal{YZ})$ *denote a set of ideal adversaries. The scheme* $(\mathsf{Auth}, \mathsf{Ver})$ ε-reduces to $\mathscr{A}$-adversaries with key leakage ℓ *iff the following holds: for all initial message states* $|\rho\rangle^{\mathcal{MZ}}$*, for all adversaries* $\mathcal{O} \in \mathrm{T}(\mathcal{YZ}, \mathcal{YZ})$*, there exists a collection of ideal adversaries* $\{\mathcal{I}_h\} \subset \mathscr{A}$*, indexed by* $h \in \{0,1\}^{\log K'}$*, such that the following (not necessarily normalized) states are* ε*-close in trace distance:*

- *(Real experiment)* $\mathbb{E}_k |k\rangle\langle k| \otimes \left[(\mathsf{Ver}_k \otimes \mathbb{I}^{\mathcal{Z}}) \circ \mathcal{O} \circ (\mathsf{Auth}_k \otimes \mathbb{I}^{\mathcal{Z}})\right] (\rho^{\mathcal{MZ}})$
- *(Ideal experiment)* $\mathbb{E}_k |k\rangle\langle k| \otimes \left[(\mathsf{Ver}_k \otimes \mathbb{I}^{\mathcal{Z}}) \circ \mathcal{I}_{\ell(k)} \circ (\mathsf{Auth}_k \otimes \mathbb{I}^{\mathcal{Z}})\right] (\rho^{\mathcal{MZ}})$.

Definition 11 (Total authentication with key leakage)**.** *Let* $K' \leq |\mathcal{K}|$ *and let* $\ell : \mathcal{K} \to \{0,1\}^{\log K'}$ *be a key leakage function. An authentication scheme* $(\mathsf{Auth}, \mathsf{Ver})$ ε-totally authenticates with key leakeage ℓ *iff it* ε*-reduces to the class of oblivious adversaries with key leakeage* ℓ*.*

This definition may seem somewhat strange: how is an ideal adversary able to learn bits $\ell(k)$ of the key k, if it doesn't act on the authenticated part of the state at all? Of course, any adversary that learns something about the key must have acted on the authenticated state, but the point is that, conditioned on successful verification, the adversary "effectively" behaved like an oblivious adversary that had access to $\ell(k)$.

In Section 8 we present a very simple scheme that achieves total authentication with some key leakage: to authenticate an arbitrary quantum state ρ, first apply the classical Carter-Wegman authentication scheme on it using key k. Then, apply $H^{\otimes n}$ to all the qubits in the authenticated state (i.e. apply the quantum Fourier transform over $\mathbb{Z}_2$). Finally, apply the classical Carter-Wegman scheme again using a fresh key h. Thus, we are authenticating the state ρ in complementary bases. We call this the "Auth-QFT-Auth" scheme.

We will show that this in fact achieves total authentication (and hence encryption of the state), but at the cost of leaking the "outer key" h:

Theorem 12 (Security of the Auth-QFT-Auth scheme)**.** *The Auth-QFT-Auth scheme is* δ*-totally authenticating with outer key leakage, where* $\delta = O(\sqrt{|\mathcal{M}|^{5/2}/|\mathcal{Y}|})$.

While this scheme leaks some bits of the outer key, it preserves the secrecy of the state ρ and the "inner key" k. Furthermore, it is much more "lightweight" than the full unitary design scheme that achieves total authentication without key leakage. It also illustrates that applying a simple classical authentication scheme in complementary bases is already enough to reduce a full quantum adversary to performing only trivial attacks. Finally, the analysis of this scheme crucially relies on the basis dependent security definition above.

We note that Hayden, Leung, and Mayers show that the authentication scheme of [BCG+02] satisfies total authentication with key leakage [HLM11], but it is unclear whether it satisfies the strongest definition of total authentication without key leakage.

5 Properties of basis-dependent authentication

5.1 Indistinguishability from measured

Here, we show that any classical scheme that authenticates relative to the computational basis implies that the authenticated state is indistinguishable from being measured in the computational basis. For concreteness we will work with the computational basis; this is without loss of generality.

Definition 13. *If* Auth *is a classical scheme that is* ε-indistinguishable from measured *in the computational basis, then for any state* $\rho^{\mathcal{MZ}}$*, the following two states are* ε *close:*

- $\mathbb{E}_k \left[\mathsf{Auth}_k \otimes \mathbb{I}^{\mathcal{Z}}\right] (\rho^{\mathcal{MZ}})$ *(the unmeasured authenticated state), and*
- $\mathbb{E}_k \left[(\mathsf{Meas} \otimes \mathbb{I}^{\mathcal{Z}}) \circ (\mathsf{Auth}_k \otimes \mathbb{I}^{\mathcal{Z}})\right] ()\rho^{\mathcal{MZ}})$ *(the measured authenticated state), where* Meas *denotes measuring in the computational basis.*

Theorem 14. *If* $(\mathsf{Auth}, \mathsf{Ver})$ ε*-authenticates relative to the computational basis, then* Auth *is* $7\sqrt{\varepsilon}$*-indistinguishable from measured.*

Proof. We prove this theorem by contradiction: assuming an adversary can distinguish from measured, we will obtain a violation of the security of authentication. Analogous to the proof that authentication implies encryption of Barnum et al. [BCG+02], our proof will proceed in two parts. First, we will reduce to the case where we assume the distinguishing adversary has very high success probability. Second, we will show that by iterating the scheme, we boost a low success probability adversary into a high success probability adversary. For this proof, we will not need the full security where the key k is considered — instead, we will invoke the authentication security by tracing out and averaging over the key as in prior works.

Let $\rho^{\mathcal{M},\mathcal{Z}}$ be a quantum state. Let D be a distinguisher violating the indistinguishability from measured property. Suppose D has very large distinguishing advantage $1-\gamma$. This means that

- $D(\mathbb{E}_k \left[\mathsf{Auth}_k \otimes \mathbb{I}^{\mathcal{Z}}\right] (\rho^{\mathcal{MZ}}))$ outputs 1 with probability at least $1-\gamma$, and
- $D(\mathbb{E}_k \left[(\mathsf{Meas} \otimes \mathbb{I}^{\mathcal{Z}}) \circ (\mathsf{Auth}_k \otimes \mathbb{I}^{\mathcal{Z}})\right] ()\rho^{\mathcal{MZ}}))$ outputs 1 with probability at most γ

Now, we set up the state $(\frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1|) \otimes \rho^{\mathcal{MZ}}$. We conditionally measure $\rho^{\mathcal{M}}$ in the computational basis based on the first bit: if 0, we measure, if 1 we leave intact. Next, we discard the first bit by tracing it out. The resulting state is $\left[\frac{1}{2}(\mathsf{Meas} + \mathbb{I}^{\mathcal{M}}) \otimes \mathbb{I}^{\mathcal{Z}}\right] (\rho^{\mathcal{MZ}})$.

Now we authenticate. Since the scheme is classical, authentication commutes with measurement in the computational basis. Therefore, the authenticated state is

$$\left[\frac{1}{2}((\mathsf{Meas} + \mathbb{I}^{\mathcal{Y}}) \circ \mathsf{Auth}_k) \otimes \mathbb{I}^{\mathcal{Z}}\right] (\rho^{\mathcal{MZ}})$$

The adversary now applies D, copying the result into its auxiliary state. Because D has high distinguishing advantage, applying D and conditioning on D giving the right answer only negligibly affects the state. Therefore, it is straightforward to show the resulting state is $4\sqrt{2\gamma}$-close to:

$$\frac{1}{2}\mathop{\mathbb{E}}_{k}\left[\mathbb{I}^{\mathcal{Y}\mathcal{Z}}\circ(\mathsf{Auth}_k\otimes\mathbb{I}^{\mathcal{Z}})\right](\rho^{\mathcal{M}\mathcal{Z}})\otimes|1\rangle\langle 1|+\frac{1}{2}\mathop{\mathbb{E}}_{k}\left[(\mathsf{Meas}\otimes\mathbb{I}^{\mathcal{Z}})\circ(\mathsf{Auth}_k\otimes\mathbb{I}^{\mathcal{Z}})\right](\rho^{\mathcal{M}\mathcal{Z}})\otimes|0\rangle\langle 0|$$

Now, this state will pass verification with probability 1, since the authentication scheme is classical. Therefore, this state is approximated by an ideal adversary that is computational basis respecting. Note that such adversaries commute with the measurement in the computational basis. Therefore, the final bit in the approximated superposition is either 0 or 1 or some mixture of the two, but the mixture is independent of whether the authenticated space is measured or not.

Therefore, the state above has a distance of $\frac{1}{2}$ from any ideal adversary, a contradiction.

Thus, we have that if the scheme $\frac{1}{2}-4\sqrt{2\gamma}$-authenticates in the computational basis, there is no distinguisher with advantage $1-\gamma$.

Next, we show how to boost a low-advantage distinguisher for a scheme $(\mathsf{Auth},\mathsf{Ver})$ into a high-advantage distinguisher for the product scheme $(\mathsf{Auth}^t,\mathsf{Ver}^t)$ which acts on message space $\mathcal{M}^t$ by applying Auth to each message component with an independent key.

A simple hybrid argument shows that, if $(\mathsf{Auth},\mathsf{Ver})$ ε-authenticates in the computational basis, then $(\mathsf{Auth}^t,\mathsf{Ver}^t)$ $t\varepsilon$-authenticates in the computational basis. Note that Barnum et al.'s proof of this required somewhat more effort; however, for us, due tot he fact that we consider side information in our definition, in our case the security of the product scheme comes essentially for free.

Next, assume D distinguishes from measured for the state $\rho^{\mathcal{M}\mathcal{Z}}$ in the scheme $(\mathsf{Auth},\mathsf{Ver})$ with advantage δ. Then we can boost the success probability to a distinguisher D^t for the state $(\rho^{\mathcal{M}\mathcal{Z}})^{\otimes t}$ in scheme $(\mathsf{Auth}^t,\mathsf{Ver}^t)$ with advantage $1-2e^{-t\delta^2/2}$. But from the above, this means that the scheme $(\mathsf{Auth}^t,\mathsf{Ver}^t)$ cannot $\frac{1}{2}-8e^{-t\delta^2/4}$-authenticate. Thus,

$$t\varepsilon>\frac{1}{2}-8e^{-t\delta^2/4}$$

Choosing $t=1/3/\varepsilon$ gives $\delta<7\sqrt{\varepsilon}$.

□

5.2 Unforgeability

In this section we show that our security definition of authentication schemes relative to a basis implies the classical security definition of authentication schemes – namely, that the adversary, after having received the authenticated message state, cannot produce two distinct authenticated message-tag pairs with non-negligible probability. This property is called **unforgeability**. Thus this shows that our security definition recovers the Boneh-Zhandry security definition for one-time MACs.

Our model for signature forgery is the following. Let $(\mathsf{Auth},\mathsf{Ver})$ be a classical authentication scheme that is $\mathcal{B}$-respecting for some basis. We will let $\mathcal{B}$ be the computational basis without loss of generality. Furthermore, we will restrict our attention to MACs where for a classical message $m\in\mathcal{M}$, $\mathsf{Auth}_k(m)=(m,\sigma(k,m))$, although our arguments extend to general classical authentication schemes.

Without loss of generality we can assume that the initial message state is a pure state $|\rho\rangle^{\mathcal{MZ}} = \sum_m \alpha_m |m\rangle^{\mathcal{M}} \otimes |\varphi_m\rangle^{\mathcal{Z}}$ where the $|\varphi_m\rangle$ are arbitary pure states held by the adversary. After signing, we have

$$\tau^{\mathcal{KYZ}} = \mathop{\mathbb{E}}_k |k\rangle\langle k| \otimes \mathsf{Auth}_k(\rho^{\mathcal{MZ}}).$$

The adversary applies some superoperator $\mathcal{E}$ on $\mathcal{YZ}$ and outputs a system on $\mathcal{Y}_1\mathcal{Y}_2\mathcal{Z}$. The spaces $\mathcal{Y}_1$ and $\mathcal{Y}_2$ are both isomorphic to $\mathcal{Y}$. Let the tampered state be denoted as

$$\widetilde{\tau}^{\mathcal{KY}_1\mathcal{Y}_2\mathcal{Z}} = \mathop{\mathbb{E}}_k |k\rangle\langle k| \otimes \mathcal{E}(\mathsf{Auth}_k(\rho^{\mathcal{MZ}})).$$

We define the **probability of forgery by $\mathcal{E}$ on input ρ** to be the probability that, upon measuring $\mathcal{K}$, $\mathcal{Y}_1$, and $\mathcal{Y}_2$ in the computational basis, we obtain a key k and two valid signed messages $(m, \sigma(k,m))$ and $(m', \sigma(k,m'))$ with $m \neq m'$.

The next theorem shows that quantum-secure authentication schemes possess the unforgeability property. The idea of the proof is as follows: suppose that there was an authentication scheme $(\mathsf{Auth}, \mathsf{Ver})$, an adversary $\mathcal{E}$ and an initial message state $\rho^{\mathcal{M}}$ such that $\mathcal{E}$ on input ρ could forge an authenticated message with non-negligible probability. Using the fact that the authentication scheme is secure, we can in fact find a *fixed* message $m \in \mathcal{M}$ and another adversary $\widehat{\mathcal{E}}$ that, when given an authentication of message m, forges a valid signed message $(m', \sigma(k,m'))$ where $m' \neq m$ with non-negligible probability. The definition of secure authentication scheme easily implies this is impossible.

Theorem 15. *Let* $(\mathsf{Auth}, \mathsf{Ver})$ *be an authentication scheme that is ε-authenticating relative to the computational basis. Let $\mathcal{E}$ be a forger. Then for all initial message states $\rho^{\mathcal{MZ}}$, the probability of forgery by $\mathcal{E}$ on input ρ is at most* 3ε.

Proof. Suppose for contradiction that the probability of forgery is at least $\delta = 3\varepsilon$. Since the scheme is ε-authenticating relative to the computational basis, we can simulate the forger by an ideal adversary $\mathcal{I}$ that respects the computational basis: on input $\tau^{\mathcal{KYZ}}$ (the authentication of ρ), it first measures the $\mathcal{Y}$ register to yield a valid signed message $(m, \sigma(k,m))$. Then, conditioned on this result, it applies an arbitrary quantum operation on the $\mathcal{Z}$ register. Since $\mathcal{E}$ is a forger, the ideal adversary $\mathcal{I}$ is also a forger: measuring $\mathcal{KYZ}$ in the computational basis will yield k, $(m, \sigma(k,m))$ and $(m', \sigma(k,m'))$ where $m \neq m'$ with probability at least $\delta - \varepsilon = 2\varepsilon$. Let E_m denote the event that measuring $\mathcal{Y}$ yields a valid signature of the message m. Let F_m denote the event that measuring $\mathcal{Z}$ yields a valid signature of a message that's distinct from m.

Thus

$$\sum_m \Pr[E_m] \cdot \Pr[F_m | E_m] \geq 2\varepsilon$$

where the probabilities are with respect to the ideal adversary $\mathcal{I}$. Thus by averaging there exists an m where $\Pr[F_m|E_m] \geq 2\varepsilon$. But notice that $\Pr[E_m]$ is independent of the key, and simply $|\alpha_m|^2$, because the ideal adversary only measures the $\mathcal{Y}$ register of τ in the computational basis. Thus, if we condition the state $\mathcal{I}(\tau)$ on the event E_m, we have the following state:

$$\mathcal{I}(\tau^{\mathcal{KYZ}})\big|_{E_m} = \mathop{\mathbb{E}}_k |k\rangle\langle k|^{\mathcal{K}} \otimes |m, \sigma(k,m)\rangle\langle m, \sigma(k,m)|^{\mathcal{Y}} \otimes \mathcal{I}_{m,\sigma(k,m)} \left(|\varphi_m\rangle\langle\varphi_m|^{\mathcal{Z}} \right)$$

where $\mathcal{I}_{m,\sigma(k,m)}$ denotes the attack that the ideal adversary performs on the side information, conditioned on reading $(m, \sigma(k,m))$ in $\mathcal{Y}$. However, $\Pr[F_m|E_m] \geq 2\varepsilon$ implies that measuring $\mathbb{E}_k |k\rangle\langle k| \otimes \mathcal{I}_{m,\sigma(k,m)} \left(|\varphi_m\rangle\langle\varphi_m|^{\mathcal{Z}}\right)$ in the computational basis yields k and a forgery $(m', \sigma(k,m'))$ where $m' \neq m$ with probability at least 2ε. However, this is impossible, as $(m, \sigma(k,m))$ should have negligible information about a valid signature of m'.

□

6 Properties of total authentication

6.1 Encryption

Analogous to the Barnum et al.'s [BCG+02] result that authentication implies encryption, we show that authentication when considering side information must encrypt the state, even to an adversary that may be entangled with the state. This result is incompatible with Barnum et al.'s: we start from a stronger property that considers side information, and end with a stronger form of authentication that also considers side information.

Definition 16. *If* Auth *is an ε-secure encryption scheme with side information, then for any two states $\rho_0^{\mathcal{MZ}}, \rho_1^{\mathcal{MZ}}$ such that $\rho_0^{\mathcal{Z}}$ and $\rho_1^{\mathcal{Z}}$ are δ-close, the following two distributions are $\delta + \varepsilon$ close:*

- $\mathbb{E}_k \left[\mathsf{Auth}_k \otimes \mathbb{I}^{\mathcal{Z}}\right] (\rho_0^{\mathcal{MZ}})$ *and*
- $\mathbb{E}_k \left[\mathsf{Auth}_k \otimes \mathbb{I}^{\mathcal{Z}}\right] (\rho_1^{\mathcal{MZ}})$

Theorem 17. *If* $(\mathsf{Auth}, \mathsf{Ver})$ *ε-authenticates, then* Auth *is an $14\sqrt{\varepsilon}$-secure encryption scheme.*

Proof. First, we observe that any scheme that gives ε secure encryption in the case $\delta = 0$ gives 2ε secure encryption in the general case. Indeed, by assumption, $\mathbb{E}_k \left[\mathsf{Auth}_k \otimes \mathbb{I}^{\mathcal{Z}}\right] (\rho_0^{\mathcal{MZ}})$ is ε-close to $\mathbb{E}_k \mathsf{Auth}_k(|0\rangle\langle 0|) \otimes \rho_0^{\mathcal{Z}}$, which is δ close to $\mathbb{E}_k \mathsf{Auth}_k(|0\rangle\langle 0|) \otimes \rho_1^{\mathcal{Z}}$, which is ε close to $\mathbb{E}_k \left[\mathsf{Auth}_k \otimes \mathbb{I}^{\mathcal{Z}}\right] (\rho_1^{\mathcal{MZ}})$.

Therefore, it suffices to prove that Auth is $7\sqrt{\varepsilon}$ secure for states with $\delta = 0$.

We prove this theorem by contradiction: assuming an adversary can distinguish from measured, we will obtain a violation of the security of authentication. Our proof will very similar to the proof of Theorem 14. First, we will reduce to the case where we assume the distinguishing adversary has very high success probability. Second, we will show that by iterating the scheme, we boost a low success probability adversary into a high success probability adversary. For this proof, we will not need the full security where the key k is considered — instead, we will invoke the authentication security by tracing out and averaging over the key as in prior works.

Let $\rho_0^{\mathcal{M},\mathcal{Z}}, \rho_1^{\mathcal{M},\mathcal{Z}}$ be quantum states. Let D be a distinguisher that distinguishes between the two with probability τ. Suppose D has very large distinguishing advantage $1 - \gamma$. This means that

- $D(\mathbb{E}_k \left[\mathsf{Auth}_k \otimes \mathbb{I}^{\mathcal{Z}}\right] (\rho_1^{\mathcal{MZ}}))$ outputs 1 with probability at least $1 - \gamma$, and
- $D(\mathbb{E}_k \left[\mathsf{Auth}_k \otimes \mathbb{I}^{\mathcal{Z}}\right] (\rho_0^{\mathcal{MZ}})$ outputs 1 with probability at most γ

Now, we set up the state $\frac{1}{2}|0\rangle\langle 0| \otimes \rho_0^{\mathcal{MZ}} + \frac{1}{2}|1\rangle\langle 1| \otimes \rho_1^{\mathcal{MZ}}$. Next, we discard the first bit by tracing it out. The resulting state is $\frac{1}{2}(\rho_0^{\mathcal{MZ}} + \rho_1^{\mathcal{MZ}})$. Then we authenticate. By the linearity of quantum operations, we have that the state is

$$\frac{1}{2}(\left[\mathsf{Auth}_k \otimes \mathbb{I}^{\mathcal{Z}}\right](\rho_0^{\mathcal{MZ}}) + \left[\mathsf{Auth}_k \otimes \mathbb{I}^{\mathcal{Z}}\right](\rho_1^{\mathcal{MZ}}))$$

The adversary now applies D, copying the result into its auxiliary state. Because D has high distinguishing advantage, applying D and conditioning on D giving the right answer only negligibly affects the state. Therefore, it is straightforward to show the resulting state is $4\sqrt{2\gamma}$-close to:

$$\frac{1}{2}\mathop{\mathbb{E}}_k\left[\mathsf{Auth}_k \otimes \mathbb{I}^{\mathcal{Z}}\right](\rho_0^{\mathcal{MZ}}) \otimes |0\rangle\langle 0| + \frac{1}{2}\mathop{\mathbb{E}}_k\left[\mathsf{Auth}_k \otimes \mathbb{I}^{\mathcal{Z}}\right](\rho_1^{\mathcal{MZ}}) \otimes |1\rangle\langle 1|$$

Now, this state will pass verification with probability 1, since each component is a valid authenticated state and authentication is linear. Therefore, this state is approximated by an ideal adversary that does nothing except forward the state as is or reject the state, and modify its auxiliary registers independently of the authenticated state. Therefore, the final bit in the approximated superposition is either 0 or 1 or some mixture of the two, and the mixture may depend on $\rho^{\mathcal{Z}}$, but not $\rho^{\mathcal{M}}$. But recall that by assumption $\rho_0^{\mathcal{Z}} = \rho_1^{\mathcal{Z}}$, and so an ideal adversary cannot distinguish the two cases. Therefore, the state above has a distance of $\frac{1}{2}$ from any ideal adversary, a contradiction.

Thus, we have that if the scheme $\frac{1}{2} - 4\sqrt{2\gamma}$-authenticates in the computational basis, there is no distinguisher with advantage $1 - \gamma$.

Next, we show how to boost a low-advantage distinguisher for a scheme $(\mathsf{Auth}, \mathsf{Ver})$ into a high-advantage distinguisher for the product scheme $(\mathsf{Auth}^t, \mathsf{Ver}^t)$ which acts on message space $\mathcal{M}^t$ by applying Auth to each message component with an independent key.

A simple hybrid argument shows that, if $(\mathsf{Auth}, \mathsf{Ver})$ ε-authenticates, then $(\mathsf{Auth}^t, \mathsf{Ver}^t)$ $t\varepsilon$-authenticates in the computational basis.

Next, assume D distinguishes from measured for the state $\rho^{\mathcal{MZ}}$ in the scheme $(\mathsf{Auth}, \mathsf{Ver})$ with advantage δ. Then we can boost the success probability to a distinguisher D^t for the state $(\rho^{\mathcal{MZ}})^{\otimes t}$ in scheme $(\mathsf{Auth}^t, \mathsf{Ver}^t)$ with advantage $1 - 2e^{-t\delta^2/2}$. But from the above, this means that the scheme $(\mathsf{Auth}^t, \mathsf{Ver}^t)$ cannot $\frac{1}{2} - 8e^{-t\delta^2/4}$-authenticate. Thus,

$$t\varepsilon > \frac{1}{2} - 8e^{-t\delta^2/4}$$

Choosing $t = 1/3/\varepsilon$ gives $\delta < 7\sqrt{\varepsilon}$.

□

6.2 Quantum Key Distribution

Suppose we have a total authentication scheme. Then as argued in the Introduction, we immediately get a simple method to perform quantum key distribution. However, the QKD scheme sketched in the Introduction is rather fragile: any small amount of tampering by the adversary will cause Alice and Bob to abort. Here we sketch a slightly more robust way of carrying out QKD using a total authentication scheme.

Suppose Alice and Bob want to generate n bits of perfectly correlated key bits. We now describe a protocol that takes 2 rounds and $O(n \log n)$ bits of communication, and tolerates the adversary attacking at most $O(n/\log n)$ qubits of communication. If this is the case, then Alice and Bob can

distill at least $\Omega(n)$ bits of shared key. Let $(\mathsf{Auth}, \mathsf{Ver})$ be a scheme that encodes single qubits as $O(\log n)$ qubits, and is ε-totally authenticating for $\varepsilon = n^{-\Omega(1)}$. The unitary design scheme is one such example.

The QKD protocol is as follows:

1. Alice prepares the maximally entangled state over $2n$ qubits i.e. $|\Phi\rangle^{\mathcal{AB}} = \frac{1}{\sqrt{2^n}} \sum_{x \in \{0,1\}^n} |xx\rangle^{\mathcal{AB}}$.
2. Alice will generate independent keys $k_1, \ldots, k_n$ for n uses of the authentication scheme $(\mathsf{Auth}, \mathsf{Ver})$. She authenticates each of the n qubits on the $\mathcal{B}$-half of $|\Phi\rangle^{\mathcal{AB}}$ using an independent key. She sends $\mathcal{B}$ to Bob.
3. Bob sends a bit to Alice acknowledging that he received some state through the quantum channel (that may have been tampered by the adversary).
4. Alice sends the keys $k_1, \ldots, k_n$ over an authenticated, but non-private, classical channel.
5. On the quantum state he received, Bob performs the verification procedure $\mathsf{Ver}_{k_1} \otimes \cdots \mathsf{Ver}_{k_n}$ on n parts of $\log n$ qubits each. He relays to Alice which parts successfully passed verification. Let $S \subset [n]$ denote the successfully unauthenticated qubits.
6. Alice and Bob measure the part of their respective states corresponding to S in the computational basis, and use these bits as their shared key.

Since $(\mathsf{Auth}, \mathsf{Ver})$ is totally authenticating, after Bob successfully unauthenticates the qubits in S, the qubits shared between Alice and Bob in S will be $\approx \varepsilon n$-close to the maximally entangled state. Thus when they both measure, they will both share a keys (x, x') that are εn-close to uniform, perfectly correlated, and private from any other system (because the maximally entangled state is in tensor product with any other quantum system). If we assume that the probability that Bob successfully verifies is not too small, then this means that Alice and Bob have successfully performed quantum key distribution.

6.3 Key Reuse

Alice reuses the key once she gets back an acknowledgement from Bob that he accepted the authenticated state. We have ε-secure Total Authentication implying

$$||\mathbb{E}_k |k\rangle\langle k| \otimes [(\mathrm{Ver}_k \otimes \mathbb{I}^{\mathcal{Z}}) \circ \mathcal{O}(\sigma^{\mathcal{YZ}})] - \mathbb{E}_k |k\rangle\langle k| \otimes [(\mathrm{Ver}_k \otimes \mathbb{I}^{\mathcal{Z}}) \circ (\mathbb{I}^{\mathcal{Y}} \otimes \mathcal{S})(\sigma^{\mathcal{YZ}})]||_1 \leq \varepsilon$$

where $\mathcal{I} = \mathbb{I}^{\mathcal{Y}} \otimes \mathcal{S}$ is the ideal adversary. As in the ideal case, adversary never touches k and the authenticated state, final state after verification is completely unentangled with the key k and distribution of k is uniform. Therefore, for a scheme satisfying total authentication, when Bob accepts, the final state (including adversary's register) is close in trace distance to an ideal state and we can reuse the key k again.

7 Quantum MACs from 3-universal hashing

In the classical setting, secure one-time MACs can be constructed via universal hashing. Let $\{h_k\}_k$ be a strongly (2-)universal hash family. Then it is well known that the classical authenticiation

protocol $\mathsf{Auth}_k(m) = (m, h_k(m))$ is secure against classical adversaries [WC81]. Here, we show that the *same* authentication protocol is also quantum-secure, provided that the hash family $\{h_k\}_k$ satisfies the following: for all distinct m_1, m_2, m_3, the distribution of $(h_k(m_1), h_k(m_2), h_k(m_3))$ for a randomly chosen $k \in \mathcal{K}$ is uniform in $\mathcal{T}^3$. Such a family is called a 3-*universal hash family*. We will overload notation and use $k(\cdot)$ to denote the function $h_k(\cdot)$.

We note that Boneh and Zhandry showed that, when authenticating classical messages in the one-time setting, pairwise independence is sufficient to ensure that a quantum adversary cannot forge a new signed message, as long as the length of the tag is longer than the message! When the tag is shorter than the message, they showed that pairwise independence is insecure, and 3-wise independence is necessary.

Our analysis of the 3-wise independent Carter-Wegman MAC requires that, in order to obtain security against quantum side information, the message tag needs to be longer than the message. Thus it is conceivable that pairwise independence is sufficient for the same guarantee; we leave this as an open question.

Theorem 18. *Let $\mathcal{K} = \{k\}$ be a 3-universal hash family. Let $\mathsf{Auth}_k(m) = (m, k(m))$ and Ver_k be the corresponding verification function. Then the authentication scheme $(\mathsf{Auth}, \mathsf{Ver})$ is $O(\sqrt{|\mathcal{M}|/|\mathcal{T}|})$-authenticating relative to the computational basis.*

Before beginning the proof we first state what the implications for key length are. Suppose we wish to guarantee that the Carter-Wegman MAC is ε-authenticating relative to the computational basis, then $|\mathcal{M}|/|\mathcal{T}| \leq O(\varepsilon^2)$, which implies that $\log|\mathcal{T}| \geq \log|\mathcal{M}| + 2\log\frac{1}{\varepsilon} + O(1)$. To ensure three-wise independence, it is sufficient for the key to have length $3\log|\mathcal{M}| + 6\log\frac{1}{\varepsilon} + O(1)$.

Proof. To prove this, we need to show that for all message states $\rho^{\mathcal{MZ}}$ and all adversaries $\mathcal{E} \in \mathrm{T}(\mathcal{YZ}, \mathcal{YZ})$, the result of the QMAC is to reduce the action of the adversary on the authenticated message to an ideal, computational basis-respecting adversary.

We will concentrate on the case of signing pure state messages – this is because we can always purify the initial message state, and give the purification to the adversary. In other words, we will show that Carter-Wegman MAC is a quantum secure MAC when the initial message state is a state $|\rho\rangle^{\mathcal{MZ}} = \sum_m \alpha_m |m\rangle^{\mathcal{M}} \otimes |\varphi_m\rangle^{\mathcal{Z}}$. The register $\mathcal{M}$ corresponds to the message, and the register $\mathcal{Z}$ is held by the adversary.

It will actually be more useful to work with the Schmidt decomposition of $|\rho\rangle$, which we write as

$$|\rho\rangle^{\mathcal{MZ}} = \sum_z \sqrt{\lambda_z} \left(\sum_m \alpha_{zm} |m\rangle^{\mathcal{M}} \right) \otimes |\varphi_z\rangle^{\mathcal{Z}}$$

where for $z \neq z'$, we have $\langle \varphi_z | \varphi_{z'} \rangle = 0$, and the λ_z's are nonnegative numbers summing to 1. Furthermore, the dimension of the span of $\{|\varphi_z\rangle\}_z$ is at most $|\mathcal{M}|$.

After signing, the state becomes

$$\sigma^{\mathcal{KYZ}} = \mathop{\mathbb{E}}_k |k\rangle\langle k| \otimes \mathsf{Auth}_k(\rho)$$

where $\mathcal{Y} = \mathcal{MT}$. Now consider an attack $\mathcal{E}$ of the adversary. By Stinespring's Dilation Theorem, the superoperator $\mathcal{E}$ can be implemented by applying a unitary V on registers $\mathcal{YZ}$, as well as

some auxiliary register $\mathcal{Z}'$ held by the adversary, followed by a projective measurement P on $\mathcal{Z}\mathcal{Z}'$, followed by tracing out $\mathcal{Z}'$.

First, we will assume that the auxiliary space $\mathcal{Z}'$ is part of the purification in $|\rho\rangle^{\mathcal{M}\mathcal{Z}}$. Secondly, we will ignore the projector P for now, and handle it later.

We specify the action of V on $\mathcal{Y}\mathcal{Z}$ as

$$V : |m,t\rangle^{\mathcal{M}\mathcal{T}} \otimes |\varphi_z\rangle^{\mathcal{Z}} \mapsto |\psi_{mtz}\rangle^{\mathcal{M}\mathcal{T}\mathcal{Z}}$$

where $\{|\psi_{mtz}\rangle\}$ are a collection of states in $\mathcal{M}\mathcal{T}\mathcal{Z}$ such that for all $(m,t,z) \neq (m',t',z')$, $\langle\psi_{mtz}|\psi_{m't'z'}\rangle = 0$. Furthermore, write the states as follows:

$$|\psi_{mtz}\rangle = \sum_{a,b} \beta_{ab}^{mtz} |a,b\rangle \otimes |\phi_{ab}^{mtz}\rangle$$

where the $\{|\phi_{a,}^{mtz}\rangle\}$ are an arbitrary collection of unit vectors residing in the space $\mathcal{Z}$, and $|a,b\rangle$ are vectors in $\mathcal{Y} = \mathcal{M}\mathcal{T}$. Therefore after the attack we have

$$\widetilde{\sigma}^{\mathcal{K}\mathcal{Y}\mathcal{Z}} = \mathop{\mathbb{E}}_{k} |k\rangle\langle k| \otimes V \,\mathsf{Auth}_k(\rho)\, V^\dagger.$$

Now we apply the verification procedure to this state to obtain τ, where we've conditioned on the procedure accepting:

$$\tau^{\mathcal{K}\mathcal{Y}\mathcal{Z}} = \mathsf{Ver}(\widetilde{\sigma}^{\mathcal{K}\mathcal{Y}\mathcal{Z}}) = \mathop{\mathbb{E}}_{k} |k\rangle\langle k| \otimes \mathsf{Ver}_k \left(V \,\mathsf{Auth}_k(\rho)\, V^\dagger \right)$$

Note that τ does not have unit trace in general (because the verification procedure Ver_k may not pass with probability 1). For a fixed key k, we can write

$$|\tau_k\rangle = \mathsf{Ver}_k \, V \,\mathsf{Auth}_k |\rho\rangle = \sum_{z,m,a} \sqrt{\lambda_z}\alpha_{zm}\beta_{ak_a}^{mk_mz} \; |a\rangle^{\mathcal{M}} \otimes |\phi_{ak_a}^{mk_mz}\rangle^{\mathcal{T}}$$

where we abbreviate $k(m)$ and $k(a)$ by k_m and k_a respectively. We can decompose the vector $|\tau_k\rangle = |\tau_{k,ideal}\rangle + |\tau_{k,err}\rangle$ where

$$|\tau_{k,ideal}\rangle^{\mathcal{M}\mathcal{T}\mathcal{Z}} = \sum_{z,m} \sqrt{\lambda_z}\alpha_{zm}\beta_{mk_m}^{mk_mz} \; |m\rangle^{\mathcal{M}} \otimes |\phi_{mk_m}^{mk_mz}\rangle^{\mathcal{Z}} \tag{2}$$

$$|\tau_{k,err}\rangle^{\mathcal{M}\mathcal{T}\mathcal{Z}} = \sum_{z,m,a:a\neq m} \sqrt{\lambda_z}\alpha_{zm}\beta_{ak_a}^{mk_mz} \; |a\rangle^{\mathcal{M}} \otimes |\phi_{ak_a}^{mk_mz}\rangle^{\mathcal{Z}} \tag{3}$$

Thus $\tau^{\mathcal{K}\mathcal{Y}\mathcal{Z}} = \tau_{ideal} + \tau_{err}$ where

$$\tau_{ideal} = \mathop{\mathbb{E}}_{k} |k\rangle\langle k| \otimes |\tau_{k,ideal}\rangle\langle\tau_{k,ideal}|,$$

and let

$$\tau_{err} = \mathop{\mathbb{E}}_{k} |k\rangle\langle k| \otimes \left(|\tau_{k,ideal}\rangle\langle\tau_{k,err}| + |\tau_{k,err}\rangle\langle\tau_{k,ideal}| + |\tau_{k,err}\rangle\langle\tau_{k,err}| \right).$$

The τ_{ideal} represents the part of τ that looks like it underwent an *ideal* attack, while the term τ_{ideal} represents the rest of τ. We will bound this error term and show that its size is small within τ, and thus this will show that τ is close to the result of an ideal attack.

To bound the size of τ_{err}, we note that

$$\begin{aligned}
\|\tau_{err}\|_1 &\le \mathop{\mathbb{E}}_k [2\| \ |\tau_{k,ideal}\rangle\langle\tau_{k,err}| \ \|_1 + \| \ |\tau_{k,err}\rangle\langle\tau_{k,err}| \ \|_1] \\
&= \mathop{\mathbb{E}}_k \left[2\sqrt{\langle\tau_{k,err}|\tau_{k,err}\rangle \cdot \langle\tau_{k,ideal}|\tau_{k,ideal}\rangle} + \langle\tau_{k,err}|\tau_{k,err}\rangle \right] \\
&\le 3 \mathop{\mathbb{E}}_k \sqrt{\langle\tau_{k,err}|\tau_{k,err}\rangle} \\
&\le 3\sqrt{\mathop{\mathbb{E}}_k \langle\tau_{k,err}|\tau_{k,err}\rangle}
\end{aligned}$$

where in the equality we used that for two pure states $|\varphi\rangle$ and $|\psi\rangle$, $\| \ |\varphi\rangle\langle\psi| \ \|_1 = \sqrt{\langle\varphi|\varphi\rangle \cdot \langle\psi|\psi\rangle}$. In the second-to-last inequality we used that $\langle\tau_{k,ideal}|\tau_{k,ideal}\rangle \le 1$, and in the last inequality we used the concavity of the square-root function. Now,

$$\mathop{\mathbb{E}}_k \langle\tau_{k,err}|\tau_{k,err}\rangle = \mathop{\mathbb{E}}_k \sum_{\substack{z,z' \\ a,m,m' : a \notin \{m,m'\}}} \sqrt{\lambda_z \lambda_{z'}} \cdot \alpha_{zm} \overline{\alpha}_{z'm'} \cdot \beta_{ak_a}^{mk_m z} \overline{\beta}_{ak_a}^{m'k_{m'}z'} \cdot \langle \phi_{ak_a}^{m'k_{m'}z'} | \phi_{ak_a}^{mk_m z} \rangle \tag{4}$$

$$= \sum_{\substack{z,z' \\ a,m,m' : a \notin \{m,m'\}}} \sqrt{\lambda_z \lambda_{z'}} \cdot \alpha_{zm} \overline{\alpha}_{z'm'} \cdot \left(\mathop{\mathbb{E}}_k \beta_{ak_a}^{mk_m z} \overline{\beta}_{ak_a}^{m'k_{m'}z'} \cdot \langle \phi_{ak_a}^{m'k_{m'}z'} | \phi_{ak_a}^{mk_m z} \rangle \right) \tag{5}$$

Observe that, for every a, m, m' such that $a \notin \{m, m'\}$, k_a is independent of k_m and $k_{m'}$ (this is where we use 3-wise independence of k). Therefore, we can write

$$\mathop{\mathbb{E}}_k \beta_{ak_a}^{mk_m z} \overline{\beta}_{ak_a}^{m'k_{m'}z'} \cdot \langle \phi_{ak_a}^{m'k_{m'}z'} | \phi_{ak_a}^{mk_m z} \rangle = \mathop{\mathbb{E}}_{k,h} \beta_{ah_a}^{mk_m z} \overline{\beta}_{ah_a}^{m'k_{m'}z'} \cdot \langle \phi_{ah_a}^{m'k_{m'}z'} | \phi_{ah_a}^{mk_m z} \rangle$$

where the expectation on the right hand side is over two independent hash families k and h. We have equality because $(k_m, k_{m'}, k_a)$ and $(k_m, k_{m'}, h_a)$ are identically distributed.

This motivates us to define

$$\xi_1 = \mathop{\mathbb{E}}_{k,h} \sum_{z,z',m,m'} \sqrt{\lambda_z \lambda_{z'}} \cdot \alpha_{zm} \overline{\alpha}_{z'm'} \cdot \beta_{mh_m}^{mk_m z} \overline{\beta}_{mh_m}^{m'k_{m'}z'} \cdot \langle \phi_{mh_m}^{m'k_{m'}z'} | \phi_{mh_m}^{mk_m z} \rangle$$

$$\xi_2 = \mathop{\mathbb{E}}_{k,h} \sum_{z,z',m} \sqrt{\lambda_z \lambda_{z'}} \cdot \alpha_{zm} \overline{\alpha}_{z'm} \cdot \beta_{mh_m}^{mk_m z} \overline{\beta}_{mh_m}^{mk_m z'} \cdot \langle \phi_{mh_m}^{mk_m z'} | \phi_{mh_m}^{mk_m z} \rangle.$$

We will momentarily show that ξ_1 and ξ_2 are small in magnitude. Assuming this, we add ξ_1 and ξ_2 to (5) to get a nicer-looking sum:

$$(5) \ + \xi_1 + \overline{\xi}_1 - \xi_2 = \sum_{\substack{z,z' \\ a,m,m'}} \sqrt{\lambda_z \lambda_{z'}} \cdot \alpha_{zm} \overline{\alpha}_{z'm'} \cdot \left(\mathop{\mathbb{E}}_{k,h} \beta_{ah_a}^{mk_m z} \overline{\beta}_{ah_a}^{m'k_{m'}z'} \cdot \langle \phi_{ah_a}^{m'k_{m'}z'} | \phi_{ah_a}^{mk_m z} \rangle \right) \tag{6}$$

$$= \frac{1}{|\mathcal{T}|} \sum_{z,z',m,m'} \sqrt{\lambda_z \lambda_{z'}} \cdot \alpha_{zm} \overline{\alpha}_{z'm'} \cdot \mathop{\mathbb{E}}_k \sum_{a,b} \beta_{ab}^{mk_m z} \overline{\beta}_{ab}^{m'k_{m'}z'} \cdot \langle \phi_{ab}^{m'k_{m'}z'} | \phi_{ab}^{mk_m z} \rangle \tag{7}$$

$$= \frac{1}{|\mathcal{T}|} \sum_{z,m} \lambda_z \cdot |\alpha_{zm}|^2 \mathop{\mathbb{E}}_k \sum_{a,b} |\beta_{ab}^{mk_m z}|^2 \tag{8}$$

$$= \frac{1}{|\mathcal{T}|}. \tag{9}$$

To go from the second line to the third line we used the orthogonality conditions

$$\langle \psi_{m't'z'} | \psi_{mtz} \rangle = \sum_{a,b} \beta_{ab}^{mtz} \overline{\beta}_{ab}^{m't'z'} \langle \phi_{ab}^{m't'z'} | \phi_{ab}^{mtz} \rangle = 0$$

whenever $(m,t,z) \neq (m',t',z')$.

Now we bound the magnitudes of ξ_1 and ξ_2. We use Cauchy-Schwarz repeatedly to bound $|\xi_1|$:

$$|\xi_1| = \frac{1}{|\mathcal{T}|} \left| \mathop{\mathbb{E}}_k \sum_{z,z',m,m'} \sqrt{\lambda_z \lambda_{z'}} \cdot \alpha_{zm} \overline{\alpha}_{z'm'} \cdot \sum_b \beta_{mb}^{mk_mz} \overline{\beta}_{mb}^{m'k_{m'}z'} \cdot \langle \phi_{mb}^{m'k_{m'}z'} | \phi_{mb}^{mk_mz} \rangle \right| \tag{10}$$

$$\leq \frac{1}{|\mathcal{T}|} \mathop{\mathbb{E}}_k \sqrt{\sum_{z,z',m,m'} |\alpha_{zm}|^2 |\alpha_{z'm'}|^2 \cdot \left| \sum_b \beta_{mb}^{mk_mz} \overline{\beta}_{mb}^{m'k_{m'}z'} \cdot \langle \phi_{mb}^{m'k_{m'}z'} | \phi_{mb}^{mk_mz} \rangle \right|^2} \tag{11}$$

$$\leq \frac{1}{|\mathcal{T}|} \mathop{\mathbb{E}}_k \sqrt{\sum_{z,z',m,m'} |\alpha_{zm}|^2 |\alpha_{z'm'}|^2 \cdot \left(\sum_b |\beta_{mb}^{mk_mz}|^2 \cdot \| |\phi_{mb}^{mk_mz} \rangle \|^2 \right) \left(\sum_b |\beta_{mb}^{m'k_{m'}z'}|^2 \cdot \| |\phi_{mb}^{m'k_{m'}z'} \rangle \|^2 \right)} \tag{12}$$

$$\leq \frac{1}{|\mathcal{T}|} \mathop{\mathbb{E}}_k \sqrt{\sum_{z,z',m,m'} |\alpha_{zm}|^2 |\alpha_{z'm'}|^2 \cdot \left(\sum_{a,b} |\beta_{ab}^{mk_mz}|^2 \cdot \| |\phi_{ab}^{mk_mz} \rangle \|^2 \right) \left(\sum_{a,b} |\beta_{ab}^{m'k_{m'}z'}|^2 \cdot \| |\phi_{ab}^{m'k_{m'}z'} \rangle \|^2 \right)} \tag{13}$$

$$\leq \frac{1}{|\mathcal{T}|} \mathop{\mathbb{E}}_k \sqrt{\sum_{z,z',m,m'} |\alpha_{zm}|^2 |\alpha_{z'm'}|^2 \cdot 2} \tag{14}$$

$$\leq \sqrt{2} \frac{|\mathcal{M}|}{|\mathcal{T}|}. \tag{15}$$

In the last line, we used the fact that the dimension of the span of the $|\varphi_z\rangle$'s is at most $|\mathcal{M}|$. A similar calculation will show that $|\xi_2| \leq \sqrt{2}|\mathcal{M}|/|\mathcal{T}|$ as well. Putting everything together, we get that

$$\left| \mathop{\mathbb{E}}_k \langle \tau_{k,err} | \tau_{k,err} \rangle \right| \leq (1 + 3\sqrt{2}) |\mathcal{M}|/|\mathcal{T}|.$$

This implies that

$$\| \tau - \tau_{ideal} \|_1 \leq 3\sqrt{6/|\mathcal{T}|}.$$

Recall that we have ignored the final projector P that the real adversary $\mathcal{E}$ may have applied after applying the unitary V. Since P acts on $\mathcal{Z}$ only, it commutes with the verification operation, and thus we have that

$$\| P\tau P^\dagger - P\tau_{ideal} P^\dagger \|_1 \leq 3\sqrt{6/|\mathcal{T}|}.$$

where $P\tau P^\dagger = \mathbb{E}_k |k\rangle\langle k| \otimes \mathsf{Ver}_k \circ \mathcal{E} \circ \mathsf{Auth}_k(\rho^{\mathcal{MZ}})$, the true final state of the protocol.

Finally, we have to argue that $P\tau_{ideal}P^\dagger$ is actually equal to $\mathbb{E}_k |k\rangle\langle k| \otimes \mathsf{Ver}_k \circ \mathcal{I} \circ \mathsf{Auth}_k(\rho^{\mathcal{MZ}})$ for some computational basis-respecting adversary $\mathcal{I}$. The ideal adversary behaves as follows when given the $\mathcal{YZ}$ registers of $\sigma^{\mathcal{KYZ}}$:

1. The adversary prepares auxiliary registers $\mathcal{M}'\mathcal{T}'\mathcal{Z}_2$ in the $|0\cdots0\rangle$ state. The $\mathcal{Y}' = \mathcal{M}'\mathcal{T}'$ registers are isomorphic to $\mathcal{MT}$, and $\mathcal{Z}_2$ is a qubit register.

2. First the ideal adversary makes a copy of the $\mathcal{MT}$ registers in the computational basis and coherently stores the copy in auxiliary registers $\mathcal{M}'\mathcal{T}'$.

3. The ideal adversary then applies the original adversary unitary V to registers $\mathcal{M}'\mathcal{T}'\mathcal{Z}$.

4. The adversary checks whether the values of the $\mathcal{MT}$ and $\mathcal{M}'\mathcal{T}'$ registers are the same in the computational basis; if so, the $\mathcal{Z}_2$ qubit is set to $|0\rangle$, and the $\mathcal{M}'\mathcal{T}'$ registers are set to $|0\cdots 0\rangle$. Otherwise, it is kept at $|1\rangle$. In other words, the basis vector $|m,t,m',t',0\rangle^{\mathcal{MTM}'\mathcal{T}'\mathcal{Z}_2}$ is mapped to $|m,t,0\cdots 0\rangle^{\mathcal{MTM}'\mathcal{T}'\mathcal{Z}_2}$ iff $m=m'$ and $t=t'$.

5. The adversary measures the $\mathcal{Z}_2$ qubit register, and the $\mathcal{Z}$ register using the POVM element $\{P, \mathbb{I}-P\}$, and accepts only on outcome $|0\rangle$ for $\mathcal{Z}_2$ and P for $\mathcal{Z}$.

Observe that this ideal attack $\mathcal{I}$ can be implemented as

$$\mathcal{I}:\sigma^{\mathcal{YZ}}\mapsto \mathrm{Tr}_{\mathcal{Y}'\mathcal{Z}_2}\left((P\otimes|0\rangle\langle 0|^{\mathcal{Z}_2})V_{ideal}\sigma^{\mathcal{YZ}}V_{ideal}^{\dagger}(P\otimes|0\rangle\langle 0|^{\mathcal{Z}_2})\right)$$

where V_{ideal} is an isometry mapping the space $\mathcal{YZ}$ to the space $\mathcal{YY}'\mathcal{ZZ}_2$, $P\otimes|0\rangle\langle 0|^{\mathcal{Z}_2}$ is a projector acting on $\mathcal{ZZ}_2$, and $\mathrm{Tr}_{\mathcal{Y}'\mathcal{Z}_2}(\cdot)$ is the partial trace over system $\mathcal{Y}'\mathcal{Z}_2$. Furthermore, V_{ideal} is an isometry that leaves the $\mathcal{MT}$ registers unchanged, and hence is a computational basis-respecting adversary. Since $P\tau_{ideal}^{\mathcal{KYZ}}P^{\dagger}=\mathsf{Ver}\left(\mathcal{I}(\sigma^{\mathcal{KYZ}})\right)$, and this holds for every adversary and every input state, this implies that $(\mathsf{Auth},\mathsf{Ver})$ is $O(\sqrt{M/T})$-authenticating relative to the computational basis. □

8 Total authentication (with key leakage) from complementary classical authentication

In the previous section, we saw how the classical Carter-Wegman message authentication scheme is still secure even when used on a superposition of messages, and even if the adversary has access to quantum side information about the messages. Here, we will show that using the Carter-Wegman scheme as a primitive, we obtain *total quantum state authentication*, which implies encryption of the quantum state.

The quantum state authentication scheme is simple: the sender authenticates the message state using the Carter-Wegman MAC in the computational basis, and then authenticates again in the Fourier basis (using a new key). To verification procedure is the reverse of this: the receiver first checks the outer authentication, performs the inverse Fourier transform, and then checks the inner authentication. We call this the "Auth-QFT-Auth" scheme. This is pleasingly analogous to the quantum one-time pad (QOTP), which encrypts quantum data using the classical one-time pad in complementary bases. However, the QOTP does not have authentication properties. Our analysis requires the 3-wise independence property of the Carter-Wegman MAC.

There is one slight caveat: we show that Auth-QFT-Auth achieves total authentication *with key leakage*. That is, we argue that conditioned on the receiver verification succeeding, the effect of an arbitrary adversary is to have ignored the authenticated state, and only act on the adversary's side information, in a manner that may depend on the key used for the second authentication (what we call the "outer key"). In other words, we sacrifice the secrecy of the outer key, but in exchange we get complete quantum state encryption.

8.1 The Auth-QFT-Auth scheme

Let $|\rho\rangle^{\mathcal{MZ}} = \sum_m \alpha_m |m\rangle^{\mathcal{M}} \otimes |\varphi_m\rangle^{\mathcal{Z}}$ be the initial message state, where $\mathcal{Z}$ is held by the adversary. Again, it will be advantageous to rewrite this state in terms of the Schmidt decomposition:

$$|\rho\rangle^{\mathcal{MZ}} = \sum_z \sqrt{\lambda_z} \left(\sum_m \alpha_{zm} |m\rangle^{\mathcal{M}} \right) \otimes |\varphi_z\rangle^{\mathcal{Z}}$$

where for $z \neq z'$, we have $\langle \varphi_z | \varphi_{z'} \rangle = 0$, and the λ_z's are nonnegative numbers summing to 1. Furthermore, the dimension of the span of $\{|\varphi_z\rangle\}_z$ is at most $|\mathcal{M}|$.

The authentication scheme is the composed operation $\mathsf{Auth}_2(H^{\otimes N}(\mathsf{Auth}_1(\rho)))$, where Auth_1 is the *inner* authentication scheme that uses key k, $H^{\otimes N}$ is the quantum Fourier transform over $\mathbb{Z}_2$, and Auth_2 is the *outer* authentication that uses key h. The keys k and h are independent.

The inner authentication scheme Auth_1 maps $\mathcal{M}$ to $\mathcal{Y}_1 = \mathcal{MT}_1$. We define $N = |\mathcal{Y}_1|$. H is the single-qubit Hadamard unitary, and the Fourier transform $H^{\otimes N}$ acts on $\mathcal{Y}_1$. The outer authentication scheme Auth_2 maps $\mathcal{Y}_1$ to $\mathcal{Y}_2 = \mathcal{MT}_1\mathcal{T}_2$. The keys k and h live in the registers $\mathcal{K}$ and $\mathcal{H}$, respectively. The evolution of the initial message state is as follows:

1. **Inner authentication**. When the inner authentication key (henceforth called the *inner key*) is k, the state becomes
$$\sum_z \sqrt{\lambda_z} \left(\sum_m \alpha_{zm} |m, k(m)\rangle^{\mathcal{Y}_1} \right) \otimes |\varphi_z\rangle^{\mathcal{Z}}$$
2. **Fourier transform over $\mathbb{Z}_2$**: Let $\{|x\rangle\}$ be a basis for $\mathcal{Y}_1$. Then:
$$\frac{1}{\sqrt{N}} \sum_z \sqrt{\lambda_z} \left(\sum_{m,x} \alpha_{zm} (-1)^{(m,k(m))\cdot x} |x\rangle^{\mathcal{Y}_1} \right) \otimes |\varphi_z\rangle^{\mathcal{Z}}.$$
3. **Outer authentication**. The outer key is denoted by h. The final authenticated state is then
$$|\sigma_{kh}\rangle^{\mathcal{Y}\mathcal{T}_2\mathcal{Z}} = \frac{1}{\sqrt{N}} \sum_z \sqrt{\lambda_z} \left(\sum_{m,x} \alpha_{zm} (-1)^{(m,k(m))\cdot x} |x, h(x)\rangle^{\mathcal{Y}_1\mathcal{T}_2} \right) \otimes |\varphi_z\rangle^{\mathcal{Z}}$$
where $\mathcal{T}_2$ is the space of the tag $h(x)$.

Let
$$\sigma^{\mathcal{KHY}_1\mathcal{T}_2\mathcal{Z}} = \mathop{\mathbb{E}}_{kh} |kh\rangle\langle kh|^{\mathcal{KH}} \otimes |\sigma_{kh}\rangle\langle\sigma_{kh}|^{\mathcal{Y}_1\mathcal{T}_2\mathcal{Z}}.$$

The adversary is then given the $\mathcal{Y}_1\mathcal{T}_2$ registers of σ, and performs a general unitary attack V that acts on $\mathcal{Y}_1\mathcal{T}_2\mathcal{Z}$:
$$\widetilde{\sigma}^{\mathcal{KHY}_1\mathcal{T}_2\mathcal{Z}} = V\sigma V^\dagger.$$

Let $\widetilde{\tau}^{\mathcal{KHMZ}} = \mathsf{Auth}_1^{-1} \circ \mathsf{Ver}_1 \circ \mathrm{QFT}^{-1} \circ \mathsf{Auth}_2^{-1} \circ \mathsf{Ver}_2(\widetilde{\sigma})$.

Let the inner authentication scheme be the 3-wise independent hashing QMAC with tag length $\log T$, and message length $\log M$. Let the outer authentication scheme be a QMAC that ε-authenticates with respect to the computational basis.

The Auth-QFT-Auth scheme can potentially leak some bits of the outer key h, but we will show that this is the *only* thing that is leaked; otherwise, it is performs total authentication (and hence encryption).

Theorem 19 (Security of the Auth-QFT-Auth scheme)**.** *The Auth-QFT-Auth scheme is δ-totally authenticating with outer key leakage, where $\delta = \varepsilon + O(\sqrt{|\mathcal{M}|^{3/2}/|\mathcal{T}_1|})$.*

Again before starting the proof we consider the key requirements. The outer authentication scheme need not be a Carter-Wegman MAC, but let's assume that it is. In order to achieve δ-total authentication, the inner MAC must be such that $|\mathcal{M}|^{3/2}/|\mathcal{T}_1| \leq O(\delta^2)$, or in other words, $\log|\mathcal{T}_1| \geq \frac{3}{2}\log|\mathcal{M}| + 2\log\frac{1}{\delta} + O(1)$. The key needed for the inner MAC must be at least $\frac{9}{2}\log|\mathcal{M}| + 6\log\frac{1}{\delta} + O(1)$. The "message length" that is given to the outer MAC is $\log|\mathcal{M}| + \log|\mathcal{T}_1| \geq \frac{5}{2}\log|\mathcal{M}| + 2\log\frac{1}{\delta} + O(1)$, and thus $\log|\mathcal{T}_2| \geq \frac{5}{2}\log|\mathcal{M}| + 4\log\frac{1}{\delta} + O(1)$. The key length for the outer MAC needs to be at least $\frac{15}{2}\log|\mathcal{M}| + 12\log\frac{1}{\delta} + O(1)$, so the total key needed is $12\log|\mathcal{M}| + 18\log\frac{1}{\delta} + O(1)$.

While the inner key can be recycled (upon successful verification), the outer key unfortunately cannot be.

Proof. We will let $M = |\mathcal{M}|$, $T = |\mathcal{T}_1|$, and $N = MT = |\mathcal{Y}_1|$. We will assume that $M^{3/2} \leq T$; otherwise the theorem statement is vacuous.

Suppose the outer authentication scheme was ε-secure. By definition, there exists an ideal computational basis adversary $\mathcal{I}$ such that $\|\mathsf{Ver}_2(\widetilde{\sigma}) - \mathsf{Ver}_2(\mathcal{I}(\sigma))\|_1 \leq \varepsilon$, where Ver_2 denotes the verification procedure for the outer authentication scheme. There exists a computational basis-respecting linear map $\Lambda \in \mathrm{L}(\mathcal{Y}_2\mathcal{Z})$ such that

$$\mathcal{I} : \sigma \mapsto \Lambda\sigma\Lambda^\dagger.$$

Since Λ is computational basis-respecting, we have for all (x, s, z):

$$\Lambda|x,s\rangle^{\mathcal{Y}_1\mathcal{T}_2} \otimes |\varphi_z\rangle^{\mathcal{Z}} = |x,s\rangle^{\mathcal{Y}_1\mathcal{T}_2} \otimes |\phi_{xsz}\rangle^{\mathcal{Z}}.$$

for some collection of (not necessarily normalized) states $\{|\phi_{xsz}\rangle\}$.

Therefore the effect of the adversary on the authenticated state (after verification) is to be close to $\mathcal{I}(\sigma) = \mathbb{E}_{k,h}\,|kh\rangle\langle kh| \otimes |\tau_{kh}\rangle\langle\tau_{kh}|$ where for fixed inner/outer keys k, h

$$|\tau_{kh}\rangle = \frac{1}{N}\sum_z \sqrt{\lambda_z}\sum_{m,x}\alpha_{zm}(-1)^{(m,k(m))\cdot x}|x\rangle \otimes |\phi_{xh_xz}\rangle.$$

Thus, the final state that Bob has, after performing full (i.e. inner and outer) verification, is ε-close to

$$\mathop{\mathbb{E}}_{k,h}\,|kh\rangle\langle kh| \otimes |\mu_{kh}\rangle\langle\mu_{kh}|$$

where

$$|\mu_{kh}\rangle = \sum_z \sqrt{\lambda_z}\sum_m \left(\frac{1}{N}\sum_{x,m'}\alpha_{zm'}(-1)^{(m+m',k(m)+k(m'))\cdot x}\right)|m\rangle \otimes |\phi_{xh_xz}\rangle.$$

Then security of Auth-QFT-Auth is established if we show that for every h,

$$\mathop{\mathbb{E}}_{k}\,\||\mu_{kh}\rangle - |\nu_h\rangle\|^2$$

is small, where

$$|\nu_h\rangle^{\mathcal{M}\mathcal{Z}} = \sum_z \sqrt{\lambda_z}\sum_m \alpha_{zm}|m\rangle^{\mathcal{M}} \otimes |\eta_{hz}\rangle^{\mathcal{Z}}$$

with $|\eta_{hz}\rangle^{\mathcal{Z}} = \frac{1}{N}\sum_x |\phi_{xh_xz}\rangle^{\mathcal{Z}}$. Assuming this, the next Lemma will show that there is an ideal oblivious, but outer key-dependent, adversary whose actions lead to the global state $\mathbb{E}_{kh} |kh\rangle\langle kh| \otimes |\nu_h\rangle\langle\nu_h|$.

Lemma 20 (Constructing the ideal oblivious adversary)**.** *For all h there exists an ideal oblivious adversary $\mathcal{I}_h$ acting on $\mathcal{Z}$ only such that*

$$|\nu_h\rangle\langle\nu_h|^{\mathcal{MZ}} = \mathcal{I}_h(|\rho\rangle\langle\rho|^{\mathcal{MZ}}).$$

Proof. We now construct an ideal adversary $\mathcal{I}_h$, derived from the computational basis adversary $\mathcal{I}$. By definition of $\mathcal{I}$, there exists a computational basis-respecting isometry $V \in \mathrm{J}(\mathcal{Y}_2\mathcal{Z}, \mathcal{Y}_2\mathcal{Z}\mathcal{Y}_2'\mathcal{Z}_2)$ where $\mathcal{Y}_2'$ is an auxiliary register isomorphic to $\mathcal{Y}_2$, and $\mathcal{Z}_2$ is an auxiliary qubit register, such that

$$\mathcal{I} : \sigma^{\mathcal{YZ}} \mapsto \mathrm{Tr}_{\mathcal{Y}'\mathcal{Z}_2}\left(\Pi V \sigma^{\mathcal{YZ}} V^\dagger \Pi\right).$$

Here $\Pi = P \otimes |0\rangle\langle 0|^{\mathcal{Z}_2}$ for some projector P acting on $\mathcal{Z}$. Furthermore, V is computational basis respecting:

$$\Pi V |x,s\rangle^{\mathcal{Y}_2} \otimes |\varphi_z\rangle^{\mathcal{Z}} = |x,s\rangle^{\mathcal{Y}_2} \otimes |\phi_{xsz}\rangle^{\mathcal{Z}} \otimes |0\cdots 0\rangle^{\mathcal{Y}_2'\mathcal{Z}_2}$$

where the $|\phi_{xsz}\rangle^{\mathcal{Z}}$ were defined above.

Now we construct the ideal general adversary $\mathcal{I}_h$ as follows:

1. First, the adversary creates the entangled state $|\Phi_h\rangle^{\mathcal{AA}'} = \frac{1}{\sqrt{N}}\sum_x |x,h(x)\rangle^{\mathcal{A}}|x,h(x)\rangle^{\mathcal{A}'}$ in new registers $\mathcal{A}\otimes\mathcal{A}'$, which are isomorphic to $\mathcal{Y}_2\otimes\mathcal{Y}_2$, and $\{|x\rangle\}$ is a basis for $\mathcal{Y}_1$.
2. It then applies the unitary V to half of $|\Phi_h\rangle^{\mathcal{AA}'}$ that resides in $\mathcal{A}$, and the $\mathcal{Z}$ part of the input state $|\rho\rangle$.
3. The adversary measures $\mathcal{AA}'\mathcal{Z}\mathcal{Z}_2$ using the projective measurement $\{Q, \mathbb{I}-Q\}$, where $Q = |\Phi_h\rangle\langle\Phi_h|^{\mathcal{AA}'}\otimes\Pi$. The adversary discards the outcome corresponding to $\mathbb{I}-Q$, and leaves the state unnormalized:
$$\frac{1}{N}\sum_{z,x,m}\sqrt{\lambda_z}\alpha_{zm}|m\rangle^{\mathcal{M}}|\phi_{xsz}\rangle^{\mathcal{Z}}|\Phi\rangle^{\mathcal{AA}'}|0\cdots 0\rangle^{\mathcal{Y}_2'\mathcal{Z}_2}$$
4. The adversary discards the $\mathcal{AA}'\mathcal{Y}_2'\mathcal{Z}_2$ registers:
$$\frac{1}{N}\sum_{z,x,m}\sqrt{\lambda_z}\alpha_{zm}|m\rangle^{\mathcal{M}}\otimes|\phi_{xsz}\rangle^{\mathcal{Z}}$$

This is precisely the state $|\nu_h\rangle$, and the $\mathcal{I}_h$ only interacts with $\mathcal{Z}$ and auxiliary registers in the adversary's control, so it is an ideal general adversary. □

We now turn to bounding $\mathbb{E}_k \||\mu_{kh}\rangle - |\nu_h\rangle\|^2$:

$$\begin{aligned}
&\mathop{\mathbb{E}}_k \||\mu_{kh}\rangle - |\nu_h\rangle\|^2 \\
&= \frac{1}{N^2}\mathop{\mathbb{E}}_k \sum_{m,z,z'}\sqrt{\lambda_z\lambda_{z'}}\sum_{\substack{x',x'',m',m'' \\ m\notin\{m',m''\}}}\overline{\alpha}_{z'm'}\alpha_{zm''}(-1)^{(m+m',k(m)+k(m'))\cdot x'}(-1)^{(m+m'',k(m)+k(m''))\cdot x''}\langle\phi^h_{x'z'}|\phi^h_{x''z}\rangle
\end{aligned}$$

$$= \frac{1}{N^2} \sum_{\substack{m,z,z' \\ x',x'',m',m'' \\ m \notin \{m',m''\}}} \sqrt{\lambda_z \lambda_{z'}} \overline{\alpha}_{z'm'} \alpha_{zm''} (-1)^{(m+m')\cdot x_1' + (m+m'')\cdot x_2''} \langle \phi^h_{x'z'} | \phi^h_{x''z} \rangle \mathop{\mathbb{E}}_k (-1)^{(k(m)+k(m'))\cdot x_2'} (-1)^{(k(m)+k(m''))\cdot x_2''}.$$

We use the abbreviation $|\phi^h_{xz}\rangle = |\phi_{xh_xz}\rangle$. In the second line, we divided x into two parts (x_1, x_2), where x_1 corresponds to $\mathcal{M}$, and x_2 corresponds to $\mathcal{T}_1$. We focus on the expectation $\chi_{m,m',m'',x_2',x_2''} = \mathbb{E}_k (-1)^{(k(m)+k(m'))\cdot x_2'} (-1)^{(k(m)+k(m''))\cdot x_2''}$. We consider two cases:

Case 1: $m' = m'', m' \neq m$**.** Then $\chi_{m,m',m'',x_2',x_2''} = 0$ if $x_2' \neq x_2''$, otherwise $\chi_{m,m',m'',x_2',x_2''} = 1$.

$$\begin{aligned}
&\frac{1}{N^2} \left| \sum_{\substack{z,z',x',x'' \\ m,m': m \neq m'}} \sqrt{\lambda_z \lambda_{z'}} \overline{\alpha}_{z'm'} \alpha_{zm'} (-1)^{(m+m')\cdot(x_1' + x_1'')} \langle \phi^h_{x'z'} | \phi^h_{x''z} \rangle \chi_{m,m',m',x',x''} \right| \\
&= \frac{1}{N^2} \left| \sum_{\substack{z,z',x',x_1'' \\ m,m': m \neq m'}} \sqrt{\lambda_z \lambda_{z'}} \overline{\alpha}_{z'm'} \alpha_{zm'} (-1)^{(m+m')\cdot(x_1' + x_1'')} \langle \phi^h_{x'z'} | \phi^h_{x_1''x_2'z} \rangle \right| \\
&\leq \frac{1}{N^2} \sum_{z,z'} \sqrt{\lambda_z \lambda_{z'}} \sqrt{\sum_{m \neq m'} \left| \sum_{x', x_1''} (-1)^{(m+m')\cdot(x_1' + x_1'')} \langle \phi^h_{x'z'} | \phi^h_{x_1''x_2'z} \rangle \right|^2} \qquad \text{(Cauchy-Schwarz)} \\
&\leq \frac{1}{N^2} \sum_{z,z'} \sqrt{\lambda_z \lambda_{z'}} \sqrt{\sum_{m,m'} \sum_{x', x_1'', \tilde{x}', \tilde{x}_1''} (-1)^{(m+m')\cdot(x_1' + x_1'' + \tilde{x}_1' + \tilde{x}_1'')} \langle \phi^h_{\tilde{x}_1''\tilde{x}_2'z} | \phi^h_{\tilde{x}'z'} \rangle \langle \phi^h_{x'z'} | \phi^h_{x_1''x_2'z} \rangle} \\
&= \frac{1}{N^2} \sum_{z,z'} \sqrt{\lambda_z \lambda_{z'}} \sqrt{M^2 \sum_{\substack{x', x_1'', \tilde{x}', \tilde{x}_1'' \\ x_1' + x_1'' + \tilde{x}_1' + \tilde{x}_1'' = 0}} \langle \phi^h_{\tilde{x}_1''\tilde{x}_2'z} | \phi^h_{\tilde{x}'z'} \rangle \langle \phi^h_{x'z'} | \phi^h_{x_1''x_2'z} \rangle} \\
&\leq \frac{1}{N^2} \sum_{z,z'} \sqrt{\lambda_z \lambda_{z'}} \sqrt{M^3 N^2} \\
&\leq \frac{M^{5/2}}{N} \qquad \text{(at most } M \text{ } z\text{'s)} \\
&= \frac{M^{3/2}}{T}.
\end{aligned}$$

Case 2: m, m', m'' **are all distinct.** Then $\chi_{m,m',m'',x_2',x_2''} = 0$ unless $x_2' = x_2'' = 0$, in which case $\chi_{m,m',m'',x_2',x_2''} = 1$. This uses the three-independence of $k(\cdot)$.

$$\begin{aligned}
&\frac{1}{N^2} \left| \sum_{\substack{z,z',x',x'' \\ m,m',m'' \text{ distinct}}} \sqrt{\lambda_z \lambda_{z'}} \overline{\alpha}_{z'm'} \alpha_{zm'} (-1)^{(m+m')\cdot x_1' + (m+m'')\cdot x_1''} \langle \phi^h_{x'z'} | \phi^h_{x''z} \rangle \chi_{m,m',m'',x',x''} \right| \\
&= \frac{1}{N^2} \left| \sum_{\substack{z,z',x_1',x_1'' \\ m,m',m'' \text{ distinct}}} \sqrt{\lambda_z \lambda_{z'}} \overline{\alpha}_{z'm'} \alpha_{zm'} (-1)^{(m+m')\cdot x_1' + (m+m'')\cdot x_1''} \langle \phi^h_{x_1'0z'} | \phi^h_{x_1''0z} \rangle \right|
\end{aligned}$$

$$\leq \frac{1}{N^2}\sum_{z,z'}\sqrt{\lambda_z\lambda_{z'}}\sqrt{\sum_{m,m',m'' \text{ distinct}}\left|\sum_{x_1',x_1''}(-1)^{(m+m')\cdot x_1'+(m+m'')\cdot x_1''}\langle\phi^h_{x_1'0z'}|\phi^h_{x_1''0z}\rangle\right|^2} \qquad \text{(Cauchy-Schwarz)}$$

$$\leq \frac{M^{9/2}}{N^2}$$

$$\leq \frac{M^{3/2}}{T}$$

where we used the fact that $M^{3/2} \leq T$. Therefore, for every h we have

$$\mathop{\mathbb{E}}_k \||\mu_{kh}\rangle - |\nu_h\rangle\|^2 = O(M^{3/2}/T)$$

as desired. Using Fact 2 and Jensen's inequality, $\mathbb{E}_{kh}\||\mu_{kh}\rangle\langle\mu_{kh}| - |\nu_h\rangle\langle\nu_h|\| \leq O(\sqrt{M^{3/2}/T})$.

Thus, the final state of Bob is $\varepsilon + O(\sqrt{M^{3/2}/T})$-close to

$$\mathop{\mathbb{E}}_{kh}|kh\rangle\langle kh| \otimes |\nu_h\rangle\langle\nu_h| = \mathop{\mathbb{E}}_{kh}|kh\rangle\langle kh| \otimes \mathcal{I}_h(|\rho\rangle\langle\rho|)$$

where $\mathcal{I}_h$ are the ideal adversaries given by Lemma 20.

□

9 Total authentication from approximate unitary designs

We now present a scheme that satisfies the strongest security definition, that of total authentication (without *any* key leakage). In particular, this implies complete reuse of the entire key. This property of complete reuse of the key was not known before; it is not known whether the entire key can be reused in the authentication scheme of Barnum, et al [BCG+02].

This scheme is based on *unitary designs*, which are in some sense the quantum analogue of t-wise independent hash functions: a t-unitary design (also simply called a *t-design*) is a distribution $\mathscr{D}$ over unitary matrices such that degree t polynomials cannot distinguish between a unitary drawn from $\mathscr{D}$ and a fully random unitary. Furthermore, there are constructions of efficient unitary designs [BHH12].

9.1 The unitary design scheme

We call this scheme the *unitary design scheme*. Let s be a security parameter. The input state is $|\rho\rangle^{\mathcal{MZ}}$, where the $\mathcal{Z}$ register is held by the adversary.

1. The sender Alice first appends s $|0\rangle$ qubits in an auxiliary $\mathcal{T}$ register.
2. Using her secret key k, Alice samples a random unitary U_k drawn from an (approximate) unitary t-design that acts jointly on $\mathcal{M} \otimes \mathcal{T}$. We will set the parameter $t = 4$.
3. Alice applies U_k to the $\mathcal{M} \otimes \mathcal{T}$ register, and sends $\mathcal{M} \otimes \mathcal{T}$ across the quantum channel to Bob.
4. Bob receives some state, and applies the inverse unitary $U_k^\dagger$ to it. He measures the last s qubits and accepts if they all measure to be 0. Otherwise he rejects.

Theorem 21. *The unitary design scheme is efficiently computable, and is $2^{-s/2}$-totally authenticating.*

This is very similar to the *non-malleable quantum encryption scheme* proposed by Ambainis, Bouda, and Winter [ABW09]. A quantum encryption scheme is non-malleable if, in addition to revealing no information about the state to an eavesdropper, the eavesdropper cannot effect any controlled modifications to the encrypted state. Ambainis, Bouda and Winter show that applying a random unitary drawn from a 2-design to a state will encrypt it, and reduces the adversary to one that either forwards the state, or replaces it with the maximally mixed state. Clearly, such a scheme does not provide any authentication, but our scheme, where one additionally appends some dummy zeroes before authenticating, provides *both* encryption and authentication. Furthermore, their analysis does not handle the case of quantum side information, and it only gives a security guarantee *on average* over the key. Here, we will show that we obtain authentication and encryption *with high probability* over the key.

The key requirements of this scheme are rather significant, as constructions of approximate unitary 4-designs acting on n qubits involve choosing a random quantum circuit of size $\Theta(n^2)$, and thus the randomness required is at least $\Omega(n^2)$ [BHH12]. Furthermore, this scheme requires a full-fledged quantum computer running for at least $\Omega(n)$ sequential time steps. However we feel that this scheme is conceptually simple ("To encrypt and authenticate quantum data, apply a random quantum circuit for a while"), and it also confers the benefit that the *entire key* can be reused (upon successful verification), something that was not known before. We also believe that our analysis of this scheme may be of independent interest.

Notation and useful lemmas. We set up some notation. We let $\mathcal{M}$ denote the message space, $\mathcal{T}$ to denote the space of the dummy zero qubits. We let $\mathcal{Y} = \mathcal{M} \otimes \mathcal{T}$. We let $M = |\mathcal{M}|$, $|\mathcal{T}| = 2^s$, and $N = M2^s = |\mathcal{Y}|$.

Let $\mathcal{E}$ be an adversary acting on $\mathcal{Y} \otimes \mathcal{Z}$. By the Stinespring representation theorem, there exists a unitary V acting on a possibly larger space $\mathcal{Y} \otimes \mathcal{Z} \otimes \mathcal{Z}'$, followed by a projection P that acts on $\mathcal{Z}\mathcal{Z}'$, followed by a partial trace over $\mathcal{Z}'$. However without loss of generality we shall simply treat this additional space $\mathcal{Z}'$ as part of $\mathcal{Z}$, and ignore the partial trace operation. Thus, the adversary's action is to perform some unitary V on $\mathcal{Y} \otimes \mathcal{Z}$, followed by a projection on P on $\mathcal{Z}$.

To analyze the behavior of this scheme, we will first analyze the case when the randomizing unitary U is drawn from the Haar measure over the unitary group $U(\mathcal{Y})$, rather from a t-design. We will show that this scheme is totally authenticating. Then, we will show that actually using a $O(1)$-unitary design will suffice.

The crucial hammer we will need is a version of *Levy's Lemma*:

Definition 22. *A function $f : U(d) \to \mathbb{R}$ is η-Lipschitz if*

$$\sup_{U_1, U_2 \in U(d)} \frac{|f(U_1) - f(U_2)|}{\|U_1 - U_2\|_2} \leq \eta.$$

Lemma 23 (Levy's Lemma [MS09])**.** *Let $f : U(d) \to \mathbb{R}$ be an η-Lipschitz function on the unitary group of dimension d with mean $\mathbb{E} f$. Then*

$$\Pr\left(|f - \mathbb{E} f| \geq \delta\right) \leq 4 \exp\left(-\frac{Cd\delta^2}{\eta^2}\right)$$

where $C = 2/9\pi^3$ and the probability is over U drawn from the Haar measure on $U(d)$.

Another useful lemma we will need is the following, giving two formulas for averaging over the (Haar measure of the) unitary group. We use δ_{ij} to denote the Dirac delta function that is 1 if $i = j$ and 0 otherwise.

Lemma 24 (Appendix B.5 of [Bee97])**.** *For a function $f : \mathrm{U}(d) \to \mathbb{R}$, we let $\langle f \rangle$ to denote $\int f(U)\, dU$, where $\int \cdot dU$ is integration over the Haar measure on $\mathrm{U}(d)$. Then*

$$\langle U_{ab} U_{ij} U^*_{a'b'} U^*_{i'j'} \rangle = \frac{1}{d^2-1}(\delta_{aa'}\delta_{bb'}\delta_{ii'}\delta_{jj'} + \delta_{ai'}\delta_{bj'}\delta_{ia'}\delta_{jb'}) - \frac{1}{d(d^2-1)}(\delta_{aa'}\delta_{bj'}\delta_{ii'}\delta_{jb'} + \delta_{ai'}\delta_{bb'}\delta_{ia'}\delta_{jj'})$$

9.2 Total authentication with Haar-random unitaries

We now prove that the unitary design scheme yields total authentication. Let $\Lambda_U = \langle 0|^{\otimes s} U^\dagger V U |0\rangle^{\otimes s}$ is be a map from $\mathcal{M} \otimes \mathcal{Z}$ to $\mathcal{M} \otimes \mathcal{Z}$.

Lemma 25. *Let $N = \dim(\mathcal{Y})$. For all $\delta > 0$, for all initial message states $|\rho\rangle^{\mathcal{MY}}$ have that*

$$\Pr_U \left(\| \Gamma_V |\rho\rangle - \Lambda_U |\rho\rangle \|_2^2 \geq 2^{-s} + \delta \right) \leq \exp(-C' N \delta^2)$$

where $\Gamma_V = \mathrm{Tr}_{\mathcal{Y}}(V)/\dim(\mathcal{Y})$, C' is a universal constant, and U is a Haar-random unitary.

Proof. First, we write $|\rho\rangle^{\mathcal{MY}} = \sum_x \rho_x |x\rangle^{\mathcal{M}} \otimes |\varphi_x\rangle^{\mathcal{Z}}$ where $\{|x\rangle\}$ is a basis for $\mathcal{M}$, and $\{|\varphi_x\rangle\}$ are arbitrary unit vectors in $\mathcal{Z}$.

Write U as the following:

$$U = \sum_{u,x} |\psi_{u,x}\rangle\langle u, x|$$

where $|u\rangle \in \mathcal{T}, |x\rangle \in \mathcal{M}$ are standard basis vectors, and $\{|\psi_{u,x}\rangle\} \subset \mathcal{T} \otimes \mathcal{M}$ is a set of orthonormal unit vectors. Then $U|0\rangle^{\otimes s}$ becomes a linear operator that accepts vectors in $\mathcal{M}$ and outputs vectors in $\mathcal{Y} = \mathcal{T} \otimes \mathcal{M}$:

$$U|0\rangle^{\otimes s} = \sum_x |\psi_{0^s,x}\rangle\langle x|$$

We will simply write $|\psi_x\rangle$ to denote $|\psi_{0^s,x}\rangle$. We can write Λ_U as

$$\Lambda_U = \sum_{x,x'} |x'\rangle\langle x| \ \langle \psi_{x'}|V|\psi_x\rangle.$$

Let's compute the average operator

$$\int \Lambda_U \, dU = \sum_{x,x'} |x\rangle\langle x'| \ \int \langle \psi_x|V|\psi_{x'}\rangle \, dU \tag{16}$$

$$= \sum_x |x\rangle\langle x| \ \int \langle \psi_x|V|\psi_x\rangle \, dU + \sum_{x \neq x'} |x\rangle\langle x'| \ \int \langle \psi_x|V|\psi_{x'}\rangle \, dU \tag{17}$$

$$= \sum_x |x\rangle\langle x| \otimes \frac{1}{\dim(\mathcal{Y})} \mathrm{Tr}_{\mathcal{Y}}(V) \tag{18}$$

$$= \mathbb{I}^{\mathcal{M}} \otimes \Gamma_V \tag{19}$$

The second term in (17) (the sum over off-diagonal elements) averages to 0, because for $x \neq x'$, the vectors $|\psi_{x'}\rangle$ and $|\psi_x\rangle$ are random orthogonal unit vectors. Conditioned on a fixing of $|\psi_x\rangle$, for any vector $|\varphi\rangle$ that is orthogonal to $|\psi_x\rangle$, $|\psi_{x'}\rangle$ is equally likely to be $|\varphi\rangle$ or $-|\varphi\rangle$, so $\int \langle\psi_{x'}|V|\psi_x\rangle \, dU = 0$.

In the last step we used the fact that given an operator X mapping $\mathcal{Y} \otimes \mathcal{Z}$ to $\mathcal{Y} \otimes \mathcal{Z}$, if we average over the unit sphere, $\int (\langle\psi|^{\mathcal{Y}} \otimes \mathbb{I}^{\mathcal{Z}}) X (|\psi\rangle^{\mathcal{Y}} \otimes \mathbb{I}^{\mathcal{Z}}) \, d\psi$ is equal to the partial trace $\mathrm{Tr}_{\mathcal{Y}}(X)/\dim(\mathcal{Y})$. We'll let N denote $\dim(\mathcal{Y})$.

Thus, this tells us that on average, this operator should act as the identity on $\mathcal{M}$ and some linear map (not necessarily unitary) Γ_V on $\mathcal{Z}$. We now prove that Λ_U behaves this way on $|\rho\rangle$ with high probability. Define

$$f(U) = \| \Gamma_V |\rho\rangle - \Lambda_U |\rho\rangle \|_2^2.$$

Bounding the average of f. Expanding f and averaging, we get

$$\int f(U) \, dU = \int \left(\langle\rho|\Gamma_V^\dagger - \langle\rho|\Lambda_U^\dagger \right) \left(\Gamma_V|\rho\rangle - \Lambda_U|\rho\rangle \right) \, dU \tag{20}$$

$$= \int \langle\rho|\Gamma_V^\dagger\Gamma_V|\rho\rangle - \langle\rho|\Lambda_U^\dagger\Gamma_V|\rho\rangle - \langle\rho|\Gamma_V^\dagger\Lambda_U|\rho\rangle + \langle\rho|\Lambda_U^\dagger\Lambda_U|\rho\rangle \, dU \tag{21}$$

$$= -\langle\rho|\Gamma_V^\dagger\Gamma_V|\rho\rangle + \int \langle\rho|\Lambda_U^\dagger\Lambda_U|\rho\rangle \, dU \tag{22}$$

where in the last line we used our calculation of $\int \Lambda_U \, dU$ above. We bound this last term. We have that

$$\Lambda_U|\rho\rangle = \sum_{x,x'} \rho_{x'} |x\rangle \, \langle\psi_x| V (|\psi_{x'}\rangle \otimes |\varphi_{x'}\rangle)$$

Thus

$$\int \langle\rho|\Lambda_U^\dagger\Lambda_U|\rho\rangle \, dU = \int \sum_{x,x',x''} \rho_{x'}\rho_{x''}^* (\langle\psi_{x''}| \otimes \langle\varphi_{x''}|) V^\dagger |\psi_x\rangle\langle\psi_x| V (|\psi_{x'}\rangle \otimes |\varphi_{x'}\rangle) \, dU \tag{23}$$

$$= \sum_{x'} |\rho_{x'}|^2 \sum_x \int (\langle\psi_{x'}| \otimes \langle\varphi_{x'}|) V^\dagger |\psi_x\rangle\langle\psi_x| V (|\psi_{x'}\rangle \otimes |\varphi_{x'}\rangle) \, dU \tag{24}$$

$$= \sum_{x'} |\rho_{x'}|^2 \sum_x \int \left\| \langle\psi_x| V (|\psi_{x'}\rangle \otimes |\varphi_{x'}\rangle) \right\|_2^2 \, dU \tag{25}$$

$$= \sum_{x'} |\rho_{x'}|^2 \sum_{x \neq x'} \int \left\| \langle\psi_x| V (|\psi_{x'}\rangle \otimes |\varphi_{x'}\rangle) \right\|_2^2 \, dU + \tag{26}$$

$$\sum_x |\rho_x|^2 \int \left\| \langle\psi_x| V (|\psi_x\rangle \otimes |\varphi_x\rangle) \right\|_2^2 \, dU. \tag{27}$$

Let $\{|z\rangle\}$ be a basis for $\mathcal{Z}$. Now notice that

$$\left\| \langle\psi_x| V (|\psi_{x'}\rangle \otimes |\varphi_{x'}\rangle) \right\|_2^2 = \left\| \sum_z |z\rangle\langle z|^{\mathcal{Z}} \, \langle\psi_x| V (|\psi_{x'}\rangle \otimes |\varphi_{x'}\rangle) \right\|_2^2 \tag{28}$$

$$= \sum_z \left| (\langle\psi_x| \otimes \langle z|) V (|\psi_{x'}\rangle \otimes |\varphi_{x'}\rangle) \right|^2. \tag{29}$$

Write $|\varphi'_x\rangle = \sum_z \beta_{x'z}|z\rangle$. Then we have

$$\left|(\langle\psi_x| \otimes \langle z|)V(|\psi_{x'}\rangle \otimes |\varphi_{x'}\rangle)\right|^2 = \left|\sum_{z'} \beta_{x'z'}(\langle\psi_x| \otimes \langle z|)V(|\psi_{x'}\rangle \otimes |z'\rangle)\right|^2 \tag{30}$$

$$= \left|\sum_{z'} \beta_{x'z'} \sum_{ij} V_{(i,z),(j,z')} U^*_{ix} U_{jx'}\right|^2 \tag{31}$$

$$= \sum_{z',z''} \beta_{x'z'}\beta^*_{x'z''} \sum_{iji'j'} V_{(i,z),(j,z')} V^*_{(i',z),(j',z'')} U_{i'x} U_{jx'} U^*_{ix} U^*_{j'x'} \tag{32}$$

where the rows and columns of V are indexed by (i,z) and (j,z'), respectively. Again, we identify $|\psi_x\rangle$ as the x'th column of U, and U_{ix} denotes the i'th entry of $|\psi_x\rangle$.

We now go back to bound the sum over $x \neq x'$ in (27). Fix x, x' such that $x \neq x'$. Substituting (32) in and using Lemma 24, we get:

$$\int \left\|\langle\psi_{x'}|V(|\psi_x\rangle \otimes |\varphi_{x'}\rangle)\right\|_2^2 \, dU \tag{33}$$

$$= \int \sum_{z,z',z''} \beta_{x'z'}\beta^*_{x'z''} \sum_{iji'j'} V_{(i,z),(j,z')} V^*_{(i',z),(j',z'')} U_{i'x} U_{jx'} U^*_{ix} U^*_{j'x'} \, dU \tag{34}$$

$$= \sum_{z',z''} \beta_{x'z'}\beta^*_{x'z''} \left[\frac{1}{N^2-1} \sum_{ijz} V_{(i,z),(j,z')} V^*_{(i,z),(j,z'')} - \frac{1}{N(N^2-1)} \sum_{ii'z} V_{(i,z),(i,z')} V^*_{(i',z),(i',z'')}\right] \tag{35}$$

$$= \frac{N}{N^2-1} - \frac{1}{N(N^2-1)} \sum_{z',z''} \beta_{x'z'}\beta^*_{x'z''} \sum_{ii'z} V_{(i,z),(i,z')} V^*_{(i',z),(i',z'')} \tag{36}$$

$$= \frac{N}{N^2-1} - \frac{1}{N(N^2-1)} \sum_{z} \left|\sum_{z',i} \beta_{x',z'} V_{(i,z),(i,z')}\right|^2 \tag{37}$$

$$\leq \frac{N}{N^2-1} \tag{38}$$

where we used the fact that V is unitary and that $\sum_{z'} |\beta_{x'z'}|^2 = 1$. Summing (38) over all $x \neq x'$, we get

$$\sum_{x'} |\rho_{x'}|^2 \sum_{x \neq x'} \int \left\|\langle\psi_x|V(|\psi_{x'}\rangle \otimes |\varphi_{x'}\rangle)\right\|_2^2 \, dU \leq \sum_{x'} |\rho_{x'}|^2 \sum_{x \neq x'} \frac{N}{N^2-1} = \frac{N(M-1)}{N^2-1}.$$

Now fix an x; we bound the second term of (27). Using Lemma 24 again, we have

$$\int \left\|\langle\psi_x|V(|\psi_x\rangle \otimes |\varphi_x\rangle)\right\|_2^2 \, dU \tag{39}$$

$$= \sum_{z,z',z''} \beta_{xz'}\beta^*_{xz''} \int \sum_{iji'j'} V_{(i,z),(j,z')} V^*_{(i',z),(j',z'')} U_{i'x} U_{jx'} U^*_{ix} U^*_{j'x'} \, dU \tag{40}$$

$$= \left[\frac{1}{N^2-1} - \frac{1}{N(N^2-1)}\right] \cdot \left[\sum_{z} \left|\sum_{z',i} \beta_{x',z'} V_{(i,z),(i,z')}\right|^2 + N\right] \tag{41}$$

$$= \frac{1}{N(N+1)} \cdot \left[\sum_{z} \left|\sum_{z',i} \beta_{x',z'} V_{(i,z),(i,z')}\right|^2 + N\right] \tag{42}$$

Putting everything together, we can bound (27) by

$$\int \langle\rho|\Lambda_U^\dagger \Lambda_U|\rho\rangle \, dU \leq \frac{N(M-1)}{N^2-1} + \frac{1}{N(N+1)} \cdot \left[\sum_z \left|\sum_{z',i} \beta_{x',z'} V_{(i,z),(i,z')}\right|^2 + N\right] \tag{43}$$

$$= \frac{NM-1}{N^2-1} + \frac{1}{N(N+1)} \sum_z \left|\sum_{z',i} \beta_{x',z'} V_{(i,z),(i,z')}\right|^2. \tag{44}$$

We have to compare this to $\langle\rho|\Gamma_V^\dagger \Gamma_V|\rho\rangle = \|\Gamma_V|\rho\rangle\|_2^2$. We expand $\Gamma_V|\rho\rangle$:

$$\Gamma_V|\rho\rangle = \frac{1}{N}\mathrm{Tr}_{\mathcal{Y}}(V)|\rho\rangle \tag{45}$$

$$= \frac{1}{N}\sum_{i,z} |z\rangle^{\mathcal{Z}} \langle i,z|V|i\rangle \left(\sum_{x,z'} \rho_x \beta_{xz'} |x\rangle^{\mathcal{M}} \otimes |z'\rangle^{\mathcal{Z}}\right) \tag{46}$$

$$= \frac{1}{N}\sum_{x,z} \rho_x |x,z\rangle^{\mathcal{M}\mathcal{Z}} \left(\sum_{i,z'} \beta_{xz'} \langle i,z|V|i,z'\rangle\right) \tag{47}$$

$$= \frac{1}{N}\sum_{x,z} \rho_x |x,z\rangle^{\mathcal{M}\mathcal{Z}} \left(\sum_{i,z'} \beta_{xz'} V_{(i,z),(i,z')}\right) \tag{48}$$

So therefore

$$\langle\rho|\Gamma_V^\dagger \Gamma_V|\rho\rangle = \frac{1}{N^2}\sum_z \left|\sum_{z',i} \beta_{x',z'} V_{(i,z),(i,z')}\right|^2.$$

This shows that our desired average of f is small:

$$\int f(U) \, dU \leq \frac{NM-1}{N^2-1}.$$

Bounding the Lipschitz constant of f. We compute the Lipschitz continuity of f in parts. Let $g(U) = \langle\rho|\Gamma_V^\dagger \Lambda_U|\rho\rangle$, where $|\rho\rangle = \sum_x \rho_x |x\rangle \otimes |\varphi_x\rangle$. Expanding, we get

$$g(U) = \langle\rho|(\mathbb{I}^{\mathcal{Y}} \otimes \Gamma_V^\dagger) \sum_{x,x'} |x\rangle\langle x'| \otimes \langle\psi_x|V|\psi_{x'}\rangle|\rho\rangle \tag{49}$$

$$= \sum_{x,x'} \rho_x^* \rho_{x'} (\langle\psi_x| \otimes \langle\varphi_x|)\Gamma_V^\dagger V(|\psi_{x'}\rangle \otimes |\varphi_{x'}\rangle) \tag{50}$$

$$= \left(\sum_x \rho_x^* \langle\psi_x| \otimes \langle\varphi_x|\right) \Gamma_V^\dagger V \left(\sum_x \rho_x |\psi_x\rangle \otimes |\varphi_x\rangle\right) \tag{51}$$

$$= \langle\theta|\Gamma_V^\dagger V|\theta\rangle \tag{52}$$

where we used that $\Gamma_V^\dagger$ is an operator that acts on $\mathcal{Z}$ only, and we define $|\theta\rangle = \sum_x \rho_x |\psi_x\rangle \otimes |\varphi_x\rangle$. Thus for two unitaries $U, \widehat{U}$, we have

$$\left|g(U) - g(\widehat{U})\right| = \left|\langle\theta|\Gamma_V^\dagger V|\theta\rangle - \langle\widehat{\theta}|\Gamma_V^\dagger V|\widehat{\theta}\rangle\right| \tag{53}$$

$$= \left|\mathrm{Tr}\left(\Gamma_V^\dagger V(|\theta\rangle\langle\theta| - |\widehat{\theta}\rangle\langle\widehat{\theta}|)\right)\right| \tag{54}$$

$$\leq \left\|\Gamma_V^\dagger V\right\|_\infty \cdot \| \, |\theta\rangle\langle\theta| - |\widehat{\theta}\rangle\langle\widehat{\theta}| \, \|_1 \tag{55}$$

where in the inequality we used Hölder's inequality for matrices: $\mathrm{Tr}(AB) \leq \|A\|_\infty \|B\|_1$. Now, the operator norm is submultiplicative, so $\|\Gamma_V^\dagger V\|_\infty \leq \|\Gamma_V^\dagger\|_\infty \cdot \|V\|_\infty \leq \|\Gamma_V^\dagger\|_\infty$, because V is a unitary and hence its operator norm is 1. But then $\|\Gamma_V^\dagger\|_\infty = \frac{1}{N}\|\sum_y \langle y|V|y\rangle\|_\infty \leq \frac{1}{N}\sum_y \|\langle y|V|y\rangle\|_\infty$, because the operator norm satisfies the triangle inequality. Here, $|y\rangle$ is a basis element of $\mathcal{Y}$, and $\langle y|V|y\rangle$ is an operator that maps $\mathcal{Z}$ to $\mathcal{Z}$. For each y, we can bound $\|\langle y|V|y\rangle\|_\infty \leq 1$. This implies that $|g(U) - g(\widehat{U})| \leq \||\theta\rangle\langle\theta| - |\widehat{\theta}\rangle\langle\widehat{\theta}|\|_1$.

Thus the Lipschitz constant of g can be bounded by

$$\eta_g \leq \sup_{U,\widehat{U}} \frac{2\||\theta\rangle - |\widehat{\theta}\rangle\|_2}{\|U - \widehat{U}\|_2}.$$

Since the columns of $U, \widehat{U}$ are $|\psi_{u,x}\rangle$ and $|\widehat{\psi}_{u,x}\rangle$, the denominator $\|U - \widehat{U}\|_2$ can be written as $\sqrt{\sum_{u,x} \||\psi_{u,x}\rangle - |\widehat{\psi}_{u,x}\rangle\|_2^2}$. Notice that the numerator only depends on the column vectors $|\psi_{0^s,x}\rangle = |\psi_x\rangle$ and $|\widehat{\psi}_{0^s,x}\rangle = |\widehat{\psi}_x\rangle$, so the denominator can be minimized to to be $\sqrt{\sum_x \||\psi_x\rangle - |\widehat{\psi}_x\rangle\|_2^2}$ without affecting the numerator. The numerator can be bounded as

$$\begin{aligned}
\|\sum_x \rho_x(|\psi_x\rangle \otimes |\varphi_x\rangle - |\widehat{\psi}_x\rangle \otimes |\varphi_x\rangle)\|_2 &\leq \sum_x |\rho_x| \cdot \||\psi_x\rangle \otimes |\varphi_x\rangle - |\widehat{\psi}_x\rangle \otimes |\varphi_x\rangle\|_2 && (56)\\
&\leq \sqrt{\sum_x |\rho_x|^2 \sum_x \||\psi_x\rangle - |\widehat{\psi}_x\rangle\|_2^2} && (57)\\
&= \sqrt{\sum_x \||\psi_x\rangle - |\widehat{\psi}_x\rangle\|_2^2} && (58)
\end{aligned}$$

where in the first line we used the triangle inequality, and in the second line we used Cauchy-Schwarz. Thus the Lipschitz constant of g is at most 2.

Now we bound the Lipschitz continuity of $h(U) = \langle\rho|\Lambda_U^\dagger \Lambda_U|\rho\rangle$. We have that

$$\begin{aligned}
h(U) &= \sum_{x,x',x''} \rho_{x'}\rho_{x''}^*(\langle\psi_{x''}| \otimes \langle\varphi_{x''}|)V^\dagger|\psi_x\rangle\langle\psi_x|V(|\psi_{x'}\rangle \otimes |\varphi_{x'}\rangle) && (59)\\
&= \sum_x \langle\theta|V^\dagger|\psi_x\rangle\langle\psi_x|V|\theta\rangle && (60)\\
&= \mathrm{Tr}\left(\sum_x |\psi_x\rangle\langle\psi_x|V|\theta\rangle\langle\theta|V^\dagger\right) && (61)
\end{aligned}$$

where $|\theta\rangle$ is the same as above. Let $\Pi_U = \sum_x |\psi_x\rangle\langle\psi_x|$. Therefore

$$\begin{aligned}
|h(U) - h(\widehat{U})| &= \left|\mathrm{Tr}\left(\Pi_U V|\theta\rangle\langle\theta|V^\dagger - \Pi_{\widehat{U}} V|\widehat{\theta}\rangle\langle\widehat{\theta}|V^\dagger\right)\right| && (62)\\
&= \left|\mathrm{Tr}\left(\Pi_U V\left(|\theta\rangle\langle\theta| - |\widehat{\theta}\rangle\langle\widehat{\theta}|\right)V^\dagger + (\Pi_U - \Pi_{\widehat{U}})V|\widehat{\theta}\rangle\langle\widehat{\theta}|V^\dagger\right)\right| && (63)\\
&\leq \left\||\theta\rangle\langle\theta| - |\widehat{\theta}\rangle\langle\widehat{\theta}|\right\|_1 + \left|\langle\widehat{\theta}|V^\dagger(\Pi_U - \Pi_{\widehat{U}})V|\widehat{\theta}\rangle\right| && (64)\\
&\leq \left\||\theta\rangle\langle\theta| - |\widehat{\theta}\rangle\langle\widehat{\theta}|\right\|_1 + \left\|\Pi_U - \Pi_{\widehat{U}}\right\|_\infty && (65)
\end{aligned}$$

where in the first inequality we use that $\Pi_U = \Pi_U \cdot \Pi_U$ is a projector, and that $\mathrm{Tr}(\Pi X) \leq \|X\|_1$ for all operators X and $-\mathbb{I} \leq \Pi \leq \mathbb{I}$. The second term can be bounded by

$$\left\|\Pi_U - \Pi_{\widehat{U}}\right\|_\infty = \sup_{|v\rangle} \left\|(\Pi_U - \Pi_{\widehat{U}})|v\rangle\right\|_2 \tag{66}$$

$$= \sup_{|v\rangle} \left\| \sum_x (|\psi_x\rangle - |\widehat{\psi}_x\rangle)(\langle\psi_x|v\rangle - \langle\widehat{\psi}_x|v\rangle) \right\|_2 \tag{67}$$

$$\leq \sup_{|v\rangle} \sum_x \left\| |\psi_x\rangle - |\widehat{\psi}_x\rangle \right\|_2 \cdot \left| \langle\psi_x|v\rangle - \langle\widehat{\psi}_x|v\rangle \right| \tag{68}$$

$$\leq \sup_{|v\rangle} \sum_x (|\langle\psi_x|v\rangle| + |\langle\widehat{\psi}_x|v\rangle|) \cdot \left\| |\psi_x\rangle - |\widehat{\psi}_x\rangle \right\|_2 \tag{69}$$

$$\leq \sup_{|v\rangle} \sqrt{\sum_x |\langle\psi_x|v\rangle|^2 \sum_x \left\| |\psi_x\rangle - |\widehat{\psi}_x\rangle \right\|_2^2} + \sqrt{\sum_x |\langle\widehat{\psi}_x|v\rangle|^2 \sum_x \left\| |\psi_x\rangle - |\widehat{\psi}_x\rangle \right\|_2^2} \tag{70}$$

$$\leq 2\sqrt{\sum_x \left\| |\psi_x\rangle - |\widehat{\psi}_x\rangle \right\|_2^2}. \tag{71}$$

Therefore the Lipschitz constant η_h of h is at most 4, so the Lipschitz constant η of f is at most 8.

Now we invoke Levy's Lemma once more, and we obtain

$$\Pr\left(\|\Gamma_V|\rho\rangle - \Lambda_U|\rho\rangle\|_2^2 \geq \delta \right) \leq 4\exp\left(-\frac{CN\delta^2}{\eta^2} \right) \tag{72}$$

$$\leq 4\exp\left(-C'M^2/N \right) \tag{73}$$

where $\delta = 2M/N$ and C' is some universal constant.

□

9.3 Constructing the ideal oblivious adversary

Now we demonstrate that the map $|\rho\rangle^{\mathcal{MZ}} \mapsto \Gamma_V|\rho\rangle^{\mathcal{MZ}}$ can be implemented by an ideal oblivious adversary.

Consider the following ideal adversary, which given a state $|\sigma\rangle^{\mathcal{YZ}}$ performs the following:

1. First, the adversary creates a maximally entangled state $|\Phi\rangle^{\mathcal{Y}'\mathcal{Y}''} = \frac{1}{\sqrt{N}} \sum_y |yy\rangle^{\mathcal{Y}'\mathcal{Y}''}$ in new registers $\mathcal{Y}' \otimes \mathcal{Y}''$.

2. It then applies the unitary V to half of $|\Phi\rangle^{\mathcal{Y}'\mathcal{Y}''}$ that resides in $\mathcal{Y}'$, and the $\mathcal{Z}$ part of $|\sigma\rangle^{\mathcal{YZ}}$. The state currently looks like:

$$\frac{1}{\sqrt{N}} \sum_y (\mathbb{I}^{\mathcal{Y}} \otimes V^{\mathcal{Z}\mathcal{Y}'})|\sigma\rangle^{\mathcal{YZ}} \otimes |yy\rangle^{\mathcal{Y}'\mathcal{Y}''} \tag{74}$$

$$= \frac{1}{\sqrt{N}} \sum_y (\mathbb{I}^{\mathcal{Y}} \otimes \mathbb{I}^{\mathcal{Y}'\mathcal{Y}''} \otimes V^{\mathcal{Z}\mathcal{Y}'})|\sigma\rangle^{\mathcal{YZ}} \otimes |yy\rangle^{\mathcal{Y}'\mathcal{Y}''} \tag{75}$$

$$= \frac{1}{\sqrt{N}} \sum_{y,y'} (\mathbb{I}^{\mathcal{Y}} \otimes \langle y'|^{\mathcal{Y}'} V^{\mathcal{Z}\mathcal{Y}'} |y\rangle^{\mathcal{Y}'})|\sigma\rangle^{\mathcal{YZ}} \otimes |y'y\rangle^{\mathcal{Y}'\mathcal{Y}''} \tag{76}$$

3. The adversary projects $\mathcal{Y}'\mathcal{Y}''$ using the projector $|\Phi\rangle\langle\Phi|^{\mathcal{Y}'\mathcal{Y}''}$ (and leaves the state unnormalized):

$$\frac{1}{N} \sum_y (\mathbb{I}^{\mathcal{Y}} \otimes \langle y|^{\mathcal{Y}'} V^{\mathcal{Z}\mathcal{Y}'} |y\rangle^{\mathcal{Y}'})|\sigma\rangle^{\mathcal{YZ}} \otimes |\Phi\rangle^{\mathcal{Y}'\mathcal{Y}''}$$

4. The adversary discards the $\mathcal{Y}'\mathcal{Y}''$ register:

$$\frac{1}{N}\sum_y (\mathbb{I}^{\mathcal{Y}} \otimes \langle y|^{\mathcal{Y}'} V^{\mathcal{Z}\mathcal{Y}'} |y\rangle^{\mathcal{Y}'}) |\sigma\rangle^{\mathcal{Y}\mathcal{Z}}$$

This is precisely the state $\Gamma_V |\sigma\rangle^{\mathcal{Y}\mathcal{Z}}$, and the adversary described above never touches the $\mathcal{Y}$ register, so it is ideal.

9.4 Derandomizing the analysis using approximate unitary designs

The analysis of this scheme is nearly complete; however, the main missing component is that the analysis above assumes that the authentication scheme uses a truly random unitary U to scramble the message state and the tag. Unfortunately, sampling a truly random unitary on n qubits and applying it is infeasible: only a vanishing fraction of unitaries are succinctly describable or are efficiently computable.

The authentication scheme instead samples a unitary from a *unitary design*, discussed earlier. These are efficiently sampleable, efficiently computable ensembles of unitaries that are *pseudorandom*: they fool polynomials of low degree.

It won't be necessary to present formal definitions of a unitary design; we will use them in a black box manner. We will appeal to a general derandomization result of Low who proved that, if one establishes a measure of concentration result for a low degree polynomial f that's evaluated on a Haar-random unitary, then it still satisfies (nearly) the same measure of concentration when f is evaluated on a unitary drawn from an approximate t-design. More formally:

Theorem 26 ([Low09])**.** *Let $f : U(N) \to \mathbb{R}$ be a polynomial of degree K. Let $f(U) = \sum_i \alpha_i M_i(U)$ where $M_i(U)$ are monomials and let $\alpha(f) = \sum_i |\alpha_i|$. Suppose that f has probability concentration*

$$\Pr_{U\sim\nu_{Haar}}(|f-\mu| \geq \delta) \leq C\exp(-a\delta^2)$$

and let μ be an ε-approximate unitary t-design. Then

$$\Pr_{U\sim\mu}(|f-\mu| \geq \delta) \leq \frac{1}{\delta^{2m}}\left(C\left(\frac{m}{a}\right)^m + \varepsilon(\alpha+|\mu|)^{2m}\right)$$

for integer m with $2mK \leq t$.

Furthermore, there exist efficient constructions of approximate t-unitary designs, for any t.

Theorem 27 ([BHH12])**.** *For every ε, t, and n, there exists a finite set of unitaries $D_{\varepsilon,t,n} \subset U(N)$ for $N = 2^n$, and a probability distribution $\mu_{\varepsilon,t,n}$ over $D_{\varepsilon,t,n}$ such that*

1. *$\mu_{\varepsilon,t,n}$ is an ε-approximate t-unitary design.*
2. *$\mu_{\varepsilon,t,n}$ can be sampled from in $\mathrm{poly}(n, t, \log 1/\varepsilon)$ time*
3. *Each unitary $U \in D_{\varepsilon,t,n}$ can be implemented by a quantum circuit acting on n qubits of size at most $O(n\log(4t)^2 t^9(2nt + \log(1/\varepsilon)))$.*

We combine these two theorems to prove our final result:

Theorem 28 (Restatement of Theorem 21)**.** *The unitary design scheme is efficiently computable, and is* $2^{-s/2}$*-totally authenticating.*

Proof. Note that $f(U)$ is a polynomial of degree 4 in the entries of U. We compute $\alpha(f)$ by computing $\alpha(f_0)$, $\alpha(g)$, and $\alpha(h)$ where $f_0 = \langle\rho|\Gamma_V^\dagger\Gamma_V|\rho\rangle$ is a constant, $g(U) = \langle\rho|\Gamma_V^\dagger\Lambda_U|\rho\rangle$, and $h(U) = \langle\rho|\Lambda_U^\dagger\Lambda_U|\rho\rangle$. Clearly, $\alpha(f) \leq \alpha(f_0) + 2\alpha(g) + \alpha(h)$.

Since f_0 is a constant function, $\alpha(f_0)$ is at most $|f_0| \leq 1$. We turn to g. Let $\{|x\rangle\}$ be a basis for $\mathcal{M}$. Then for x, x', define the operator $T^{xx'} = \langle\varphi_x|\Gamma_V^\dagger V|\varphi_{x'}\rangle$ to be the linear operator that maps $\mathcal{Y}$ to $\mathcal{Y}$ (recall that $|\rho\rangle = \sum_x \rho_x|x\rangle \otimes |\varphi_x\rangle$). Then,

$$g(U) = \sum_{x,x'} \rho_x^* \rho_{x'} \langle\psi_x|T^{xx'}|\psi_{x'}\rangle \tag{77}$$

$$= \sum_{x,x',y,y'} \rho_x^* \rho_{x'} T_{yy'}^{xx'} U_{yx}^* U_{y'x'} \tag{78}$$

For every x, x', y, y', we have a distinct monomial $U_{yx}^* U_{y'x'}$, and the corresponding coefficient is $\rho_x^* \rho_{x'} T_{yy'}^{xx'}$, which has absolute value at most 1. Therefore $\alpha(g) \leq M^2N^2$.

Now we turn to $h(U)$. Recall that

$$h(U) = \sum_{x',x''} \rho_{x'}^* \rho_{x''} \sum_x (\langle\psi_{x'}| \otimes \langle\varphi_{x'}|)V^\dagger|\psi_x\rangle\langle\psi_x|V(|\psi_{x''}\rangle \otimes |\varphi_{x''}\rangle) \tag{79}$$

$$= \sum_{i,j,x',x''} \rho_{x'}^* \rho_{x''} U_{ix'}^* U_{jx''} \sum_x (\langle i| \otimes \langle\varphi_{x'}|)V^\dagger|\psi_x\rangle\langle\psi_x|V(|j\rangle \otimes |\varphi_{x''}\rangle) \tag{80}$$

where $|\psi_x\rangle = \sum_i U_{ix}|i\rangle$, $|\psi_{x'}\rangle = \sum_i U_{ix'}|i\rangle$ and $|\psi_{x''}\rangle = \sum_j U_{jx''}|j\rangle$. Define $|\tau^{ix'}\rangle = V|i\rangle \otimes |\varphi_{x'}\rangle$ and $|\tau^{jx''}\rangle = V|j\rangle \otimes |\varphi_{x''}\rangle$. Then we have

$$h(U) = \sum_{i,j,i',j'} \sum_{x,x',x''} U_{ix'}^* U_{jx''} U_{i'x} U_{j'x}^* \rho_{x'}^* \rho_{x''} \langle\tau^{ix'}|i'\rangle\langle j'|\tau^{jx''}\rangle \tag{81}$$

$$= \sum_{i,j,i',j'} \sum_{x,x',x''} U_{ix'}^* U_{jx''} U_{i'x} U_{j'x}^* \rho_{x'}^* \rho_{x''} \sum_z (\tau_{i'z}^{ix'})^* \ \tau_{j'z}^{jx''} \tag{82}$$

where we alternatively write $|\tau^{ix'}\rangle = \sum_z \tau_{i'z}^{ix'}|i', z\rangle$ and $|\tau^{jx''}\rangle = \sum_z \tau_{j'z}^{jx''}|j', z\rangle$. For every choice of i, j, i', j', x, x', x'', we have a distinct monomial, and the associated coefficient has norm at most

$$\left|\rho_{x'}^* \rho_{x''} \sum_z (\tau_{i'z}^{ix'})^* \ \tau_{j'z}^{jx''}\right|^2 \leq \left(\sum_z |\tau_{i'z}^{ix'}|^2\right) \cdot \left(\sum_z |\tau_{j'z}^{jx''}|^2\right) \leq 1.$$

Thus $\alpha(h)$ is at most M^3N^4. This implies that $\alpha(f) \leq O(N^7)$.

Now we are ready to leverage Theorems 26 and 27. in Lemma 25 we proved that function $f(U) = \|\Gamma_V|\rho\rangle - \Gamma_U|\rho\rangle\|_2^2$ has probability concentration

$$\Pr_{U \sim \nu_{Haar}} (|f - \mu| \geq \delta) \leq 4\exp(-CN\delta^2)$$

where C is a universal constant. Thus our parameters are:

1. (Average of f) $\mu = M/N$

2. (Error in probability concentration) $\delta = \sqrt{M/N}$

3. (Degree of f) $K = 4$

4. (Probability concentration exponent) $a = CN$

5. (Norm of f) $\alpha(f) = O(N^7)$

We will set $m = 1$, $\varepsilon = N^{-17}$, and $t = 8$.

By Theorem 27, there exists a distribution $\mu_{\varepsilon,t,n}$ over unitaries acting on n qubits that forms an efficiently computable ε-approximate t-unitary design. Then, plugging everything into Theorem 26, we have that

$$\Pr_{U\sim\mu_{\varepsilon,t,n}}(f \geq M/N + \sqrt{M/N}) \leq O(1/M) \tag{83}$$

Note that $M/N = 2^{-s}$. □

10 Open problems

We close with some open problems:

1. Is three-wise independence necessary for the Carter-Wegman scheme to be quantumly secure?

2. We showed that the Auth-QFT-Auth scheme achieves total authentication (with outer key leakage) when the inner authentication scheme is instantiated with the Carter-Wegman scheme using threewise-independent hashing. Can one show that Auth-QFT-Auth achieves total authentication when both inner and outer authentication schemes are *arbitrary* authentication schemes secure relative to the computational basis?

3. We showed that the scheme based on unitary 8-design achieves total authentication. Can one show the same for unitary 2-designs? Does random Clifford circuit [ABOE10], which is a unitary 3-design [Web15] achieve total authentication?

4. Under what circumstances can the key be reused in any of the protocols presented in this paper, when the receiver rejects the state? For example, we conjecture that in the unitary design protocol, much of the key can be reused.

5. Our security definitions are specific to "one-time" authentication schemes (although the key reuse properties allow multiple uses). Are there natural "many-time" versions of our security definitions?

Acknowledgments. We thank Debbie Leung for kindly sharing a manuscript of [HLM11], and thank Debbie, Patrick Hayden, and Fang Song for useful discussions. H.Y. was supported by Simons Foundation grant 360893, and National Science Foundation grant 1218547.

References

[ABOE10] Dorit Aharonov, Michael Ben-Or, and Elad Eban. Interactive proofs for quantum computations. In *Proceedings of Innovations in Computer Science*. Tsinghua University Press, 2010.

[ABW09] Andris Ambainis, Jan Bouda, and Andreas Winter. Nonmalleable encryption of quantum information. *Journal of Mathematical Physics*, 50(4):042106, 2009.

[BCG+02] Howard Barnum, Claude Crépeau, Daniel Gottesman, Adam Smith, and Alain Tapp. Authentication of quantum messages. In *The Proceedings of the 43rd Annual IEEE Foundations of Computer Science, 2002.*, pages 449–458. IEEE, 2002.

[BDF+11] Dan Boneh, Özgür Dagdelen, Marc Fischlin, Anja Lehmann, Christian Schaffner, and Mark Zhandry. Random Oracles in a Quantum World. In *Proceedings of ASIACRYPT*, 2011.

[Bee97] Carlo WJ Beenakker. Random-matrix theory of quantum transport. *Reviews of modern physics*, 69(3):731, 1997.

[BGS13] Anne Broadbent, Gus Gutoski, and Douglas Stebila. Quantum one-time programs. In *Advances in Cryptology–CRYPTO 2013*, pages 344–360. Springer, 2013.

[BHH12] Fernando GSL Brandao, Aram W Harrow, and Michal Horodecki. Local random quantum circuits are approximate polynomial-designs. *arXiv preprint arXiv:1208.0692*, 2012.

[BW16] Anne Broadbent and Evelyn Wainewright. Efficient simulation for quantum message authentication. *arXiv preprint arXiv:1607.03075*, 2016.

[BZ13a] Dan Boneh and Mark Zhandry. Quantum-secure message authentication codes. In *Advances in Cryptology–EUROCRYPT 2013*, pages 592–608. Springer, 2013.

[BZ13b] Dan Boneh and Mark Zhandry. Secure signatures and chosen ciphertext security in a quantum computing world. In *Advances in Cryptology–CRYPTO 2013*, pages 361–379. Springer, 2013.

[DFNS13] Ivan Damgård, Jakob Funder, Jesper Buus Nielsen, and Louis Salvail. Superposition attacks on cryptographic protocols. In *Information Theoretic Security*, pages 142–161. Springer, 2013.

[DNS12] Frédéric Dupuis, Jesper Buus Nielsen, and Louis Salvail. Actively secure two-party evaluation of any quantum operation. In *Advances in Cryptology–CRYPTO 2012*, pages 794–811. Springer, 2012.

[DPS05] Ivan Damgård, Thomas Brochmann Pedersen, and Louis Salvail. A quantum cipher with near optimal key-recycling. In *Proceedings of the 25th Annual International Conference on Advances in Cryptology*, CRYPTO'05, pages 494–510, Berlin, Heidelberg, 2005. Springer-Verlag.

[HLM11] Patrick M. Hayden, Debbie W. Leung, and Dominic Mayers. The universal composable security of quantum message authentication with key recycling. *In preparation*, 2011.

[KLLNP16] Marc Kaplan, Gaëtan Leurent, Anthony Leverrier, and María Naya-Plasencia. Breaking symmetric cryptosystems using quantum period finding. *arXiv preprint arXiv:1602.05973*, 2016.

[Low09] Richard A Low. Large deviation bounds for k-designs. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 465, pages 3289–3308. The Royal Society, 2009.

[MS09] Vitali D Milman and Gideon Schechtman. *Asymptotic Theory of Finite Dimensional Normed Spaces: Isoperimetric Inequalities in Riemannian Manifolds*, volume 1200. Springer, 2009.

[WC81] Mark N Wegman and J Lawrence Carter. New hash functions and their use in authentication and set equality. *Journal of computer and system sciences*, 22(3):265–279, 1981.

[Web15] Zak Webb. The clifford group forms a unitary 3-design. *arXiv preprint arXiv:1510.02769*, 2015.

[Zha12] Mark Zhandry. How to Construct Quantum Random Functions. In *Proceedings of the 53rd IEEE Symposium on Foundations of Computer Science (FOCS)*, 2012.