New security notions and feasibility results for authentication of quantum data

Sumegha Garg[1⋆], Henry Yuen[2⋆⋆], and Mark Zhandry[1⋆⋆⋆]

[1] Princeton
[2] UC Berkeley

Abstract. We give a new class of security definitions for authentication in the quantum setting. These definitions capture and strengthen existing definitions of security against quantum adversaries for both *classical* message authentication codes (MACs) as well as full quantum state authentication schemes. The main feature of our definitions is that they precisely *characterize* the effective behavior of any adversary when the authentication protocol accepts, including correlations with the key. Our definitions readily yield a host of desirable properties and interesting consequences; for example, our security definition for full quantum state authentication implies that the entire secret key can be re-used if the authentication protocol succeeds.

Next, we present several protocols satisfying our security definitions. We show that the classical Wegman-Carter authentication scheme with 3- universal hashing is secure against superposition attacks, as well as adversaries with quantum side information. We then present conceptually simple constructions of full quantum state authentication.

Finally, we prove a lifting theorem which shows that, as long as a protocol can securely authenticate the maximally entangled state, it can securely authenticate any state, even those that are entangled with the adversary. Thus, this shows that protocols satisfying a fairly weak form of authentication security automatically satisfy a stronger notion of security (in particular, the definition of Dupuis, et al (2012)).

1 Introduction

Authenticating messages is a fundamental operation in classical cryptography. A sender Alice wishes to send a message m over an insecure channel to a receiver Bob, with the guarantee that the message was not tampered with in transit. To accomplish this, Alice appends a "signature" σ to m using a shared secret key k and sends the message/signature pair (m, σ) to Bob. Bob receives some

⋆ sumeghag@cs.princeton.edu
⋆⋆ hyuen@cs.berkeley.edu
⋆⋆⋆ mzhandry@princeton.edu

potentially altered pair (m', σ'), and then verifies that σ' is a valid signature of m' under key k. If verification passes, Bob accepts m', and if verification fails, Bob ignores the message and discards it. A secure authentication protocol guarantees the following: even if the adversary has arbitrarily tampered with the communication channel, as long as the adversary does not know the secret key k, then either Bob rejects with high probability, or the message he receives is m. 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 adversary – even a computationally unbounded one – can modify the message without detection [WC81].

Just as authentication is a fundamental operation in 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, and consider quantum attacks on both *classical* authentication protocols, as well as full-fledged *quantum* protocols for authenticating quantum data. What kinds of security guarantees can we hope for in the quantum setting? Various notions of security for authentication schemes against quantum attacks have been considered in the literature. However, as we will discuss below, these existing definitions do not fully capture security properties we would expect of a secure authentication scheme.

The contribution of our paper is three-fold: first, we present new security definitions for authentication in a quantum setting that strengthen previous definitions and address their limitations. Second, we prove interesting consequences of our stronger security definition for quantum authentication, such as information-theoretic key recycling and an easy protocol for quantum key distribution. Finally, we prove that several natural authentication protocols satisfy our security definitions.

1.1 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 setting of MACs, an adversary in such an attack is able to trick the sender into signing a superposition of messages.[3] That is, the sender computes the map

[3] One motivation for studying superposition attacks comes from the "Frozen Smart-Card" example [GHS15]: real-world classical authentication systems are frequently implemented on small electronic devices such as RFID tags or a smart-cards. A determined and sophisticated attacker in possession of such a smart-card could try to perform a quantum "side-channel attack" on it: he places the device in a very low temperature environment, and attempts to query the device in quantum superposition. One would like to guarantee that even then the attacker is unable to, say, extract the secret key.

$|m\rangle \mapsto |m, \sigma_m\rangle$ 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. The usual classical security notion asks that an adversary, after seeing a signed message, cannot produce a different message with a valid signature. The natural way to translate this into our setting is to require that the adversary, after seeing a signed superposition, cannot produce a different forged quantum state with valid signature. For classical authentication schemes, this goal is unfortunately impossible. The adversary can tamper with the signed state by measuring 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 very 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 simultaneously produce two valid signed messages (m, σ_m) and $(m', \sigma_{m'})$ for $m \neq m'$. Note that in the classical setting, this requirement is equivalent to the traditional MAC security definition, so it appears to be a reasonable requirement for any quantum security notion. More generally, given q signed states, the Boneh-Zhandry definition says that the adversary should not be able to produce $q + 1$ distinct valid signed messages.

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 disallows an adversary from 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 is reasonable to insist as a security requirement 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 in the quantum setting.

Boneh and Zhandry's definition 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.

1.2 Quantum Authentication of Quantum Data.

We turn to the setting of schemes for authenticating quantum states. Barnum et al. [BCG$^+$02] was the first to study this, and they present a definition of non-interactive quantum authentication where, conditioned on the protocol succeeding, the sender has effectively teleported a quantum state to the receiver (provided that the probability of success is not too small). They then give a scheme (called the *purity testing scheme*) which attains this definition. Interestingly, they also show that quantum state authentication also implies quantum state *encryption*.[4] Subsequent works [BCG$^+$06,ABE10,DNS12,BW16] presented some stronger security definitions that we will discuss momentarily.

Roughly speaking, a (private-key) quantum authentication scheme is a pair of keyed quantum operations $(\mathsf{Auth}_k, \mathsf{Ver}_k)$, where k is a secret key shared by the sender and receiver, where Auth_k is a map that takes in a quantum message state ρ, and outputs an authenticated state σ. The map Ver_k is a verification operation that takes in a (possibly) tampered state $\widetilde{\sigma}$ and outputs a state $\widetilde{\rho}$, along with a flag ACC or REJ indicating whether the verification succeeded or failed. These maps are such that for all input states ρ and all keys k, we have $\mathsf{Ver}_k(\mathsf{Auth}_k(\rho)) = \rho \otimes |\mathrm{ACC}\rangle\langle\mathrm{ACC}|$.

Informally, Barnum, et al. define a secure authentication scheme to be such that, for all adversaries $\mathcal{O}$, either the receiver rejects the state $\mathsf{Ver}_k(\mathcal{O}(\mathsf{Auth}_k(\rho)))$ with high probability, or it is close to the original state ρ. However it has the shortcoming that it does not consider the possibility that the adversary is *entangled* with the original message ρ, and thus may act on the entanglement to tamper with the state in an undetectable manner. Thus, the security definition of [BCG$^+$02] is not *composable*.

In many situations we would like to use authentication not as a stand along task, but as a primitive in a larger protocol – indeed, quantum authentication

[4] By contrast, in the classical setting, message authentication does *not* imply message encryption.

has been used as a primitive in schemes for delegated quantum computation, e.g., [ABE10,BGS13]. Here, the "adversary" (which may be other components of the protocol) may generate the inputs to the authentication scheme, and thus be entangled with the message that is supposed to be authenticated. If an authentication scheme satisfied a composable security definition, then we may use the security of the authentication primitive in a black box manner to argue the proper functioning of the larger protocol.

Correlations between final state and the key -
Recently, several works [HLM16,DNS12,BW16] have proposed composable security definitions for quantum authentication – that is, they handle adversaries with quantum side information. However, their security definitions do not explicitly consider *correlations* between the key and the final state of the protocol.

Suppose Alice sends Bob the authenticated state $\sigma_k = \mathsf{Auth}_k(\rho)$ using key k. Bob receives a (possibly tampered) state $\widetilde{\sigma}_k$, and proceeds to verify the authentication. Let τ_k denote Bob's state *conditioned* on successful verification. Roughly speaking, the security definitions of [BCG$^+$02,DNS12,BW16] refers to the *average* state $\mathbb{E}_k \tau_k$; in particular, it states that $\mathbb{E}_k \tau_k$ is close to the original state ρ. This statement does not by itself imply that τ_k is close to the original state ρ *with high probability* over k. In other words, the state of the key is traced out in their security definitions.

Later, we will show how taking into account the correlations between the key and the final state of the protocol yields interesting consequences – such as the ability to reuse the key upon successful verification.

2 Our contributions

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 demanding 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 follows 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 ex-

ecution of an ideal 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 the behavior of any adversary (not necessarily ideal) 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, and the quantum side information held by the adversary *conditioned on successful verification* must be indistinguishable from the same joint state arising from the actions of *some* ideal adversary from the class $\mathscr{A}$. Since our notion of simulation is so general, this implies that our security definitions satisfy security under *sequential composition*; that is, the authentication protocols that realize our security definition can be securely composed with arbitary cryptographic protocols in a sequential fashion.

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

2.1 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.

More generally, we think of the protocol acting on message 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 $|\varphi_m\rangle^{\mathcal{Z}}$ 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. This would not be detected by the classical verification procedure, but the final message would have been changed.

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 (partial) measurement of the message in the computational basis, and controlled on the outcome of the partial measurement, perform some quantum operation on his own private qubits. 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 it 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[5]. Finally, we show that the classical Wegman-Carter MAC that uses three-universal hashing is sufficient for achieving this strong security definition. This improves on Boneh-Zhandry in two ways, as they show that *four*-wise independence gives their weaker security notion.

2.2 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.

As mentioned above, prior works have put forth composable security definitions for quantum authentication [DNS12,BW16], who consider quantum side information held by the adversary. Our definition builds upon these definitions: not only do we we consider side information about the plaintext state, we also allow 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

[5] 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

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 significant information about the key. This fact has interesting consequences:

1. **Key reuse.** For example, our definition immediately implies that, 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 essentially independent of the adversary and can therefore be used again in the same protocol. This is in contrast to the classical setting: in general keys cannot be recycled without computational assumptions. 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.

   Previous works have explored partial key reuse in various quantum protocols [OH05,DPS05,HLM16]. However, to our knowledge, our work is the first to establish that the *entire key* can be recycled upon successful verification.[6]
2. **A simple quantum key distribution protocol.** Our definition also gives a conceptually simple quantum key distribution (QKD) protocol[7]. 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 "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.

[6] The work of Dåmgard et al. [DPS05] argue that the key can be recycled entirely when authenticating *classical messages*, but their protocol does not appear to extend to handling quantum messages.

[7] The observation that quantum authentication implies a form of QKD is due to Charlie Bennett and also observed by Gottesman [Got02].

A protocol satisfying total authentication. 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.

Total authentication with key leakage. 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 a 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 [HLM16] essentially shows that the Barnum et al. protocol satisfies total authentication with (minor) key leakage. We also give a simple protocol that achieves this, based on the *classical* Wegman-Carter authentication scheme.

A lifting theorem. Finally, we prove an intriguing *equivalence* between a very weak form of authentication security and a stronger notion. Specifically, this weak form of authentication security only guarantees that an authentication scheme is able to authenticate a *single state*: a Bell state. Furthermore, this Bell state is unentangled with the adversary, and the security guarantee holds with the key traced out (i.e. correlations with the key are not kept track of).

We prove a *lifting theorem* that "lifts" this weak security to a much stronger one that shows the same authentication scheme, when augmented with a Pauli randomization step, is actually secure when authenticating *arbitrary* messages, which might be entangled with the adversary! This stronger security notion still traces out the key, so it does not achieve total authentication. Nonetheless, we find it conceptually very interesting that such a lifting theorem holds.

We believe that our work contributes to broadening our understanding of *what security definitions are possible* for various primitives in the quantum world. In classical cryptography an eavesdropper can be correlated with the secret key simply by copying the ciphertext; thus it does not make sense for a security definition to keep track of the correlations between an adversary's private memory and the key. Our results demonstrate that it *is* meaningful to do so in the quantum setting. This is the motivation behind the name "total authentication": protocols satisfying total authentication are achieving the "best possible" security *within the framework used for the definition*.

2.3 Subsequent Work

Subsequent to the initial posting of our work, there have been several very interesting developments in quantum authentication. Portman [Por17] uses the

Abstract Cryptography (AC) framework to model quantum authentication with key recycling. He shows that the Barnum et al. [BCG$^+$02] is secure in this setting, thus demonstrating that the Barnum et al protocol satisfies complete key recycling. Moreover, he shows that authentication based on unitary 2-designs is secure with key recycling; our analysis requires 8-designs to demonstrate key recycling. Alagic and Majenz [AM16] independently show that total authentication can be achieved with unitary 2-designs. Fehr and Salvail [FS16] examine quantum authentication of *classical* messages and demonstrate a scheme that admits key recycling as well.

Outline. In the next section we cover some preliminaries and notation. In Section 4 we formally present the fundamental security definitions used in our paper. In Sections 5.1 and 5.2 we present several properties satisfied by our definitions. In Section 6, we analyze the security of the Wegman-Carter MAC with 3-universal hashing within our security framework. In Section 7 we present and analyze the Auth-QFT-Auth scheme. In Section 8 we present and analyze the unitary design scheme. In Section 9 we prove the lifting theorem.

3 Preliminaries

3.1 Notation

Quantum information. We assume basic familiarity with quantum computing concepts, such as states, measurements, and unitary operations. We will use calligraphic 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 they 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 superscripts 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 operator $(U^{\mathcal{Y}} \otimes \mathbb{I}^{\mathcal{Z}}) V^{\mathcal{Y}\mathcal{Z}}$; 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[8]:

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 $|\psi\rangle$.

[8] A seasoned veteran of quantum information may notice that this departs slightly from the convention in quantum information theory where physically realizable quantum operations are CPTP maps. Here the difference is that we consider maps that can possibly *decrease* the trace of an operator, which corresponds to post-selection.

The most important matrix norm 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 $\|\rho - \sigma\|_1$ is proportional to 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 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 $\theta = |\theta\rangle\langle\theta|$.

Proof. It is well known that $\|\psi - \theta\|_1 \leq 2\sqrt{1 - |\langle\psi|\theta\rangle|^2}$ (see, e.g., [NC10]). But now notice that $1 - x^2 \leq 2(1 - x)$ for all x. Therefore the trace distance is at most $2\sqrt{2(1 - |\langle\psi|\theta\rangle|)} \leq 2\||\psi\rangle - |\theta\rangle\|_2$.

3.2 Basic definitions for authentication

Spaces. We let $\mathcal{K}$ denote the **key space**, $\mathcal{M}$ denote the **message space**, $\mathcal{Y}$ denote the **authenticated space**, and $\mathcal{F}$ denote the **flag space**. The flag space $\mathcal{F}$ is a two-dimensional Hilbert space spanned by orthogonal states $|\mathrm{ACC}\rangle$ and $|\mathrm{REJ}\rangle$. The space $\mathcal{Z}$ is the **private space of the adversary**. We will let $\mathcal{S}$ denote the registers held by the sender and receiver that, during the execution of the authentication protocol, are not communicated nor acted upon by the sender, receiver, or adversary.

Authentication scheme. An 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 $\mathrm{D}(\mathcal{M} \otimes \mathcal{F})$.

satisfying the correctness requirements that for any quantum state $\rho \in \mathrm{D}(\mathcal{M})$, for all keys $k \in \mathcal{K}$, $\mathsf{Ver}_k(\mathsf{Auth}_k(\rho)) = \rho \otimes |\mathrm{ACC}\rangle\langle\mathrm{ACC}|$.[9]

We will also use Auth and Ver to denote the operators

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

[9] One can also discuss schemes where the correctness requirements hold *approximately* (e.g., the state $\mathsf{Ver}_k(\mathsf{Auth}_k(\rho))$ is within trace distance δ of $\rho \otimes |\mathrm{ACC}\rangle\langle\mathrm{ACC}|$); using this correctness condition does not significantly affect the discussion in this paper.

Some simplifying assumptions. This definition of authentication scheme is more general than we need in this paper. Throughout this work, we shall work with a simplified model of authentication schemes: first, we will assume that Auth_k behaves as an isometry taking $\mathcal{M}$ to $\mathcal{Y}$ (i.e. it isn't probabilistic). Let $\mathcal{V}_k$ denote the subspace of the Hilbert space $\mathcal{Y}$ that is the image of Auth_k, let $\Pi_{\mathcal{V}_k}$ denote the projector onto the space $\mathcal{V}_k$, and let Auth_k^{-1} denote the inverse isometry that maps $\mathcal{V}_k$ to $\mathcal{M}$. In this case, a canonical way to define the Ver_k superoperator is as follows:

$$\rho \mapsto \left(\mathsf{Auth}_k^{-1} \circ \Pi_{\mathcal{V}_k}\right)(\rho) \otimes |\mathrm{ACC}\rangle\langle\mathrm{ACC}|^{\mathcal{F}} + \mathrm{Tr}((\mathbb{I} - \Pi_{\mathcal{V}_k})\,\rho)\,\frac{\mathbb{I}_{\mathcal{M}}}{|\mathcal{M}|} \otimes |\mathrm{REJ}\rangle\langle\mathrm{REJ}|^{\mathcal{F}} \tag{1}$$

Here, $\mathsf{Auth}_k^{-1} \circ \Pi_{\mathcal{V}_k}$ denotes the operation that first applies the projection $\Pi_{\mathcal{V}_k}$ to the state, followed by the inverse isometry Auth_k^{-1}. The state $\frac{\mathbb{I}_{\mathcal{M}}}{|\mathcal{M}|}$ is the maximally mixed state on the message space. In other words, the verification procedure first checks that the received state (which resides in $\mathcal{Y}$) is supported on the subspace of valid signed states $\mathcal{V}_k$. If so, then it inverts the authentication isometry to obtain an unsigned message state, and sets the $\mathcal{F}$ register to $|\mathrm{ACC}\rangle$. Otherwise, it replaces the state with a uniformly random message state, and sets the $\mathcal{F}$ register to $|\mathrm{REJ}\rangle$.

However in this paper we are mostly concerned with the output of the Ver_k procedure in the *accepting* case. For technical convenience then, throughout this paper we will treat Ver_k as the following superoperator mapping $\mathrm{D}(\mathcal{Y})$ to $\mathrm{D}(\mathcal{M})$:

$$\mathsf{Ver}_k(\rho) = \left(\mathsf{Auth}_k^{-1} \circ \Pi_{\mathcal{V}_k}\right)(\rho).$$

In other words, it only outputs the $|\mathrm{ACC}\rangle$ part of (1), and does not output a ACC or REJ flag. Furthermore, notice that this superoperator is not trace preserving; the trace of $\mathsf{Ver}_k(\rho)$ is equal to the probability that ρ was accepted by the verification procedure defined in (1). Thus one can view Ver_k as a "filter" that only accepts states that were properly authenticated.

We stress, however, that these simplifying assumptions are not crucial to our results – it is mostly for notational convenience that we treat Ver_k as a filter.

Classical Authentication. In a classical authentication protocol, the authentication operator Auth_k is specified by a classical (reversible) 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}.

Message authentication codes. A message authentication code (or MAC) is a 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:

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

Adversaries. We model adversaries in the following way: the adversary prepares the initial message state $|\rho\rangle^{\mathcal{MSZ}}$, where we can assume that the adversary possesses the purification of $\rho^{\mathcal{MS}}$. 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 unauthenticated 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 (which implies that the success probability of the protocol 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.

A remark about the sender and receiver's private register $\mathcal{S}$. The reader may wonder why we do not allow the sender, receiver, nor adversary to act upon the $\mathcal{S}$ register during the execution of the authentication protocol. The register $\mathcal{S}$ is supposed to model entanglement the sender and receiver may keep during the protocol. The important aspect of it is that the adversary does *not* have access to this side information.

If, when analyzing the authentication scheme in the context of a larger protocol in which the sender/receiver *do* act upon the register $\mathcal{S}$, we can assume that during the authentication phase, the sender and receiver do not touch $\mathcal{S}$, but wait until the authentication protocol is over. Thus we can analyze the behavior of the authentication protocol without this action.

4 Security Framework for Quantum Authentication

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 $\mathcal{Y}$.

Definition 1. *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* $(\mathsf{Auth}, \mathsf{Ver})$ ε-reduces to $\mathscr{A}$-adversaries

iff the following holds: for all initial message states $|\rho\rangle^{\mathcal{MSZ}}$*, 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)* $\mathbb{E}_k \, |k\rangle\langle k| \otimes [\mathsf{Ver}_k \circ \mathcal{O} \circ \mathsf{Auth}_k] \, (\rho^{\mathcal{MSZ}})$
- *(Ideal experiment)* $\mathbb{E}_k \, |k\rangle\langle k| \otimes [\mathsf{Ver}_k \circ \mathcal{I} \circ \mathsf{Auth}_k] \, (\rho^{\mathcal{MSZ}})$

where Auth_k *acts on* $\mathcal{M}$*,* Ver_k *acts on* $\mathcal{Y}$*, and both act as the identity on* $\mathcal{SZ}$*.*

Intuitively, our security definition states that for an authentication scheme $(\mathsf{Auth}, \mathsf{Ver})$ that is $\mathscr{A}$-secure, for all initial message states $\rho^{\mathcal{MSZ}}$, an *arbitrary* adversary that acts on an authenticated state $\mathsf{Auth}_k(\rho^{\mathcal{MSZ}})$ 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 $\mathscr{A}$. We allow the real adversary to prepare the message state $\rho^{\mathcal{MSZ}}$ and hence, allow the ideal adversary to depend on it.

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.

4.1 Basis-dependent authentication

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

Definition 2 (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 $\sigma \in \mathrm{D}(\mathcal{Y}\mathcal{Z})$, where Π is a projector acting on $\mathcal{Z}\mathcal{Z}'$, and $V \in \mathrm{J}(\mathcal{Y}\mathcal{Z}, \mathcal{Y}\mathcal{Z}\mathcal{Z}')$ 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, $V_\psi \in \mathrm{J}(\mathcal{Z}, \mathcal{Z}\mathcal{Z}')$ 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 performs some (partial) measurement on 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 3 (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 5.1, we establish consequences of our definition of basis-dependent security, including the property of 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 6 we show that the classical Wegman-Carter 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 3. *The Wegman-Carter 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

In this section we formally define our notion of total authentication. First, we define *oblivious adversaries*.

Definition 4 (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 5 (Total authentication). *An authentication scheme* $(\mathsf{Auth}, \mathsf{Ver})$ ε-totally authenticates *iff it* ε*-reduces to the class of oblivious adversaries.*

[DNS12]'s security definition is similar, except it traces out the key register. Therefore, it does not keep track of potential correlations between the adversary and the key. [DNS12] considers what happens in the reject case, while total authentication only makes requirements when the verifier accepts. A subsequent work by Alagic et al. [AM16] indeed showed that total authentication implies [DNS12] with a slight modification of decryption outputting $\perp$ whenever it rejects. We will argue shortly that our definition of total authentication is strictly stronger than the definition of [DNS12]; that is, there are protocols which satisfy the security definition of [DNS12], but do not satisfy total authentication.

In Section 5.2 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). [HLM16] 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 8 we present a scheme, called the *unitary design scheme*, that achieves total authentication, and to our knowledge this is the first scheme that achieves such security.

Theorem 4. *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; [HLM16] 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 $\mathcal{K}'$ be such that $|\mathcal{K}'| \leq |\mathcal{K}|$. Define a *key leakage function* $\ell : \mathcal{K} \mapsto \mathcal{K}'$. If $|\mathcal{K}'|$ is strictly smaller than $|\mathcal{K}|$, then $\ell(k)$ must necessarily lose information about the key $k \in \mathcal{K}$, but it will also 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 6. *Let* $(\mathsf{Auth}, \mathsf{Ver})$ *be an authentication scheme. Let $\mathcal{K}'$ be some domain such that $|\mathcal{K}'| \leq |\mathcal{K}|$ and let $\ell : \mathcal{K} \to \mathcal{K}'$ 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 $|\rho\rangle^{\mathcal{MSZ}}$, 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 $h \in \mathcal{K}'$, such that the following (not necessarily normalized) states are ε-close in trace distance:*

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

Definition 7 (Total authentication with key leakage). *Let $\mathcal{K}'$ be some domain such that $|\mathcal{K}'| \leq |\mathcal{K}|$ and let $\ell : \mathcal{K} \to \mathcal{K}'$ be a key leakage function. An authentication scheme* $(\mathsf{Auth}, \mathsf{Ver})$ ε-totally authenticates with key leakage ℓ *iff it ε-reduces to the class of oblivious adversaries with key leakage 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 7 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 Wegman-Carter 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 Wegman-Carter 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 5 (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 [HLM16], but it is unclear whether it satisfies the strongest definition of total authentication without key leakage.

4.4 A remark about efficiency

Recently, Broadbent and Wainewright [BW16] study the *efficiency* of simulating ideal adversaries in the security proofs of two authentication schemes, the Clifford scheme and the trap code scheme. Specifically, they show that if the adversary in the authentication protocol is a quantum computer that runs in time T, then the ideal adversary which simulates it also runs in time $O(T)$. This efficiency-preservation is important for notions of composable security.

We note that the constructions of the ideal adversary in our analysis of the Wegman-Carter scheme, the Auth-QFT-Auth scheme, and the unitary design scheme are also efficiency preserving, and hence if the arbitrary adversary runs in polynomial time, then the simulating adversary also runs in polynomial time.

4.5 Comparison with security definition in [DNS12]

Similarly to our definition, the security definition of message authentication [DNS12] implies that essentially all the adversary can do is act on its own private workspace.

However, it traces out the key register, and thus it does not keep track of correlations between the adversary and the secret key. It is a natural question to ask whether the security definition of [DNS12] *implies* our definition of total authentication. Here we show that it cannot, because there are protocols that satisfy the [DNS12] definition, but not ours.[10] [AM16] gives a formal proof of the fact that any protocol satisfying total authentication satisfies [DNS12] definition (with a slight modification).

Consider a protocol $(\mathsf{Auth}, \mathsf{Ver})$ that satisfies the [DNS12] definition. Let k denote the secret key used in the protocol. Now consider the following modified protocol $(\mathsf{Auth}', \mathsf{Ver}')$: to authenticate a message state ρ, it produces $\mathsf{Auth}_k(\rho)$, but then appends an independently random bit b, where (k, b) is the secret key register of $(\mathsf{Auth}', \mathsf{Ver}')$. To verify, the receiver just applies the Ver_k operation, and ignores the last bit. This new protocol still satisfies the [DNS12] definition, because the extra bit b is independent of the $(\mathsf{Auth}, \mathsf{Ver})$ process, and thus final state of the protocol can be simulated by an ideal adversary that generates its own b bit – as long as we're tracing out the key. However, this protocol does not satisfy total authentication. This is because an adversary can simply copy the bit b into its private workspace; but this cannot be simulated by an ideal adversary that is unentangled with the (k, b) register.

Furthermore, any authentication scheme satisfying [DNS12]'s security definition also satisfies "total authentication with key leakage" for some key leakage function ℓ and any authentication scheme satisfying "total authentication with key leakage" satisfies the key-averaged security definition (with slight modification of decryption outputting $\perp$ in reject case). Hence, these two security definitions are equivalent (up to some error).

5 Properties of security definitions

5.1 Properties of basis-dependent authentication

Unforgeability Our security definition of authentication schemes relative to a basis implies the standard, classical security definition of authentication schemes called EUF-CMA. Namely, this says 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 (quantum) security definition for one-time MACs.

For detailed discussion and proof, refer to the full version of the paper [GYZ16][11].

[10] See Section 9 for a formal statement of the [DNS12] definition.

[11] `https://arxiv.org/abs/1607.07759`

5.2 Properties of total authentication

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 compatible with Barnum et al.'s: we start from a stronger property that considers side information, and end with a stronger form of encryption that also considers side information. For proof, refer to the full version [GYZ16].

Quantum Key Distribution As mentioned in introduction, and noticed by previous works, a total authentication scheme gives a simple method to perform quantum key distribution. For details of the protocol, refer to the full version [GYZ16].

Key Reuse It is easy to see that our definition of total authentication implies that, conditioned on successful verification of an authentication scheme (satisfying total authentication), the key can be reused by the sender and receiver for some other purpose. This is because conditioned on acceptance, the final state of the adversary is within ε/α trace distance of being independent of the key, where α is the probability of acceptance in the authentication protocol.

6 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 authentication 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 Wegman-Carter 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 6. *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.*

We first state what the implications for key length are. Suppose we wish to guarantee that the Wegman-Carter 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 $3\log|\mathcal{M}| + 6\log\frac{1}{\varepsilon} + O(1)$.

For proof of the above theorem, refer to the full version [GYZ16][12].

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

In Section 6, we saw how the classical Wegman-Carter 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 Wegman-Carter 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 Wegman-Carter MAC in the computational basis, and then authenticates again in the Fourier basis (using a new key). The 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 Wegman-Carter 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 acted 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.

[12] `https://arxiv.org/abs/1607.07759`

7.1 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.

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

$$|\sigma_{kh}\rangle^{\mathcal{YT}_2\mathcal{Z}} = \frac{1}{\sqrt{N}} \sum_z \sqrt{\lambda_z} \left(\sum_{m,x} \alpha_{zm} (-1)^{(m,k(m))\cdot x} |x, h(x)\rangle^{\mathcal{Y}_1\mathcal{T}_2} \right) \otimes |\varphi_z\rangle^{\mathcal{Z}}$$

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 $\mathcal{Y}_1\mathcal{T}_2\mathcal{Z}$:

$$\widetilde{\sigma}^{\mathcal{K}\mathcal{H}\mathcal{Y}_1\mathcal{T}_2\mathcal{Z}} = V\sigma V^\dagger.$$

Let $\widetilde{\tau}^{\mathcal{K}\mathcal{H}\mathcal{M}\mathcal{Z}} = \mathsf{Ver}_1 \circ \mathrm{QFT}^{-1} \circ \mathsf{Ver}_2(\widetilde{\sigma})$. 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 7 (Security of the Auth-QFT-Auth scheme). *The Auth-QFT-Auth scheme is δ-totally authenticating with outer key leakage, where $\delta = \varepsilon + O(\sqrt{|\mathcal{M}|^{3/2}/|\mathcal{T}_1|})$.*

Firstly, we consider the key requirements. The outer authentication scheme need not be a Wegman-Carter 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 Sketch: We will omit mention of the sender/receiver's private space $\mathcal{S}$, and discuss how our proof generalizes to the case of non-empty $\mathcal{S}$ later. 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

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

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

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

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

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

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 $|\Phi_h\rangle^{\mathcal{A}\mathcal{A}'} = \frac{1}{\sqrt{N}} \sum_x |x, h(x)\rangle^{\mathcal{A}} |x, h(x)\rangle^{\mathcal{A}'}$ 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 $|\Phi_h\rangle^{\mathcal{A}\mathcal{A}'}$ that resides in $\mathcal{A}$, and the $\mathcal{Z}$ part of the input state $|\rho\rangle$.
3. The adversary measures $\mathcal{A}\mathcal{A}'\mathcal{Z}\mathcal{Z}_2$ using the projective measurement $\{Q, \mathbb{I} - Q\}$, where $Q = |\Phi_h\rangle\langle\Phi_h|^{\mathcal{A}\mathcal{A}'} \otimes \Pi$. The adversary discards the outcome corresponding to $\mathbb{I} - Q$, and leaves the state unnormalized:

$$\frac{1}{N} \sum_{z,x,m} \sqrt{\lambda_z} \alpha_{zm} |m\rangle^{\mathcal{M}} |\phi_{xsz}\rangle^{\mathcal{Z}} |\Phi\rangle^{\mathcal{A}\mathcal{A}'} |0 \cdots 0\rangle^{\mathcal{Y}_2' \mathcal{Z}_2}$$

4. The adversary discards the $\mathcal{A}\mathcal{A}'\mathcal{Y}_2'\mathcal{Z}_2$ 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 bound $\mathbb{E}_k \| |\mu_{kh}\rangle - |\nu_h\rangle \|^2$ by $O(M^{3/2}/T)$. Refer to the full version [GYZ16] for the proof. 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 1.

To conclude the theorem, we now observe that when $\mathcal{S}$ is non-empty, we can use the same analysis as above where we bundle together $\mathcal{S}$ and $\mathcal{Z}$ as a joint adversary register, and the ideal adversary given by Lemma 1 will act as the identity on the $\mathcal{S}$ register. This establishes that $(\mathsf{Auth}, \mathsf{Ver})$ is a total authentication scheme with outer key leakage.

8 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. For a precise definition of unitary designs and efficient constructions, please see, e.g., [BHH12].

8.1 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 $t = 8$.
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 8. *The unitary design scheme is efficiently computable, and is* $2^{-s/2}$*-totally authenticating.*

This scheme is inspired by the *Clifford code authentication scheme*, first proposed by Aharonov, et al. [ABE10], and further analyzed in [DNS12,BW16]. Our protocol is exactly the same, except the ensemble of unitaries, instead of being an approximate 8-design, is the *Clifford group*, which is a well-studied set of unitaries that are central to quantum error-correction, simulation, and more. It was also recently shown that the Clifford group is a 3-unitary design [Web15,Zhu15][13]. [DNS12,BW16] show that the Clifford authentication scheme is secure even against entangled adversaries; however, as mentioned before, their security guarantee does not take into account the key.

Our unitary design scheme is also very similar to the *non-malleable quantum encryption scheme* proposed by Ambainis, Bouda, and Winter [ABW09], wherein a unitary 2-design is used to encrypt a quantum state. However, non-malleable quantum encryption does not imply authentication.

We now remark upon the key requirements of the unitary design scheme. Constructions of approximate unitary 8-designs acting on n qubits involve choosing a random quantum circuit of size $\Theta(n^2)$, and thus the randomness required is $\Theta(n^2)$ [BHH12]. This asymptotically matches the randomness requirements required of the Clifford scheme described above, but is much larger than the randomness requirements of the purity-testing-based protocol of [HLM16], which uses $\Theta(n)$ bits of key to authenticate an n-qubit quantum state.

[13] However, it is not an 8-design

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.

Formally, we first prove the following lemma to get total authentication when unitary U is drawn from the Haar measure over the unitary group $U(\mathcal{Y})$.

Lemma 2. *Let $N = \dim(\mathcal{Y})$. For all $\delta > 0$, for all initial message states $|\rho\rangle^{\mathcal{M}\mathcal{Z}}$ 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.

Here, Γ_V would be the ideal adversary corresponding to the real adversary V.

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

Definition 8. *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 3 (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*

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

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

We define $f(U) = \|\Gamma_V |\rho\rangle - \Lambda_U |\rho\rangle\|_2^2$, and bound the average and Lipshcitz constant of f to use Levy's Lemma to prove Lemma 2. For proof details, refer to the full version [GYZ16].

We then appeal to a general derandomization result of Low [Low09] 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.

For detailed proof of unitary 8-designs being totally authenticating, refer to the full version [GYZ16].

9 A lifting theorem for authentication

We will prove a *lifting theorem* which shows that a weak form of authentication security that doesn't take into account quantum side information actually implies stronger security against quantum side information. The initial weak form of security is very weak indeed: as long as the authentication scheme can securely authenticate a *single* state (namely, one half of the maximally entangled state), in a key-averaged manner, then we can actually obtain an authentication scheme that can authenticate all states — even those that are entangled with the adversary.

Specifically, we show that this weak authentication security implies the security definition of [DNS12], which we reproduce here:

Definition 9 ([DNS12] security definition). *An authentication scheme* $(\mathsf{Auth}, \mathsf{Ver})$ *is ε-secure according to the [DNS12] definition iff for all initial message states* $|\rho\rangle^{\mathcal{MSZ}}$*, for all adversaries* $\mathcal{O} \in \mathrm{T}(\mathcal{YZ}, \mathcal{YZ})$*, there exists an oblivious adversary* $\mathcal{I}$ *such that the following are ε-close in trace distance:*

- *(Real experiment)* $\mathbb{E}_k \left[\mathsf{Ver}_k \circ \mathcal{O} \circ \mathsf{Auth}_k\right] (\rho^{\mathcal{MSZ}})$
- *(Ideal experiment)* $\mathcal{I}(\rho^{\mathcal{MSZ}})$

where Auth_k *acts on* $\mathcal{M}$*,* Ver_k *acts on* $\mathcal{Y}$*, and both act as the identity on* $\mathcal{SZ}$*.*

Unlike total authentication, the key is averaged over in this security definition.

There is a minor caveat: we do not prove this implication for all authentication schemes. Instead, we prove it for authentication schemes *composed* with a Pauli randomization step. If $(\mathsf{Auth}, \mathsf{Ver})$ is an authentication scheme, we call this composed scheme $\mathsf{Pauli} + (\mathsf{Auth}, \mathsf{Ver})$, and it behaves as follows:

The secret key for $\mathsf{Pauli} + (\mathsf{Auth}, \mathsf{Ver})$ consists of the key k for $(\mathsf{Auth}, \mathsf{Ver})$, as well as a new, independent key k'. The procedure to authenticate a message register $\mathcal{M}$ behaves as follows: first, the key k' is used to choose a random unitary from the Pauli group that acts on the space $\mathcal{M}$.[14] We call this the *Pauli random-*

[14] For simplicitly let us think of $\mathcal{M}$ as $(\mathbb{C}^2)^{\otimes n}$ (i.e., n qubits). Then the Pauli group consists of all operators of the form $X^p Z^q$, where $p, q \in \{0,1\}^n$. Here, the operator X^p is defined to be the tensor product of $X_j^{p_j}$, where X_j is the X Pauli operator acting on the j'th qubit. Z^q is defined similarly.

ization step. Next, the key k is used to apply Auth_k to the register $\mathcal{M}$ to produce a state in the $\mathcal{Y}$ register. This is the authenticated state, which is then subject to attack by the adversary.

To un-authenticate, the Ver_k procedure is applied. Note that this is not a unitary operation, but includes the projection on the receiver's acceptance (see the Preliminaries for a discussion of this). Finally, the Pauli randomization is undone using the key k'.

Theorem 9 (Lifting weak authentication to total authentication). *Let* $(\mathsf{Auth}, \mathsf{Ver})$ *be an authentication scheme, and suppose the composed scheme* $\mathsf{Pauli} + (\mathsf{Auth}, \mathsf{Ver})$ *satisfies the following security guarantee: for all adversaries* $\mathcal{O} \in \mathrm{T}(\mathcal{Y}\mathcal{Z}, \mathcal{Y}\mathcal{Z})$*, for all adversary ancilla qubits* $|\theta\rangle^{\mathcal{Z}\mathcal{Z}'}$*, there exists an oblivious adversary* $\mathcal{I}$ *acting on* $\mathcal{Z}$ *only such that the following are* ε*-close in trace distance:*

- *(Real experiment)* $\mathbb{E}_{k,k'}\left[\mathsf{Pauli}_{k'}^{\dagger} \circ \mathsf{Ver}_k \circ \mathcal{O} \circ \mathsf{Auth}_k \circ \mathsf{Pauli}_{k'}\right](|\Phi\rangle\langle\Phi|^{\mathcal{M}\mathcal{B}} \otimes |\theta\rangle\langle\theta|^{\mathcal{Z}\mathcal{Z}'})$
- *(Ideal experiment)* $|\Phi\rangle\langle\Phi|^{\mathcal{M}\mathcal{B}} \otimes \mathcal{I}(|\theta\rangle\langle\theta|^{\mathcal{Z}\mathcal{Z}'})$

where $\mathcal{B}$ *is a Hilbert space isomorphic to* $\mathcal{M}$*, and* $|\Phi\rangle^{\mathcal{M}\mathcal{B}}$ *is the maximally entangled state.*

Then, the composed scheme $\mathsf{Pauli} + (\mathsf{Auth}, \mathsf{Ver})$ *is a* ε*-secure according to the [DNS12] definition.*

For proof of the above theorem, refer to the full version [GYZ16][15].

10 Open problems

We close with some open problems:

1. We showed that the Auth-QFT-Auth scheme achieves total authentication (with outer key leakage) when the inner authentication scheme is instantiated with the Wegman-Carter 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?
2. 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.
3. 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?
4. Does total authentication satisfy *Universally Composable* security (as defined in [BHL+05,Unr10])?

[15] `https://arxiv.org/abs/1607.07759`

References

ABE10. 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.

AM16. Gorjan Alagic and Christian Majenz. Quantum non-malleability and authentication. *arXiv preprint arXiv:1610.04214*, 2016.

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.

BCG+06. Michael Ben-Or, Claude Crépeau, Daniel Gottesman, Avinatan Hassidim, and Adam Smith. Secure multiparty quantum computation with (only) a strict honest majority. In *2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*, pages 249–260. IEEE, 2006.

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.

BHL+05. Michael Ben-Or, Michał Horodecki, Debbie W Leung, Dominic Mayers, and Jonathan Oppenheim. The universal composable security of quantum key distribution. In *Theory of Cryptography Conference*, pages 386–406. Springer, 2005.

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.

FS16. Serge Fehr and Louis Salvail. Quantum authentication and encryption with key recycling. *arXiv preprint arXiv:1610.05614*, 2016.

GHS15. Tommaso Gagliardoni, Andreas Hülsing, and Christian Schaffner. Semantic security and indistinguishability in the quantum world. *arXiv preprint arXiv:1504.05255*, 2015.

Got02. Daniel Gottesman. Uncloneable encryption. *arXiv preprint quant-ph/0210062*, 2002.

GYZ16. Sumegha Garg, Henry Yuen, and Mark Zhandry. New security notions and feasibility results for authentication of quantum data. *arXiv preprint arXiv:1607.07759*, 2016.

HLM16. Patrick Hayden, Debbie W Leung, and Dominic Mayers. The universal composable security of quantum message authentication with key recyling. *arXiv preprint arXiv:1610.09434*, 2016.

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.

NC10. Michael A Nielsen and Isaac L Chuang. *Quantum computation and quantum information*. Cambridge university press, 2010.

OH05. Jonathan Oppenheim and Michał Horodecki. How to reuse a one-time pad and other notes on authentication, encryption, and protection of quantum information. *Physical Review A*, 72(4):042309, 2005.

Por17. Christopher Portman. Quantum authentication with key recycling. In *Advances in Cryptology–EUROCRYPT 2017*. Springer, 2017.

Unr10. Dominique Unruh. Universally composable quantum multi-party computation. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 486–505. Springer, 2010.

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.

Zhu15. Huangjun Zhu. Multiqubit clifford groups are unitary 3-designs. *arXiv preprint arXiv:1510.02619*, 2015.