

# Why cut-and-choose quantum state verification cannot be both efficient and secure

Fabian Wiesner<sup>1</sup> , Ziad Chaoui<sup>1</sup> , Diana Kessler<sup>2</sup> , Anna Pappa<sup>1</sup>  and Martti Karvonen<sup>3</sup> 

<sup>1</sup> Technische Universität Berlin, Germany

<sup>2</sup> Tallinn University of Technology, Estonia

<sup>3</sup> University College London, UK

**Abstract.** Quantum state verification plays a vital role in many quantum cryptographic protocols, as it allows the use of quantum states from untrusted sources. While some progress has been made in this direction, the question of whether the most prevalent type of quantum state verification, namely cut-and-choose verification, can be efficient and secure, is still not answered in full generality. In this work, we show a fundamental limit for quantum state verification for all cut-and-choose approaches used to verify arbitrary quantum states. We provide a no-go result showing that the cut-and-choose techniques cannot lead to quantum state verification protocols that are both efficient in the number of rounds and secure. We show this trade-off for stand-alone and composable security, where the scaling of the lower bound for the security parameters renders cut-and-choose quantum state verification effectively unusable.

**Keywords:** Quantum state verification · Security limitations.

## 1 Introduction

For much of cryptography’s history, security has been assumed but not proven. Even today, we rely on protocols whose security proofs are based on conjectured hardness assumptions [KL14]. In the comparably young field of quantum cryptography, many protocols claim provable security under the assumption that the devices used in these protocols are trustworthy. Although they offer a real advantage in tackling modern cryptographic challenges [GRTZ02, PAB<sup>+</sup>20], they often come with two caveats:

1. Quantum hardware is expensive and difficult to operate and maintain. This is particularly true for quantum computers and their main building blocks, such as implementations of entangling gates [Pre18].
2. The devices might not be trustworthy. To assume otherwise might in fact be a very strong assumption; someone untrusted could be operating the device, or there could be a hardware-based attack that leaks important information, as was done in the past for quantum key distribution systems [LWW<sup>+</sup>10].

Interestingly, these two issues are connected. Indeed, one way to address the first issue is to delegate some complex tasks to other parties while ensuring they execute them as required. Quantum correlations provide a way to check that the operations and tasks at

---

E-mail: [f.wiesner@tu-berlin.de](mailto:f.wiesner@tu-berlin.de) (Fabian Wiesner), [ziad.chaoui@tu-berlin.de](mailto:ziad.chaoui@tu-berlin.de) (Ziad Chaoui), [diana-maria.kessler@taltech.ee](mailto:diana-maria.kessler@taltech.ee) (Diana Kessler), [anna.pappa@tu-berlin.de](mailto:anna.pappa@tu-berlin.de) (Anna Pappa), [martti.karvonen@ucl.ac.uk](mailto:martti.karvonen@ucl.ac.uk) (Martti Karvonen)



hand are executed correctly. In the most general case, this is done through a framework called “Device-Independence” [ABG<sup>+</sup>07], where the parties involved in the protocol can verify that the operations performed are correct, without putting any trust in the hardware. In this paper, we focus on one specific task: quantum state verification. Quantum state verification is crucial to a variety of applications, such as network parameter estimation [SHM22], in which certain parties (the clients) cannot prepare a required quantum state themselves<sup>1</sup> and therefore need to query a potentially dishonest source. Usually, in quantum state verification protocols, the untrusted source prepares quantum states and distributes them among the clients who are sometimes considered honest. If the source is honest, it always prepares the target state, i.e. the state the clients desire to hold, and the clients accept the result. However, if the source is dishonest, it might not always send the target state, and the clients should ideally reject it. Due to the no-cloning theorem, clients cannot simply measure and use the quantum states sent by the source. Hence, some other form of verification is necessary. A typical way to verify quantum states is for the source to send several copies of the state, some of which are then measured by the clients. If sufficiently many measurements match the expected state, the clients are convinced the source is honest. This *cut-and-choose* type of verification is used in many applications, such as anonymous conference key agreement [HdJP20], network parameter estimation [SHM22], and blind quantum computing [HM15]. With the emergence of quantum technologies and the effort to build a secure quantum internet [WEH18], these applications are of increasing importance. For cut-and-choose quantum state verification to be a viable subroutine to all these protocols, we expect it to be composable secure. While composable security is usually defined with respect to a certain framework (e.g. abstract cryptography), one can prove negative results in a more general way. Indeed, we present a no-go result for such verification techniques, which is valid for composable security (independent of the framework) as well as for stand-alone security.

### 1.1 Our contribution and related work

Many protocols implement quantum state verification for different types of states, e.g. [MTH17, TM18, PCW<sup>+</sup>12, UM22]. However, all protocols we are aware of solely rely on cut-and-choose and suffer from the same efficiency vs. security trade-off: a quantum state verification protocol cannot be both secure and efficient. We investigate this trade-off in a more general setting and prove the following no-go result for quantum state verification.

**Theorem 1** (Main result (informal)). *Let  $\pi$  be a cut-and-choose protocol for quantum state verification in which the clients output the state without performing any map on it, if they evaluate the source’s behavior as honest. At least one of the following statements about  $\pi$  with security parameter  $\lambda$  is false:*

1.  $\pi$  rejects the target state with a probability negligible in  $\lambda$ .
2. If the source is dishonest, either the probability to accept or the distance to the target state is negligible in  $\lambda$ .
3. The number of rounds  $N$  is polynomial in  $\lambda$ .

*Regarding composable security, we find with  $\varepsilon_H$  being the distinguishability to the idealized process if the source is honest and  $\varepsilon_D$  if it is not*

$$\varepsilon_H + \varepsilon_D \geq \frac{\sqrt{\eta_1}}{4\sqrt{N}},$$

---

<sup>1</sup>Note that if the clients could prepare the required quantum state, they would not query the untrusted source and would not use quantum state verification

where  $\eta_1$  is the highest eigenvalue of the target state (i.e. 1 if the target state is pure). Regarding stand-alone security, we find with  $\kappa_H$  being the fidelity between the honest output and  $\phi$  and  $\kappa_D$  being the fidelity of the dishonest output and  $p\phi \oplus (1-p)^2$  maximized over  $p$

$$\kappa_H + \kappa_D \geq \frac{1}{7N}.$$

We state and prove these results more formally in Section 3. We show that for a generic cut-and-choose protocol described in Algorithm 1 the inequalities of Theorem 1 hold. These inequalities imply a trade-off between correctness, security, and efficiency: A protocol with high correctness, that is with high  $\varepsilon_H$  or  $\kappa_H$ , has low security, i.e. low  $\varepsilon_D$  or  $\kappa_D$ , and vice versa. The efficiency of the protocol is given by the number of rounds  $N$ , and we can see that the lower bounds for the sums of correctness and security scale inversely with the round numbers. Similar trade-offs have been proven in other works before in the hypothesis testing framework<sup>3</sup>. More specifically, the previous works showing similar trade-offs [PLM18, ZH19] consider a protocol to be  $(\varepsilon, \delta)$ -secure, if the probability that the client accepts the behavior of the source is upper bounded by  $\delta$  when the deviation from the honest behavior is intolerable. This deviation is defined to be intolerable if the fidelity between the average output state of the protocol  $\rho$  and the target state  $\phi$  is upper bounded by  $1 - \varepsilon$ , where the average output state is obtained by the composition of any attack of the source and the protocol of the client, i.e.

$$\Pr[\text{accept} | F(\rho, \phi) \leq 1 - \varepsilon] \leq \delta.$$

Intuitively, this means that the probability that the client outputs a state that is too different from the target state should be low for all attacks the source might use. Using this definition and the assumption that the acceptance probability if the source is honest is 1, the authors of [PLM18] and [ZH19] showed that for a quantum verification protocol for any pure target state  $\phi$  with a fixed number of  $N + 1$  rounds, it holds that<sup>4</sup>

$$\begin{aligned} \Pr[\text{accept} | F(\rho, \phi) \leq 1 - \varepsilon] &= (1 - c\varepsilon)^N \leq \delta \\ \Rightarrow N &\geq \frac{\ln(\delta)}{\ln(1 - c\varepsilon)} \end{aligned}$$

where  $c = 1$  in [PLM18] and  $c \leq 1$  is a constant in [ZH19] which depends on the verification strategy.

Both, [PLM18] and [ZH19], assume that the clients perform single-round tests, i.e., do not use coherent measurements, although in [PLM18] the authors argue that this is not a restriction. However, our work differs from [PLM18, ZH19] in many aspects. First, the assumptions differ: we allow for *coherent measurements*, we derive a bound for *mixed target states* as well, do not require *perfect correctness*, and, very importantly, we do not consider a *fixed number of rounds*. While we note that single-state measurements suffice to optimally distinguish pure states and that one could use Uhlmann's theorem to derive a bound for mixed states as well, a fixed round number is a strong assumption for the protocol. Especially if the client does not use coherent but uncorrelated measurements, one can randomize the number of rounds, e.g., as in [PCW<sup>+</sup>12], to obtain protocols outside of the scope of previous results. Indeed, there is an intuitive attack against protocols with a fixed round number: An attacker always sends the target state except for one round, for which it sends an orthogonal state, guessing that this is the output round. Such an intuitive attack is generally not available if the round number is randomized or the attacker is i.i.d.-restricted.

<sup>2</sup>Note, that we consider the abort probability  $1 - p$  in a one-dimensional space using the direct sum, see the Section 2 for more details.

<sup>3</sup>See [YSG22] for a review on quantum state verification focused on the hypothesis testing approach.

<sup>4</sup>Note that dividing by  $\ln(1 - c\varepsilon)$  changes the direction of the inequality since  $\ln(1 - c\varepsilon) \leq 0$ .

However, our perspective on the topic also differs from that of previous works. In contrast to [PLM18, ZH19], we do not use the hypothesis testing framework, which is not common in many areas of quantum cryptography, despite being useful for quantum state verification. We argue, in line with [YDK21, CMY24], that quantum state verification should be viewed as a building block of larger protocols and hence investigate composability as well. Because of this difference, we developed a framework-agnostic proof technique by refuting an implication of security that all composable security frameworks share, i.e. we find a lower bound on the trace distance between ideal and real output states of a general quantum state verification protocol, which implies a no-go result in the different frameworks for composable cryptography. For both types, composable and stand-alone security, we present i.i.d. attacks that break cut-and-choose quantum state verification. While we find, for a fixed round number, that the intuitive attack is optimal for stand-alone security, our presented i.i.d. attack achieves a higher violation of composable security. We expect that our proofs can be adapted for other functionalities for which hypothesis testing is less common.

So while a direct comparison of the different bounds is not reasonable due to the differences in the security definitions and assumptions, a summary and comparison of the assumptions with the previous results is presented in Table 1.

**Table 1:** Summary of results regarding the trade-off. Note that while [ZH19] considers arbitrary attacks, in [PLM18] only i.i.d. attacks are possible.

Assumptions about the protocols as discussed above: 1) Fixed round number, 2) pure target state, 3) perfect correctness, 4) No coherent measurements for verification.

	Security type	Bound	Assumptions
[PLM18]	Hypothesis testing	$N \geq \frac{\ln(\delta)}{\ln(1-\varepsilon)}$	1,2,3,4
[ZH19]	Hypothesis testing	$N \geq \frac{\ln(\delta)}{\ln(1-c\varepsilon)}$	1,2,3,4
This Work, Lem. 2, 3 and Thm. 7	Fidelity-based	$\varepsilon_H + \varepsilon_D \geq 1/N+1$	1,2,4
This Work, Thm. 8	Fidelity-based	$\varepsilon_H + \varepsilon_D \geq 1/7N$	-
This Work, Thm. 9	Composable	$\varepsilon_H + \varepsilon_D \geq 1/4\sqrt{N}$	-

Finally, our results provide bounds for self-testing as well [vB20]. Self-testing is slightly different from quantum state verification, since there is a single client that does not trust any of their devices, including the preparation and measurement apparatus. Self-testing can therefore be seen as a stricter case of quantum state verification. Hence, any attack on quantum state verification implies an attack on self-testing.

## 1.2 Structure

Our work is structured as follows: In Section 2, we first present preliminaries that we need for our security proofs. In Section 3, we provide first the no-go result for a fidelity-based security definition and then for generic composable security with an i.i.d. restriction for the attacker. In Section 4, we investigate optimal attacks outside of the i.i.d. setting. Finally, we discuss open questions and possible implications of our work in Section 5. In the appendix, we present a generalization of our no-go result for protocols with a probabilistic round number, and the security proof for a specific protocol, which provides guidance regarding the tightness of the bounds we prove and the advantage of more advanced attacks.

## 2 Preliminaries

In the following we present some preliminaries – mainly on quantum information theory – that we need for our security analyses in the subsequent sections.

We denote by  $D(\mathcal{X})$  the space of density operators on the Hilbert space  $\mathcal{X}$ . For a density operator  $\rho \in D(\mathcal{X})$ , we define the *trace norm* to be  $\|\rho\|_1 := \text{Tr}(\sqrt{\rho\rho^\dagger})$ .  $P(\mathcal{X})$  is the space of all positive semidefinite operators, and we define a (binary) measurement to be a function of the form  $\mu : \{0, 1\} \rightarrow P(\mathcal{X})$ , satisfying  $\mu(0) + \mu(1) = \mathbb{1}_{\mathcal{X}}$ . For a density operator  $\rho \in D(\mathcal{X})$ ,  $\langle \mu(b) | \rho \rangle := \text{Tr}(\mu(b)^\dagger \rho)$  is then the probability to obtain measurement outcome  $b \in \{0, 1\}$  when measuring  $\rho$  with  $\mu$ . The trace distance yields a bound on the achievable distinguishing advantage of a measurement between two density operators given by the Holevo-Helstrom Theorem.

**Theorem 2** (Holevo-Helstrom Theorem). *Let  $\rho_0, \rho_1 \in D(\mathcal{X})$  be density operators, and let  $\lambda \in [0, 1]$ . For any measurement  $\mu : \{0, 1\} \rightarrow P(\mathcal{X})$  it then holds*

$$\lambda \langle \mu(0) | \rho_0 \rangle + (1 - \lambda) \langle \mu(1) | \rho_1 \rangle \leq \frac{1}{2} + \frac{1}{2} \|\lambda \rho_0 - (1 - \lambda) \rho_1\|_1. \quad (2.1)$$

Moreover there exists a projective measurement  $\mu : \{0, 1\} \rightarrow P(\mathcal{X})$  for which equality is achieved in (2.1).

To see that this actually gives a bound on the distinguishing advantage, we set  $\lambda = \frac{1}{2}$  in (2.1) and we obtain

$$\begin{aligned} \frac{1}{2} \langle \mu(0) | \rho_0 \rangle + \frac{1}{2} \langle \mu(1) | \rho_1 \rangle &\leq \frac{1}{2} + \frac{1}{4} \|\rho_0 - \rho_1\|_1 \\ \Leftrightarrow (\langle \mu(1) | \rho_1 \rangle - 1) + \langle \mu(0) | \rho_0 \rangle &= \langle \mu(0) | \rho_0 \rangle - \langle \mu(0) | \rho_1 \rangle \leq \frac{1}{2} \|\rho_0 - \rho_1\|_1. \end{aligned} \quad (2.2)$$

If  $\langle \mu(0) | \rho_0 \rangle \geq \langle \mu(0) | \rho_1 \rangle$  holds, then (2.2) gives a bound on the distinguishing advantage  $|\langle \mu(0) | \rho_0 \rangle - \langle \mu(0) | \rho_1 \rangle|$  which is the absolute value of the difference of the probabilities of the outcome 0. If, however,  $\langle \mu(0) | \rho_0 \rangle < \langle \mu(0) | \rho_1 \rangle$ , we define the measurement operator  $\gamma(0) := \mathbb{1}_{\mathcal{X}} - \mu(0)$  and find  $\langle \gamma(0) | \rho_0 \rangle \geq \langle \gamma(0) | \rho_1 \rangle$ . Since (2.1) holds for every measurement we then find

$$\langle \gamma(0) | \rho_0 \rangle - \langle \gamma(0) | \rho_1 \rangle = (1 - \langle \mu(0) | \rho_0 \rangle) - (1 - \langle \mu(0) | \rho_1 \rangle) = \langle \mu(0) | \rho_1 \rangle - \langle \mu(0) | \rho_0 \rangle \leq \frac{1}{2} \|\rho_0 - \rho_1\|_1$$

Hence, either way it holds

$$|\langle \mu(0) | \rho_0 \rangle - \langle \mu(0) | \rho_1 \rangle| \leq \frac{1}{2} \|\rho_0 - \rho_1\|_1. \quad (2.3)$$

Another important quantity is the fidelity. The fidelity between two density operators  $\rho_0, \rho_1$  is given by<sup>5</sup>

$$F(\rho_0, \rho_1) := \text{Tr} \left( \sqrt{\sqrt{\rho_0} \rho_1 \sqrt{\rho_0}} \right)^2.$$

Although the fidelity is not a metric, it allows to quantify how close or similar two density operators are: the higher the fidelity, the closer the states. We also define the infidelity to be  $1 - F(\rho_0, \rho_1)$ . For all pure states  $\rho_0 = |\psi_0\rangle\langle\psi_0|$  and  $\rho_1 = |\psi_1\rangle\langle\psi_1|$  it holds that

$$F(\rho_0, \rho_1) = \langle \psi_0 | \rho_1 | \psi_0 \rangle = |\langle \psi_0 | \psi_1 \rangle|^2, \quad (2.4)$$

<sup>5</sup>Note that other authors define the square root of this expression as the fidelity.

where the first equality holds even if  $\rho_1$  is not pure. Further, we have the following properties for all density operators  $\rho_0, \sigma_0 \in D(\mathcal{X})$ ,  $\rho_1, \sigma_1 \in D(\mathcal{Y})$ , and  $\lambda \geq 0$

$$F(\rho_0 \otimes \rho_1, \sigma_0 \otimes \sigma_1) = F(\rho_0, \sigma_0)F(\rho_1, \sigma_1). \quad (2.5)$$

$$\sqrt{F(\rho_0 \oplus \rho_1, \sigma_0 \oplus \sigma_1)} = \sqrt{F(\rho_0, \sigma_0)} + \sqrt{F(\rho_1, \sigma_1)} \quad (2.6)$$

$$F(\lambda \rho_0, \sigma_0) = \lambda F(\rho_0, \sigma_0), \quad (2.7)$$

where  $R \oplus Q$  denotes the direct sum for linear operators, i.e.

$$R \oplus Q = \begin{pmatrix} R_{1,1} & \dots & R_{1,b} \\ \vdots & \vdots & \vdots \\ R_{a,1} & \dots & R_{a,b} \end{pmatrix} \oplus \begin{pmatrix} Q_{1,1} & \dots & Q_{1,d} \\ \vdots & \vdots & \vdots \\ Q_{c,1} & \dots & Q_{c,d} \end{pmatrix} := \begin{pmatrix} R_{1,1} & \dots & R_{1,b} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ R_{a,1} & \dots & R_{a,b} & 0 & \dots & 0 \\ 0 & \dots & 0 & Q_{1,1} & \dots & Q_{1,d} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & Q_{c,1} & \dots & Q_{c,d} \end{pmatrix}.$$

The Fuchs-van de Graaf inequalities link the trace distance to the fidelity [Wat18].

**Theorem 3** (Fuchs-van de Graaf Inequalities). *Let  $\rho_0, \rho_1 \in D(\mathcal{X})$  be density operators, it holds that*

$$1 - \sqrt{F(\rho_0, \rho_1)} \leq \frac{1}{2} \|\rho_0 - \rho_1\|_1 \leq \sqrt{1 - F(\rho_0, \rho_1)}, \quad (2.8)$$

$$\left(1 - \frac{1}{2} \|\rho_0 - \rho_1\|_1\right)^2 \leq F(\rho_0, \rho_1) \leq 1 - \frac{1}{4} \|\rho_0 - \rho_1\|_1^2. \quad (2.9)$$

Using these inequalities and the properties of fidelity, we can easily derive a simple bound on  $k$ -tuples of density operators  $\{(\rho_i, \sigma_i)\}_{i=1}^k$ :

$$\frac{1}{2} \left\| \bigotimes_{i=1}^k \rho_i - \bigotimes_{i=1}^k \sigma_i \right\|_1 \stackrel{(2.8)}{\leq} \sqrt{1 - F\left(\bigotimes_{i=1}^k \rho_i, \bigotimes_{i=1}^k \sigma_i\right)} \stackrel{(2.5)}{=} \sqrt{1 - \prod_{i=1}^k F(\rho_i, \sigma_i)}. \quad (2.10)$$

For pure states  $\rho = |\psi\rangle\langle\psi|$ ,  $\sigma = |\phi\rangle\langle\phi|$ , it holds that

$$\| |\psi\rangle\langle\psi| - |\phi\rangle\langle\phi| \|_1 = 2\sqrt{1 - |\langle\psi|\phi\rangle|^2}, \quad (2.11)$$

which implies for  $\rho_i = |\psi_i\rangle\langle\psi_i|$ ,  $\sigma_i = |\phi_i\rangle\langle\phi_i|$

$$\frac{1}{2} \left\| \bigotimes_{i=1}^k \rho_i - \bigotimes_{i=1}^k \sigma_i \right\|_1 = \sqrt{1 - \prod_{i=1}^k |\langle\psi_i|\phi_i\rangle|^2}. \quad (2.12)$$

\* \* \*

For our proofs, we will also use an important result from probability theory: Jensen's inequality for concave functions. We use the standard notation  $\mathbb{E}(X)$  for the expected value of a random variable  $X$ .

**Theorem 4** (Jensen's inequality). *Suppose  $X$  is a random variable and  $f : \mathbb{R} \rightarrow \mathbb{R}$  a concave function. It holds that*

$$f(\mathbb{E}(X)) \geq \mathbb{E}(f(X)).$$

*In particular for a random variable  $X$  with binomial distribution  $B(n, p)$ , we have*

$$f(np) \geq \sum_{i=0}^n \binom{n}{i} p^i (1-p)^{n-i} f(i). \quad (2.13)$$

We will use the following result repeatedly.

**Lemma 1.** *Let  $g : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$  given by  $g : x \mapsto \sqrt{1 - a^x}$ .  $g$  is concave on  $\mathbb{R}_{\geq 0}$  if  $a \in [0, 1]$ .*

*Proof.* To prove the claim, consider the first derivative

$$\frac{dg}{dx}(y) = \frac{\ln(a^{-1})a^y}{2\sqrt{1 - a^y}}.$$

The first derivative is non-negative, the numerator is non-increasing, and the denominator is non-decreasing in  $\mathbb{R}_{\geq 0}$  since  $a \in [0, 1]$ . Hence, the first derivative is non-increasing, which implies that  $g$  is concave.  $\square$

\* \* \*

Note that we will consider a different but equivalent notation for the mixture of the actual quantum output of the verification protocol and the abort state. Many works denote such a state as  $p\sigma + (1 - p)|\perp\rangle\langle\perp|$ , where  $p$  is the acceptance probability, and require that  $\sigma$  lives in a subspace that is orthogonal to  $|\perp\rangle\langle\perp|$ . In contrast, we make this orthogonality more explicit by writing  $p\sigma \oplus (1 - p)$ . Hence, the state in our notation obeys a direct sum structure of the space of the quantum output and a one-dimensional abort space in which  $1 - p$  is a sub-normalized state.

We further shall explain the concept of an average state: The average state is the outcome of the protocol averaged over all branches (i.e., all sampling processes and measurement outcomes), where each branch is weighted with its probability. By the commutation of the sum and the direct sum, it holds that the average state for a verification protocol is given by

$$\rho_{\text{av}} = \left( \sum_{e \in E} q_e p_e \rho_e \right) \oplus \sum_{e \in E} q_e (1 - p_e), \quad (2.14)$$

where  $E$  is the set of all branches,  $q_e$  is the probability of this branch,  $p_e$  is the acceptance probability and  $\rho_e$  is the output for each branch.

### 3 No-go results with i.i.d. attacks

We consider a generic protocol with  $N + 1$  rounds, where in each round, the source sends one register to the clients. The clients sample according to a distribution  $\omega$  which register  $i$  they use for the output and perform a measurement  $\mu_i = \{\mu_i(0), \mathbb{1} - \mu_i(0)\}$  on the other registers. If the outcome is 0, the clients output the remaining register, otherwise they output the abort signal  $|\perp\rangle\langle\perp|$ .

---

**Algorithm 1** Generic cut-and-choose protocol.

---

- $\phi \in D(\mathcal{X})$  is the target state,
- $N + 1$  is the number of rounds,
- $\omega$  is a probability distribution on  $\{1, \dots, N + 1\}$ ,
- $\{\mu_i(0) \in P(\mathcal{X}^{\otimes N})\}_{1 \leq i \leq N + 1}$  is a set of measurement operators.

- 1: The source sends  $N + 1$  registers  $\rho_1, \dots, \rho_{N+1}$ .
- 2: The clients sample which round is used for the output:  $k \leftarrow_{\omega} \{1, \dots, N + 1\}$
- 3: The clients measure  $\rho_1, \dots, \rho_{k-1}, \rho_{k+1}, \dots, \rho_{N+1}$  using  $\{\mu_k(0), \mathbb{1}_{\mathcal{X}^{\otimes N}} - \mu_k(0)\}$ , the outcome is  $r \in \{0, 1\}$ .
- 4: **if**  $r = 0$  **then**
- 5:     The clients output  $\rho_k$ .
- 6: **else**
- 7:     The clients output  $|\perp\rangle\langle\perp|$ .
- 8: **end if**

---

### 3.1 Stand-Alone security

In this section, we first present the no-go result for stand-alone security with a fidelity-based security definition. The security definition we consider is rather simple: If all parties are honest, the output of the protocol should have a high fidelity with the target state. If, however, the source is dishonest, the clients can abort, which is why we consider the optimal abort probability and compute the fidelity of the output with the target state probabilistically mixed with the abort state. As mentioned in Section 2, we expand the space  $\phi$  lives in by the one-dimensional abort space using the direct sum. The average states of the clients when the source is honest or dishonest,  $\rho_H$  and  $\rho_D$ , have the same structure: The first part of the direct sum represents the actual quantum state the clients get if they accept and the second part is the probability that they reject. To simplify the notation we don't write out the implicit one dimensional abort state, and only write the abort probability.

**Definition 1.** Consider a quantum state verification protocol with a target state  $\phi$ . Denote with  $\rho_H$  the average state of the clients if all parties are honest. We define the protocol to be  $\varepsilon_H$ -correct if

$$F(\rho_H, \phi \oplus 0) \geq 1 - \varepsilon_H. \quad (3.1)$$

In the case where the source is dishonest, we denote with  $\rho_D$  the average state of the clients. We define a quantum state verification protocol to be  $\varepsilon_D$ -secure against a dishonest source if

$$\max_{p \in [0,1]} F(\rho_D, p\phi \oplus (1-p)) \geq 1 - \varepsilon_D. \quad (3.2)$$

Intuitively, (3.2) demands that the average output state for any attack the source conducts is at most  $\varepsilon_D$  different from the closest state the client could get if there was the possibility to abort (with probability  $1 - p$ ) without replacing the actual quantum register. So while (3.1) punishes a high abort probability if the source is honest, a high abort probability should not be disadvantageous if the source is dishonest. Further, note that the maximization over the ideal acceptance probability  $p \in [0, 1]$  in (3.2) is only of notational relevance and does not imply a deviation from the security definition as used in e.g. [CMY24]. In fact, it holds that

$$\left( \max_{p \in [0,1]} F(\rho_D, p\phi \oplus (1-p)) \geq 1 - \varepsilon_D \right) \Leftrightarrow \left( \exists p \in [0,1] : F(\rho_D, p\phi \oplus (1-p)) \geq 1 - \varepsilon_D \right).$$

And since  $F(\rho_D, p\phi \oplus (1-p))$  is continuous in  $p$ , as one can see using (2.6), (2.7) and (2.14), and  $[0, 1]$  is a compact set, the maximum always exists. However, providing an upper bound one has to consider the maximization over  $p \in [0, 1]$  either way.

Now we are equipped to prove one of our main results; the trade-off for cut-and-choose quantum state verification protocols with regard to stand-alone security as defined in Definition 1. Theorem 5 formalizes this trade-off, i.e., that cut-and-choose quantum state verification cannot be simultaneously correct, secure, and efficient. More specifically, (3.3) states that a low incorrectness, corresponding to a low  $\varepsilon_H$ , implies a high insecurity  $\varepsilon_D$ . The efficiency of the protocol is given by the number of rounds, and the lower bound scales inversely to the round number: The fewer rounds the protocol requires, the higher the lower bound for the incorrectness and insecurity.

**Theorem 5.** *Let  $\pi = (\pi_C, \pi_S)$  be a protocol as described in 1. If  $\pi$  is  $\varepsilon_H$ -correct and  $\varepsilon_D$ -secure according to definition 1, it holds*

$$\varepsilon_H + \varepsilon_D \geq \frac{1}{7N}. \quad (3.3)$$

*Proof.* In the proof, we use the property of the fidelity under the direct sum for both settings, the honest one and the dishonest one, and simplify and add the resulting equations. Finally, the i.i.d. property of the attack allows us to discard the probabilities  $\omega(n)$ , and known inequalities and optimized choices for the parameter of the attack yield the result. We start by denoting

$$p_A^H := \sum_{i=1}^{N+1} \omega(i) \langle \mu_i(0) | \phi^{\otimes N} \rangle,$$

$$p_A^D := \sum_{i=1}^{N+1} \omega(i) \langle \mu_i(0) | \psi^{\otimes N} \rangle,$$

where  $\psi$  is the state of which a dishonest source sends  $N + 1$  copies in an i.i.d. attack. Then  $p_A^H$  and  $p_A^D$  are the average probabilities that the clients accept the verification in the honest and dishonest case, respectively. We find

$$\rho_H = p_A^H \phi \oplus (1 - p_A^H),$$

which implies by (2.6)

$$F(\rho_H, \phi \oplus 0) = p_A^H \Rightarrow \varepsilon_H \geq 1 - p_A^H. \quad (3.4)$$

If the source is dishonest and sends  $N + 1$  copies of  $\psi$  we find

$$\rho_D = p_A^D \psi \oplus (1 - p_A^D),$$

which implies again by (2.6)

$$\max_{p \in [0,1]} F(\rho_D, p\phi \oplus (1 - p)) = \max_{p \in [0,1]} \left( \sqrt{p_A^D p} \sqrt{F(\phi, \psi)} + \sqrt{(1 - p_A^D)(1 - p)} \right)^2$$

$$\Rightarrow \varepsilon_D \geq \min_{p \in [0,1]} \left( 1 - \left( \sqrt{p_A^D p} \sqrt{F(\phi, \psi)} + \sqrt{(1 - p_A^D)(1 - p)} \right)^2 \right).$$

When considering

$$f(p) := \sqrt{pa} + \sqrt{1 - pb}$$

we find for  $a, b \geq 0$  and  $0 < p' < 1$ :

$$\frac{d}{dp} f(p') = \frac{a}{2\sqrt{p'}} - \frac{b}{2\sqrt{1 - p'}} = 0 \Leftrightarrow \frac{a^2}{p'} = \frac{b^2}{1 - p'} \Leftrightarrow p' = \frac{a^2}{a^2 + b^2}.$$

Using this for maximizing  $f^2$  yields

$$\max_{p \in [0,1]} f(p)^2 = \left( \frac{a^2}{\sqrt{a^2 + b^2}} + \frac{b^2}{\sqrt{a^2 + b^2}} \right)^2 = a^2 + b^2.$$

Using  $a = \sqrt{p_A^D F(\phi, \psi)}$  and  $b = \sqrt{1 - p_A^D}$  gives

$$\varepsilon_D \geq 1 - \max_{p \in [0,1]} F(\rho_D, p\phi \oplus (1 - p)) = 1 - (p_A^D F(\phi, \psi) + 1 - p_A^D) = p_A^D (1 - F(\phi, \psi)),$$

which is smaller than  $1 - F(\rho_D, \phi \oplus 0) = 1 - p_A^D F(\phi, \psi)$  and  $1 - F(\rho_D, (0 \cdot \phi) \oplus 1) = p_A^D$ . Combined with the bound for  $\varepsilon_H$ , we find that

$$\begin{aligned} \varepsilon_H + \varepsilon_D &\geq 1 - p_A^H + p_A^D (1 - F(\phi, \psi)) \geq (1 - F(\phi, \psi)) (1 - p_A^H) + (1 - F(\phi, \psi)) p_A^D \\ &= (1 - F(\phi, \psi)) (1 - (p_A^H - p_A^D)) \geq (1 - F(\phi, \psi)) (1 - |p_A^H - p_A^D|). \end{aligned}$$

$|p_A^H - p_A^D|$  is the distinguishing advantage, i.e. the absolute value of the difference of the measurement probabilities, when distinguishing  $\phi^{\otimes N}$  and  $\psi^{\otimes N}$  using  $\mu_i$  averaged over  $i$ . Therefore, using the Holevo-Helstrom theorem (in particular (2.3)) and (2.10) we find

$$\begin{aligned} |p_A^H - p_A^D| &\leq \sum_{i=1}^{N+1} \omega(i) |\langle \mu_i(0) | \phi^{\otimes N} \rangle - \langle \mu_i(0) | \psi^{\otimes N} \rangle| \\ &\leq \frac{1}{2} \|\phi^{\otimes N} - \psi^{\otimes N}\| \sum_{i=1}^{N+1} \omega(i) \leq \sqrt{1 - F(\phi, \psi)^N}. \end{aligned} \quad (3.5)$$

Combined with the above, this means

$$\varepsilon_H + \varepsilon_D \geq (1 - F(\phi, \psi)) \left( 1 - \sqrt{1 - F(\phi, \psi)^N} \right).$$

Now we define  $\tau := (1 - F(\phi, \psi))$ , i.e.  $F(\phi, \psi) = (1 - \tau)$ :

$$\varepsilon_H + \varepsilon_D \geq \tau \left( 1 - \sqrt{1 - (1 - \tau)^N} \right).$$

With  $\alpha \in [0, 1]$ , we choose  $\tau = \alpha/N$ , and using  $(1 - \alpha/N)^N \geq 1 - \alpha$  we get

$$\varepsilon_H + \varepsilon_D \geq \frac{\alpha}{N} \left( 1 - \sqrt{1 - \left( 1 - \frac{\alpha}{N} \right)^N} \right) \geq \frac{\alpha}{N} (1 - \sqrt{\alpha}) =: h_N(\alpha), \quad (3.6)$$

which is maximized for  $\alpha = 4/9$  since

$$\frac{dh_N}{d\alpha}(\alpha_0) = N - \frac{3N}{2} \sqrt{\alpha_0} = 0 \Rightarrow \alpha_0 = \frac{4}{9}$$

and  $h_N(0) = h_N(1) = 0$ .

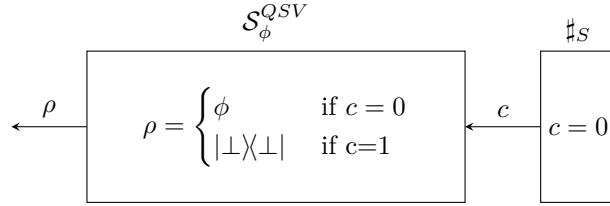
This in turn yields

$$\varepsilon_H + \varepsilon_D \geq \frac{4}{27N} \geq \frac{1}{7N}.$$

□

## 3.2 Composable security

We consider composable security definitions following the ‘real world vs. ideal world’ paradigm. Such security definitions have been presented for *universal composable* (UC) [Can20], *abstract cryptography* (AC) [MR11], and *categorical composable cryptography* [BK22, BK23] (CCC); for every attack on the implementation, there has to be an attack on the ideal resource which makes the two settings indistinguishable up to some  $\varepsilon \geq 0$ . While in AC and UC, a simulator translates attacks on the implementation into attacks on the ideal resource, in CCC, the user chooses the ideal attack more freely. We omit the actual security definitions and focus on finding a lower bound on the trace distance between ideal and real output states, which translates into a no-go result for the above-mentioned frameworks for composable cryptography. Agnostic of the actual framework, we consider the ideal resource in Fig. 1 for quantum state verification for a target state  $\phi$ .



**Figure 1:** Ideal quantum state verification: The source chooses whether to be honest or not by inputting a bit  $c \in \{0, 1\}$  to the ideal resource  $S_\phi^{QSV}$ . If it is honest,  $c$  is 0 and the ideal resource sends the target state  $\phi$  to the clients. If the source is dishonest,  $c$  can be set to 1 and the clients receive an abort signal. The right box  $\#_S$  represents a filter that enforces honest behavior of the ideal resource if the source is honest, i.e.  $\#_S$  fixes the input  $c = 0$ . In case the source is dishonest, we ignore the filter which allows the source to input  $c = 1$ .

Regarding correctness, the Holevo-Helstrom theorem (Theorem 2) states that the distinguishing advantage when given  $\phi$  from the ideal resource or  $\rho_H$  from the implementation, is given by the trace distance, i.e.  $\varepsilon_H$ -correctness implies

$$\frac{1}{2} \|\rho_H - (\phi \oplus 0)\|_1 \leq \varepsilon_H.$$

For security, if the source is dishonest, we need to consider all possible ideal attacks or simulators, respectively. However, all an ideal attack or simulator can do at the interface of the ideal resource is to input  $c = 0$  with some probability  $p$  which might depend on the attack on the implementation. While a distinguisher (or attacker) has the freedom to use a private register to increase the distinguishing advantage, it suffices for our no-go result to omit such register, i.e. composable security up to  $\varepsilon_D$  implies with  $\rho_D$  being the output state of the implementation

$$\min_{p \in [0,1]} \frac{1}{2} \|\rho_D - (p\phi \oplus (1-p))\|_1 \leq \varepsilon_D$$

but not necessarily vice-versa. Note that the minimization over all probabilities comes from the choice of the ideal attack or simulator, respectively.

**Theorem 6.** *Let  $\pi = (\pi_C, \pi_S)$  be a protocol as described in Protocol 1. If*

$$\frac{1}{2} \|\rho_H - (\phi \oplus 0)\|_1 \leq \varepsilon_H$$

and

$$\min_{p \in [0,1]} \frac{1}{2} \|\rho_D - (p\phi \oplus (1-p))\|_1 \leq \varepsilon_D$$

*It holds  $\varepsilon_H + \varepsilon_D \geq \frac{\sqrt{\eta_1}}{4\sqrt{N}}$  if  $\phi$  is mixed and  $\eta_1$  is the largest eigenvalue of  $\phi$  and  $\varepsilon_H + \varepsilon_D \geq \frac{1}{4\sqrt{N}}$  if  $\phi$  is pure.*

*Proof.* Similar to Theorem 5, we use the properties of the trace distance for each setting, honest and dishonest source, separately, and add the resulting lower bounds. Next, we fix that type of states that the source uses for the attack and derive in an intricate analysis, a lower bound for the trace distance of  $N$  copies of the target state and the state the dishonest source sends, which is similar to the trace distance of  $N$  pure states. Finally, known inequalities and optimized parameter choices prove the lower bound claimed in the

theorem.

We first note that for any positive semidefinite operators  $P_0, P_1, Q_0$  and  $Q_1$  it holds

$$\|(P_0 \oplus P_1) - (Q_0 \oplus Q_1)\|_1 = \|P_0 - Q_0\|_1 + \|P_1 - Q_1\|_1.$$

As before, we denote

$$\begin{aligned} p_A^H &:= \sum_{i=1}^{N+1} \omega(i) \langle \mu_i(0) | \phi^{\otimes N} \rangle \\ p_A^D &:= \sum_{i=1}^{N+1} \omega(i) \langle \mu_i(0) | \psi^{\otimes N} \rangle. \end{aligned}$$

We find

$$\rho_H = p_A^H \phi \oplus (1 - p_A^H),$$

which implies

$$\varepsilon_H \geq \frac{1}{2} \|\rho_H - (\phi \oplus 0)\| = 1 - p_A^H.$$

If the source is dishonest and sends  $N + 1$  copies of  $\psi$  we find

$$\rho_D = p_A^D \psi \oplus (1 - p_A^D),$$

which implies, using the triangle inequality

$$\begin{aligned} \varepsilon_D &\geq \min_{p \in [0,1]} \frac{1}{2} \|\rho_D - (p\phi \oplus (1-p))\|_1 = \min_{p \in [0,1]} \frac{1}{2} \|p_A^D \psi - p\phi\|_1 + \frac{1}{2} |p - p_A^D| \\ &= \min_{p \in [0,1]} \frac{1}{2} \|p_A^D \psi - p\phi\|_1 + \frac{1}{2} \|p\phi - p_A^D \phi\|_1 \geq \frac{p_A^D}{2} \|\psi - \phi\|_1. \end{aligned}$$

Now, we can combine both settings and find

$$\varepsilon_H + \varepsilon_D \geq \frac{p_A^D}{2} \|\psi - \phi\|_1 + 1 - p_A^H \geq \frac{1}{2} \|\psi - \phi\|_1 (1 - |p_A^H - p_A^D|).$$

Using (3.5), we find

$$\varepsilon_H + \varepsilon_D \geq \frac{1}{2} \|\psi - \phi\|_1 \left( 1 - \frac{1}{2} \|\phi^{\otimes N} - \psi^{\otimes N}\|_1 \right) \quad (3.7)$$

In the next step, we fix  $\psi$ . We write  $\phi$  in spectral decomposition as  $\phi = \sum_{i=1}^d \eta_i |\phi_i\rangle\langle\phi_i|$  with  $d = \text{rank}(\phi)$  and  $\eta_i \geq \eta_{i+1}$  for  $1 \leq i < d$ . For some pure state  $|\chi\rangle$ , we fix  $\psi = \eta_1 |\chi\rangle\langle\chi| + \sum_{i=2}^d \eta_i |\phi_i\rangle\langle\phi_i|$ . This implies

$$\frac{1}{2} \|\phi - \psi\|_1 = \eta_1 \sqrt{1 - |\langle\phi_1|\chi\rangle|^2}. \quad (3.8)$$

In the case of  $N$  copies, things become more complicated. We first note that

$$\phi^{\otimes N} = \sum_{A \in \{1, \dots, d\}^N} \bigotimes_{i \in A} \eta_i |\phi_i\rangle\langle\phi_i|,$$

and similarly with  $|\chi_i\rangle\langle\chi_i| := |\phi_i\rangle\langle\phi_i|$  for  $i \neq 1$  and  $|\chi_1\rangle\langle\chi_1| := |\chi\rangle\langle\chi|$

$$\psi^{\otimes N} = \sum_{A \in \{1, \dots, d\}^N} \bigotimes_{i \in A} \eta_i |\chi_i\rangle\langle\chi_i|.$$

Hence, we find

$$\phi^{\otimes N} - \psi^{\otimes N} = \sum_{A \in \{1, \dots, d\}^N} \left[ \bigotimes_{i \in A} \eta_i |\phi_i\rangle\langle\phi_i| - \bigotimes_{i \in A} \eta_i |\chi_i\rangle\langle\chi_i| \right].$$

Now consider the trace norm of this difference. We find an upper bound using the triangle inequality in a way, such that we group the terms for which the positions of the 1s in  $A$  are the same, i.e. we get  $2^N$  trace differences. Next, using the unitary invariance of the trace norm, we can move all positions corresponding to 1 in  $A$  to the front. Next, we use

$$\|\rho_1^{\otimes i} \otimes \nu^{\otimes(N-i)} - \rho_2^{\otimes i} \otimes \nu^{\otimes(N-i)}\|_1 = \|\rho_1^{\otimes i} - \rho_2^{\otimes i}\|_1 \|\nu^{\otimes(N-i)}\|_1,$$

with  $\rho_1 = \eta_1 |\phi_1\rangle\langle\phi_1|$ ,  $\rho_2 = \eta_1 |\chi_1\rangle\langle\chi_1|$  and  $\nu = \sum_{i=2}^N \eta_i |\phi_i\rangle\langle\phi_i|$ . Now, we note that  $\|\nu^{\otimes(N-i)}\|_1 = (1 - \eta_1)^{N-i}$ . So we find

$$\frac{1}{2} \|\phi^{\otimes N} - \psi^{\otimes N}\|_1 \leq \sum_{i=0}^N \binom{N}{i} \frac{1}{2} \left\| |\phi_1\rangle\langle\phi_1|^{\otimes i} - |\chi\rangle\langle\chi|^{\otimes i} \right\|_1 \eta_1^i (1 - \eta_1)^{N-i}.$$

Using the expression for the trace distance of pure states (2.11) and Jensen's inequality for concave functions with a binomial distribution (2.13), with  $f : x \rightarrow \sqrt{1 - |\langle\phi_1|\chi\rangle|^{2x}}$  (cf. Lemma 1) and binomial distribution  $B(N, \eta_1)$ , we find

$$\begin{aligned} \frac{1}{2} \|\phi^{\otimes N} - \psi^{\otimes N}\|_1 &\leq \sum_{i=0}^N \binom{N}{i} \sqrt{1 - |\langle\phi_1|\chi\rangle|^{2i}} \eta_1^i (1 - \eta_1)^{N-i} \\ &\leq \sqrt{1 - |\langle\phi_1|\chi\rangle|^{2\eta_1 N}}. \end{aligned}$$

So for  $\varepsilon_H + \varepsilon_D$  we find

$$\varepsilon_H + \varepsilon_D \geq \eta_1 \sqrt{1 - |\langle\phi_1|\chi\rangle|^2} \left( 1 - \sqrt{1 - |\langle\phi_1|\chi\rangle|^{2\eta_1 N}} \right).$$

Writing  $\tau := \sqrt{1 - |\langle\phi_1|\chi\rangle|^2}$  yields

$$\varepsilon_H + \varepsilon_D \geq \eta_1 \tau \left( 1 - \sqrt{1 - (1 - \tau^2)^{\eta_1 N}} \right).$$

Choosing  $\tau = \alpha/\sqrt{\eta_1 N}$  gives us using  $(1 - a/N)^N \geq 1 - a$

$$\varepsilon_H + \varepsilon_D \geq \alpha \sqrt{\eta_1 N} / \sqrt{N} (1 - \alpha).$$

The optimal value for  $\alpha$  is  $1/2$ , hence we find

$$\varepsilon_H + \varepsilon_D \geq \frac{\sqrt{\eta_1}}{4\sqrt{N}}.$$

In particular, for pure states this gives us

$$\varepsilon_H + \varepsilon_D \geq \frac{1}{4\sqrt{N}}.$$

□

This provides directly a lower bound for composable security definitions as used in [MR11, BK22]. The distinguisher can implement the presented attack without an auxiliary register. By design of the ideal resource, the simulator can only accept or reject with some probability; hence, the relevant measure for this attack is in fact the trace distance.

## 4 Optimal attacks

With a similar proof technique, we now prove our no-go result for more complex attacks without the i.i.d. restriction. We study optimal attacks, attacks leading to the highest security violation, i.e. that maximize  $\varepsilon_H + \varepsilon_D$ . By dropping the i.i.d. restriction, we study the case where the dishonest source is free to send an arbitrary state in every round. Nevertheless, we still consider the attacker, i.e. the dishonest source, to act in the class of separable state attacks. More precisely, we demand that there is no entanglement between states from different rounds, i.e., an attacker sends separable states  $\{\psi_j\}_{j=1}^{N+1}$ . This class, in fact, contains a naive attack that could lead to the clients sharing a state that is orthogonal to the desired output, which is the maximal deviation possible. In this naive attack the source guesses which round will be used for the output and sends for this particular round an orthogonal state, but for all other rounds it sends the actual target state to minimize the risk of getting caught.

This naive attack and its comparison to the i.i.d. attacks motivate this investigation.

**Lemma 2.** *If  $\phi$  in Protocol 1 is pure and an attacker sends separable states  $\{\psi_j\}_{j=1}^{N+1}$ , it holds*

$$F(\rho_H, \phi) = \sum_{i=1}^{N+1} \omega(i) p_H(i)$$

$$\max_{p \in [0,1]} F(\rho_D, p\phi \oplus (1-p)) = 1 - \sum_{i=1}^{N+1} \omega(i) p_D(i)(1 - F(\psi_i, \phi)),$$

where  $\rho_H$  is the output state in the honest setting,  $\rho_D$  the one if the source cheats and

$$p_H(i) := \left\langle \mu_i(0) \middle| \bigotimes_{j=1, j \neq i}^{N+1} \phi \right\rangle$$

$$p_D(i) := \left\langle \mu_i(0) \middle| \bigotimes_{j=1, j \neq i}^{N+1} \psi_j \right\rangle.$$

*Proof.* The proof uses the property of the fidelity under the direct sum and results which we introduced in Theorem 5.

In Protocol 1, we find if the source is honest:

$$\rho_H = \left( \sum_{i=1}^{N+1} \omega(i) p_H(i) \right) \phi \oplus \left( 1 - \sum_{i=1}^{N+1} \omega(i) p_H(i) \right),$$

and

$$F(\rho_H, \phi \oplus 0) = \sum_{i=1}^{N+1} \omega(i) p_H(i),$$

which is already the first part of the statement. In the dishonest setting, we find

$$\rho_D = \sum_{i=1}^{N+1} \omega(i) p_D(i) \psi_i \oplus \left( 1 - \sum_{i=1}^{N+1} \omega(i) p_D(i) \right).$$

Similarly, as in the previous section, this implies

$$\max_{p \in [0,1]} F(\rho_D, p\phi \oplus 1-p) = \left( \sqrt{p} \sqrt{F \left( \sum_{i=1}^{N+1} \omega(i) p_D(i) \psi_i, \phi \right)} + \sqrt{1-p} \sqrt{1 - \sum_{i=1}^{N+1} \omega(i) p_D(i)} \right)^2.$$

Using  $\max_{p \in [0,1]} (\sqrt{p}a + \sqrt{1-p}b)^2 = a^2 + b^2$ , we find

$$\max_{p \in [0,1]} F(\rho_D, p\phi \oplus 1 - p) = F\left(\sum_{i=1}^{N+1} \omega(i)p_D(i)\psi_i, \phi\right) + \left(1 - \sum_{i=1}^{N+1} \omega(i)p_D(i)\right).$$

The assumption that  $\phi$  is pure implies

$$\begin{aligned} \max_{p \in [0,1]} F(\rho_D, p\phi \oplus 1 - p) &= \sum_{i=1}^{N+1} \omega(i)p_D(i) \langle \phi | \psi_i | \phi \rangle + \left(1 - \sum_{i=1}^{N+1} \omega(i)p_D(i)\right) \\ &= \sum_{i=1}^{N+1} \omega(i)p_D(i)F(\psi_i, \phi) + \left(1 - \sum_{i=1}^{N+1} \omega(i)p_D(i)\right) = 1 - \sum_{i=1}^{N+1} \omega(i)p_D(i)(1 - F(\psi_i, \phi)). \end{aligned}$$

□

**Lemma 3.** *Let  $\pi$  be a quantum state verification protocol as described in Protocol 1 for a pure target state  $\phi$ ,  $\varepsilon_H$  be the lowest value that fulfills (3.1) and  $\varepsilon_D$  be the lowest value that fulfills (3.2), i.e., the best possible security parameter in fidelity-based security (Definition 1). If an attacker sends separable states  $\{\psi_j\}_{j=1}^{N+1}$ , it holds*

$$\varepsilon_H + \varepsilon_D \geq \sum_{i=1}^{N+1} \omega(i)(1 - F(\psi_i, \phi)) \left(1 - \sqrt{1 - \prod_{j=1, j \neq i}^{N+1} F(\phi, \psi_j)}\right), \quad (4.1)$$

where the inequality is saturated if the protocol uses optimal measurements to distinguish the target state from any other state, and the source sends pure states.

*Proof.* The result follows from the previous lemma, the Fuchs-van de Graaf inequalities, and the Holevo Helstrom theorem. We start by using Lemma 2 and find

$$\begin{aligned} \varepsilon_H + \varepsilon_D &= \sum_{i=1}^{N+1} \omega(i)(1 - p_H(i)) + \sum_{i=1}^{N+1} \omega(i)p_D(i)(1 - F(\psi_i, \phi)) \\ &= \sum_{i=1}^{N+1} \omega(i)((1 - p_H(i)) + p_D(i)(1 - F(\psi_i, \phi))). \end{aligned}$$

If the protocol uses optimal measurements to distinguish a pure target state from any other state, then  $p_H(i) = 1$ , which implies  $(1 - F(\psi_i, \phi))(1 - p_H(i)) = 1 - p_H(i)$ . If the protocol uses other measurements it holds that  $(1 - F(\psi_i, \phi))(1 - p_H(i)) \leq 1 - p_H(i)$ . Hence,

$$\begin{aligned} \varepsilon_H + \varepsilon_D &= \sum_{i=1}^{N+1} \omega(i)(1 - p_H(i)) + \sum_{i=1}^{N+1} \omega(i)p_D(i)(1 - F(\psi_i, \phi)) \\ &\geq \sum_{i=1}^{N+1} \omega(i)(1 - F(\psi_i, \phi))(1 - (p_H(i) - p_D(i))), \end{aligned}$$

where the inequality is saturated if the protocol uses optimal measurements. One finds by (2.1), (2.10) and Theorem 3 that

$$p_H(i) - p_D(i) \leq |p_H(i) - p_D(i)| \leq \frac{1}{2} \left\| \bigotimes_{j=1, j \neq i}^{N+1} \psi_j - \bigotimes_{j=1, j \neq i}^{N+1} \phi \right\|_1 \leq \sqrt{1 - \prod_{j=1, j \neq i}^{N+1} F(\phi, \psi_j)},$$

where the inequalities are saturated if the measurements are optimal, and the states the source sent are pure. Hence, one finds that

$$\varepsilon_H + \varepsilon_D \geq \sum_{i=1}^{N+1} \omega(i)(1 - F(\psi_i, \phi)) \left( 1 - \sqrt{1 - \prod_{j=1, j \neq i}^{N+1} F(\phi, \psi_j)} \right),$$

where under the mentioned assumptions the inequality is saturated.  $\square$

We show that an attack in which the attacker sends an orthogonal state in one round and the target state in all other rounds maximizes this bound.

**Theorem 7.** *Assume  $\phi$  is pure, hence, there exists at least one state  $\phi^\perp$  such that  $F(\phi, \phi^\perp) = 0$ . This allows for an attack where  $F(\psi_j, \phi) = 1 - \delta_{j,\ell}$  with  $\max_{1 \leq j \leq N+1} \omega(j) = \omega(\ell)$ . This attack is optimal in the class of separable state attacks with regard to the bound presented in (4.1).*

*Proof.* We first derive an inequality which implies that the bound presented in (4.1) cannot exceed  $\omega_\ell$ . We prove this inequality using the non-negativity and concavity of  $\log(x+1)$  and eventually demonstrate that the attack described in the claim yields exactly  $\omega_\ell$  inserted in the bound presented in (4.1).

We denote the right hand side of (4.1) as the function  $B_N : [0, 1]^{N+1} \rightarrow [0, 1]$  and first prove that for  $\mathbf{f} \in (0, 1]^{N+1}$  we have

$$B_N(\mathbf{f}) := \sum_{i=1}^{N+1} \omega(i)(1 - f_i) \left( 1 - \sqrt{1 - \prod_{j=1, j \neq i}^{N+1} f_j} \right) \leq \omega(\ell). \quad (4.2)$$

We use  $1 - \sqrt{1 - \prod_{j=1, j \neq i}^{N+1} f_j} \leq \prod_{j=1, j \neq i}^{N+1} f_j$  and find

$$B_N(\mathbf{f}) \leq \sum_{i=1}^{N+1} \omega(i)(1 - f_i) \left( \prod_{j=1, j \neq i}^{N+1} f_j \right) = \left( \prod_{j=1}^{N+1} f_j \right) \sum_{i=1}^{N+1} \omega(i) \frac{1 - f_i}{f_i} \leq \omega(\ell) \left( \prod_{j=1}^{N+1} f_j \right) \sum_{i=1}^{N+1} \frac{1 - f_i}{f_i}.$$

So (4.2) follows from

$$\sum_{i=1}^{N+1} \frac{1 - f_i}{f_i} = \sum_{i=1}^{N+1} \left( \frac{1}{f_i} - 1 \right) \leq \prod_{i=1}^{N+1} \frac{1}{f_i}, \quad (4.3)$$

which we now prove. First note that  $\log(x+1)$  is non-negative and is concave, i.e. for  $a, b \geq 0$

$$\log(a + b + 1) \leq \log(a + 1) + \log(b + 1),$$

meaning  $\log(x+1)$  is sub-additive. Further,  $\log$  is strictly increasing and thus preserves inequalities. Hence, we find for (4.3)

$$\begin{aligned} \sum_{i=1}^{N+1} \left( \frac{1}{f_i} - 1 \right) &\leq \prod_{i=1}^{N+1} \frac{1}{f_i} \Leftarrow \log \left( 1 + \sum_{i=1}^{N+1} \left( \frac{1}{f_i} - 1 \right) \right) \leq \sum_{i=1}^{N+1} \log \left( \frac{1}{f_i} \right) \\ &= \log \left( \prod_{i=1}^{N+1} \frac{1}{f_i} \right) \leq \log \left( 1 + \prod_{i=1}^{N+1} \frac{1}{f_i} \right), \end{aligned}$$

which implies (4.2).

We now consider  $\mathbf{f} \in [0, 1]^{N+1}$  and define the set  $A_{\mathbf{f}} = \{i \mid 1 \leq i \leq N+1, f_i = 0\}$ . We note that  $B_N(\mathbf{f})$  is non-zero only when  $A$  is empty or a singleton. Indeed, if two or more  $f_k = 0$  it holds that  $1 - \sqrt{1 - \prod_{j=1, j \neq k}^{N+1} f_j} = 0$  for all  $i \in \{1, \dots, N+1\}$ . However, if  $A_{\mathbf{f}}$  is a singleton, i.e.  $A_{\mathbf{f}} = \{k\}$  we find  $B_N(\mathbf{f}) = \omega(k) \left(1 - \sqrt{1 - \prod_{j=1, j \neq k}^{N+1} f_j}\right)$ . The case where  $A_{\mathbf{f}}$  is empty is covered in the first half of the proof. Hence, we find that  $B_N(\mathbf{f}) \leq \omega(\ell)$  and get equality for the described attack.  $\square$

The described attack in Theorem 7 maximizes the bound proved in Lemma 3. This Lemma also proves that if the protocol uses optimal measurements, the target state is pure and the source sends pure states, the bound is saturated. Hence, for such an optimal protocol for a pure target state, the presented attack is optimal in the class of separable state attacks, i.e., no other separable state attack can achieve higher violation of security. Indeed, Lemma 5 in the appendix even implies that a larger security violation is not always possible, since it proves that a specific choice of parameters in Protocol 1 yields 0-correctness and  $1/(N+1)$ -security for stand-alone security where  $\omega(i) = 1/(N+1)$  for all  $0 < i \leq N+1$  was chosen. While the presented attack is the intuitive and naive approach to break verification, it was unknown if it was optimal in the class of separable pure-state attacks. However, while this all holds with respect to the fidelity-based security definition, interestingly, the statement is false for composable security definitions, as we show in the remainder of this section. We start with the composable version of Lemma 2.

**Lemma 4.** *If  $\phi$  in Protocol 1 is pure and an attacker sends separable states  $\{\psi_j\}_{j=1}^{N+1}$ , it holds*

$$\begin{aligned} \frac{1}{2} \|\rho_H - \phi \oplus 0\|_1 &= 1 - \sum_{i=1}^{N+1} \omega(i)p_H(i), \\ \frac{1}{2} \min_{p \in [0,1]} \|\rho_D - p\phi \oplus (1-p)\|_1 &= \frac{1}{2} \left\| \sum_{i=1}^{N+1} \omega(i)p_D(i)(\psi_i - \phi) \right\|_1, \end{aligned}$$

where  $\rho_H$  is the output state in the honest setting,  $\rho_D$  the one if the source cheats and  $p_H(i)$  and  $p_D(i)$  are defined as in Lemma 2.

*Proof.* We start with the honest setting. When the source is honest, the output state is

$$\rho_H = \sum_{i=1}^{N+1} \omega(i)p_H(i)\phi \oplus \left(1 - \sum_{i=1}^{N+1} \omega(i)p_H(i)\right).$$

This implies

$$\frac{1}{2} \|\rho_H - \phi \oplus 0\|_1 = \frac{1}{2} \left\| \sum_{i=1}^{N+1} \omega(i)p_H(i)\phi - \phi \right\|_1 + \frac{1}{2} \left| 1 - \sum_{i=1}^{N+1} \omega(i)p_H(i) \right| = 1 - \sum_{i=1}^{N+1} \omega(i)p_H(i).$$

If the source is dishonest, the output state of the protocol is

$$\rho_D = \sum_{i=1}^{N+1} \omega(i)p_D(i)\psi_i \oplus \left(1 - \sum_{i=1}^{N+1} \omega(i)p_D(i)\right),$$

which implies

$$\min_{p \in [0,1]} \frac{1}{2} \|\rho_D - p\phi \oplus (1-p)\|_1 = \min_{p \in [0,1]} \frac{1}{2} \left\| \sum_{i=1}^{N+1} \omega(i)p_D(i)\psi_i - p\phi \right\|_1 + \frac{1}{2} \left| p - \sum_{i=1}^{N+1} \omega(i)p_D(i) \right|.$$

On the one hand, with  $p = \sum_{i=1}^{N+1} \omega(i)p_D(i)$  we find

$$\min_{p \in [0,1]} \frac{1}{2} \|\rho_D - p\phi \oplus (1-p)\|_1 \leq \frac{1}{2} \left\| \sum_{i=1}^{N+1} \omega(i)p_D(i)(\psi_i - \phi) \right\|_1.$$

On the other hand, we find using the triangle inequality and multiplication by  $1 = \|\phi\|_1$ :

$$\begin{aligned} \min_{p \in [0,1]} \frac{1}{2} \|\rho_D - p\phi \oplus (1-p)\|_1 &= \min_{p \in [0,1]} \frac{1}{2} \left\| \sum_{i=1}^{N+1} \omega(i)p_D(i)\psi_i - \phi \right\|_1 + \frac{1}{2} \left\| p\phi - \sum_{i=1}^{N+1} \omega(i)p_D(i)\phi \right\|_1 \\ &\geq \frac{1}{2} \left\| \sum_{i=1}^{N+1} \omega(i)p_D(i)(\psi_i - \phi) \right\|_1, \end{aligned}$$

which proves the claim.  $\square$

Now we can easily show that the naive approach of sending an orthogonal state for the round with the highest output probability and the target state in all other rounds is not always optimal. If the target state is pure and the protocol uses optimal measurements, we find for this specific attack that

$$\begin{aligned} \min_{p \in [0,1]} \frac{1}{2} \|\rho_D - p\phi \oplus (1-p)\|_1 &= \frac{1}{2} \left\| \sum_{i=1}^{N+1} \omega(i)p_D(i)(\psi_i - \phi) \right\|_1 = \frac{\omega(\ell)}{2} \|\psi_\ell - \phi\|_1 = \omega(\ell), \\ \frac{1}{2} \|\rho_H - \phi \oplus 0\|_1 &= 0 \end{aligned}$$

since for all  $i \neq \ell$  it holds  $\psi_i = \phi$  which implies  $p_D(\ell) = 1$  for optimal measurements and  $\psi_\ell = \phi^\perp$  which implies  $\frac{1}{2}\|\psi_\ell - \phi\|_1 = 1$ . With the specifications of the protocol being  $\omega(i) = 1/(N+1)$ , we find for this specific attack on this protocol  $\varepsilon_H + \varepsilon_D = 1/(N+1) < 1/N$ . However, we already know that there is an i.i.d. attack such that  $\varepsilon_H + \varepsilon_D \geq 1/4\sqrt{N}$  if  $\phi$  is pure by virtue of Theorem 6; hence, we find that the naive approach is not optimal if  $N > 16$ .

## 5 Discussion

Quantum state verification is of utmost importance for quantum cryptography. We demonstrate that implementations using the popular cut-and-choose approach cannot succeed in a desirable parameter regime even if the attacker is restricted to i.i.d. attacks. Further, we find that the naive approach for protocols with a fixed number of rounds is optimal for the fidelity-based security definition but exhibits a suboptimal scaling for composable security. Indeed, the bound of  $1/4\sqrt{N}$ , which we prove for composable security, has a tight scaling as implicitly shown in the appendix (cf. Lemma 5). This tight scaling and the tightness of the naive approach (cf. Theorem 7 and Lemma 5) for fixed round numbers pose a crucial problem for quantum state verification, especially in the context of composed protocols, even against rather limited attacks, such as i.i.d. attacks. While we restrict the protocol type in the main part of this work to a fixed number of rounds, we prove the same bounds for a probabilistic number of rounds in the appendix (cf. Theorems 8 and 9), further closing loopholes to circumvent this no-go result. Importantly, this protocol type was not affected by previous works, and intuitive approaches to break verification are only available for a few combinations of distributions for the number of rounds and output round.

Our proofs furthermore shed light on an interesting trade-off in the attack between the acceptance probability and the deviation from the target state. While the intuitive

attack discussed in Section 4 yields a constant fidelity of 0 accepted with a probability  $1/N+1$  for a uniform choice of the output round, the acceptance probability in the i.i.d. attacks converges to a constant and the trace distance decreases or fidelity increases in the number of rounds. For example, we find for composable security that the acceptance probability for a protocol that projects onto a pure target state is lower bounded by  $1/2$  and the trace distance between the sent state and the target state is  $1/2\sqrt{N}$ . Hence, even with  $10^4$  rounds, the probability of accepting a state that is (at least)  $1/200$  apart from the target state is at least  $1/2$ .

Nevertheless, we note that quantum state verification is not a lost cause; although the cut-and-choose approach cannot yield desirable security using composition theorems or in a stand-alone fashion, one might prove security of a composition with cut-and-choose quantum state verification in a non-modular fashion. Further, other verification mechanisms, such as error detection, might allow for better security but fall out of the scope of this work, which was about the direct cut-and-choose approach for verification. Finally, further research is needed to evaluate the potential of quantum state verification in general, and we need to find out where techniques similar to ours provide further limitations and where positive results with negligible distinguishing advantage can be found.

## 6 Acknowledgements

F. W., Z.C. and A. P. acknowledge support from the DFG via the Emmy Noether grant No. 41829458 and the Hector Fellow Academy. This work was funded by the European Union’s Horizon Europe research and innovation programme under grant agreement No. 101102140 – QIA Phase 1. This project was financially supported by BERLIN QUANTUM, an initiative endowed by the Innovation Promotion Fund of the city of Berlin. M. K. is supported by EPSRC grant EP/V040944/1 Resources in Computation.

## References

- [ABG<sup>+</sup>07] Antonio Acín, Nicolas Brunner, Nicolas Gisin, Serge Massar, Stefano Pironio, and Valerio Scarani. Device-independent security of quantum cryptography against collective attacks. *Phys. Rev. Lett.*, 98(23):230501, June 2007. [doi:10.1103/physrevlett.98.230501](https://doi.org/10.1103/physrevlett.98.230501).
- [BK22] Anne Broadbent and Martti Karvonen. Categorical composable cryptography. In *Foundations of Software Science and Computation Structures (FoSSaCS)*, volume 13242 of *Lecture Notes in Computer Science*. Springer, 2022. [doi:10.1007/978-3-030-99253-8\\_9](https://doi.org/10.1007/978-3-030-99253-8_9).
- [BK23] Anne Broadbent and Martti Karvonen. Categorical composable cryptography: extended version. *Logical Methods in Computer Science*, 19:30:1–30:46, 2023. [doi:10.46298/LMCS-19\(4:30\)2023](https://doi.org/10.46298/LMCS-19(4:30)2023).
- [Can20] Ran Canetti. Universally composable security. *J. ACM*, 67(5), September 2020. [doi:10.1145/3402457](https://doi.org/10.1145/3402457).
- [CMY24] Léo Colisson, Damian Markham, and Raja Yehia. All graph state verification protocols are compositely secure, 2024. [arXiv:2402.01445](https://arxiv.org/abs/2402.01445).
- [GRTZ02] Nicolas Gisin, Grégoire Ribordy, Wolfgang Tittel, and Hugo Zbinden. Quantum cryptography. *Rev. Mod. Phys.*, 74:145–195, Mar 2002. [doi:10.1103/RevModPhys.74.145](https://doi.org/10.1103/RevModPhys.74.145).

[HdJP20] Frederik Hahn, Jarn de Jong, and Anna Pappa. Anonymous quantum conference key agreement. *PRX Quantum*, 1:020325, Dec 2020. [doi:10.1103/PRXQuantum.1.020325](https://doi.org/10.1103/PRXQuantum.1.020325).

[HM15] Masahito Hayashi and Tomoyuki Morimae. Verifiable Measurement-Only Blind Quantum Computing with Stabilizer Testing. *Phys. Rev. Lett.*, 115(22):220502, November 2015. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.115.220502>, doi:10.1103/PhysRevLett.115.220502.

[KL14] Jonathan Katz and Yehuda Lindell. *Introduction to Modern Cryptography, Second Edition*. Chapman & Hall/CRC, 2nd edition, 2014.

[LWW<sup>+</sup>10] Lars Lydersen, Carlos Wiechers, Christoffer Wittmann, Dominique Elser, Johannes Skaar, and Vadim Makarov. Hacking commercial quantum cryptography systems by tailored bright illumination. *Nature Photonics*, 4:686–689, 2010. [doi:10.1038/nphoton.2010.214](https://doi.org/10.1038/nphoton.2010.214).

[MR11] Ueli Maurer and Renato Renner. Abstract cryptography. In *The Second Symposium on Innovations in Computer Science, ICS 2011*. Tsinghua University Press, 1 2011.

[MTH17] Tomoyuki Morimae, Yuki Takeuchi, and Masahito Hayashi. Verification of hypergraph states. *Phys. Rev. A*, 96:062321, Dec 2017. [doi:10.1103/PhysRevA.96.062321](https://doi.org/10.1103/PhysRevA.96.062321).

[PAB<sup>+</sup>20] S. Pirandola, U. L. Andersen, L. Banchi, M. Berta, D. Bunandar, R. Colbeck, D. Englund, T. Gehring, C. Lupo, C. Ottaviani, J. L. Pereira, M. Razavi, J. Shamsul Shaari, M. Tomamichel, V. C. Usenko, G. Vallone, P. Villoresi, and P. Wallden. Advances in quantum cryptography. *Adv. Opt. Photon.*, 12(4):1012–1236, Dec 2020. [doi:10.1364/AOP.361502](https://doi.org/10.1364/AOP.361502).

[PCW<sup>+</sup>12] Anna Pappa, André Chailloux, Stephanie Wehner, Eleni Diamanti, and Iordanis Kerenidis. Multipartite entanglement verification resistant against dishonest parties. *Phys. Rev. Lett.*, 108(26):260502, June 2012. [doi:10.1103/physrevlett.108.260502](https://doi.org/10.1103/physrevlett.108.260502).

[PLM18] Sam Pallister, Noah Linden, and Ashley Montanaro. Optimal verification of entangled states with local measurements. *Phys. Rev. Lett.*, 120(17):170502, April 2018. [doi:10.1103/physrevlett.120.170502](https://doi.org/10.1103/physrevlett.120.170502).

[Pre18] John Preskill. Quantum computing in the nisq era and beyond. *Quantum*, 2:79, August 2018. [doi:10.22331/q-2018-08-06-79](https://doi.org/10.22331/q-2018-08-06-79).

[SHM22] Nathan Shettell, Majid Hassani, and Damian Markham. Private network parameter estimation with quantum sensors, July 2022. arXiv:2207.14450 [quant-ph]. URL: <http://arxiv.org/abs/2207.14450>.

[TM18] Yuki Takeuchi and Tomoyuki Morimae. Verification of many-qubit states. *Phys. Rev. X*, 8(2):021060, 2018. [doi:10.1103/PhysRevX.8.021060](https://doi.org/10.1103/PhysRevX.8.021060).

[UM22] Anupama Unnikrishnan and Damian Markham. Verification of graph states in an untrusted network. *Phys. Rev. A*, 105(5):052420, May 2022. [doi:10.1103/physreva.105.052420](https://doi.org/10.1103/physreva.105.052420).

[vB20] Ivan Šupić and Joseph Bowles. Self-testing of quantum systems: a review. *Quantum*, 4:337, September 2020. [doi:10.22331/q-2020-09-30-337](https://doi.org/10.22331/q-2020-09-30-337).

---

- [Wat18] John Watrous. *The Theory of Quantum Information*. Cambridge University Press, 1 edition, April 2018. [doi:10.1017/9781316848142](https://doi.org/10.1017/9781316848142).
- [WEH18] Stephanie Wehner, David Elkouss, and Ronald Hanson. Quantum internet: A vision for the road ahead. *Science*, 362(6412):eaam9288, October 2018. URL: <https://www.science.org/doi/10.1126/science.aam9288>, doi: [10.1126/science.aam9288](https://doi.org/10.1126/science.aam9288).
- [YDK21] Raja Yehia, Eleni Diamanti, and Iordanis Kerenidis. Composable security for multipartite entanglement verification. *Phys. Rev. A*, 103(5), May 2021. [doi:10.1103/physreva.103.052609](https://doi.org/10.1103/physreva.103.052609).
- [YSG22] Xiao-Dong Yu, Jiangwei Shang, and Otfried Gühne. Statistical methods for quantum state verification and fidelity estimation. *Advanced Quantum Technologies*, 5(5):2100126, March 2022. [doi:10.1002/qute.202100126](https://doi.org/10.1002/qute.202100126).
- [ZH19] Huangjun Zhu and Masahito Hayashi. Efficient verification of pure quantum states in the adversarial scenario. *Phys. Rev. Lett.*, 123:260504, Dec 2019. [doi:10.1103/PhysRevLett.123.260504](https://doi.org/10.1103/PhysRevLett.123.260504).

## A A more general protocol type

While we already provide our no-go result for a rather general type of protocols, one might want to go a few steps further in generalization. The most important aspect that is missing in the main matter is a probabilistic round number: one might choose not to fix the round number in advance but during the protocol. While we do not model the communication between the source and the clients in detail, we introduce the following parameters:

- $\Omega : \mathbb{N} \rightarrow [0, 1]$  is the probability distribution that governs the number of verification rounds. The total number of rounds is one plus the number of verification rounds.
- $(\omega_n : \{0, \dots, n\} \rightarrow [0, 1])_{n \in \mathbb{N}_+}$  is a sequence of probability distributions for the output round.
- $\mu_{n,i}(0) \in Pos(\mathcal{X}^{\otimes n})$  is the measurement operator associated with acceptance in the protocol when the protocol uses  $n$  rounds and outputs in round  $1 \leq i \leq n$ .

We now derive the bound for this protocol for stand-alone security.

**Theorem 8.** *Let  $\pi = (\pi_C, \pi_S)$  be a protocol as described above. If  $\pi$  is  $\varepsilon_H$ -correct and  $\varepsilon_D$  secure according definition 1, it holds*

$$\varepsilon_H + \varepsilon_D \geq \frac{1}{7N},$$

where  $N$  is the expected number of verification rounds.

*Proof.* The proof follows the same ideas as the proof of Theorem 5 but uses Jensen's inequality to discard the probabilistic round number. We first set

$$\begin{aligned} p_{n,i}^H &:= \langle \mu_{n,i}(0) | \phi^{\otimes n} \rangle \\ p_{n,i}^D &:= \langle \mu_{n,i}(0) | \psi^{\otimes n} \rangle. \end{aligned}$$

where  $\psi$  is the state the dishonest source sends for each round. Now, we can define

$$\begin{aligned} p_A^H &:= \sum_{n=0}^{\infty} \Omega(n) \sum_{i=0}^n \omega_n(i) p_{n,i}^H \\ p_A^D &:= \sum_{n=0}^{\infty} \Omega(n) \sum_{i=0}^n \omega_n(i) p_{n,i}^D \end{aligned}$$

and find

$$\begin{aligned} \varepsilon_H &\geq 1 - p_A^H \\ \varepsilon_D &\geq p_A^D (1 - F(\phi, \psi)) \end{aligned}$$

following the same manipulations as in the proof of Theorem 5 with an i.i.d. attack using  $\psi$ .

Hence, we find for the sum of  $\varepsilon_H$  and  $\varepsilon_D$

$$\varepsilon_H + \varepsilon_D \geq 1 - p_A^H + p_A^D (1 - F(\phi, \psi)) \geq (1 - F(\phi, \psi)) (1 - |p_A^H - p_A^D|).$$

Now the differences between the protocol types come into play, as  $|p_A^H - p_A^D|$  is the expected distinguishing advantage with a probabilistic number of states, i.e., using (2.10):

$$|p_A^H - p_A^D| = \sum_{n=0}^{\infty} \Omega(n) \sum_{i=0}^n \omega_n(i) |p_{n,i}^H - p_{n,i}^D| \leq \sum_{n=0}^{\infty} \Omega(n) \frac{1}{2} \|\phi^{\otimes n} - \psi^{\otimes n}\| \leq \sum_{n=0}^{\infty} \Omega(n) \sqrt{1 - F(\phi, \psi)^n}$$

As  $f(n) = \sqrt{1 - a^n}$  is concave for  $0 \leq a \leq 1$  (cf. Lemma 1) we can use Jensen's inequality and find

$$|p_A^H - p_A^D| \leq \sqrt{1 - F(\phi, \psi)^N},$$

where  $N$  is the expected number of verifications. Combined with the above, this means

$$\varepsilon_H + \varepsilon_D \geq (1 - F(\phi, \psi)) \left( 1 - \sqrt{1 - F(\phi, \psi)^N} \right).$$

Now we set  $\tau := (1 - F(\phi, \psi))$ , i.e.  $F(\phi, \psi) = (1 - \tau)$ :

$$\varepsilon_H + \varepsilon_D \geq \tau \left( 1 - \sqrt{1 - (1 - \tau)^N} \right).$$

We choose  $\tau = \alpha/N$  and get

$$\varepsilon_H + \varepsilon_D \geq \frac{\alpha}{N} (1 - \sqrt{\alpha}),$$

which is optimal for  $\alpha = 4/9$ , as was shown in the proof of Theorem 5, which gives

$$\varepsilon_H + \varepsilon_D \geq \frac{4}{27N} \geq \frac{1}{7N}$$

□

We also find the same result as for the simpler protocol type with regard to composable security.

**Theorem 9.** *Let  $\pi = (\pi_C, \pi_S)$  be a protocol as above. If*

$$\frac{1}{2} \|\rho_H - (\phi \oplus 0)\|_1 \leq \varepsilon_H$$

and

$$\min_{p \in [0,1]} \frac{1}{2} \|\rho_D - (p\phi \oplus (1-p))\|_1 \leq \varepsilon_D.$$

*It holds  $\varepsilon_H + \varepsilon_D \geq \frac{\sqrt{\eta_1}}{4\sqrt{N}}$  if  $\phi$  is mixed and  $\eta_1$  is the largest eigenvalue of  $\phi$  and  $\varepsilon_H + \varepsilon_D \geq \frac{1}{4\sqrt{N}}$  if  $\phi$  is pure.*

*Proof.* As in the previous theorem, we can use the result for a fixed round number (Theorem 6) by using Jensen's inequality. Again we denote

$$\begin{aligned} p_{n,i}^H &:= \langle \mu_{n,i}(0) | \phi^{\otimes n} \rangle \\ p_{n,i}^D &:= \langle \mu_{n,i}(0) | \psi^{\otimes n} \rangle \\ p_A^H &:= \sum_{n=0}^{\infty} \Omega(n) \sum_{i=0}^n p_{n,i}^H \\ p_A^D &:= \sum_{n=0}^{\infty} \Omega(n) \sum_{i=0}^n p_{n,i}^D. \end{aligned}$$

Following the same line as for the proof of Theorem 6, we find

$$\varepsilon_H + \varepsilon_D \geq \frac{p_A^D}{2} \|\psi - \phi\|_1 + 1 - p_A^H \geq \frac{1}{2} \|\psi - \phi\|_1 (1 - |p_A^H - p_A^D|).$$

Again, we fix  $\psi$  now. We write  $\phi$  in spectral decomposition  $\phi = \sum_{i=1}^d \eta_i |\phi_i\rangle\langle\phi_i|$  with  $\eta_{i+1} \leq \eta_i$  for  $1 \leq i < d$ . We fix  $\psi = \eta_1 |\chi\rangle\langle\chi| + \sum_{i=2}^d \eta_i |\phi_i\rangle\langle\phi_i|$  for some pure state  $|\chi\rangle$ . This implies

$$\frac{1}{2} \|\phi - \psi\|_1 = \eta_1 \sqrt{1 - |\langle\phi_1|\chi\rangle|^2}. \quad (\text{A.1})$$

In the case of  $n$  copies, we utilize the result from the proof of Theorem 6 and find

$$\frac{1}{2} \|\phi^{\otimes n} - \psi^{\otimes n}\|_1 \leq \sqrt{1 - |\langle\phi_1|\chi\rangle|^{2\eta_1 n}}.$$

This implies that

$$\begin{aligned} |p_A^H - p_A^D| &= \sum_{n=0}^{\infty} \Omega(n) \sum_{i=0}^n \omega_n(i) |p_{n,i}^H - p_{n,i}^D| \leq \frac{1}{2} \sum_{n=0}^{\infty} \Omega(n) \|\phi^{\otimes n} - \psi^{\otimes n}\|_1 \\ &\leq \sum_{n=0}^{\infty} \Omega(n) \sqrt{1 - |\langle\phi_1|\chi\rangle|^{2\eta_1 n}}. \end{aligned}$$

We use concavity and Jensen's inequality and Lemma 1 again and find

$$|p_A^H - p_A^D| \leq \sqrt{1 - |\langle\phi_1|\chi\rangle|^{2\eta_1 N}},$$

where  $N$  is the expected number of verification rounds. So for  $\varepsilon_H + \varepsilon_D$  we find

$$\varepsilon_H + \varepsilon_D \geq \eta_1 \sqrt{1 - |\langle\phi_1|\chi\rangle|^2} \left( 1 - \sqrt{1 - |\langle\phi_1|\chi\rangle|^{2\eta_1 N}} \right).$$

We know from the proof of the bound for the simpler protocol that we can choose  $|\chi\rangle$  such that

$$\varepsilon_H + \varepsilon_D \geq \frac{\sqrt{\eta_1}}{4\sqrt{N}}.$$

In particular, for pure states this gives us

$$\varepsilon_H + \varepsilon_D \geq \frac{1}{4\sqrt{N}}.$$

□

## B Tightness of scaling

In order to investigate how tight the lower bounds that we provide are, we analyze the correctness and security of a protocol of the type shown in Protocol 1, in which the target state is pure, every round has the same probability to be the output round, and the measurement is always a projection onto  $N$  copies of the target state. We find that this choice of parameters yields a protocol which is perfectly correct in both, stand-alone and composable security,  $1/N+1$ -secure in stand-alone and  $2/\sqrt{N+1}$ -secure in composable security, respectively.

These results are in line with the i.i.d. bound of  $1/7N$  given for stand-alone security in Theorem 5, the i.i.d. bound of  $1/4\sqrt{N}$  given for composable security in Theorem 6 and the bound for stand-alone security of  $1/N+1$  given by the combination of Lemma 2, Lemma 3 and Theorem 7 in Section 4. The last of these three bounds was shown to be optimal in the class of separable state attacks. Indeed, the following lemma proves that it is optimal,

since it is tight, in the class of all attacks for the protocol considered in this section. For the two i.i.d. bounds we can use  $1/N+1 \leq 1/N$  and  $2/\sqrt{N+1} \leq 2/\sqrt{N}$ , which implies that the protocol is also  $1/N$ -secure in stand-alone and  $2/\sqrt{N}$ -secure in composable security. Hence, the lower bounds for  $\varepsilon_H + \varepsilon_D$  with the i.i.d. restriction for the attacker we presented before exhibit the same scaling in  $N$  as the upper bounds for the same quantities for this specific protocol without restrictions on the attacks. Therefore, the following lemma implies that the scaling of the i.i.d. bounds is tight and no attack can violate security for all choices of parameters in Protocol 1 with a better scaling in  $N$ . Nevertheless, there might be tighter general bounds with the same scaling or bounds with better scaling for specific choices of parameters in Protocol 1.

**Lemma 5.** *Let  $\pi = (\pi_C, \pi_S)$  be a protocol as described in Protocol 1 where the parameters are chosen so that*

- $\phi = |\phi_0\rangle\langle\phi_0| \in D(\mathcal{X})$  is the pure target state,
- for all  $i \in \{1, \dots, N+1\}$  it holds  $\omega(i) = 1/N+1$ , i.e., every round has the same probability to be the output round,
- and for all  $i \in \{1, \dots, N+1\}$  the clients use  $\mu_i(0) = |\phi_0^{\otimes N}\rangle\langle\phi_0^{\otimes N}|$ .

This implies

1.  $\pi$  is 0-correct and  $1/N+1$ -secure with respect to Definition 1.
2.  $\pi$  is 0-composable correct with respect to  $\sharp_S(\mathcal{S}_\phi^{QSV})$  and  $2/\sqrt{N+1}$ -composable secure with respect to  $\mathcal{S}_\phi^{QSV}$ .

*Proof.* While correctness follows immediately, we use a symmetry argument for security since the clients choose the output round randomly and bound a sum of diagonal elements by the trace. Finally, we find the composable security from a known result proven in [CMY24].

The correctness in the first statement follows from  $\langle|\phi_0\rangle\langle\phi_0|^{\otimes N}| |\phi_0\rangle\langle\phi_0|^{\otimes N}\rangle = |\langle\phi_0|\phi_0\rangle|^N = 1$ . Hence, the protocol is perfectly correct with respect to Definition 1, since the probability to reject the behavior of an honest source is 0.. If the source is dishonest, we assume that it sends a state  $\psi \in D(\mathcal{X}^{\otimes N+1})$ , which might be entangled across the rounds, consider the completion of  $|\phi_0\rangle$  to an orthonormal basis  $\{|\phi_i\rangle\}_{i=0}^{\dim(\mathcal{X})-1}$  of  $\mathcal{X}$  and define  $A_\ell, B_\ell \in \mathcal{U}(\mathcal{X}^{\otimes N+1})$  for  $\ell \in \{1, \dots, N+1\}$  as

$$\begin{aligned} \ell \in \{1, \dots, N\} : A_\ell &:= \sum_{i,j=0}^{\dim(\mathcal{X})-1} (\mathbb{1}_\mathcal{X}^{\otimes \ell-1} \otimes |\phi_i\rangle \otimes |\phi_j\rangle \otimes \mathbb{1}_\mathcal{X}^{\otimes N-\ell}) (\mathbb{1}_\mathcal{X}^{\otimes \ell-1} \otimes \langle\phi_j| \otimes \langle\phi_i| \otimes \mathbb{1}_\mathcal{X}^{\otimes N-\ell}), \\ A_{N+1} &:= \mathbb{1}_\mathcal{X}^{\otimes N+1} \\ \ell \in \{1, \dots, N+1\} : B_\ell &:= A_N A_{N-1} \dots A_{\ell+1} A_\ell, \end{aligned}$$

hence,  $A_\ell$  is the unitary that exchanges the  $\ell$ -th and the  $(\ell+1)$ -th register, and  $B_\ell$  pushes the  $\ell$ -th register to the last position. Equipped with these definitions, we find that if the source is dishonest, the average state is given by

$$\begin{aligned} \rho_D = & \left( \langle\phi_0|^{\otimes N} \otimes \mathbb{1}_\mathcal{X} \right) \left( \frac{1}{N+1} \sum_{\ell=1}^{N+1} B_\ell \psi B_\ell^\dagger \right) \left( |\phi_0\rangle^{\otimes N} \otimes \mathbb{1}_\mathcal{X} \right) \oplus \\ & \left( 1 - \text{Tr} \left( \left( \langle\phi_0|^{\otimes N} \otimes \mathbb{1}_\mathcal{X} \right) \left( \frac{1}{N+1} \sum_{\ell=1}^{N+1} B_\ell \psi B_\ell^\dagger \right) \left( |\phi_0\rangle^{\otimes N} \otimes \mathbb{1}_\mathcal{X} \right) \right) \right), \end{aligned}$$

where we again used  $\langle |\phi_0\rangle\langle\phi_0|^{\otimes N}|\chi\rangle = \langle\phi_0|^{\otimes N}\chi|\phi_0\rangle^{\otimes N}$ . Using the behavior of the fidelity under direct sums and the optimal ideal acceptance probability  $p$  we find

$$\begin{aligned} \max_{p \in [0,1]} F(\rho_D, p\phi \oplus (1-p)) &= F\left(\phi, \left(\langle\phi_0|^{\otimes N} \otimes \mathbb{1}_{\mathcal{X}}\right) \left(\frac{1}{N+1} \sum_{\ell=1}^{N+1} B_{\ell}\psi B_{\ell}^{\dagger}\right) \left(|\phi_0\rangle^{\otimes N} \otimes \mathbb{1}_{\mathcal{X}}\right)\right) + \\ &\quad 1 - \text{Tr}\left(\left(\langle\phi_0|^{\otimes N} \otimes \mathbb{1}_{\mathcal{X}}\right) \left(\frac{1}{N+1} \sum_{\ell=1}^{N+1} B_{\ell}\psi B_{\ell}^{\dagger}\right) \left(|\phi_0\rangle^{\otimes N} \otimes \mathbb{1}_{\mathcal{X}}\right)\right). \end{aligned}$$

Next, we use that  $F(\rho, |\gamma\rangle\langle\gamma|) = \langle\gamma|\rho|\gamma\rangle$ , express the trace with the basis  $\{|\phi_i\rangle\}_{i=0}^{\dim(\mathcal{X})-1}$  and apply  $B_{\ell}$  on the vectors instead of the operator and find

$$\begin{aligned} \max_{p \in [0,1]} F(\rho_D, p\phi \oplus (1-p)) &= \langle\phi_0|^{\otimes N+1} \left(\frac{1}{N+1} \sum_{\ell=1}^{N+1} B_{\ell}\psi B_{\ell}^{\dagger}\right) |\phi_0\rangle^{\otimes N+1} \\ &+ 1 - \sum_{i=0}^{\dim(\mathcal{X})-1} \left(\langle\phi_0|^{\otimes N} \otimes \langle\phi_i|\right) \left(\frac{1}{N+1} \sum_{\ell=1}^{N+1} B_{\ell}\psi B_{\ell}^{\dagger}\right) \left(|\phi_0\rangle^{\otimes N} \otimes |\phi_i\rangle\right) \\ &= 1 - \sum_{i=1}^{\dim(\mathcal{X})-1} \left(\langle\phi_0|^{\otimes N} \otimes \langle\phi_i|\right) \left(\frac{1}{N+1} \sum_{\ell=1}^{N+1} B_{\ell}\psi B_{\ell}^{\dagger}\right) \left(|\phi_0\rangle^{\otimes N} \otimes |\phi_i\rangle\right) \\ &= 1 - \frac{1}{N+1} \sum_{i=1}^{\dim(\mathcal{X})-1} \sum_{\ell=1}^{N+1} \left(\langle\phi_0|^{\otimes \ell-1} \otimes \langle\phi_i| \otimes \langle\phi_0|^{\otimes N-\ell+1}\right) \psi \left(|\phi_0\rangle^{\otimes \ell-1} \otimes |\phi_i\rangle \otimes |\phi_0\rangle^{\otimes N-\ell+1}\right). \end{aligned}$$

To prove the first claim of the lemma, it is now sufficient to recognize that the sum in the above equation adds up some of the diagonal elements of  $\psi$ . However, since  $\psi \in D(\mathcal{X}^{N+1})$ , this sum cannot exceed the trace of  $\psi$ , which is 1, i.e.

$$\max_{p \in [0,1]} F(\rho_D, p\phi \oplus (1-p)) \geq 1 - \frac{1}{N+1}.$$

Composable correctness follows immediately from the impossibility of rejecting the behavior of the source if it is honest. For the proof of composable security, we refer to [CMY24], which states (with our definition of fidelity) that

$$\max_{p \in [0,1]} F(\rho_D, p\phi \oplus (1-p)) \geq \sqrt{1-\kappa},$$

implies  $2\sqrt{2\kappa - \kappa^2}$ - composable security. Note that if all clients are honest, the necessity that  $\phi$  is a graph state does not apply, and the ideal resources in [CMY24] are identical to  $\mathcal{S}_{\phi}^{QSV}$ . Since it holds

$$2\sqrt{2\kappa - \kappa^2} = 2\sqrt{1 - (1 - \kappa)^2},$$

and we just proved  $\max_{p \in [0,1]} F(\rho_D, p\phi \oplus (1-p)) \geq 1 - \frac{1}{N+1}$  we can set  $(1 - \kappa)^2 = 1 - \frac{1}{N+1}$ , which proves  $2/\sqrt{N+1}$  composable security.  $\square$