

ORIGINAL RESEARCH OPEN ACCESS

# Approaches to Quantum Remote Memory Attestation

 Jesse Laeuchli<sup>1</sup>  | Rolando Trujillo Rasua<sup>2</sup> 
<sup>1</sup>School of Computer Science and Engineering, UNSW Sydney, Sydney, Australia | <sup>2</sup>Departament d'Enginyeria Informàtica i Matemàtiques, Rovira i Virgili University, Tarragona, Spain

**Correspondence:** Jesse Laeuchli ([j.laeuchli@unsw.edu.au](mailto:j.laeuchli@unsw.edu.au))

**Received:** 3 June 2025 | **Revised:** 7 September 2025 | **Accepted:** 15 September 2025

**Handling Editor:** Tianyu Ye

**Funding:** Rolando Trujillo Rasua was supported by a Ramón y Cajal grant (RYC2020-028954-I) from the Spanish Ministry of Science and Innovation and the EU, as well as by projects PROVTOPIA (PID2023-150098OB-I00), funded by MICIU/AEI/10.13039/501100011033 and FEDER (EU) and HERMES, funded by INCIBE and the EU's NextGenerationEU/PRTR.

**Keywords:** quantum communication | quantum information | telecommunication security

## ABSTRACT

In this article, we uncover flaws and pitfalls of a quantum-based remote memory attestation procedure for Internet of things devices. We also show limitations of quantum memory that suggests the attestation problem for quantum memory is fundamentally different to the attestation problem for classical memory, even when the devices can perform quantum computation. The identified problems are of interest for quantum-based security protocol designers in general, particularly those dealing with corrupt devices. Finally, we make use of the lessons learnt to design a quantum-based attestation system for classical memory with improved communication efficiency and security.

## 1 | Introduction

Determining whether a remote embedded device is in a secure state is an important problem in computer security. A solution to this problem is *remote memory attestation*: a two-party communication protocol whereby a device is asked to prove that its memory is in a given state. If the device successfully passes the protocol, then the analyst can be confident that it has not been compromised.

Developing secure protocols for remote attestation is inherently difficult due to the assumption that the proving device is compromised. This assumption breaks conventional security protocols, which require the communicating parties to either share a cryptographic secret, follow the protocol specification or securely connect to a trusted-third-party. Because the remote device is potentially infected with malware, the state-of-the-practice in remote memory attestation is to deploy secure

hardware on the proving device [1–4] in such a way that at least one of the three above requirements is satisfied.

Trusted hardware, however, increases the manufacturing cost, which typically cannot be afforded by resource-constrained IoT systems [5], and depends on engineering and supply chain guarantees rather than on information-theoretic security. If a security vulnerability is later found with the installed hardware, then all deployed devices must be recalled. Unlike software vulnerabilities, hardware vulnerabilities cannot be hot fixed.

Orthogonal to hardware-based approaches are those [6–12] that rely on side-channel information to authenticate the prover and to ensure that prover and verifier run the protocol in a clean environment, that is, without the interference of a network attacker. These approaches are called *software-based*. In between hardware and software-based approach lies the *hybrid* approach, whose goal is either to emulate secure hardware or minimises its use [13–18].

This is an open access article under the terms of the [Creative Commons Attribution-NoDerivs](https://creativecommons.org/licenses/by-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited and no modifications or adaptations are made.

© 2025 The Author(s). *IET Quantum Communication* published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology.

## 1.1 | Problem Statement

In classical computing, it is still an open problem how to achieve remote memory attestation in the presence of a network attacker without assuming trusted hardware on the remote device [6, 12, 19]. That is to say, existing software-based attestation protocols whose participants are modelled as Turing machines do not satisfy (cryptographic) authentication nor resist *proxy attacks*: an attack whereby the prover outsources the attestation task to a third-device. That is an important limitation, which forces the verifier to run the protocol in a clean environment aided by the use of out-of-band channels, such as visual inspection. In quantum computing, however, the problem has been claimed to be successfully addressed by the remote memory attestation protocol *Soteria* [20], which does assume a quantum physically uncloneable function (QPUF) on the remote device, but does not require trusted hardware to prevent malware from interfering with the attestation routine.

## 1.2 | Contributions

This article makes three contributions.

1. First, we show that *Soteria* does not achieve its intended security goals, thereby leaving open the problem of remote attestation without secure hardware in the presence of a network attacker, even within the realm of quantum computing. The flaws and pitfalls we report on *Soteria* are non-trivial, some of them arising from counterintuitive effects of quantum mechanics. Therefore, we believe they are useful for a more general audience, including, but not limited to, the community working on quantum-based security protocols.
2. We prove an impossibility result on the use hashing for quantum memory, which is a standard practice in the classical memory attestation problem. In addition, we show a counterintuitive result on the attestation of quantum memory that forces any protocol (where the prover sends its memory content back to verifier) to use a number of messages that quadratically increases with the inverse of the size of the malware. Those two proofs indicate that approaches to the attestation of quantum memory might need to be fundamentally different to existing approaches for classical memory.
3. On a more positive note, we show that quantum computing do offer features that have no counterpart in classical computing, and that can enhance the security of remote attestation protocols. One of those features is *teleportation*, which is used in Ref. [21] to prevent a remote device from trivially outsourcing the attestation routing onto a colluding (more powerful) device. The other feature is *superdense coding*, which we show in this article can be used by the verifier to agree with the prover on a random quantum circuit at runtime. This leads to an improvement over the design introduced in Ref. [21] in terms of resistance to proxy attacks. More precisely, we introduce a protocol that is secure against an attacker that colludes with the prover from a sufficiently long distance. To the best our knowledge, this is the highest level of security

remote attestation protocols have achieved so far, in both classical and quantum computing, without resorting into trusted hardware.

## 1.3 | Organisation

In Section 2, we review the fundamentals of quantum information science (QIS) necessary to understand out theoretical results. In Section 3, we discuss how the authors in Ref. [20] approached the problem of remote memory attestation. Because this design illustrates many of the issues that may derail proposed quantum implementations of memory attestation, we note the correctness and security flaws of this method here, proving several theorems that will impact any potential approach. Limitations on the attestation of quantum memory that go beyond *Soteria* are reported in Section 4. Section 5 discuss the proposal of the authors in Rep. [21], and taking into account some of the limitations and flaws reported in previous sections, we design a new system which improves on both approaches. We note that a brief review of the literature on remote memory attestation is provided in Section 1, whereas details on the two existing quantum-based memory attestation protocols are given elsewhere in the paper.

## 2 | Preliminaries

Here we introduce the notation and tools we intend to make use throughout the article. We begin by reviewing the basics of quantum states and quantum teleportation. For a complete introduction to these ideas, the reader is recommended to consult Ref. [22].

### 2.1 | Quantum Notation

We represent qubits in the standard Bra-ket notation, being  $|0\rangle$  and  $|1\rangle$  the *computational basis states* of a qubit. Unlike bits in classical computing, which are either 0 or 1, a qubit can be in superposition of its basis states, denoted by the following equation:

$$|\psi\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle \quad (1)$$

In the notation above,  $\psi$  is the *name* of the qubit, whereas  $\alpha_0$  and  $\alpha_1$  are complex numbers satisfying  $|\alpha_0|^2 + |\alpha_1|^2 = 1$ . Because the values  $\alpha_0$  and  $\alpha_1$  are complex, a qubit lies on a unit sphere known as the Bloch sphere. Considering  $(\alpha_0, \alpha_1)$  a vector on the Bloch sphere with polar angle  $\theta$ , and azimuthal angle  $\phi$ , we can represent any qubit as  $\cos(\frac{\theta}{2})|1\rangle + \sin(\frac{\theta}{2})e^{i\phi}|0\rangle$ .

We define an  $n$ -qubit state by the following equation:

$$|\psi\rangle = \sum_{i=0}^{2^n-1} \alpha_i|i\rangle \quad (2)$$

where  $\alpha_0, \dots, \alpha_{2^n-1} \in \mathbb{C}$ ,  $\sum_{i=0}^{2^n-1} |\alpha_i|^2 = 1$ , and, for every  $i \in \{0, \dots, 2^n - 1\}$ ,  $|i\rangle$  denotes the combination of the basis states

$|b_0\rangle\cdots|b_{n-1}\rangle = |b_0\cdots b_{n-1}\rangle$  where  $b_0\cdots b_{n-1}$  is  $i$ 's binary representation. According to the Born rule, the probability of obtaining any basis state  $|i\rangle$  on measurement on the computational basis is  $|\alpha_i|^2$ . That is, measuring a quantum state in superposition will make it collapse into one of its basis states.

At this point it is worth introducing the *Bell basis*, which we will use later to describe quantum teleportation. The Bell basis for 2-qubit systems is represented, in terms of the computational basis, in the following way.

$$\begin{aligned} |\Phi^+\rangle &= \frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|11\rangle \\ |\Phi^-\rangle &= \frac{1}{\sqrt{2}}|00\rangle - \frac{1}{\sqrt{2}}|11\rangle \\ |\Psi^+\rangle &= \frac{1}{\sqrt{2}}|01\rangle + \frac{1}{\sqrt{2}}|10\rangle \\ |\Psi^-\rangle &= \frac{1}{\sqrt{2}}|01\rangle - \frac{1}{\sqrt{2}}|10\rangle \end{aligned}$$

We can represent a 2-qubit state  $|\psi\rangle$  in the Bell basis by writing  $\alpha_0|\Phi^+\rangle + \alpha_1|\Phi^-\rangle + \alpha_2|\Psi^+\rangle + \alpha_3|\Psi^-\rangle$ . And, when measuring  $|\psi\rangle$  on the Bell basis, we would obtain the basis state  $|\Phi^+\rangle$  with probability  $\alpha_0^2$ , and so on. If, instead, we measure  $|\psi\rangle$  on the computational basis, we would obtain  $|00\rangle$  with probability  $\frac{(\alpha_0 + \alpha_1)^2}{2}$ , and so on. Again, measuring on a given basis, say the computational basis or the Bell basis, allows us to collapse a quantum state into one of the corresponding basis states.

## 2.2 | Entanglement

We say a state  $|\psi\rangle$  is *separable* if it can be written as the tensor product of other states, namely  $|\psi\rangle = |\psi_1\rangle \otimes |\psi_2\rangle$ . States which are not separable are *entangled*. As an example,  $\alpha_1\alpha_2|01\rangle$  is separable because it can be written as  $\alpha_1|0\rangle \otimes \alpha_2|1\rangle$ , whereas all Bell basis states are entangled.

A key property of entangled two-qubit states is that, if we provide one half of the entangled pair to one party, and the other half to another, if either party measures  $|0\rangle$  from their half of the pair, the other party will also measure  $|0\rangle$  and similarly for  $|1\rangle$ . For example, if we have a two-qubit state  $\frac{1}{\sqrt{2}}|0\rangle_a|0\rangle_b + \frac{1}{\sqrt{2}}|1\rangle_a|1\rangle_b$ , where  $|0\rangle_a$  (resp.  $|1\rangle_b$ ) denotes a qubit held by a quantum-enabled computational device  $a$  (resp.  $b$ ), then when  $a$  measures on the computational basis and obtains  $|0\rangle$ , which occurs with probability one-half, so does  $b$ , and vice versa. This property of entanglement is the basis for the power of quantum computing in general.

Notice how in the previous paragraph we wrote  $\frac{1}{\sqrt{2}}|0\rangle_a|0\rangle_b + \frac{1}{\sqrt{2}}|1\rangle_a|1\rangle_b$  instead of  $\frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|11\rangle$ . Those two-qubit quantum states are equivalent in terms of their behaviour; the subscripts are merely used to make it explicit that  $|0\rangle_a$  and  $|0\rangle_b$  are two different particles, possibly held by two different devices.

## 2.3 | Quantum Teleportation

Quantum teleportation is a method for teleporting a qubit from one party to another. The state to be teleported is transmitted instantly over any distance, but cannot be extracted without two classical bits, which can of course be transmitted no faster than the speed of light.

Suppose Alice wishes to transmit a qubit  $|\psi\rangle_c = \alpha_0|0\rangle_c + \alpha_1|1\rangle_c$  to Bob. Here, the subscript  $c$  is used to distinguish the particle to be transmitted from two other particles necessary for teleportation. Those two other particles are assumed to be entangled in any of the Bell basis states, say  $|\Phi^+\rangle_{ab} = \frac{1}{\sqrt{2}}|0\rangle_a|0\rangle_b + \frac{1}{\sqrt{2}}|1\rangle_a|1\rangle_b$ , where the subscript  $a$  (resp.  $b$ ) denotes a particle  $a$  (resp.  $b$ ) held by Alice (resp. Bob). The three-qubit system state formed by  $|\psi\rangle_c$  and  $|\Phi^+\rangle_{ab}$  is thus as follows:

$$\begin{aligned} |\psi\rangle_c \otimes |\Phi^+\rangle_{ab} &= \\ &= (\alpha_1|0\rangle_c + \alpha_2|1\rangle_c) \otimes \left( \frac{1}{\sqrt{2}}|0\rangle_a|0\rangle_b + \frac{1}{\sqrt{2}}|1\rangle_a|1\rangle_b \right) \\ &= \frac{\alpha_1}{\sqrt{2}}|0\rangle_c|0\rangle_a|0\rangle_b + \frac{\alpha_1}{\sqrt{2}}|0\rangle_c|1\rangle_a|1\rangle_b \\ &\quad + \frac{\alpha_2}{\sqrt{2}}|1\rangle_c|0\rangle_a|0\rangle_b + \frac{\alpha_2}{\sqrt{2}}|1\rangle_c|1\rangle_a|1\rangle_b \\ &= \frac{\alpha_1}{\sqrt{2}}|00\rangle_{ca}|0\rangle_b + \frac{\alpha_1}{\sqrt{2}}|01\rangle_{ca}|1\rangle_b \\ &\quad + \frac{\alpha_2}{\sqrt{2}}|10\rangle_{ca}|0\rangle_b + \frac{\alpha_2}{\sqrt{2}}|11\rangle_{ca}|1\rangle_b \end{aligned}$$

Notice that  $|\psi\rangle_c \otimes |\Phi^+\rangle_{ab}$  is written above in the computational basis. The goal next is to rewrite it in the Bell basis. We can do so by using the following identities.

$$\begin{aligned} |00\rangle_{ca} &= \frac{1}{\sqrt{2}}(|\Phi^+\rangle_{ca} + |\Phi^-\rangle_{ca}) \\ |01\rangle_{ca} &= \frac{1}{\sqrt{2}}(|\Psi^+\rangle_{ca} + |\Psi^-\rangle_{ca}) \\ |10\rangle_{ca} &= \frac{1}{\sqrt{2}}(|\Psi^+\rangle_{ca} - |\Psi^-\rangle_{ca}) \\ |11\rangle_{ca} &= \frac{1}{\sqrt{2}}(|\Phi^+\rangle_{ca} - |\Phi^-\rangle_{ca}) \end{aligned}$$

The total state of the system can thus be rewritten as follows.

$$\begin{aligned} |\psi\rangle_c \otimes |\Phi^+\rangle_{ab} &= \frac{1}{\sqrt{2}}(|\Phi^+\rangle_{ca} \otimes (\alpha_1|0\rangle_b + \alpha_2|1\rangle_b) \\ &\quad + |\Phi^-\rangle_{ca} \otimes (\alpha_1|0\rangle_b - \alpha_2|1\rangle_b) \\ &\quad + |\Psi^+\rangle_{ca} \otimes (\alpha_1|1\rangle_b + \alpha_2|0\rangle_b) \\ &\quad + |\Psi^-\rangle_{ca} \otimes (\alpha_1|1\rangle_b - \alpha_2|0\rangle_b)) \end{aligned}$$

Now Alice can measure her two qubits  $a$  and  $c$  in the Bell basis. This leads, with equal probability, the system into one of the following states.

$$\begin{aligned} & |\Phi^+\rangle_{ca} \otimes (\alpha_1|0\rangle_b + \alpha_2|1\rangle_b) \\ & |\Phi^-\rangle_{ca} \otimes (\alpha_1|0\rangle_b - \alpha_2|1\rangle_b) \\ & |\Psi^+\rangle_{ca} \otimes (\alpha_1|1\rangle_b + \alpha_2|0\rangle_b) \\ & |\Psi^-\rangle_{ca} \otimes (\alpha_1|1\rangle_b - \alpha_2|0\rangle_b) \end{aligned}$$

Alice then inform Bob which of the four states the total system is in, based on the result of her measurement by transmitting 2 classical bits. If the system is in the first state, Bob does nothing, otherwise he performs a rotation on his qubit to return it to the form  $\alpha_1|0\rangle_b + \alpha_2|1\rangle_b$ . Thus, as long as Alice can transmit two classical bits and share a Bell basis state, she can teleport an arbitrary quantum state to Bob.

## 2.4 | Superdense Coding

Whereas quantum teleportation allows us to transmit a qubit using two bits of classical information, *superdense coding* allows us to transmit two classical bits of information by transmitting one pair of an entangled qubit [23]. If the adversary obtains either of the entangled qubits separately, this provides no information about which classical pair of bits is being transmitted. The recipient, and thereby any man-in-the-middle adversary, must have both qubits in hand in order to obtain the classical information.

Like in teleportation, the protocol starts with Alice and Bob sharing a Bell state, say  $|\Phi^+\rangle_{ab} = \frac{1}{\sqrt{2}}|0\rangle_a|0\rangle_b + \frac{1}{\sqrt{2}}|1\rangle_a|1\rangle_b$ . Once Alice decides on the two bits she wishes to send to Bob, denoted  $b_0b_1$ , she rotates her qubit in such a way that the state  $|\Phi^+\rangle_{ab}$  is transformed into the  $i$ th Bell basis state where  $i$  is the integer whose binary representation is  $b_0b_1$ . Notably, if  $b_0b_1 = 00$ , no rotation is performed and  $|\Phi^+\rangle_{ab}$  remains as initially shared. After performing the rotation (if necessary), Alice sends her qubit to Bob. Upon reception of Alice's qubit, Bob will apply the Controlled NOT gate, with Alice's qubit as the control qubit, and his original qubit as the target qubit. Afterwards, he will apply the Hadamard gate on Alice's qubit. On the completion of this process he will measure either  $|00\rangle$ ,  $|01\rangle$ ,  $|10\rangle$ , or  $|11\rangle$ , depending on which Bell state Alice created before sending her qubit to Bob. Alice has thus transferred two bits of classical information, by transmitting one qubit.

## 2.5 | Amplitude Encoding and Quantum Tomography

Amplitude encoding is a mechanism to represent classical information within a quantum state. Given a vector  $X = (x_1, \dots, x_n)$  of reals, amplitude encoding codifies  $X$  in a quantum state  $|\psi_X\rangle$  as follows.

$$\frac{1}{N} [x_1|0\rangle + \dots + x_n|n-1\rangle]$$

where  $N = \sqrt{x_1^2 + \dots + x_n^2}$  is called the *normalising factor* of the encoding, which is necessary to ensure that the sum of squares of the amplitudes of the quantum state equals one.

The decoding process, that of extracting  $X$  out of  $|\psi_X\rangle$ , is known by *quantum tomography* [24]. Quantum tomography measures a

series of the same qubit over and over to attempt to extract the value of the qubit. Due to the probabilistic nature of measurement, the convergence rate of quantum tomography for an unknown qubit requires  $O(\frac{n}{\epsilon})$  number of samples, where  $\epsilon$  is the desired error.

## 3 | Flaws and Pitfalls on Soteria

This section examines *Soteria* and reports on its flaws and pitfalls. We shall introduce the two versions of *Soteria*, one addressing classical memory and the other one quantum memory. The analysis here focuses on the classical version, though, as the analysis of the quantum version is subsumed by the more general analysis we give in the next section.

### 3.1 | Specification of Soteria

The *Soteria* protocol aims at allowing a verifier to attest the memory of an IoT device, called a prover. It does so by assuming that both, prover and verifier, can perform quantum computation. In particular, the prover is manufactured with an *ideal* quantum physically uncloneable function (QPUF).

A QPUF has domain and range the universe of quantum states. A QPUF is termed *ideal* if we assume that, for every input quantum state  $|\psi\rangle$ , its output cannot be predicted unless the QPUF has been previously called on  $|\psi\rangle$ . In other words, an ideal QPUF behaves like an ideal hash function (modelled as a random oracle) where each possible query is mapped to a (fixed) random response from its output domain.

The last piece of the prover's architecture necessary to understand *Soteria* is its memory, which is split into  $m$  words of memory.<sup>1</sup> Notably, *Soteria* can be adapted to work on both quantum and classical memory. In the classical memory setting, an index is an integer in  $\{0, \dots, m-1\}$  that refers to the position of the memory word, and a word is a binary string of fixed length, typically 16 or 32 bits. In the quantum memory setting, an index is a quantum state  $\alpha_0|0\rangle + \dots + \alpha_{m-1}|m-1\rangle$  (possibly) in superposition, and so is a word. Let  $\{|\psi_0\rangle, \dots, |\psi_{m-1}\rangle\}$  be the ordered set of words in the quantum memory. Then, given an index  $\alpha_0|0\rangle + \dots + \alpha_{m-1}|m-1\rangle$ , the memory content is calculated by  $\alpha_0|\psi_0\rangle + \dots + \alpha_{m-1}|\psi_{m-1}\rangle$ . In particular, if  $\alpha_i = 1$ , then the index gives exactly the word  $|\psi_i\rangle$  rather than the superposition of many words.

The *Soteria* protocol splits into the following four phases.

#### 3.1.1 | Setup Phase

During setup, the verifier generates a random bitstring  $c$  of length  $\kappa$ , where  $\kappa$  is a security parameter, and uses amplitude encoding to create a quantum state  $|\psi_c\rangle$  (also of length  $\kappa$ ). The verifier then queries the prover's QPUF on input  $|\psi_c\rangle$  to obtain  $|\psi_r\rangle$ . This process is performed for a sufficiently large number of random inputs. Let  $\mathcal{I}_P$  denote all pairs  $(|\psi_c\rangle, |\psi_r\rangle)$  stored during the setup phase between the verifier and a prover  $P$ .

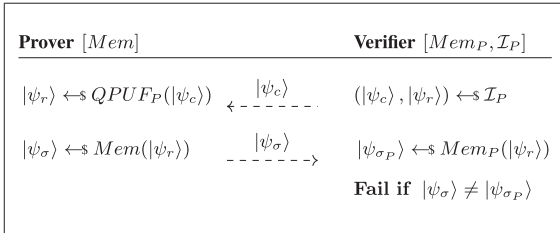
### 3.1.2 | Challenge Phase

To attest the memory of a prover  $P$ , the verifier randomly picks a pair  $(|\psi_c\rangle, |\psi_r\rangle)$  out of  $\mathcal{I}_P$  which shall determine the words of memory to be attested (see Figures 1 and 2). In the classical memory setting, the verifier also generates a nonce  $R_w$ .<sup>2</sup> The challenge  $|\psi_c\rangle$  is sent to the prover over a quantum channel, whereas  $R_w$  is sent (if necessary) over a classical channel.

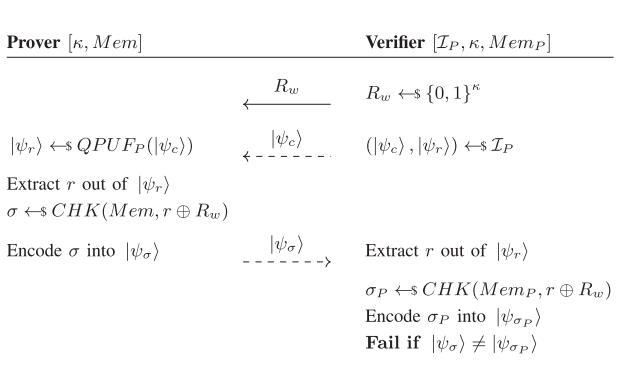
### 3.1.3 | Response Phase

The prover calls its QPUF on input  $|\psi_c\rangle$  to obtain the corresponding output  $|\psi_r\rangle$ , as previously agreed with the verifier. The next step depends on the memory setting we are in.

- In the quantum memory setting,  $|\psi_r\rangle$  is used directly as a memory index. Let  $|\psi_\sigma\rangle$  be the quantum state corresponding to the memory indexed at  $|\psi_r\rangle$ .
- In the classical memory setting, the prover extracts a bit-string  $r$  of length  $\kappa$  from  $|\psi_r\rangle$ . Soteria then follows the approach [25] to compute a checksum  $\sigma$  of the memory by using  $r \oplus R_w$  as the seed of a pseudo-random function to randomly sample memory indexes. Lastly, the prover uses amplitude encoding to encode  $\sigma$  into a quantum state  $|\psi_\sigma\rangle$ , which is sent to the verifier.



**FIGURE 1** | Soteria for quantum memory where  $Mem$  denotes the memory content of the prover and  $Mem(|\psi_r\rangle)$  denotes the memory value at index  $|\psi_r\rangle$ .



**FIGURE 2** | Soteria for classical memory where  $CHK(Mem, r \oplus R_w)$  is a (possibly partial) checksum on the prover's memory  $Mem$  with seed  $r \oplus R_w$ . Solid and dashed arrows are used to denote classical and quantum communication, respectively.

### 3.1.4 | Verification Phase

Upon reception of the quantum state  $|\psi_\sigma\rangle$ , the verifier goes through the same steps as the prover to obtain its own attestation value of the provers memory based on the challenge  $|\psi_c\rangle$ . Note that the verifier has all the information to calculate a state equal to  $|\psi_\sigma\rangle$ , including the content of the prover's memory and the challenge–response pairs obtained from the QPUF. Let  $|\psi_{\sigma_P}\rangle$  be the quantum state obtained by the verifier. The verifier then proceeds to check whether  $|\psi_\sigma\rangle = |\psi_{\sigma_P}\rangle$ . If the verifier cannot obtain the same quantum state sent by the prover, then the verifier halts the protocol and raises an alarm.

## 3.2 | Correctness Analysis

Correctness in memory attestation means that, if the prover has not been corrupted, then it should pass the protocol with a verifier. In the case of Soteria, and that of most memory attestation protocols, the prover can be modelled as a probabilistic algorithm  $P$  with domain the universe of challenges  $\mathcal{C}$  and range the universe of responses  $\mathcal{R}$ . The verifier, instead, is a probabilistic algorithm  $V$  that, on input  $P, c \in \mathcal{C}, r \in \mathcal{R}$ , outputs *pass* if  $P(c) = r$ , *fail* otherwise. This leads to the following definition of correctness.

**Definition 1.** (Correctness). *A memory attestation protocol is correct if for every honest prover  $P$ , honest verifier  $V$  and challenge  $c \in \mathcal{C}$ , the probability of  $V(P, c, P(c)) = \text{pass}$  is equal to 1.*

Notice that correctness assume neither noise nor adversarial interference in the protocol, making correctness a minimum requirement for attestation protocols. Correctness is also a necessary condition for a meaningful security analysis, as a protocol that rejects all provers has a 100% detection rate.

We show next that either Soteria is incorrect or it relies in a  $qPUF$  that is predictable hence insecure.

**Lemma 1.** *The classical version of Soteria either does not satisfy correctness or uses a  $qPUF$  implementation that outputs a basis state for every input.*

*Proof.* In Soteria, the prover receives a single challenge  $|\psi_c\rangle$  from the verifier. When the prover calls  $|\psi_r\rangle = qPUF(|\psi_c\rangle)$ , the quantum state  $|\psi_c\rangle$  rotates in a random manner, meaning that  $|\psi_c\rangle$  can only be used once and, therefore,  $r$  should be extracted out of a single copy of  $|\psi_r\rangle$ .

Suppose  $|\psi_r\rangle = \alpha_0|0\rangle + \dots + \alpha_\kappa|\kappa\rangle$ , where  $\kappa$  is the security parameter. Now, after measuring  $|\psi_r\rangle$ , the prover will obtain  $|i\rangle$  with probability  $\alpha_i^2$ . Likewise, the verifier will obtain  $|j\rangle$  with probability  $\alpha_j^2$ . This means that the probability of the prover and verifier obtaining the same basis state is  $\sum_{i=0}^{\kappa} \alpha_i^4$ . Such a probability is equal to 1 only when  $\alpha_i = 1$  for some  $i \in \{0, \dots, \kappa\}$ . Therefore, either the  $qPUF$  outputs a predictable basis state  $|i\rangle$ , which contradicts the security properties of  $qPUFs$ , or the protocol does not satisfy correctness.  $\square$

### 3.3 | Analysing Potential Fixes for Correctness

Based on the proof above, naively it appears that *Soteria* can be fixed by letting the verifier sends many copies of the challenge state. We show, however, that the classical version of *Soteria* requires at least a quadratic number of messages in terms of the security parameter  $\kappa$ , rendering the protocol unviable for reasonable security values.

**Theorem 1.** Consider a version of *Soteria* where the classical prover is able to correctly decode the bitstring  $r$ . That version requires  $\mathcal{O}(\kappa^2)$  copies of the challenge  $|\psi_c\rangle$ .

*Proof.* As mentioned in the preliminaries, the only way to extract the amplitudes from a qubit is the process of quantum tomography, which requires measuring the same qubit over and over to estimate the amplitudes value of the qubit. The question then is how many measurements *Soteria* needs to extract  $r$  out of  $|\psi_r\rangle$ .

Suppose we perform quantum tomography on  $|\psi_r\rangle$  up to accuracy  $\epsilon$ , that is  $\epsilon$  is the maximum distance from the original amplitudes  $\{\alpha_1, \dots, \alpha_\kappa\}$  and the estimate amplitudes  $\{\alpha'_1, \dots, \alpha'_\kappa\}$ , which is to say that  $\sum |\alpha_i - \alpha'_i| = \epsilon$ . The next step is to map the vector of the real values  $\{\alpha_1, \dots, \alpha_\kappa\}$  into a bitstring  $r = r_1 \dots r_\kappa$ . *Soteria* does not give a procedure to make such mapping. However, because they use amplitude encoding for transforming bitstring into state amplitudes, it is fair to assume that they divide by the normalisation factor to recover the bitstring out of the amplitudes of a quantum state. That is to say, there exists a normalisation factor  $N$  known by both the prover and the verifier satisfying that  $r_i = N\alpha_i$  with  $N = \sqrt{r_1^2 + \dots + r_\kappa^2}$ , which would allow the devices to obtain  $r_1, \dots, r_\kappa$  out of the amplitudes of  $|\psi_r\rangle$ .

Now, let  $\{\alpha_1^p, \dots, \alpha_\kappa^p\}$  and  $\{\alpha_1^v, \dots, \alpha_\kappa^v\}$  be the estimated amplitudes obtained by the prover and the verifier, respectively. Then, the memory indexes calculated by the prover and the verifier are, respectively,  $(r_1^p, \dots, r_\kappa^p) = (\lfloor N\alpha_1^p \rfloor, \dots, \lfloor N\alpha_\kappa^p \rfloor)$  and  $(r_1^v, \dots, r_\kappa^v) = (\lfloor N\alpha_1^v \rfloor, \dots, \lfloor N\alpha_\kappa^v \rfloor)$ . Because  $r$  is used as a seed of a pseudo-random function, *Soteria* needs  $r_1^p \dots r_\kappa^p = r_1^v \dots r_\kappa^v$  to function properly.

Recall that  $\alpha_i^p = \alpha_i + \epsilon_i^p$  for every  $i$  and that  $\sum |\epsilon_i^p| = \epsilon$ . Likewise, considering that the verifier would estimate the amplitudes up to the same accuracy as the prover, we obtain that  $\alpha_i^v = \alpha_i + \epsilon_i^v$  with  $\sum |\epsilon_i^v| = \epsilon$ . So the worst case for ensuring that  $r_i^p = r_i^v$  is when all the error is concentrated on the  $i$ th component for, say the prover, whereas there is no error on the  $i$ th component of the verifier. That is, when  $\epsilon_i^v = 0$  and  $\epsilon_i^p = \epsilon$  or vice versa. In that case, we have  $r_i^p = r_i^v \iff \lfloor N\alpha_i \rfloor = \lfloor N(\alpha_i + \epsilon) \rfloor$ , which implies that *Soteria* needs  $N\epsilon < 0.5$ , otherwise the protocol would incorrectly believe that the prover has been compromised.

Once we have established that  $\epsilon < \frac{1}{2N}$ , the next proof step is to use the result in Ref. [24], which states that the convergence rate of quantum tomography for an unknown qubit  $|\psi_r\rangle$  requires  $\mathcal{O}(\frac{\kappa}{\epsilon^2})$  number of samples, where  $\epsilon$  is the desired error. That is to say, *Soteria* requires  $\mathcal{O}(\kappa N^2)$  measurements of  $|\psi_r\rangle$ . Because of the

no-cloning theorem, we cannot make additional copies of  $|\psi_r\rangle$ , we must produce new ones. To do this, *Soteria* will need new copies of  $|\psi_c\rangle$  from the verifier. In other words, the prover in *Soteria* requires a quadratic number of copies from the verifier.

We end the proof by noticing that, because  $r_i$  is a bit,  $N$  is bounded by  $\sqrt{\kappa}$ , which gives a communication complexity  $\mathcal{O}(\kappa^2)$  for the response phase of *Soteria*.  $\square$

We remark that, if there is prior knowledge of the value of the qubit to be decoded, then one can use Bayesian approaches to attempt to speed the convergence of this method. However, the output of the QPUF must be unknown to the prover (and thus also to the attacker). Thus these methods cannot be employed.

### 3.4 | Security Analysis Assuming an Isolated Execution Environment

Security analysis requires an adversary model. The standard adversary for software-based memory attestation consists of a malicious prover with corrupt memory trying to convince the verifier that its memory content is in a safe state. Notably, in this model, the prover receives no help from a third-device or col-luder, known as the *isolation assumption*.

We model the adversary based on a security experiment  $\text{EXP}_{P,d,\delta}^A$  where  $d$  is an unspecified distance function on memory content and  $\delta$  a distance threshold (see Figure 3). In the setup phase, the adversary is given a memory content  $Mem_P$  such that  $d(Mem_P, Mem_P) < \delta$ . By giving the adversary a random memory at a  $\delta$ -distance from the correct memory  $Mem_P$ , we are avoiding attacks whereby the adversary reconstructs  $Mem_P$  from  $Mem_P$ . In the challenge phase of the experiment, the adversary outputs a malicious prover  $P'$  with memory  $Mem_{P'}$ . The adversary wins the experiment if, given a random challenge  $c \in \mathcal{C}$ , it holds that  $P'(c) = P(c)$ .

**Definition 2.** ( $\epsilon$ -secure attestation). A software-based attestation protocol is said to be  $\epsilon$ -secure if for every prover  $P$  the probability of the adversary winning  $\text{EXP}_P^A$  is lower than  $\epsilon$ .

The definition above ensures that, whenever a prover has been corrupt in an irreversible manner, that is, in a way that the corrupt prover cannot recover its original state, then an  $\epsilon$ -secure attestation protocol will detect the corruption with probability at least  $1 - \epsilon$ . We note this is a fairly conservative definition of security. Nonetheless, one that *Soteria* cannot meet when it is assumed correct, as we show next.

$$\begin{array}{l} \text{EXP}_{P,d,\delta}^A \\ 1 : Mem_{P'} \leftarrow \{x | d(x, Mem_P) < \delta\} \\ 2 : P' \leftarrow \mathcal{A}(Mem_{P'}) \\ 3 : |\psi_c\rangle \leftarrow \mathcal{C} \\ 4 : \text{return } P'(|\psi_c\rangle) = P(|\psi_c\rangle) \end{array}$$

FIGURE 3 | The security experiment.

**Theorem 2.** Assuming the destructive version of the SWAP test, Soteria cannot detect attacks with probability higher than  $\frac{1}{2}$ .

*Proof.* For the proof we assume a correct version of Soteria, that is, one where the prover can correctly decode the bitstring  $r$  even at the cost of (at least) a quadratic number of messages. Recall that incorrect protocols can achieve arbitrary levels of security. We also assume the destructive version of the SWAP test, as the non-destructive version requires high depth circuits and extremely low noise levels to keep the states in super position across many runs. This approach is, therefore, not suitable for near term quantum computing.

The SWAP test [26] takes two input states  $|\phi\rangle, |\psi\rangle$  and outputs a Bernoulli random variable that is 0 with probability  $\frac{1}{2} + \frac{1}{2}\langle\phi, \psi\rangle^2$ , where  $\langle\phi, \psi\rangle$  is the inner product of the two states. If two states  $|\psi\rangle$  and  $|\phi\rangle$  are identical, the test will return 0 with probability 1. This means that, to satisfy correctness, the verifier ought to accept responses for which the SWAP tests output 0. Now, being that the case, then any attacker  $P'$  wins the security experiment with probability  $\frac{1}{2} + \frac{1}{2}\langle P'(|\psi_c\rangle), P(|\psi_c\rangle)\rangle$ . In the worst case, that is, when the two states are orthogonal, the attacker still has a  $\frac{1}{2}$  chance of passing the protocol.  $\square$

To enhance the security of Soteria, one would need to run the SWAP test many times, which forces the prover to create and send many copies of  $|\sigma\rangle$ . In general, if the square of the inner product of two states differ by  $\epsilon$ , then  $n$  tests will be required to detect an attack with probability  $1 - (1 - (\frac{1}{2} - (\frac{1-\epsilon}{2})))^n$ . In other words, the verifier needs  $O(\frac{1}{\epsilon^2})$  responses (copies of the quantum state) from the prover to detect an attack that deviates from the correct response by an  $\epsilon$  error. This requirement greatly increases the actual number of runs of the proposed algorithm needed to attest the prover's memory beyond what is reported in Ref. [20].

### 3.5 | Analysing the Impact of the SWAP Test on Memory Attestation

We note another counter-intuitive implication of the SWAP test to this application. In classical communication, the prover has incentive to increase the length of vector of memory bits  $\sigma$  sent to the prover. In quantum communication, however, this is not necessarily the case. Consider the example  $|\psi\rangle_\sigma = \frac{1}{\sqrt{3}}|0\rangle + 0|1\rangle + \frac{1}{\sqrt{3}}|2\rangle + \frac{1}{\sqrt{3}}|3\rangle + 0|4\rangle$ , whereas the true memory is encoded on the verifier as  $|\psi\rangle_{\sigma^*} = 0|0\rangle + 0|1\rangle + \frac{1}{\sqrt{2}}|2\rangle + \frac{1}{\sqrt{2}}|3\rangle + 0|4\rangle$ , we have that  $\langle\psi_\sigma|\psi_{\sigma^*}\rangle = 0.8166$ , showing we have a  $p = \frac{1}{2} - \frac{0.8166^2}{2}$  probability to detect the attack after the swap test. On the other hand if we expand our sampled bits by one, assuming we sample an uncorrupted bit, we obtain  $|\psi\rangle_\sigma = \frac{1}{\sqrt{4}}|0\rangle + 0|1\rangle + \frac{1}{\sqrt{4}}|2\rangle + \frac{1}{\sqrt{4}}|3\rangle + 0|4\rangle + \frac{1}{\sqrt{4}}|5\rangle$ ,  $|\psi\rangle_{\sigma^*} = 0|0\rangle + 0|1\rangle + \frac{1}{\sqrt{3}}|2\rangle + \frac{1}{\sqrt{3}}|3\rangle + 0|4\rangle + \frac{1}{\sqrt{3}}|5\rangle$ , which yields  $\langle\psi_\sigma|\psi_{\sigma^*}\rangle = 0.8661$ , showing we have a  $p = \frac{1}{2} - \frac{0.8661^2}{2}$  probability to detect the attack after a single swap test. Paradoxically, even though we have tested more bits of memory we have a lower chance to detect the attack.

We can analyse this protocol to decide what the average probability of detecting the attack is, based on the length of bits sampled.

**Lemma 2.** We show that increasing the number of bits sampled does not on average increase the probability of detecting an attack.

*Proof.* Let  $b$  be the number of bits sampled from memory. Let  $\sigma$  be the sampled memory on the prover and  $\sigma_*$  be the sampled memory on the verifier. Let  $m$  be the number of positive bits sampled, which in expectation are  $m = \mathbf{E}\left[\frac{b}{2}\right]$ . Let  $\epsilon$  be the probability that a bit is flipped, chosen by the attacker. Let  $|\psi\rangle_\sigma = \sum_{i=0, \sigma_i \neq 0}^b \frac{1}{\sqrt{m}}|i\rangle$  be the encoded sampled memory, and let  $|\psi\rangle_{\sigma^*} = \sum_{i=0, \sigma_i(i) \neq 0}^b \frac{1}{\sqrt{m}}|i\rangle$  be the encoded true memory that the verifier is using to compare with the prover's memory. Let  $S$  be the number of bits where the memory matches in both the prover's and verifier's sample. We can compute this as  $\mathbf{E}[S] = m(1 - \epsilon)$ . We note that in expectation both the prover and verifier will have the same number of positive bits in their sample.

We want to compute  $\mathbf{E}[|\langle\psi_\sigma|\psi_{\sigma^*}\rangle|^2]$ . Where either element of the vector is 0, that element does not contribute to the dot product. We then have  $\mathbf{E}[|\langle\psi_\sigma|\psi_{\sigma^*}\rangle|^2] = \left|S\left(\frac{1}{\sqrt{m}} \frac{1}{\sqrt{m}}\right)\right|^2 = |m(1 - \epsilon)\left(\frac{1}{m}\right)|^2 = |1 - \epsilon|^2$ .

Because this is independent of  $m$ , increasing the length of bits sampled does not increase the probability of detecting the attack.  $\square$

## 4 | On the Limitations of Quantum Memory Attestation

In this section we discuss limitations of quantum memory that make the memory attestation problem challenging. These limitations are applicable to all challenge-response protocols aiming at attesting quantum memory, including the quantum version of Soteria displayed in Figure 1.

### 4.1 | Overview of Quantum Memory

Quantum memory [27], like regular memory, allows for the storage of words of memory in the system, each containing a qubit or collection of qubits. However, unlike classical memory, the system does not have to restrict itself to returning only a single word. A quantum memory system can return a superposition of these words. Given a memory index  $|\psi\rangle_R$ , the system returns a super-position of each memory word weighted by the amplitudes of  $|\psi\rangle_R$ .

Since we can return a super-position of quantum memory words, it might seem like it would be easier to design a remote memory attestation procedure than a classical memory attestation procedure, because one can sample from many words of memory simultaneously. However, our analysis has revealed three issues that in practice any quantum remote memory

attestation system would have to overcome that make this a challenging problem.

## 4.2 | Issue #1: Quantum Memory Cannot Be Copied

The no cloning theorem states that in general quantum states cannot be copied [28]. This is a serious problem for quantum remote memory attestation, because it implies that the state of the quantum memory cannot be copied and distributed to a remote verifier. The memory must moved from the prover to the verifier. This means that either the process is a destructive one from the prover's standpoint, where some amount of memory must be sacrificed to attest the remainder, or the verifier must test the memory in a non-destructive way, then return it to the prover.

Keeping the memory in super-position, transmitting it to the verifier, then comparing it to another quantum state using the swap test and then returning it to the prover, all without collapsing the super-position, is not a practical proposal for any near term quantum devices. This suggests that the more practical approach would be for the prover to sacrifice some number of qubits of its quantum memory for the attestation protocol. Even in that case, we show next that a large number of qubits will be required to attest the memory, in the presence of an attacker.

## 4.3 | Issue #2: Quadratic Communication From the Prover Required

Even if the prover is willing to sacrifice some subset of its memory in order to satisfy a remote attestation protocol, we find that if the attacker keeps the number of changes small, the prover will need to transmit a larger number of qubits to the verifier. In fact, because the memory elements can be complex numbers, the attacker could theoretically make changes which are undetectable. Consider the case where the attacker makes the prover's memory to hold  $|\psi\rangle_\sigma = \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle$ , where the true memory is  $|\psi\rangle_\sigma = \frac{i}{\sqrt{2}}|0\rangle + \frac{i}{\sqrt{2}}|1\rangle$ . Here we have  $|\langle\psi_\sigma|\psi_{\sigma^*}\rangle|^2 = 1$ , meaning the swap test will never detect a difference between these two vectors.

The main issue is the comparison between the prover's state and the verifier's state. To detect a difference between the two, we must again resort to the swap test. Again, counter-intuitively our probability of detecting the attack does not increase by increasing by sampling more words of the devices memory.

**Lemma 3.** *Suppose that for each qubit of quantum memory the attacker makes a change of  $|\epsilon| \leq 1$ , with probability  $p$ . We then show that the probability of detecting the attack is proportional to  $O(|\epsilon|^2)$ , regardless of the number of qubits sampled.*

*Proof.* Let  $|\psi\rangle_R$  be a state used to index into the quantum memory. Suppose we have  $m$  qubits of memory,  $|\rho_i\rangle$ , which with probability  $p$  has been modified to be  $|\rho_i\rangle + \epsilon$ . After indexing on the prover we have  $|\psi\rangle_\sigma = |\psi\rangle_R \otimes (|\rho_i\rangle + \epsilon)$ , whereas on the verifier we have  $|\psi\rangle_{\sigma^*} = |\psi\rangle_R \otimes (|\rho_i\rangle)$ . Thus state  $|\psi\rangle_\sigma =$

$$[|\psi\rangle_{R_1} \otimes (|\rho_1\rangle + \epsilon), |\psi\rangle_{R_2} \otimes (|\rho_2\rangle + \epsilon), \dots] \quad \text{and} \quad |\psi\rangle_{\sigma^*} = [|\psi\rangle_{R_1} \otimes (|\rho_1\rangle), |\psi\rangle_{R_2} \otimes (|\rho_2\rangle), \dots].$$

We have

$$\begin{aligned} E[|\psi\rangle_\sigma - |\psi\rangle_{\sigma^*}|_2] &= \\ &= [|\psi\rangle_{R_1} \otimes (|\rho_1\rangle + \epsilon), |\psi\rangle_{R_2} \otimes (|\rho_2\rangle + \epsilon), \dots] - \\ &= [|\psi\rangle_{R_1} \otimes (|\rho_1\rangle), |\psi\rangle_{R_2} \otimes (|\rho_2\rangle), \dots]|_2 = \\ &= [|\psi\rangle_{R_1} p\epsilon, |\psi\rangle_{R_2} p\epsilon, \dots]|_2 = \\ &= \sqrt{\sum_{i=0}^m |\psi\rangle_{R_i} p\epsilon|^2} = p\epsilon \end{aligned}$$

This implies

$$\begin{aligned} \sqrt{|\psi\rangle_\sigma|^2 + |\psi\rangle_{\sigma^*}|^2 - 2|\langle\psi_\sigma|\psi_{\sigma^*}\rangle|} &\leq |\psi\rangle_\sigma - |\psi\rangle_{\sigma^*}|_2 \\ \sqrt{|\psi\rangle_\sigma|^2 + |\psi\rangle_{\sigma^*}|^2 - 2|\langle\psi_\sigma|\psi_{\sigma^*}\rangle|} &\leq \\ |\psi\rangle_R \otimes (|\rho_i\rangle + \epsilon, \dots) - |\psi\rangle_R \otimes (|\rho_i\rangle, \dots)|_2 & \\ \sqrt{|\psi\rangle_\sigma|^2 + |\psi\rangle_{\sigma^*}|^2 - 2|\langle\psi_\sigma|\psi_{\sigma^*}\rangle|} &\leq |p\epsilon|_2 \\ |1|^2 + |1|^2 - 2|\langle\psi_\sigma|\psi_{\sigma^*}\rangle| &\leq |p\epsilon|_2^2 \\ 2|\langle\psi_\sigma|\psi_{\sigma^*}\rangle| &\geq 2 - |p\epsilon|_2^2 \\ 4|\langle\psi_\sigma|\psi_{\sigma^*}\rangle|^2 &\geq 4 - 4|p\epsilon|_2^2 + |p\epsilon|_2^4 \\ 4|\langle\psi_\sigma|\psi_{\sigma^*}\rangle|^2 - 4 &\geq -4|p\epsilon|_2^2 + |p\epsilon|_2^4 \\ 4 - 4|\langle\psi_\sigma|\psi_{\sigma^*}\rangle|^2 &\leq 4|p\epsilon|_2^2 - |p\epsilon|_2^4 \\ \frac{1}{2} - \frac{|\langle\psi_\sigma|\psi_{\sigma^*}\rangle|^2}{2} &\leq \frac{|p\epsilon|_2^2}{2} - \frac{|p\epsilon|_2^4}{8} \end{aligned}$$

□

We see then that if the attacker keeps  $|\epsilon|$  small then the number of qubits needed to detect the attack with high probability may overwhelm the communication channel between the prover and the verifier. We also observe that as in the case of classical memory increasing the number of words that the device indexes does not increase the chance to detect the attack, a very counter intuitive result.

## 4.4 | Issue #3: Impossibility of Hashing Quantum States

To avoid the problem of an attacker being able to force a large number of rounds needed to detect a change in the memory, we might hope to design a hash of quantum states (this is distinct from a quantum hash of classical states, algorithms for which do exist). That is, we could attempt to design a quantum circuit  $U$ , such that given two qubits containing the true and the modified memory samples  $|\psi\rangle_\sigma, |\psi\rangle_{\sigma^*}$ , with  $|\langle\psi_\sigma|\psi_{\sigma^*}\rangle|^2 \leq \epsilon$  we could apply  $U$  to pry these two vectors apart, that is,  $U|\psi\rangle_{\sigma^*} = |\psi'\rangle_{\sigma^*}$ ,  $U|\psi\rangle_\sigma = |\psi'\rangle_\sigma$  would have  $|\langle\psi'_\sigma|\psi'_{\sigma^*}\rangle|^2 \geq \epsilon$ . Unfortunately, it is straightforward to show that this is impossible.

**Lemma 4.** *It is impossible to design a circuit  $U$  where given two memory samples  $|\psi\rangle_{\sigma^*}, |\psi\rangle_\sigma$  such that  $|\langle\psi_\sigma|\psi_{\sigma^*}\rangle|^2 \leq \epsilon$ ,  $U|\psi\rangle_{\sigma^*} = |\psi'\rangle_{\sigma^*}$ ,  $U|\psi\rangle_\sigma = |\psi'\rangle_\sigma$ ,  $|\langle\psi'_\sigma|\psi'_{\sigma^*}\rangle|^2 \geq \epsilon$ .*

*Proof.* Given that  $U$  is a quantum circuit,  $U$  is a unitary matrix, which follows from the well-known linearity of quantum mechanics. This implies  $U^\dagger U = I$ .  $\langle \psi'_\sigma | \psi'_{\sigma^*} \rangle^2 = \langle \psi_\sigma U^\dagger | U \psi_{\sigma^*} \rangle^2 = \langle \psi_\sigma | \psi_{\sigma^*} \rangle^2 \leq \epsilon$ .  $\square$

Given the issues we have outlined above, we believe a fundamentally new approach is needed for attesting quantum memory; one that has no counterpart in the classical setting. We leave this problem for future work.

## 5 | Leveraging Quantum Effects to Counteract Network Attacks

We argue that the most important contribution of quantum theory to the field of remote memory attestation is the possibility of authenticating the prover without assuming trusted hardware. In particular, it seems possible by means of quantum effects to ensure that the prover has been active during the execution of the protocol; an authentication property known by *aliveness*. In Ref. [20], the prover is forced to be alive in the challenge phase by relying on a PUF implementation in the prover. As we just showed, though, their approach suffers from various flaws. Hence, this section is dedicated to introducing a software-based remote attestation protocol that forces the prover to be alive during both the challenge phase and the response phase with minimal communication overhead.

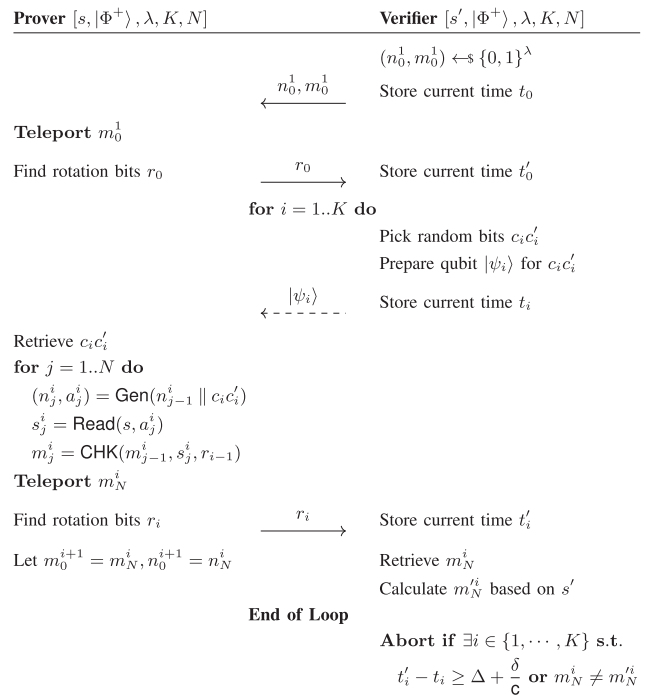
### 5.1 | The Protocol Description

We present a protocol that conservatively extends the design in Ref. [21]. This allows us to illustrate the use of aliveness during the challenge phase without sacrificing security nor having to redo the security analysis given in Ref. [21]. That is to say, we improve the protocol security by ensuring the prover is alive during the challenge phase without modifying existing security features of the protocol. The improvement lies specifically in the resistant of the protocol to proxy attacks, whereby the prover outsource the attestation task to a third-party device. In terms of communication, our conservative extension requires the transmission of a few classical bits via superdense coding. We, however, retain roughly the same communication complexity by transmitting only a few classical via superdense coding. Concretely, we transmit  $2K$  classical bits where  $K$  is a security parameter.

The protocol, depicted in Figure 4, consists of a setup and initial proximity-checking phase, multiple challenge–response phases, and a verification phase. We describe each of these phases next.

#### 5.1.1 | Setup and Initial Proximity-Checking Phase

The protocol starts with the verifier sending two nonces,  $n_0^1$  and  $m_0^1$ , to the prover. The former will be used in the next phase to feed a pseudo-random function Gen, which will create random memory addresses, the latter to calculate a checksum CHK of the prover's memory. An important feature of this phase is that



**FIGURE 4** | A software-based memory attestation protocol that employs superdense coding and teleportation to ensure the prover is alive during both the challenge and the response phase.

the prover quickly teleports  $m_0^1$  back to the verifier. This allows the verifier to check whether the prover is close.

#### 5.1.2 | Challenge–Response Phase

The challenge–response phase is executed  $\kappa$  times, where  $K$  is a security parameter. At the  $i$ th loop cycle, the verifier picks two random bits  $c_i c'_i$ , prepares the qubit  $|\psi_i\rangle$  to transmit via superdense coding, stores the current time  $t_i$ , and transmits  $|\psi_i\rangle$ . Upon reception, the prover retrieves  $c_i c'_i$  and proceeds to attest its memory by calculating a checksum of  $N$  randomly chosen memory blocks.

The calculation of the checksum is performed in  $N$  steps, where  $N$  is a security parameter. At the  $j$ th step, the function  $\text{Gen}(n_{j-1}^i || c_i c'_i)$  is used to generate a pair of nonces  $(n_j^i, a_j^i)$ ,  $s_j^i$  is used to store the memory word at the address  $a_j^i$ , and  $m_j^i$  is used to store the checksum of  $m_{j-1}^{i-1}, s_j^i$  and  $r_{i-1}$ . These checksums are encoded into a series of angles  $\theta$ , via normalisation into  $[0, \pi]$ . The prover then teleports the qubit  $|\Psi\rangle = \sin\left(\frac{\theta}{2}\right)|0\rangle + \cos\left(\frac{\theta}{2}\right)|1\rangle$ . The teleportation is carried out by entangling  $|\Psi\rangle$  with the qubit to be teleported, then measuring in the Bell basis.

#### 5.1.3 | Verification Phase

The verifier checks that, for every round  $i \in \{1, \dots, N\}$ , the checksum  $m_N^i$  is correct and the round-trip-time  $t'_i - t_i$  is below a time threshold  $\Delta$ . To verify that the checksum is correct, the verifier rotates the result based on the rotation bits ( $r$ ) provided

by the prover in its response and angle  $\theta$ . Let  $|\Psi\rangle$  be the resulting qubit. The verifier checks that  $|\Psi\rangle = |\Psi'\rangle$  by performing an inverse rotation  $|\Psi\rangle$  with angle  $\theta$ . After measuring the verifier should obtain  $|1\rangle$ . If the verifier measures  $|0\rangle$ , then the memory has been compromised.

It is worth noting that the round-trip-time  $t'_0 - t_0$  is lower than the other round-trip times  $t'_i - t_i$  with  $i > 1$ , because the first round-trip time consists of a prover reflecting back the challenge sent by the verifier while the others include the time required to compute a partial memory checksum. Therefore, in practice, the time threshold for  $t'_0 - t_0$  should be based purely on the speed of the communication channel, like in classical distance bounding. The time threshold for  $t'_i - t_i$  with  $i > 1$ , however, will need to account for the computational time taken by attestation task.

## 5.2 | A Note on the Security of the Protocol

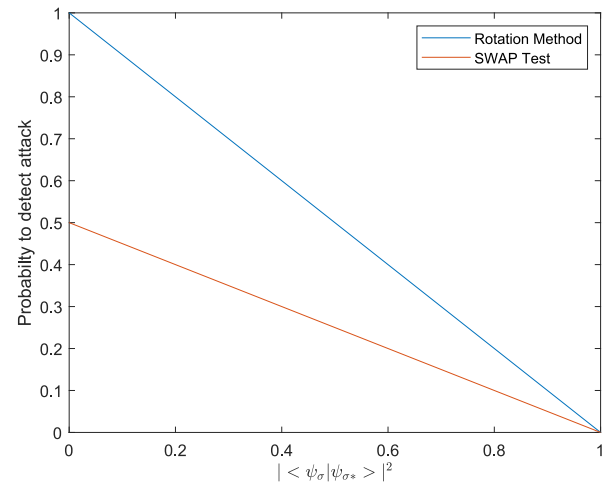
Because the introduced protocol is a conservative extension of the protocol in Ref. [21], we claim it preserves the security bounds of the latter. In practice, however, the introduced protocol is more effective at counteracting proxy attacks. The reason being that it forces the prover to interact with the network attacker during both the challenge phase and the response phase, while the protocol in Ref. [21] *only* forces the prover to participate in the response phase. The additional prover-attacker interaction increases the round-trip time measured by the verifier, thereby forcing the attacker to be closer to the verifier.

## 5.3 | A Note on the Communication Cost of Our Algorithm

While our method avoids the cost of quadratic communication between the Verifier and the Prover at the start of the algorithm, we still have a quadratic amount of communication between the Prover and Verifier during the verification phase. However, we do manage a constant improvement over the swap test required by the algorithms in Ref. [20]. The reason being that the probability of success for this method is  $|\langle \psi_\sigma | \psi_{\sigma^*} \rangle|^2$ , compared to  $\frac{1}{2} - \frac{|\langle \psi_\sigma | \psi_{\sigma^*} \rangle|^2}{2}$ . The improved efficiency is from the fact that the algorithm makes use of knowledge of what state the qubit is meant to be in, knowledge that is not exploited by the swap test. This result is in the ideal case, where there is no noise involved in the implementation of the quantum circuits involved, an unrealistic assumption in practice (Figure 5).

## 5.4 | A Note on the Impact of Noise

A complete treatment of the impact of noise on the different algorithms discussed here requires a particular physical implementation of the quantum primitives being used to implement the algorithm, since they may have different types and sources of noise. The common ways of modelling such error include depolarisation error, which models imperfect gates, thermal



**FIGURE 5** | Comparison of the two different methods of detecting an attack.

relaxation, which models decoherence and readout error, which models error in the quantum tomography.

Noise does not mean that the method does not work in practice, simply that the method must be run more times to average out the error introduced by the noise.

We have implemented a simplified circuit model of Soteria and our proposed method in Qiskit and applied noise models to the circuits. In particular, we have applied one and two qubit depolarisation error, thermal relaxation, and readout error.

The first result is below. We have modelled the PUF as applying a random rotation, and then applying the standard CSWAP circuit. Our circuit can be seen in Figure 6.

Interestingly, the noise actually has a higher impact when there is little difference between the two qubits (Figure 7). This means that the noise causes a higher reporting rate of false positives for attacks. When an attack is underway the noise plays relatively little role. However, due to the nature of the swap test, an individual run still has a 0.5 probability to fail.

Next, we have implemented our algorithm. We have implemented this as a described previously in the paper, where a random rotation is caused by the attacker's memory corruption, the qubit is teleported, and then the verifier applies a different rotation which will lead to the qubit being at an angle to  $|1\rangle$ . The standard circuit for this can be seen in Figure 8.

Interestingly, our approach has a slightly higher chance to produce a false positive than the swap test, when no attack is taking place. Overall, the noise impact is higher for this method of detection, but because of the higher ceiling of this approach, it still has a higher chance to detect an ongoing attack than the SWAP test (Figure 9).

We also note that a complete analytical treatment of noise would involve the defender's priors on both the presence of an attacker and knowledge of the noise involved. This naturally leads to a Bayesian formulation of the problem. Because of the lengthy nature of such a treatment, we leave this for future work.

### 5.5 | A Note on the Physical Implementation

There are a variety of proposals for implementing quantum communication algorithms of this nature. One attractive proposal is the use of photons as the underlying qubit. Photons are easy to transmit, which is critical for real world communication protocols. There are commercially available sources of entangled photons created through a variety of methods (e.g., spontaneous parametric down-conversion). There are also commercially available single photon sources (quantum dots), which if used in a synchronised fashion can be used to create entanglement probabilistically.

We have simulated such an approach using a popular quantum optics library SOQCS [29]. This simulator allows us to test our security protocol under a variety of assumptions with respect to different forms of error, such as noise, not inefficient detectors, etc. (Figure 10).

Simulation of our protocol using only optical methods reveals the downside of this approach. The approach is highly

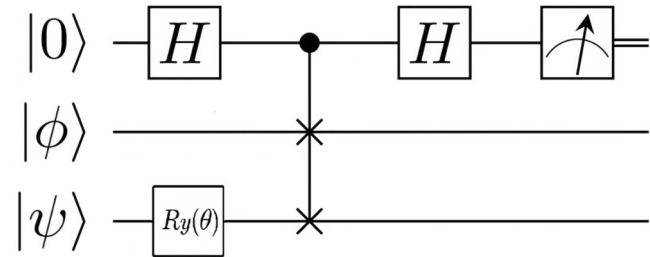


FIGURE 6 | Quantum circuit diagram of Soteria with SWAP test.

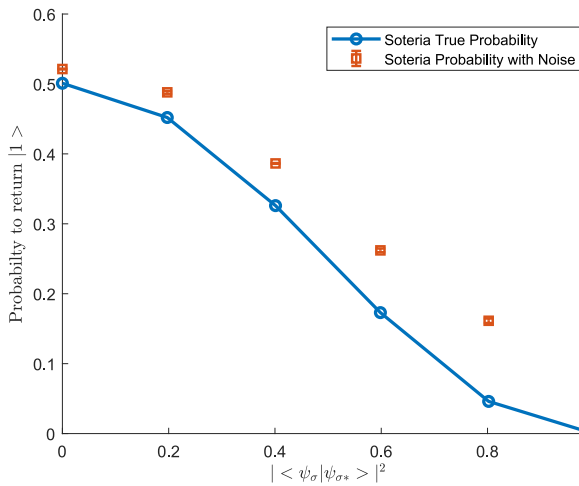


FIGURE 7 | Implementation of Soteria with noise.

non-deterministic, relying on post-selection to reveal successful runs. The simplest methods approaches of working with photonics is to use only linear optics; however, this approach cannot measure all of the possible Bell states, leading to a requirement for increased transmission. Photons are an inherently noisy communication medium, leading to the requirement to run many more rounds of the security protocol than would be needed with a more reliable and scalable qubit implementation.<sup>3</sup>

A natural candidate for a more scalable approach would be cavity quantum electrodynamics. Cavity QED traps photons in a confined space (the optical cavity) so that they can interact with more stable types of qubits such as superconducting or atomic qubits. This proven to be a successful approach to this problem when high fidelity and long term coherence is desired [30–35]. All of the quantum primitives used in this work such as teleportation and super-dense coding have been implemented using this approach. Additionally they are well suited for implementing quantum memory, which is a necessity for long term quantum computing. The disadvantage of this approach is that it requires more complex equipment. We leave an experimental implementation of our protocol using these methods to future work.

### 6 | Conclusions

In this paper we have presented a method for remote memory attestation using quantum effects. This method combines the strengths of the proposals previously made in [20, 21]. To achieve this we have analysed the methods proposed in Ref. [20] extensively, and identified previously unknown and counter-intuitive limitations of these types of approaches, which we

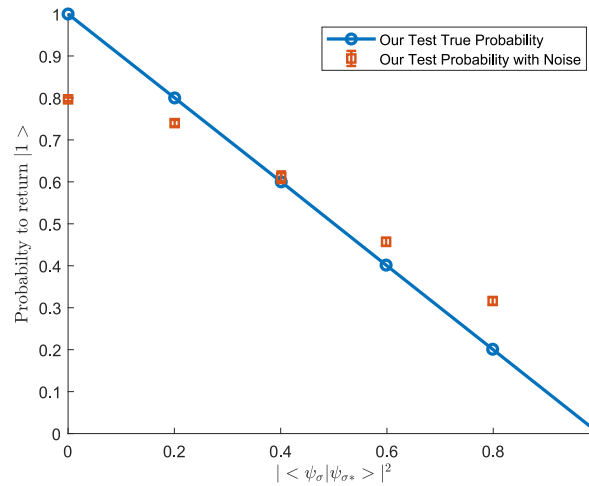


FIGURE 9 | Implementation of our algorithm with noise.

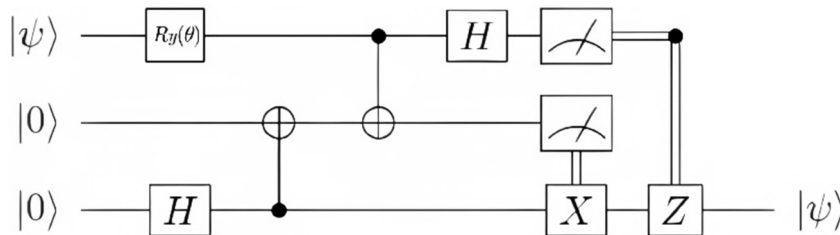
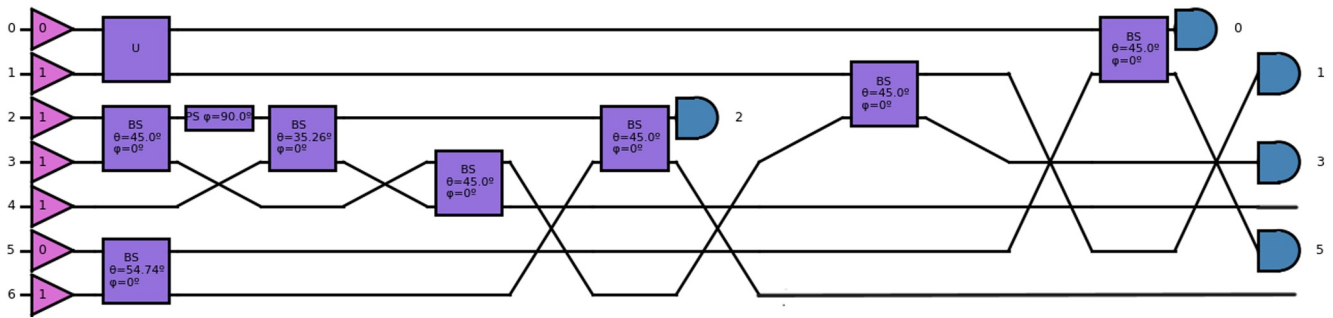


FIGURE 8 | Quantum circuit diagram of teleportation with rotation test.



**FIGURE 10** | Implementation of our security protocol using quantum optics.

believe will be useful for anyone designing these types of protocols. We have used these observations to design a new protocol that makes it harder for the prover to outsource the attestation task to a far-away colluder, as the protocol forces the prover to be involved during both: the reception of the verifier's challenge and the sending of the response. Additionally, we remove the need for special purpose qPUFs while retaining the strength they provide in this application area. This yields a device which is cheaper to produce and more easily implementable with near-term quantum computers which have a low circuit depth.

#### Author Contributions

**Jesse Laeuchli:** conceptualization, formal analysis, investigation, methodology, project administration, resources, software, validation, writing – original draft, writing – review and editing. **Rolando Trujillo Rasua:** conceptualization, formal analysis, investigation, methodology, project administration, resources, validation, visualization, writing – review and editing.

#### Acknowledgements

Rolando Trujillo Rasua was supported by a Ramón y Cajal grant (RYC2020-028954-I) from the Spanish Ministry of Science and Innovation and the EU, as well as by projects PROVTOPIA (PID2023-150098OB-I00), funded by MICIU/AEI/10.13039/501100011033 and FEDER (EU), and HERMES, funded by INCIBE and the EU's NextGenerationEU/PRTR. Open access publishing facilitated by University of New South Wales, as part of the Wiley - University of New South Wales agreement via the Council of Australian University Librarians.

#### Conflicts of Interest

The authors declare no conflicts of interest.

#### Data Availability Statement

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

#### Endnotes

<sup>1</sup> The original article assumes words of memory to be grouped into blocks. Because they play no crucial role in the security or efficiency of the protocol, we have omitted the notion of blocks.

<sup>2</sup> The original article uses an additional nonce for memory blocks, which isn't necessary here.

<sup>3</sup> <https://github.com/actionable-wallet/quantum-mem-attest>.

#### References

1. E. Dushku, M. M. Rabbani, M. Conti, L. V. Mancini, and S. Ranise, "SARA: Secure Asynchronous Remote Attestation for IoT Systems," *IEEE Transactions on Information Forensics and Security* 15 (2020): 3123–3136, <https://doi.org/10.1109/tifs.2020.2983282>.
2. M. M. Rabbani, E. Dushku, J. Vliegen, A. Braeken, N. Dragoni, and N. Mentens, "Reserve: Remote Attestation of Intermittent IoT Devices," in *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems, Ser. SenSys '21* (Association for Computing Machinery, 2021), 578–580, <https://doi.org/10.1145/3485730.3493364>.
3. R. Román, R. Arjona, and I. Baturone, "A Lightweight Remote Attestation Using PUFs and Hash-Based Signatures For Low-End IoT Devices," *Future Generation Computer Systems* 148 (2023): 425–435, <https://doi.org/10.1016/j.future.2023.06.008>.
4. M. Conti, E. Dushku, and L. V. Mancini, "RADIS: Remote Attestation of Distributed IoT Services," in *2019 Sixth International Conference on Software Defined Systems (SDS)* (IEEE, 2019), 25–32.
5. B. Parno, J. M. McCune, and A. Perrig, "Bootstrapping Trust in Commodity Computers," in *2010 IEEE Symposium on Security and Privacy* (IEEE, 2010), 414–429.
6. A. Francillon, Q. Nguyen, K. B. Rasmussen, and G. Tsudik, "A Minimalist Approach to Remote Attestation," in *2014 Design, Automation & Test in Europe Conference & Exhibition (DATE)* (IEEE, 2014), 1–6.
7. J. Cao, T. Zhu, R. Ma, Z. Guo, Y. Zhang, and H. Li, "A Software-Based Remote Attestation Scheme for Internet of Things Devices," *IEEE Transactions on Dependable and Secure Computing* 20, no. 2 (2023): 1422–1434, <https://doi.org/10.1109/tdsc.2022.3154887>.
8. A. Seshadri, A. Perrig, L. van Doorn, and P. Khosla, "SWATT: Software-Based Attestation for Embedded Devices," in *IEEE Symposium on Security and Privacy, 2004. Proceedings* (IEEE, 2004), 272–282.
9. A. Seshadri, M. Luk, E. Shi, A. Perrig, L. van Doorn, and P. Khosla, "Pioneer: Verifying Code Integrity and Enforcing Untampered Code Execution on Legacy Systems," *SIGOPS Operating Systems Review* 39, no. 5 (October 2005): 1–16, <https://doi.org/10.1145/1095809.1095812>.
10. Q. Yan, J. Han, Y. Li, R. H. Deng, and T. Li, "A Software-Based Root-of-Trust Primitive on Multicore Platforms," in *Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security, Ser. ASIACCS '11* (Association for Computing Machinery, 2011), 334–343, <https://doi.org/10.1145/1966913.1966957>.
11. X. Kovah, C. Kallenberg, C. Weathers, A. Herzog, M. Albin, and J. Butterworth, "New Results for Timing-Based Attestation," in *2012 IEEE Symposium on Security and Privacy* (IEEE, 2012), 239–253.
12. F. Armknecht, A.-R. Sadeghi, S. Schulz, and C. Wachsmann, "A Security Framework for the Analysis and Design of Software Attestation," in *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security, Ser. CCS '13* (Association for Computing Machinery, 2013), 1–12, <https://doi.org/10.1145/2508859.2516650>.

13. S. Schulz, A. Schaller, F. Kohnhäuser, and S. Katzenbeisser, "Boot Attestation: Secure Remote Reporting With Off-the-Shelf IoT Sensors," in *Computer Security – ESORICS 2017*, ed. S. N. Foley, D. Gollmann, and E. Sneekenes (Springer International Publishing, 2017), 437–455.
14. M. Ammar, B. Crispo, and G. Tsudik, "Simple: A Remote Attestation Approach for Resource-Constrained IoT Devices," in *2020 ACM/IEEE 11th International Conference on Cyber-Physical Systems (ICCPs)* (IEEE, 2020), 247–258.
15. X. Carpent, K. Eldefrawy, N. Rattanavipanon, A.-R. Sadeghi, and G. Tsudik, "Reconciling Remote Attestation and Safety-Critical Operation on Simple IoT Devices," in *Ser. DAC '18* (Association for Computing Machinery, 2018), <https://doi.org/10.1145/3195970.3199853>.
16. W. Feng, Y. Qin, S. Zhao, and D. Feng, "AAoT: Lightweight Attestation and Authentication of Low-Resource Things in IoT and CPS," *Computer Networks* 134 (2018): 167–182, <https://doi.org/10.1016/j.comnet.2018.01.039>.
17. S. Schulz, A.-R. Sadeghi, and C. Wachsmann, "Short Paper: Lightweight Remote Attestation Using Physical Functions," in *Proceedings of the Fourth ACM Conference on Wireless Network Security, Ser. WiSec '11* (Association for Computing Machinery, 2011), 109–114, <https://doi.org/10.1145/1998412.1998432>.
18. J. Kong, F. Koushanfar, P. K. Pendyala, A.-R. Sadeghi, and C. Wachsmann, "PUFatt: Embedded Platform Attestation Based on Novel Processor-Based PUFs," in *2014 51st ACM/EDAC/IEEE Design Automation Conference (DAC)* (ACM, 2014), 1–6.
19. R. Trujillo-Rasua, "Secure Memory Erasure in the Presence of Man-in-the-Middle Attackers," *Journal of Information Security and Applications* 57 (2019): 102730, <https://doi.org/10.1016/j.jisa.2020.102730>.
20. M. Khan, M. Aman, and B. Sikdar, "Soteria: A Quantum-Based Device Attestation Technique for the Internet of Things," *IEEE Internet of Things Journal* 11, no. 9 (2023): 15320–15333, <https://doi.org/10.1109/JIOT.2023.3346397>.
21. J. Laeuchli and R. Trujillo-Rasua, "Software-Based Remote Memory Attestation Using Quantum Entanglement," *Quantum Information Processing* 23, no. 6 (2024): 208, <https://doi.org/10.1007/s11128-024-04421-x>.
22. C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, "Teleporting an Unknown Quantum State via Dual Classical and Einstein-Podolsky-Rosen Channels," *Physical Review Letters* 70, no. 13 (March 1993): 1895–1899, <https://doi.org/10.1103/physrevlett.70.1895>.
23. C. H. Bennett and S. J. Wiesner, "Communication Via One- and Two-Particle Operators on Einstein-Podolsky-Rosen States," *Physical Review Letters* 69, no. 20 (November 1992): 2881–2884, <https://doi.org/10.1103/physrevlett.69.2881>.
24. J. Haah, A. W. Harrow, Z. Ji, X. Wu, and N. Yu, "Sample-Optimal Tomography of Quantum States," in *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing, Ser. STOC '16* (Association for Computing Machinery, 2016), 913–925, <https://doi.org/10.1145/2897518.2897585>.
25. M. N. Aman, M. H. Basheer, S. Dash, et al., "HAtt: Hybrid Remote Attestation for the Internet of Things With High Availability," *IEEE Internet of Things Journal* 7, no. 8 (2020): 7220–7233, <https://doi.org/10.1109/jiot.2020.2983655>.
26. H. Buhrman, R. Cleve, J. Watrous, and R. de Wolf, "Quantum Fingerprinting," *Physical Review Letters* 87, no. 16 (September 2001): 167902, <https://doi.org/10.1103/PhysRevLett.87.167902>.
27. M. Bashkansky, A. T. Black, J. M. Kwolek, and A. Kuzmich, "Quantum Memory," *Wiley Encyclopedia of Electrical and Electronics Engineering* (2021): 1–17, <https://doi.org/10.1002/047134608X.W8412>.
28. W. K. Wootters and W. H. Zurek, "A Single Quantum Cannot Be Cloned," *Nature* 299, no. 5886 (October 1982): 802–803, <https://doi.org/10.1038/299802a0>.
29. J. Osca and J. Vala, "SOQCS: A stochastic optical quantum circuit simulator," *SoftwareX* 25 (2024): 101603, <https://doi.org/10.1016/j.softx.2023.101603>.
30. S. S. Hasan, S. M. A. Anis, M. Imran, R.-U. Islam, S. Al-Kuwari, and T. Abbas, "Entanglement Swapping Using Hyperentangled Pairs of Two-Level Neutral Atoms," *IET Quantum Communication* 6, no. 1 (2025): e12121, <https://doi.org/10.1049/qtc2.12121>.
31. S. M. Arslan, S. Al-Kuwari, and T. Abbas, "Superdense Coding Using Bragg Diffracted Hyperentangled Atoms," *IEEE Journal on Selected Areas in Communications* 42, no. 7 (2024): 1950–1959, <https://doi.org/10.1109/jsac.2024.3380098>.
32. M. A. Khan, S. Shahmir, M. T. Rahim, S. Al-Kuwari, and T. Abbas, "Quantum Repeater Chains via Cavity Magnon for Scalable Quantum Networks," *arXiv preprint arXiv:2507.04499* (2025).
33. L. Ali, M. Ikram, T. Abbas, and I. Ahmad, "Teleportation of Atomic External States on the Internal Degrees of Freedom," *Quantum Information Processing* 21, no. 2 (2022): 55, <https://doi.org/10.1007/s11128-021-03400-w>.
34. L. Ali, M. Ikram, T. Abbas, I. Ahmad, "Hyperentanglement Teleportation Through External Momenta States," *Journal of Physics B: Atomic, Molecular and Optical Physics* 54, no. 23 (2022): 235501, <https://doi.org/10.1088/1361-6455/ac42dc>.
35. S. A. Anis, T. Abbas, and M. Imran, "Engineering Quantum Networks Through Bragg Diffracted Hyperentangled Atoms," *Physica Scripta* 96, no. 12 (2021): 125102, <https://doi.org/10.1088/1402-4896/ac1c17>.