A Black-box Attack on Fixed-Unitary Quantum Encryption Schemes

Cezary Pilaszewicz[1*], Lea R. Muth[1] and Marian Margraf[1]

[1]Department of Mathematics and Computer Science, Freie Universität Berlin, Takustr. 9, Berlin, 14195, Germany.

*Corresponding author(s). E-mail(s): cezary.pilaszewicz@gmail.com;
Contributing authors: lea.muth@fu-berlin.de;
marian.margraf@fu-berlin.de;

Abstract

We show how fixed-unitary quantum encryption schemes can be attacked in a black-box setting. We use an efficient technique to invert a unitary transformation on a quantum computer to retrieve an encrypted secret quantum state $|\psi\rangle$. This attack has a success rate of 100% and can be executed in constant time. We name a vulnerable scheme and suggest how to improve it to invalidate this attack. The proposed attack highlights the importance of carefully designing quantum encryption schemes to ensure their security against quantum adversaries, even in a black-box setting.

Keywords: quantum cryptography,black-box attack, quantum cryptanalysis, quantum circuits

1 Introduction

The constant development of quantum computers has made encryption methods increasingly relevant in this field. Particularly, quantum-based synergy effects are presenting new security challenges to classical methods. In this paper, an attack on a previously proposed quantum encryption scheme (QES) is carried out to demonstrate the cryptographic insecurity of this scheme. Note that this is not quantum encryption in the classical sense, e.g., Quantum Key Distribution (QKD), as the scheme is a new approach by the authors of [1]. In [1], the authors proposed the problem of sending a qubit in a secret state $|\psi\rangle$ from one entity to another. The key point here is that

no entity other than the desired receiver should have access to the qubit in the secret state $|\psi\rangle$. They propose a quantum public-key encryption scheme to accomplish this. In our attack, we abuse the deterministic part of the QES protocol to create an oracle for the private key application performed by the receiver (Alice). We apply the method of [2] to invert Alice's transformation and retrieve the qubit $|\psi\rangle$. The attack has a constant runtime and a success probability of 100%.

The paper is structured as follows: First, we explain the encryption scheme proposed by [1]. In Section 3, a technique described in [2] to invert a black-box unitary is proposed. In Section 4, we use the technique to attack the QES scheme. Finally, we mention how the protocol could be improved to prevent this attack.

2 Encryption Scheme

This chapter introduces the QES proposed in [1]. The scheme's aim is to secretly transmit a single qubit in an arbitrary state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ from the sender (Bob) to the receiver (Alice). To accomplish this, Alice implements the public-key cryptosystem QES.

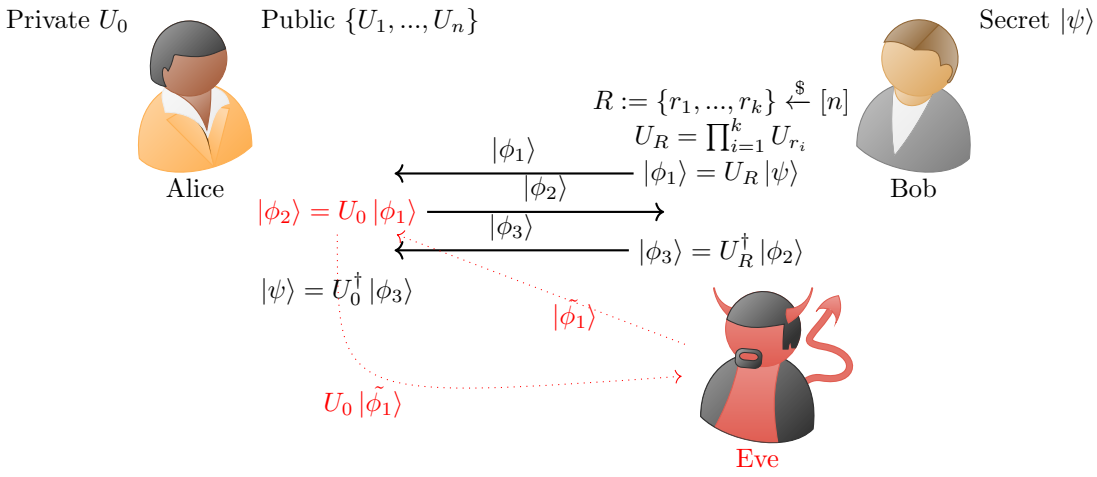

Fig. 1 The quantum encryption scheme. In red, we denoted the dangerous part.

The scheme starts with Alice generating her public and private keys. To build the private key, Alice chooses the random numbers $a, b \in \mathbb{C}, \varphi \in \mathbb{R}$ with $|a|^2 + |b|^2 = 1$. She uses those numbers to build a unitary matrix $U_0 = \begin{pmatrix} a & b \\ -e^{i\varphi}b^* & e^{i\varphi}a^* \end{pmatrix}$ [1]. She then generates n-many t-bit numbers $p_1, ..., p_n$ and computes:

$$U_i = U_0^{p_i} \quad \forall 1 \leq i \leq n.$$

The authors of [1] consider n and t as security parameters. The exponentiated matrices build the public key PubK $:= \{U_1, ..., U_n\}$, while the secret key PrivK $:= U_0$ is the original matrix. The protocol is presented as follows:

1. Bob begins by generating an arbitrary valid single qubit state

$$|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle\,,$$

with $\alpha, \beta \in \mathbb{C}$ and $|\alpha|^2 + |\beta|^2 = 1$ which he wants to send to Alice. He then chooses a subset $R \subset [n]$, and uses Alice's PubK to construct the first transformation:

$$U_R = \prod_{i \in R} U_i.$$

He applies U_R to $|\psi\rangle$ to get the encrypted state, $|\phi_1\rangle = U_R\,|\psi\rangle$ which he then transmits to Alice.

2. Alice applies U_0 to $|\phi_1\rangle$ to obtain $|\phi_2\rangle$. In the original work ([1]), the authors suggest that Alice applies $U_T = \prod_{i=0}^{n} U_i$ instead. However, we want to emphasize that there is no benefit in using U_T over U_0. For all $i \neq 0$, U_i is public knowledge, and an attacker can easily build the inverse $U_i^\dagger$. Alice then sends $|\phi_2\rangle$ back to Bob.
3. Bob uncomputes his transformation U_R to get the state $|\phi_3\rangle$. He can produce the inverse transformation by simply combining the inverses of the partial matrices $U_R^\dagger = \prod_{i \in R} U_i^\dagger$. Based on the fact that all U_i's are multiples of U_0, we know that U_i's commute. This allows the following:

$$|\phi_3\rangle = U_R^\dagger \cdot U_0 \cdot U_R\,|\psi\rangle = U_R^\dagger \cdot U_R \cdot U_0\,|\psi\rangle = U_0\,|\psi\rangle$$

The state $|\phi_3\rangle$ is then sent to Alice.

4. Alice can uncompute U_0 by simply applying $U_0^\dagger$:

$$U_0^\dagger\,|\phi_3\rangle = U_0^\dagger \cdot U_0\,|\psi\rangle = |\psi\rangle$$

With this, Alice recovered the original secret quantum state $|\psi\rangle$.

3 Inverting Black-box Unitaries

A matrix U is called unitary iff:

$$U \cdot U^\dagger = U^\dagger \cdot U = Id,$$

where $U^\dagger$ is the conjugate transpose of the matrix U. Thus, in a white-box setting, finding the inverse of the matrix U is trivial, and the runtime of the inversion depends only on the size of U. Also, in a classical black-box setting, with access to a chosen plaintext U-oracle, determining the matrix U (and therefore also $U^\dagger$) is rather simple. For an $N \times N$ matrix, one can just query N-many unit vectors $(e_i)_{i=1,...,N}$ and

reconstruct U as:

$$U = \begin{pmatrix} | & | & & | \\ U(e_1) & U(e_2) & \cdots & U(e_N) \\ | & | & & | \end{pmatrix}$$

The problem becomes more challenging in the quantum setting. Assuming U was applied to a quantum state $|\psi\rangle$, we cannot determine the amplitudes of $U|\psi\rangle$. In fact, we cannot even differentiate between the states $|0\rangle$ and $i|0\rangle$ since the global amplitude has no impact on the result of the measurement (cf. [3, p. 87]). With this in mind, the problem of finding a pre-image of a quantum state under a matrix U comes into play:

Problem 1. *Given a quantum state $|\tilde{\psi}\rangle$ and a black-box access to a unitary matrix U, find $|\psi\rangle$ such that:*

$$U|\psi\rangle = |\tilde{\psi}\rangle$$

In other words, we want to find the state $U^\dagger|\tilde{\psi}\rangle$. It is important to differentiate between two very close cases. To solve Problem 1, we do not expect the attacker to determine the amplitudes of the quantum states. Rather, he has to have access to a qubit in state $|\psi\rangle$. To achieve this, [4] proposed an exact protocol, with runtime dependent on the matrix's size. Another approach is to perform process tomography [5].

In this paper, we are not interested in inverting arbitrary unitaries. Instead, we focus on 2×2 matrices as present in Section 2. One general expression for 2×2 unitary matrices is the form already mentioned above:

$$U = \begin{pmatrix} a & b \\ -e^{i\varphi}b^* & e^{i\varphi}a^* \end{pmatrix},$$

with $a, b \in \mathbb{C}, \varphi \in \mathbb{R}$ and $|a|^2 + |b|^2 = 1$. In [2], the authors describe how to reverse an arbitrary single-qubit gate in constant time. The procedure calls the oracle U four times and applies two unitary operations V^1 and V^2. V^1 and V^2 are constructed using Clebsch-Gordan transforms (for detail, see [6]). The circuit can be seen in Figure 2 and it outputs the state $U^{-1}|\tilde{\psi}\rangle$ for an arbitrary 2×2 unitary U and an arbitrary initial state $|\tilde{\psi}\rangle$. Additionally, [2] provides an implementation of the method for a random unitary matrix and a random initial single qubit state $|\tilde{\psi}\rangle$ in Qiskit[1] code.

4 Black-box Attack

In this Section, we will explain how to attack the QES described in Section 2 with the technique from Section 3. The QES protocol's aim is to secretly transfer a qubit $|\psi\rangle$ from Bob to Alice. We assume the attack is successful whenever the attacker can obtain the qubit $|\psi\rangle$.

The attack begins with Eve intercepting the qubit $|\phi_3\rangle = U_0|\psi\rangle$ being transmitted from Bob to Alice. This takes place within step 3 of the QES protocol. At this point, the U_R transformation of Bob has already been uncomputed (cf. Figure 1).

[1] https://qiskit.org/

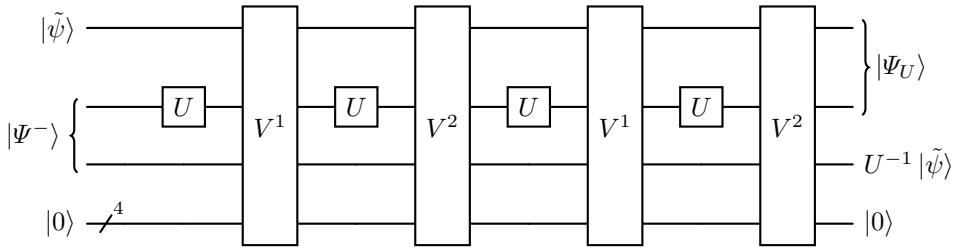

Fig. 2 The algorithm to revert an arbitrary unitary, as proposed in [2]. The state $|\Psi^-\rangle := \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$ and $|\Psi_U\rangle$ is defined as $|\Psi_U\rangle := U \otimes Id\,|\Psi^-\rangle$ and can be further reused.

Next, we use $|\Phi\rangle = |\phi_3\rangle\,|\Psi^-\rangle\,|0\rangle^{\otimes 4}$ as input to the algorithm described in Figure 2. At this point, we need to specify how Eve will achieve access to the unitary U. We observe that Alice, when being sent a qubit $|\phi_1\rangle$, in step 1 of the QES protocol, is not able to differentiate between a valid qubit of form $U_R\,|\psi\rangle$, and a qubit in an arbitrary state $|\tilde{\phi_1}\rangle$. This means she will apply U to any qubit that is being sent to her. We will abuse this fact and use Alice as an oracle for the function U. Whenever the algorithm from Figure 2 needs to apply U, we send the second qubit of $|\Phi\rangle$ to Alice, pretending it is a valid initial message of the QES protocol (cf. Eve message in Figure 1).

Finally, the transformations V^1 and V^2 are fixed, therefore, not dependent on U, and can be easily implemented in the quantum framework (cf. [2] for Qiskit code). The whole attack consists of a fixed amount of steps (four protocol calls to Alice and four applications of fixed unitary matrices V^1 and V^2). The success rate is 100%. The desired secret state $|\psi\rangle$ is now the third qubit of $|\Phi\rangle$.

5 Design Criteria for Quantum Encryption Schemes

In this Section, we want to investigate which part of the QES protocol leads to the faulty security properties. A well-established property of security protocols, in general, is the need for randomness. Here, [1] incorporates randomness in the process of Bob selecting the U_R. This approach is similar to classical schemes such as OAEP or PKCS#1, where the party which encrypts the message has to include the randomness in the encryption process to get a probabilistic encryption scheme.

In the case of QES, there is, however, the second part of the encryption process, which Alice performs. This is the step which is vulnerable to the attack mentioned in this paper. The deterministic nature of Alice's encryption is the property which we use to attack and break the scheme. We point to the fact that the attacker needs to restart the protocol four times after he obtains the state $|\phi_3\rangle$. If there would be randomness used on Alice's side, the transformation she performs would differ in each protocol call. One suggestion to prevent this vulnerability is to alter the map which Alice applies. Instead of just applying U_0, similar to Bob, Alice also picks a random subset of PubK and applies it to the qubit. In the last step, she remembers the used randomness and can uncompute each rotation. The randomness guarantees that the oracle can perform only a single operation before it becomes unusable. Here, we want to highlight an essential result from [7]. They mention a no-go theorem which states,

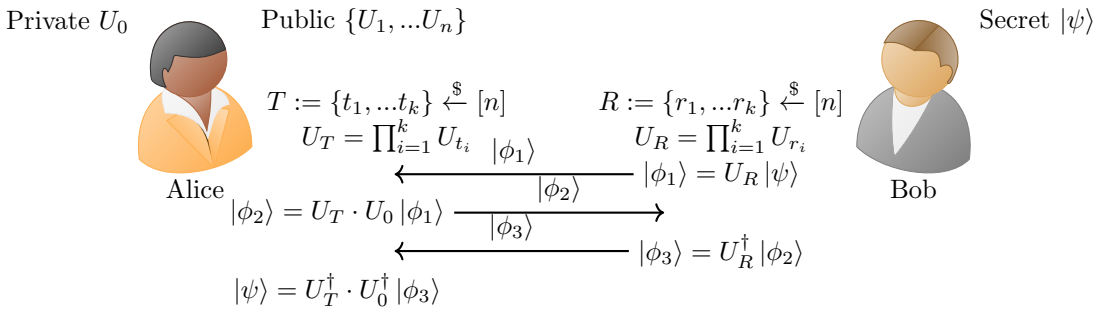

Fig. 3 The adjusted quantum encryption scheme.

that it is not possible to implement the inverse operation U^{-1} deterministically and exactly with a single call of the U-oracle, invalidating this attack. The updated scheme is presented in Figure 3.

References

[1] Plesa, M.-I., Togan, M.: A New Quantum Encryption Scheme. Advanced Journal of Graduate Research **4**, 59–67 (2018) https://doi.org/10.21467/ajgr.4.1.59-67

[2] Yoshida, S., Soeda, A., Murao, M.: Reversing Unknown Qubit-Unitary Operation, Deterministically and Exactly. Physical Review Letters **131**(12), 120602 (2023) https://doi.org/10.1103/PhysRevLett.131.120602

[3] Nielsen, M.A.: Quantum Computation and Quantum Information / Michael A. Nielsen & Isaak L. Chuang., 10th anniversary ed. edn. Cambridge Univversity Press, Cambridge [u.a (2010)

[4] Quintino, M.T., Dong, Q., Shimbo, A., Soeda, A., Murao, M.: Reversing Unknown Quantum Transformations: Universal Quantum Circuit for Inverting General Unitary Operations. Physical Review Letters **123**(21), 210502 (2019) https://doi.org/10.1103/PhysRevLett.123.210502 arxiv:1810.06944 [quant-ph]

[5] Chuang, I.L., Nielsen, M.A.: Prescription for experimental determination of the dynamics of a quantum black box. Journal of Modern Optics **44**(11-12), 2455–2467 (1997) https://doi.org/10.1080/09500349708231894 arxiv:quant-ph/9610001

[6] Yoshida, S., Soeda, A., Murao, M.: Supplemental Material for "Reversing Unknown Qubit-Unitary Operation, Deterministically and Exactly"

[7] Chiribella, G., Ebler, D.: Optimal quantum networks and one-shot entropies. New Journal of Physics **18**(9), 093053 (2016) https://doi.org/10.1088/1367-2630/18/9/093053