White-Box Encryption Scheme Using a Quantum Memory

Hidenori Kuwakado[1], Shoichi Hirose[2], and Masahiro Mambo[3]

[1] Kansai University, 2-1-1 Ryozenji-cho Takatsuki-shi, Osaka 569-1095, Japan
[2] The University of Fukui, 3-9-1 Bunkyo, Fukui-shi, Fukui 910-8507, Japan
[3] Kanazawa University, Kakuma, Kanazawa-shi, Ishikawa 920-1192, Japan

Abstract. White-box cryptography is often used in embedded applications. Although white-box cryptography with provable security has been proposed recently, the circuit size is much larger than that of usual block ciphers. We address this problem in a different way from previous works. In particular, we propose a white-box symmetric cipher using quantum memory. The size of our cipher is a polynomial in input-length and output-length of an underlying function. The security against classical attacks is reduced to the security of the underlying classical pseudorandom function. We show that quantum attacks using the generalized Grover algorithm to our cipher are ineffective.

Keywords: white-box cryptography · quantum memory · Grover's algorithm · symmetric cipher

1 Introduction

White-box cryptography provides implementations of symmetric ciphers that offer cryptographic security against an adversary who makes access to the implementation. Specifically, white-box cryptography protects the confidentiality of keys that is used in underlying symmetric ciphers. Applications of white-box cryptography include IC cards and digital rights management (DRM). Although the white-box implementation of block ciphers such as DES and AES was initially studied, no implementation successfully hides the key. The current trend of white-box cryptography is to define the security to be achieved by white-box cryptography and to show dedicated constructions satisfying the definitions. As an example of the security goal, no adversary having access to the implementation can produce a functionally equivalent circuit that is significantly smaller than the original implementation. This notion is called incompressibility [1], weak white-box [2], or space-hardness [3]. The circuit size of the constructions satisfying the notions above is much larger than that of usual block ciphers such as AES [3, 4].

In fact, there exists a secure white-box implementation of symmetric ciphers. The table of all the pairs of a plaintext and its ciphertext (e.g., the encryption with AES) is stored in the memory of a device. However, this implementation is impractical because it requires $\ell 2^{\ell+1}$-bit memory where ℓ is a block length. We solve the problem of the memory size by using quantum memory. The number of qubits in our scheme is a polynomial in $L+\ell$ where L is input-length and ℓ is output-length of an underlying pseudo-random function. Although our scheme requires quantum memory for storing the table, the plaintext and its ciphertext are classical (i.e., digital data).

The disadvantage of our scheme is that the quantum state of the table is non-reusable because of wave function collapse by measurement. However, after some secret information was shared by both sides, their communication can be made confidential by using usual symmetric ciphers. It is also allowed to request the quantum state of the same table again. Hence, the non-reusability is not a fatal flaw.

This paper is organized as follows: Section 2 describes the proposed scheme that is based on the classical pseudo-random function. Section 3 analyzes the security of the scheme on the classical channel. It is shown that the classical security is reduced to the security of the underlying pseudo-random function. Section 4 analyzes the security against attacks using the generalized Grover algorithm. It is shown that such attacks fail because of the lowness of success probability or the largeness of the number of operations. Section 5 concludes this paper.

2 XOR Scheme Using a Quantum State

2.1 Encryption Scheme

Let us consider the following scenario of white-box cryptography. Alice manufactures the same encryption devices that output a ciphertext for a given message, and she gives them to all the partners. When Bob, who is one of her partners, wants to send a message securely, Bob encrypts the message using the device and sends the ciphertext to Alice. From the viewpoint of confidentiality, partners excluding Bob cannot obtain any information about the message from the ciphertext even if the given device is used.

Suppose that partners have full access to implementation of the devices. Hence, an encryption key in the device must be hidden from partners. The obvious method to hide the encryption key was described in Sect. 1. Although the obvious method is classically infeasible, this section shows that the obvious method is implementable using quantum memory.

Let F be a function from $\{0,1\}^{\lambda} \times \{0,1\}^{L}$ to $\{0,1\}^{\ell}$ where λ, L, and ℓ denote a key-length, an input-length, and an output-length, respectively. We call the following scheme a XORQS$[F]$ *scheme.*

Alice prepares the following state $|\phi_1\rangle$.

$$\begin{aligned} |\phi_1\rangle &= I_\ell \otimes H_L \left|0^\ell\right\rangle \left|0^L\right\rangle \\ &= \frac{1}{\sqrt{2^L}} \sum_{r\in\{0,1\}^L} \left|0^\ell\right\rangle |r\rangle \,, \end{aligned}$$

where I_ℓ and H_L are the 2^ℓ-dimensional identity matrix and the 2^L-dimensional Hadamard matrix, respectively. Alice chooses a λ-bit key $\hat{v}$ according to the uniform distribution on $\{0,1\}^\lambda$. Let U_f be a unitary operator for computing f where $f = F(\hat{v}, \cdot)$.

$$\begin{aligned}|\psi\rangle &= U_f\,|\phi_1\rangle \\ &= \frac{1}{\sqrt{2^L}} \sum_{r\in\{0,1\}^L} |f(r)\rangle\,|r\rangle \\ &= \sum_{(c,r)\in\{0,1\}^\ell\times\{0,1\}^L} a_{c,r}\,|c,r\rangle \qquad (1)\end{aligned}$$

where

$$a_{c,r} = \begin{cases} \frac{1}{\sqrt{2^L}} & \text{if } c = f(r) \\ 0 & \text{otherwise.} \end{cases} \qquad (2)$$

Alice gives $|\psi\rangle$ to all the partners including Bob. Suppose that Bob wants to send an ℓ-bit message m to Alice in secret. Bob performs the projective measurement to $|\psi\rangle$. Let $(\hat{c}, \hat{r})$ be the measurement result where $\hat{c} = f(\hat{r})$ because of Eq. (2). Bob sends $(\hat{c} \oplus m, \hat{r})$ to Alice using a classical (digital) channel. After receiving it, Alice can obtain the message m by computing $(\hat{c} \oplus m) \oplus f(\hat{r})$ because Alice knows $\hat{v}$.

Since we focus on white-box applications, we assume that devices (e.g., commercial products) in which $|\psi\rangle$ is embedded are securely sent from Alice to all the partners. Unlike BB84 [5], the channel for sending $|\psi\rangle$ is not quantum channel. Furthermore, since all the partners protect devices securely, an adversary cannot make access to someone else's (e.g., Bob's) device. However, there may exist the adversary in the partners. For example, Bob himself may analyze the device. Theses assumptions are probably acceptable in white-box applications.

2.2 Properties of the XORQS[F] Scheme

The unitary operator U_f is used once to produce $|\psi\rangle$, which essentially requires $L + \ell$ qubits. The state $|\psi\rangle$ is considered as the table that describes pairs of an input and its output. If such a table is produced with digital (classical) information, then f is performed 2^L times and the table size is $(L + \ell)2^L$ bits.

The measurement result $(\hat{c}, \hat{r})$ is one-time. In some applications, it may be sufficient for Alice and Bob to share the message m once. For example, they can use m as a common-key for encrypting messages using common-key block ciphers. If Bob wants to send several messages, then Alice may give as many devices in which the same state $|\psi\rangle$ is embedded as Bob needs.

3 Classical Security Analysis

This section shows the security of the XORQS[F] scheme against a classical eavesdropper that can obtain $(\hat{c} \oplus m, \hat{r})$. From the viewpoint of classical communication, the XORQS[F] scheme is equivalent to the following scheme XOR[F].

Alice and Bob share the λ-bit key $\hat{v}$. Bob chooses $\hat{r}$ according to the uniform distribution on $\{0,1\}^L$. For an ℓ-bit message m, Bob sends $(F_{\hat{v}}(\hat{r}) \oplus m, \hat{r})$ to Alice where $F_{\hat{v}} = F(\hat{v}, \cdot)$.

We formally analyze the security of the XOR[F] scheme in terms of the left-or-right indistinguishability. Theorem 1 below means that if F is a pseudorandom function, then the security of the XOR[F] scheme mainly depends on the output-length ℓ of F.

Theorem 1. *Let* $\mathrm{Adv}^{\mathrm{lor-cpa}}_{\mathrm{XOR}[F]}(\lambda, t, q_e, \mu_e)$ *be the advantage function of* XOR[F] *in terms of the left-or-right indistinguishability for time complexity t, at most q_e queries to the oracle of* XOR[F] *and the amount of ciphertexts μ_e when any chosen-plaintext attack is performed. Let* $\mathrm{Adv}^{\mathrm{prf}}_{F}(t, q)$ *be the advantage function of F for time complexity t and at most q queries to the oracle of F in terms of the indistinguishability from a random function from $\{0,1\}^L$ to $\{0,1\}^\ell$. Then, the following inequality holds.*

$$\mathrm{Adv}^{\mathrm{lor-cpa}}_{\mathrm{XOR}[F]}(\cdot, t, q_e, \mu_e) \leq 2\mathrm{Adv}^{\mathrm{prf}}_{F}(t, q_e) + \frac{q_e(q_e - 1)}{2^{\ell+1}} \tag{3}$$

Since there is no asymptotics on the key-length, the first argument in the inequality above is written in the symbol '$\cdot$'.

4 Quantum Security Analysis

This section analyzes the security of XORQS[F] against quantum attacks. This section also assumes that if $\hat{v}$ is unknown, then $\Pr[c = f(r)] = 2^{-\ell}$ for a given $c \in \{0,1\}^\ell$ and $r \in \{0,1\}^L$. In other words, an adversary considers f as a random function.

4.1 Attacks Using the Projective Measurement

Suppose that Eve is given $|\psi\rangle$ by Alice and knows $(\hat{c} \oplus m, \hat{r})$ that Bob sent to Alice on the classical channel. To obtain $\hat{c}$, Eve measures $|\psi\rangle$ via the orthonormal computational basis. Specifically, Eve prepares a collection of measurement operators $\{\hat{R}, I_{\ell+L} - \hat{R}\}$ where

$$\hat{R} = \sum_{c \in \{0,1\}^\ell} |c, \hat{r}\rangle \langle c, \hat{r}| \,.$$

The probability that the result corresponding to $\hat{R}$ occurs is given by

$$\begin{aligned} \Pr\left[\hat{R}\right] &= \langle\psi| \hat{R}^\dagger \hat{R} |\psi\rangle \\ &= a^2_{\hat{c},\hat{r}} \end{aligned}$$

where $\hat{R}^\dagger$ denotes the Hermitian conjugate of $\hat{R}$. Hence, the probability of obtaining $\hat{c}$ is 2^{-L}. When the input length L is a typical value (e.g., 128, 256), the success probability of this attack is negligible.

4.2 Attacks Using the Generalized Grover Algorithm

Generalized Grover Algorithm We here summarize the generalized Grover algorithm given in article [8] to analyze the XORQS[F] scheme. The difference from (usual) Grover's algorithm is that an initial state is arbitrary.

Let $|\psi\rangle$ [4] be an initial state that is a superposition of any N states $|\psi_i\rangle$ $(i = 1, 2, \ldots, N)$. Suppose that $|\psi\rangle$ is a superposition of a *marked state* $|\psi^{(\mathrm{m})}\rangle$ and an *unmarked sate* $|\psi^{(\mathrm{u})}\rangle$ as

$$|\psi\rangle = |\psi^{(\mathrm{m})}\rangle + |\psi^{(\mathrm{u})}\rangle\,,$$

where

$$|\psi^{(\mathrm{m})}\rangle = \sum_{i=1}^{M} k_i\,|\psi_i\rangle\,,\quad |\psi^{(\mathrm{u})}\rangle = \sum_{i=M+1}^{N} l_i\,|\psi_i\rangle\,,$$
$$\sum_{i=1}^{M} |k_i|^2 + \sum_{i=M+1}^{N} |l_i|^2 = 1.$$

Note that ℓ and l are different symbols. We assume that there exists a unitary operator U such that

$$U\,|x\rangle = \begin{cases} -\,|x\rangle & \text{if } x \in \{\psi_1, \psi_2, \ldots, \psi_M\} \\ |x\rangle & \text{if } x \in \{\psi_{M+1}, \psi_{M+2}, \ldots, \psi_N\}. \end{cases} \tag{4}$$

Let us consider time evolution by applying Grover's iteration to $|\psi\rangle$. Let $k_i(t)$ $(i = 1, 2, \ldots, M)$ be marked amplitudes k_i at time t and let $l_i(t)$ $(i = M + 1, r + 2, \ldots, N)$ denote unmarked amplitudes l_i at time t. The average of $k_i(t)$ and that of $l_i(t)$ are denoted by $\bar{k}(t)$ and $\bar{l}(t)$, respectively, which are given by

$$\bar{k}(t) = \frac{1}{M}\sum_{i=1}^{M} k_i(t)$$
$$\bar{l}(t) = \frac{1}{N-M}\sum_{i=M+1}^{N} l_i(t).$$

Let $C(t)$ be a weighted average that is defined as

$$C(t) = -\frac{2}{N}\left(\sum_{i=1}^{M} k_i(t) - \sum_{i=M+1}^{N} l_i(t)\right).$$

According to article [8], the average marked and unmarked amplitudes are given by

$$\bar{k}(t+1) = C(t) + \bar{k}(t),$$
$$\bar{l}(t+1) = C(t) - \bar{l}(t).$$

[4] The state $|\psi\rangle$ is not limited to $|\psi\rangle$ defined as Eq. (1).

The average marked and unmarked amplitudes can be expressed concisely as follows:

$$\begin{aligned} \bar{k}(t) &= \alpha \sin(\omega t + \phi) \\ \bar{l}(t) &= \beta \cos(\omega t + \phi) \end{aligned} \tag{5}$$

where

$$\begin{aligned} \omega &= \arccos\left(1 - \frac{2M}{N}\right), \\ \phi &= \arctan\left(\frac{\bar{k}(0)}{\bar{l}(0)}\sqrt{\frac{M}{N-M}}\right), \\ \alpha^2 &= \bar{k}(0)^2 + \frac{\bar{l}(0)^2 \cdot (N-M)}{M}, \\ \beta^2 &= \bar{l}(0)^2 + \frac{\bar{k}(0)^2 \cdot M}{N-M}. \end{aligned} \tag{6}$$

The time dependence of amplitudes is given by

$$\begin{aligned} k_i(t) &= \bar{k}(t) + (k_i(0) - \bar{k}(0)), \\ l_i(t) &= \bar{l}(t) + (-1)^t(l_i(0) - \bar{l}(0)). \end{aligned} \tag{7}$$

As observed above, marked amplitudes that are related with success probability depend on the marked amplitudes at time $t = 0$. According to article [8], the value of $\sum_{i=1}^{M} |k_i(t)|^2$ is maximized when t is

$$\begin{aligned} t &= \frac{\frac{\pi}{2} - \phi}{\omega} \\ &= -\frac{\bar{k}(0)}{2\bar{l}(0)} + \frac{\pi}{4}\sqrt{\frac{N}{M}} - \frac{\pi}{24}\sqrt{\frac{M}{N}} + \mathrm{O}\left(\frac{M}{N}\right). \end{aligned} \tag{8}$$

A Key Recovery Attack Using the Generalized Grover Algorithm We consider an adversary that finds the key v for a given $(\hat{c}, \hat{r})$ such that $\hat{c} = F_v(\hat{r})$. This attack does not require $|\psi\rangle$ produced with Eq. (1). Recall that F is a pseudorandom function from $\{0,1\}^\lambda \times \{0,1\}^L$ to $\{0,1\}^\ell$ where $\lambda \geq \ell$. Let G be a function defined as

$$G(v) = \begin{cases} 1 & \text{if } F_v(\hat{r}) = \hat{c} \\ 0 & \text{otherwise.} \end{cases}$$

Since the expected number of v such that $G(v) = 1$ is $2^{\lambda-\ell}$, they are denoted by $v_1, v_2, \ldots, v_{2^{\lambda-\ell}}$. Although the key $\hat{v}$ used by Alice is one of them, we consider that the attack succeeds if any of them is found. In addition, v's such that $G(v) = 0$ are denoted by $v_{2^{\lambda-\ell}+1}, v_{2^{\lambda-\ell}+2}, \ldots, v_{2^\lambda}$. Let U_G be the following unitary operator that corresponds to Eq. (4).

$$U_G \left|v\right\rangle = \begin{cases} -\left|v\right\rangle & \text{if } v \in \{v_1, v_2, \dots, v_{2^{\lambda-\ell}}\} \\ \left|v\right\rangle & \text{otherwise.} \end{cases}$$

Parameters of the generalized Grover algorithm are defined as follows:

$$\begin{aligned} N &= 2^{\lambda} \\ M &= 2^{\lambda-\ell} \\ k_i(0) &= \frac{1}{\sqrt{2^{\lambda}}} \quad \text{for } i = 1, 2, \dots, M \\ l_i(0) &= \frac{1}{\sqrt{2^{\lambda}}} \quad \text{for } i = M+1, M+2, \dots, N. \end{aligned}$$

Then, the inital state $\left|\psi\right\rangle$ is defined as

$$\begin{aligned} \left|\psi\right\rangle &= \left|\psi^{(\mathrm{m})}\right\rangle + \left|\psi^{(\mathrm{u})}\right\rangle \\ &= \sum_{i=1}^{M} k_i(0) \left|v_i\right\rangle + \sum_{i=M+1}^{N} l_i(0) \left|v_i\right\rangle, \end{aligned}$$

and the average of amplitudes are given by

$$\bar{k}(0) = \bar{l}(0) = \frac{1}{\sqrt{2^{\lambda}}}.$$

Since the probability to find the marked state is

$$\begin{aligned} \sum_{i=1}^{M} \left|k_i(t)\right|^2 &= M \left|\bar{k}(t)\right|^2 \quad (\because \text{Eq. (7)}) \\ &= M \left|\alpha \sin(\omega t + \phi)\right|^2 \quad (\because \text{Eq. (5)}), \end{aligned}$$

its maximal value is given by

$$\begin{aligned} M\alpha^2 &= M \left(\bar{k}(0)^2 + \frac{\bar{l}(0)^2 (N-M)}{M} \right) \quad (\because \text{Eq. (6)}) \\ &= 1 \end{aligned}$$

when t is

$$t = -\frac{1}{2} + \frac{\pi}{4}\sqrt{2^{\ell}} - \frac{\pi}{24}\frac{1}{\sqrt{2^{\ell}}} + \mathrm{O}\left(2^{-\ell}\right) \quad (\because \text{Eq. (8)}).$$

Thus, the number of repetitions depends on the output length ℓ, not the key length λ when $\lambda \geq \ell$. Although the success probability of this attack is close to 1, this attack is infeasible when the output length ℓ is a typical value (e.g., 256, 512).

A Left-or-Right Distingushing Attack Using the Generalized Grover Algorithm Suppose that an adversary knows that the message m is either m_1 or m_2 with $\Pr[m = m_1] = \Pr[m = m_2] = 1/2$. The goal of the adversary is to guess m for given $(\hat{c} \oplus m, \hat{r})$ and $|\psi\rangle$ with the probability larger than 1/2. Note that if $|\psi\rangle$ is not given to the adversary, this attack is the left-or-right distinguishing attack described in Sect. 3.

The following is one of the left-or-right distinguishing attacks. Since $m = m_1$ or m_2, an adversary A considers the following state as the marked state.

$$\begin{aligned} |\psi^{(\mathrm{m})}\rangle &= \sum_{i=1}^{2} k_i \, |\psi_i\rangle \\ &= \sum_{i=1}^{2} k_i \, |(\hat{c} \oplus m) \oplus m_i, \hat{r}\rangle \end{aligned}$$

where

$$k_i = \begin{cases} \frac{1}{\sqrt{2^L}} & \text{if } m_i = m \\ 0 & \text{if } m_i \neq m. \end{cases} \tag{9}$$

The unmarked state is given as

$$\begin{aligned} |\psi^{(\mathrm{u})}\rangle &= \sum_{i=3}^{2^L+1} l_i \, |\psi_i\rangle + \sum_{i=2^L+2}^{2^{L+\ell}} l_i \, |\psi_i\rangle \\ &= \sum_{\mathcal{F}\setminus(\hat{c},\hat{r})} \frac{1}{\sqrt{2^L}} \cdot |c, r\rangle + \sum_{\bar{\mathcal{F}}\setminus(\hat{c}\oplus m_1 \oplus m_2, \hat{r})} 0 \cdot |c, r\rangle \end{aligned}$$

where

$$\begin{aligned} \mathcal{F} &= \{(c, r) | c = F(\hat{v}, r) \text{ for } r \in \{0,1\}^L\} \\ \bar{\mathcal{F}} &= (\{0,1\}^\ell \times \{0,1\}^L) \setminus \mathcal{F}. \end{aligned}$$

The sum of $|\psi^{(\mathrm{m})}\rangle$ and $|\psi^{(\mathrm{u})}\rangle$ is equivalent to $|\psi\rangle$ in Eq. (1), but they only differ in the order of indexes. Let $U_{F_{\hat{v}}}$ be the following unitary operator that is available to the adversary.

$$U_{F_{\hat{v}}} = \begin{cases} -|c, r\rangle & \text{if } |c, r\rangle \in \{|\psi_1\rangle, |\psi_2\rangle\} \\ |c, r\rangle & \text{if } |c, r\rangle \in \{|\psi_3\rangle, |\psi_4\rangle, \ldots, |\psi_{2^{L+\ell}}\rangle\} \end{cases}$$

Suppose that the adversary A performs the generalized Grover algorithm using $U_{F_{\hat{v}}}$. The adversary measures the state after applying the Grover iteration t times. Let (c', r') be the measurement result. The adversary computes $m_g = c' \oplus (\hat{c} \oplus m)$. If $m_g \in \{m_1, m_2\}$ and $r' = \hat{r}$, then the adversary outputs m_g as a guessed message. Otherwise (i.e., if one of states except for $|\psi_1\rangle$ and $|\psi_2\rangle$ is observed), the adversary randomly outputs m_1 or m_2 as a guessed message m_g.

The guessed message m_g is correct (i.e., the attack succeeds) if the state $|\psi_i\rangle$ where i is the index such that $m = m_i$ is observed. Let us explain its reason. First, suppose that $m = m_1$. Then, the marked state $|\psi^{(\mathrm{m})}\rangle$ is given by

$$\begin{aligned}|\psi^{(\mathrm{m})}\rangle &= \frac{1}{\sqrt{2^L}} \cdot |\psi_1\rangle + 0 \cdot |\psi_2\rangle \\ &= \frac{1}{\sqrt{2^L}} \cdot |\hat{c}, \hat{r}\rangle + 0 \cdot |(\hat{c} \oplus m_1) \oplus m_2, \hat{r}\rangle .\end{aligned}$$

After applying the Grover iteration, if $|\psi_1\rangle$ is observed, then the adversary outputs the correct message m_1 because

$$\begin{aligned}c' \oplus (\hat{c} \oplus m) &= \hat{c} \oplus (\hat{c} \oplus m) \\ &= m_1.\end{aligned}$$

If $|\psi_2\rangle$ is observed, then the adversary outputs the incorrect message m_2 because

$$\begin{aligned}c' \oplus (\hat{c} \oplus m) &= (\hat{c} \oplus m_1 \oplus m_2) \oplus (\hat{c} \oplus m) \\ &= (\hat{c} \oplus m_1 \oplus m_2) \oplus (\hat{c} \oplus m_1) \\ &= m_2.\end{aligned}$$

Next, suppose that $m = m_2$. Then, the marked state $|\psi^{(\mathrm{m})}\rangle$ is given by

$$\begin{aligned}|\psi^{(\mathrm{m})}\rangle &= 0 \cdot |\psi_1\rangle + \frac{1}{\sqrt{2^L}} \cdot |\psi_2\rangle \\ &= 0 \cdot |(\hat{c} \oplus m_2) \oplus m_1, \hat{r}\rangle + \frac{1}{\sqrt{2^L}} \cdot |\hat{c}, \hat{r}\rangle .\end{aligned}$$

With similar considerations to the above, if $|\psi_2\rangle$ is observed, then the adversary outputs the correct message m_2. Otherwise, the adversary outputs the incorrect message m_1.

The advantage of the adversary A at time t is defined as

$$\mathrm{Adv}(A, t) = \left| \Pr\left[m = m_g\right] - \frac{1}{2} \right| .$$

Without loss of generality, we can assume that $m = m_1$ to evaluate $\mathrm{Adv}(A, t)$ because of the symmetry of $|\psi_1\rangle$ and $|\psi_2\rangle$. Then, $\mathrm{Adv}(A, t)$ is written as

$$\mathrm{Adv}(A, t) = \left| \Pr\left[m = m_g | m = m_1\right] - \frac{1}{2} \right| . \tag{10}$$

For $i = 1, 2$, let $\mathsf{E}_i(t)$ be the event such that $|\psi_i\rangle$ is observed after applying the Grover iteration t times and let $\overline{\mathsf{E}_1(t) \vee \mathsf{E}_2(t)}$ denote the complement of the event $\mathsf{E}_1(t) \vee \mathsf{E}_2(t)$. Let $k_i(t)$ denote the amplitude of $|\psi_i\rangle$ after performing the Grover iteration t times. Since $m = m_1$, $k_1(0)$ and $k_2(0)$ are given as

$$k_1(0) = \frac{1}{\sqrt{2^L}}, \quad k_2(0) = 0 \quad (\because \text{Eq. (9)}). \tag{11}$$

Since $\Pr[\mathsf{E}_i(t)] = |k_i(t)|^2$, the conditional probability of Eq. (10) is transformed as follows:

$$\begin{aligned}\Pr[m = m_g | m = m_1] &= \Pr[m = m_g | m = m_1 \wedge \mathsf{E}_1(t)] \cdot \Pr[\mathsf{E}_1(t)] \\ &+ \Pr[m = m_g | m = m_1 \wedge \mathsf{E}_2(t)] \cdot \Pr[\mathsf{E}_2(t)] \\ &+ \Pr\left[m = m_g | m = m_1 \wedge (\overline{\mathsf{E}_1(t) \vee \mathsf{E}_2(t)})\right] \cdot \Pr\left[\overline{\mathsf{E}_1(t) \vee \mathsf{E}_2(t)}\right] \\ &= 1 \cdot |k_1(t)|^2 + 0 \cdot |k_2(t)|^2 + \frac{1}{2} \cdot \left(1 - \sum_{i=1}^{2} |k_i(t)|^2\right) \\ &= \frac{1}{2} + \frac{|k_1(t)|^2 - |k_2(t)|^2}{2}.\end{aligned}$$

Hence, the advantage of A is simplified as

$$\mathrm{Adv}(A, t) = \frac{1}{2} \cdot \left| |k_1(t)|^2 - |k_2(t)|^2 \right|.$$

Supposing that $m = m_1$ without loss of generality, we evaluate

$$\mathrm{Dif}(t) = |k_1(t)|^2 - |k_2(t)|^2. \quad (12)$$

Substituting

$$\begin{aligned}\bar{k}(0) &= \frac{1}{2} \sum_{i=1}^{2} k_i(0) \\ &= \frac{1}{2\sqrt{2^L}} \quad (\because \text{Eq. (11)})\end{aligned}$$

into Eq. (7) yields

$$k_1(t) = \bar{k}(t) + \frac{1}{2\sqrt{2^L}}, \quad k_2(t) = \bar{k}(t) - \frac{1}{2\sqrt{2^L}}. \quad (13)$$

Substituting the equations above into Eq. (12) gives

$$\begin{aligned}\mathrm{Dif}(t) &= |k_1(t)|^2 - |k_2(t)|^2 \\ &= \frac{2}{\sqrt{2^L}} \left|\bar{k}(t)\right|.\end{aligned}$$

It follows that $\mathrm{Dif}(t)$ is maximized when $\left|\bar{k}(t)\right|$ is maximum.

Let us consider $\Pr[\mathsf{E}_1(t) \vee \mathsf{E}_2(t)]$ that is the probability such that either $|\psi_1\rangle$ or $|\psi_2\rangle$ is observed after performing the Grover iteration t times. In other words, $\Pr[\mathsf{E}_1(t) \vee \mathsf{E}_2(t)]$ is the probability such that the marked state is observed after performing the Grover iteration t times. The value of $\Pr[\mathsf{E}_1(t) \vee \mathsf{E}_2(t)]$ is computed as

$$\begin{aligned}\Pr\left[\mathsf{E}_1(t) \vee \mathsf{E}_2(t)\right] &= \sum_{i=1}^{2} \left|k_i(t)\right|^2 \\ &= \left|k_1(t)\right|^2 + \left|k_2(t)\right|^2 \\ &= 2\left|\bar{k}(t)\right|^2 + \frac{1}{2^{L+1}} \quad (\because \text{Eq.}\,(13)).\end{aligned}$$

The probability above is maximized when $\left|\bar{k}(t)\right|$ is maximum. Hence, the probability above and $\mathrm{Dif}(t)$ are maximized at the same t of Eq. (8), that is,

$$t = -\frac{2^{L+\ell-1}-1}{2^L-1} + \frac{\pi}{4}\sqrt{\frac{2^{L+\ell}}{2}} - \frac{\pi}{24}\sqrt{\frac{2}{2^{L+\ell}}} + \mathrm{O}\left(\frac{2}{2^{L+\ell}}\right). \tag{14}$$

The maximum value of $\mathrm{Dif}(t)$ is given by

$$\max_t \mathrm{Dif}(t) = \frac{2}{\sqrt{2^L}}\,|\alpha| \quad (\because \text{Eq.}\,(5)). \tag{15}$$

Since $\bar{l}(0)$ is computed as

$$\begin{aligned}\bar{l}(0) &= \frac{1}{2^{L+\ell}-2}\sum_{i=3}^{2^{L+\ell}} l_i(0) \\ &= \frac{2^L-1}{(2^{L+\ell}-2)\sqrt{2^L}},\end{aligned}$$

substituting $\bar{k}(0)$ and $\bar{l}(0)$ into Eq. (6) gives

$$\begin{aligned}\alpha^2 &= \frac{1}{4\cdot 2^L} + \frac{\frac{(2^L-1)^2}{(2^{L+\ell}-2)^2 2^L}\cdot(2^{L+\ell}-2)}{2} \\ &= \frac{1}{2^{L+2}} + \frac{(2^L-1)^2}{2^{L+1}(2^{L+\ell}-2)}.\end{aligned} \tag{16}$$

Since the maximum value of $\mathrm{Dif}(t)$ is given by

$$\max_t \mathrm{Dif}(t) = \frac{2}{\sqrt{2^L}}\sqrt{\frac{1}{2^{L+2}} + \frac{(2^L-1)^2}{2^{L+1}(2^{L+\ell}-2)}} \quad (\because \text{Eqs.}\,(15), (16)),$$

we obtain the maximum advantage of A as follows:

$$\max_t \mathrm{Adv}(A,t) = \frac{1}{\sqrt{2^L}}\sqrt{\frac{1}{2^{L+2}} + \frac{(2^L-1)^2}{2^{L+1}(2^{L+\ell}-2)}}.$$

When L and ℓ are typical values (e.g., 256, 512), the advantage of A is negligibly small and t of Eq. (14) is too large to perform the attack.

5 Conclusions

White-box (classical) cryptography with provable security has been proposed recently. There exists a problem in the constructions of white-box cryptography: the size of memory is larger than that of usual symmetric ciphers. To address this problem, we have shown the XORQS[F] scheme that uses quantum memory in white-box cryptography. The size of qubits required by the XORQS[F] scheme is a polynomial in the sum of input-length and output-length of the underlying function.

We have analyzed the confidentiality of the XORQS[F] scheme against both of classical attacks and quantum attacks using the generalized Grover algorithm. The left-or-right distinguishing attack using the generalized Grover algorithm is not efficient in the sense that the success probability is small. This is because amplitudes of the initial state of the XORQS[F] scheme is not uniform.

Acknowledgments This work was supported by JSPS KAKENHI Grant Number 20K21798.

References

1. C. Delerablée, T. Lepoint, P. Paillier, and M. Rivain, "White-box security notions for symmetric encryption schemes," Selected Areas in Cryptography – SAC 2013, ed. T. Lange, K. Lauter, and P. Lisoněk, Berlin, Heidelberg, pp.247–264, 2014.
2. A. Biryukov, C. Bouillaguet, and D. Khovratovich, "Cryptographic schemes based on the ASASA structure: Black-box, white-box, and public-key," Advances in Cryptology – ASIACRYPT 2014, pp.63–84, 2014.
3. A. Bogdanov and T. Isobe, "White-box cryptography revisited: Space-hard ciphers," Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, pp.1058–1069, 2015.
4. P.A. Fouque, P. Karpman, P. Kirchner, and B. Minaud, "Efficient and provable white-box primitives," Advances in Cryptology - ASIACRYPT 2016, Lecture Notes in Computer Science, vol.10031, pp.159–188, 2016.
5. C.H. Bennett and G. Brassard, "Quantum cryptography: Public key distribution and coin tossing," Proceedings of IEEE International Conference on Computers Systems and Signal Processing, pp.175–179, 1984.
6. M. Bellare, A. Desai, E. Jokipii, and P. Rogaway, "A concrete security treatment of symmetric encryption: Analysis of the DES modes of operation," `http://www-cse.ucsd.edu/users/mihir/papers/sym-enc.html`, pp.1–31, 2000.
7. L.K. Grover, "A fast quantum mechanical algorithm for database search," Proceedings of The 28th ACM Symposium on the Theory of Computing, pp.212–219, 1996.
8. D. Biron, O. Biham, E. Biham, M. Grassl, and D.A. Lidar, "Generalized Grover search algorithm for arbitrary initial amplitude distribution." arXiv: quant-ph/9801066, 1998.

A Proof of Theorem 1

In this Appendix, the key-length λ of F is replaced with k to obtain consistency with notations of article [6]. Note that it differs from the definition of k in Sect. 4.2. Accordingly, F is a family of functions: $\{0,1\}^k \times \{0,1\}^L \to \{0,1\}^\ell$.

A scheme XOR$[F]$ is defined as follows: Alice and Bob have shared a k-bit key v that was chosen according to the uniform distribution on $\{0,1\}^k$. Let $f = F(v, \cdot)$. Since f is the function from $\{0,1\}^L$ to $\{0,1\}^\ell$, Bob chooses a nonce r according to the uniform distribution on $\{0,1\}^L$. For an ℓ-bit message m, Bob sends a ciphertext $(c, r) = (f(r) \oplus m, r)$. The decryption done by Alice is omitted.

Let us consider the following experiment $\mathrm{Exp}^{\text{lor-cpa-}b}_{\mathrm{XOR}[F],A}(k)$ for XOR$[F]$. Let b be a bit chosen according to the uniform distribution on $\{0,1\}$. An adversary A makes queries a pair of ℓ-bit plaintexts $(m^{(0)}, m^{(1)})$ to a left-or-right encryption oracle that returns a ciphertext as $(f(r) \oplus m^{(b)}, r)$. Eventually, if A guesses $b = 1$, A outputs $d = 1$. Otherwise A outputs $d = 0$. The output of the experiment is d. The advantage of the adversary A is defined by

$$\begin{aligned}&\mathrm{Adv}^{\text{lor-cpa}}_{\mathrm{XOR}[F],A}(k)\\ &= \Pr\left[\mathrm{Exp}^{\text{lor-cpa-1}}_{\mathrm{XOR}[F],A}(k) = 1\right] - \Pr\left[\mathrm{Exp}^{\text{lor-cpa-0}}_{\mathrm{XOR}[F],A}(k) = 1\right].\end{aligned}$$

The advantage function of XOR$[F]$ is defined by

$$\mathrm{Adv}^{\text{lor-cpa}}_{\mathrm{XOR}[F]}(k, t, q_e, \mu_e) = \max_A \mathrm{Adv}^{\text{lor-cpa}}_{\mathrm{XOR}[F],A}(k)$$

where t is time complexity, q_e is the number of queries to the left-or-right encryption oracle, and μ_e is the amount of ciphertexts.

Let R be the set of all the functions: $\{0,1\}^L \to \{0,1\}^\ell$. When f is chosen according to the uniform distribution on R, a scheme XOR$[R]$ can be constructed in a way similar to XOR$[F]$. The advantage of the adversary A, $\mathrm{Adv}^{\text{lor-cpa}}_{\mathrm{XOR}[R],A}(\cdot)$, and the advantage function of XOR$[R]$, $\mathrm{Adv}^{\text{lor-cpa}}_{\mathrm{XOR}[R]}(\cdot, t, q_e, \mu_e)$, are defined in a similar way. The argument on the key-length are unnecessary (denoted by '$\cdot$') because R is not a family of keyed functions. Lemma 1, which is a simplified version of Lemma 10 in article [6], shows the upper bound of the advantage function of XOR$[R]$.

Lemma 1. *For any t, q_e, and $\mu_e = q_e\ell$, the following inequality holds.*

$$\mathrm{Adv}^{\text{lor-cpa}}_{\mathrm{XOR}[R]}(\cdot, t, q_e, \mu_e) \le \frac{q_e(q_e-1)}{2^{\ell+1}} \tag{17}$$

Since R is not a keyed function, the first argument is written in the symbol '$\cdot$'.

Proof. Consider an adversary A with time complexity t, the number of queries q_e, and the amount of ciphertexts μ_e. Let r_i be the nonce that is used by the left-or-right encryption oracle for the i-th query $(m_i^{(0)}, m_i^{(1)})$. When b is fixed, let $\mathrm{Col}^{\text{lor-cpa-}b}$ be an event that there exists a collision such that $r_i = r_j$ where $i \neq j$.

The complementary event of $\mathrm{Col}^{\text{lor-cpa-}b}$ (i.e., there is no collision) is denoted by $\overline{\mathrm{Col}}^{\text{lor-cpa-}b}$. Since the choice of r_i is independent of that of b and queries of A, the equalities below hold.

$$\Pr\left[\mathrm{Col}^{\text{lor-cpa-1}}\right] = \Pr\left[\mathrm{Col}^{\text{lor-cpa-0}}\right], \tag{18}$$

$$\Pr\left[\overline{\mathrm{Col}}^{\text{lor-cpa-1}}\right] = \Pr\left[\overline{\mathrm{Col}}^{\text{lor-cpa-0}}\right]. \tag{19}$$

Since the distribution of $(f(r_i) \oplus m_i^{(1)}, r_i)$ is the same of that of $(f(r_i) \oplus m_i^{(0)}, r_i)$ if there is no collision in nonces, the following equation holds.

$$\begin{aligned} &\Pr\left[\mathrm{Exp}^{\text{lor-cpa-1}}_{\mathrm{XOR}[R],A}(\cdot) = 1 \middle| \overline{\mathsf{Col}}^{\text{lor-cpa-1}}\right] \\ &\quad = \Pr\left[\mathrm{Exp}^{\text{lor-cpa-0}}_{\mathrm{XOR}[R],A}(\cdot) = 1 \middle| \overline{\mathsf{Col}}^{\text{lor-cpa-0}}\right]. \end{aligned} \tag{20}$$

The advantage of the adversary A is evaluated as follows:

$$\begin{aligned} &\mathrm{Adv}^{\text{lor-cpa}}_{\mathrm{XOR}[R],A}(\cdot) \\ &\quad = \Pr\left[\mathrm{Exp}^{\text{lor-cpa-1}}_{\mathrm{XOR}[R],A}(\cdot) = 1\right] - \Pr\left[\mathrm{Exp}^{\text{lor-cpa-0}}_{\mathrm{XOR}[R],A}(\cdot) = 1\right] \end{aligned}$$

where for $b = 1, 0$,

$$\begin{aligned} &\Pr\left[\mathrm{Exp}^{\text{lor-cpa-}b}_{\mathrm{XOR}[R],A}(\cdot) = 1\right] \\ &= \Pr\left[\mathrm{Exp}^{\text{lor-cpa-}b}_{\mathrm{XOR}[R],A}(\cdot) = 1 \middle| \mathsf{Col}^{\text{lor-cpa-}b}\right] \Pr\left[\mathsf{Col}^{\text{lor-cpa-}b}\right] \\ &\quad + \Pr\left[\mathrm{Exp}^{\text{lor-cpa-}b}_{\mathrm{XOR}[R],A}(\cdot) = 1 \middle| \overline{\mathsf{Col}}^{\text{lor-cpa-}b}\right] \Pr\left[\overline{\mathsf{Col}}^{\text{lor-cpa-}b}\right]. \end{aligned} \tag{21}$$

Substituting Eqs. (18)–(20) into Eq. (21) gives

$$\begin{aligned} \mathrm{Adv}^{\text{lor-cpa}}_{\mathrm{XOR}[R],A}(\cdot) &= \Big(\Pr\left[\mathrm{Exp}^{\text{lor-cpa-1}}_{\mathrm{XOR}[R],A}(\cdot) = 1 \middle| \mathsf{Col}^{\text{lor-cpa-1}}\right] \\ &\quad - \Pr\left[\mathrm{Exp}^{\text{lor-cpa-0}}_{\mathrm{XOR}[R],A}(\cdot) = 1 \middle| \mathsf{Col}^{\text{lor-cpa-0}}\right]\Big) \\ &\quad \cdot \Pr\left[\mathsf{Col}^{\text{lor-cpa-1}}\right] \\ &\le \Pr\left[\mathsf{Col}^{\text{lor-cpa-1}}\right] \\ &\le \frac{q_e(q_e-1)}{2^{\ell+1}}. \end{aligned}$$

The discussion above depends on only t and q_e of A, and does not depend on the algorithm of A. The value of μ_e is equal to $q_e\ell$ regardless of b. Hence, Eq. (17) holds.

Let us consider the following experiment $\mathrm{Exp}^{\text{prf-}b}_{F,D}$. Determine b according to the uniform distribution on $\{0, 1\}$. If $b = 0$, then a function f is chosen according to the uniform distribution on R. Otherwise, f is chosen according to the uniform distribution on $\{0,1\}^k$ (i.e., $f = F(v,\cdot)$). For a query $x \in \{0,1\}^L$ of a distinguisher D, an oracle returns $f(x)$. Eventually, if D guesses $b = 0$, then D outputs $d = 0$. Otherwise D outputs $d = 1$. The output of the experiment is d. The advantage of D is defined by

$$\mathrm{Adv}^{\mathrm{prf}}_{F,D} = \Pr\left[\mathrm{Exp}^{\text{prf-1}}_{F,D} = 1\right] - \Pr\left[\mathrm{Exp}^{\text{prf-0}}_{F,D} = 1\right]$$

and the advantage function of F is defined by

$$\mathrm{Adv}^{\mathrm{prf}}_{F}(t,q) = \max_{D} \mathrm{Adv}^{\mathrm{prf}}_{F,D}$$

where t is time complexity of D and q is the number of queries to the oracle. The notation of $\mathrm{Adv}^{\mathrm{prf}}_{F}$ does not explicitly take the key-length k of F as a parameter. It means that the resource depending on k is included by t or q_e. Theorem 1 is proven below.

Proof. Consider the experiment $\mathrm{Exp}^{\text{prf-}b}_{F,D}$, that is, after the distinguisher D makes queries to the oracle f, D guesses the value b. Suppose that there exists an adversary A with the advantage $\mathrm{Adv}^{\text{lor-cpa}}_{\mathrm{XOR}[F],A}(k)$. The algorithm of D is as follows:

1. Determine b_D according to the uniform distribution on $\{0,1\}$.
2. Run A. When A makes a query $(m^{(0)}, m^{(1)})$, D chooses r according to the uniform distribution on $\{0,1\}^\ell$. After D makes a query r to the oracle f, D obtains $f(r)$ as the answer of the oracle. Then, D returns $(f(r) \oplus m^{(b_D)}, r)$ to A.
3. Eventually, A outputs b_A as the guess of b_D. Output 1 as the guess of b if $b_D = b_A$. Otherwise, output 0.

The time complexity of D, t, is equal to that of A, the number of queries by D, q_e, is that by A, and the amount of ciphertexts obtained by A is $\mu_e = q_e\ell$. The advantage of D is given by

$$\begin{aligned}\mathrm{Adv}^{\mathrm{prf}}_{F,D} &= \Pr\left[\mathrm{Exp}^{\text{prf-1}}_{F,D} = 1\right] - \Pr\left[\mathrm{Exp}^{\text{prf-0}}_{F,D} = 1\right] \\ &= \Pr\left[b_D = b_A | b = 1\right] - \Pr\left[b_D = b_A | b = 0\right]. \end{aligned} \tag{22}$$

The second term above is transformed as follows:

$$\begin{aligned}&\Pr\left[b_D = b_A | b = 0\right] \\ &= \Pr\left[b_D = b_A | b = 0 \wedge b_D = 1\right] \cdot \Pr\left[b_D = 1\right] \\ &\quad + \Pr\left[b_D = b_A | b = 0 \wedge b_D = 0\right] \cdot \Pr\left[b_D = 0\right] \\ &= \frac{1}{2}\left(\Pr\left[b_D = b_A | b = 0 \wedge b_D = 1\right]\right. \\ &\quad \left.+\Pr\left[b_D = b_A | b = 0 \wedge b_D = 0\right]\right) \\ &= \frac{1}{2}\left(\Pr\left[\mathrm{Exp}^{\text{lor-cpa-1}}_{\mathrm{XOR}[R],A} = 1\right] + \Pr\left[\mathrm{Exp}^{\text{lor-cpa-0}}_{\mathrm{XOR}[R],A} = 0\right]\right) \\ &= \frac{1}{2}\left(1 + \Pr\left[\mathrm{Exp}^{\text{lor-cpa-1}}_{\mathrm{XOR}[R],A} = 1\right] - \Pr\left[\mathrm{Exp}^{\text{lor-cpa-0}}_{\mathrm{XOR}[R],A} = 1\right]\right) \\ &= \frac{1}{2}\left(1 + \mathrm{Adv}^{\text{lor-cpa}}_{\mathrm{XOR}[R],A}(\cdot)\right) \end{aligned} \tag{23}$$

where '$\cdot$' means that R is not a keyed function. The first term is transformed in a similar way.

$$\Pr\left[b_D = b_A | b = 1\right] = \frac{1}{2}\left(1 + \mathrm{Adv}^{\text{lor-cpa}}_{\mathrm{XOR}[F],A}(\cdot)\right) \tag{24}$$

Since there is no asymptotics on the key-length, the first argument in the equation above is written in the symbol '$\cdot$'. Substituting Eq. (23) and Eq. (24) into Eq. (22) gives

$$\mathrm{Adv}^{\mathrm{prf}}_{F,D} = \frac{1}{2}\left(\mathrm{Adv}^{\text{lor-cpa}}_{\mathrm{XOR}[F],A}(\cdot) - \mathrm{Adv}^{\text{lor-cpa}}_{\mathrm{XOR}[R],A}(\cdot)\right).$$

From the definition, the following inequality holds.

$$\mathrm{Adv}^{\mathrm{prf}}_{F}(t, q_e) \geq \frac{1}{2}\left(\mathrm{Adv}^{\text{lor-cpa}}_{\mathrm{XOR}[F],A}(\cdot) - \mathrm{Adv}^{\text{lor-cpa}}_{\mathrm{XOR}[R],A}(\cdot)\right)$$

Applying Lemma 1 to the inequality above yields

$$\mathrm{Adv}^{\mathrm{prf}}_{F}(t, q_e) \geq \frac{1}{2}\left(\mathrm{Adv}^{\text{lor-cpa}}_{\mathrm{XOR}[F],A}(\cdot) - \frac{q_e(q_e-1)}{2^{\ell+1}}\right),$$

that is,

$$\mathrm{Adv}^{\text{lor-cpa}}_{\mathrm{XOR}[F],A}(\cdot) \leq 2\mathrm{Adv}^{\mathrm{prf}}_{F}(t, q_e) + \frac{q_e(q_e-1)}{2^{\ell+1}}.$$

The discussion above does not depend on an algorithm of A, that is, it only depends on resources required by A. When t, q_e, and μ_e denote the time complexity, the number of queries to the oracle of XOR$[F]$ and the amount of ciphertexts of A, respectively, the inequality below holds.

$$\mathrm{Adv}^{\text{lor-cpa}}_{\mathrm{XOR}[F]}(\cdot, t, q_e, \mu_e) \leq 2\mathrm{Adv}^{\mathrm{prf}}_{F}(t, q_e) + \frac{q_e(q_e-1)}{2^{\ell+1}}$$