From Plaintext-extractability to IND-CCA Security

Ehsan Ebrahimi

SnT & Department of Computer Science, University of Luxembourg

Abstract. We say a public-key encryption is plaintext-extractable in the random oracle model if there exists an algorithm that given access to all inputs/outputs queries to the random oracles can simulate the decryption oracle. We argue that plaintext-extractability is enough to show the indistinguishably under chosen ciphertext attack (IND-CCA) of OAEP+ transform (Shoup, Crypto 2001) when the underlying trapdoor permutation is one-way.
We extend the result to the quantum random oracle model (QROM) and show that OAEP+ is IND-CCA secure in QROM if the underlying trapdoor permutation is quantum one-way.
Keywords. Post-quantum Security, OAEP+, Quantum Random Oracle Model

1 Introduction

The OAEP transform was proposed by Bellare and Rogaway [4] to transfer a trapdoor permutation into a public-key encryption scheme using two random oracles. It was believed that the OAEP-cryptosystem is provable secure in the random oracle model based on the one-wayness of trapdoor permutation, but Shoup [14] showed it is an unjustified belief.

In [4], the technique to show the IND-CCA security is introducing a notion of plaintext-awareness (PA1) that guarantees the existence of an algorithm Ext that can simulate the decryption oracle given access to the inputs/outputs of the random oracles. Then, given this extractor Ext, we may reduce the IND-CCA security to the one-wayness of the underlying trapdoor permutation. A stronger definition of plaintext-awareness (PA2) was introduced in the random oracle model [1], in which, the adversary is able to eavesdrop some valid ciphertexts (through an oracle $\mathcal{E}^H_{\mathsf{pk}}$) and the extractor, given access to these ciphertexts and the random oracle queries made by the adversary (and not the random oracle queries used by $\mathcal{E}^H_{\mathsf{pk}}$), should be able to decrypt any ciphertext outputted by the adversary. In [1], it was shown that an encryption scheme that is PA2 and IND-CPA secure, is IND-CCA secure.

However, Shoup [14] argued that PA1 might not be sufficient to show the IND-CCA security of OAEP because the adversary might be able to turn the challenge ciphertext c^* into a new valid ciphertext for which Ext is not able to decrypt (since Ext does not have access to the random oracles queries used

to obtain c^*). And it has not been proven that the OAEP transform is PA2. Therefore, the IND-CCA security of OAEP under the one-wayness assumption of the underlying permutation remains an open question. Shoup [14] even made an argument that the existence of a IND-CCA security proof is unlikely under the one-wayness assumption. Alternatively, Shoup [14] presented the OAEP+ transform along with a IND-CCA security proof based on the one-wayness of the permutation.

The IND-CCA security of the OAEP transform was proven in [11], however, based on a stronger assumption, namely, the partial-domain one-wayness of the underlying permutation. This result is extended to the quantum random oracle model [9, 15] under the quantum partial-domain one-wayness of the underlying permutation. Since in the real world applications, a random oracle will be substituted with a cryptographic hash function and the code of this hash function is public, to claim the post-quantum security, one needs to prove the security in the quantum random oracle model in which a quantum adversary is able to make superposition queries to the random oracles. To date, the post-quantum security of OAEP+ has not been investigated. In fact, this post-quantum security proof is needed since the existence of a quantum partial-domain one-way trapdoor permutation implies the existence of a quantum one-way trapdoor permutation and not other way around. To use the result in [9,15], one needs a quantum-secure trapdoor permutation with a stronger security requirement than the quantum one-wayness.

Note that this has not been problematic so far since this does not affect the instantiation of OAEP with the RSA function (RSA-OAEP) [4]. In more details, since partial-domain one-wayness of the RSA function is equivalent to its (full-domain) one-wayness, it follows that the security of RSA–OAEP can actually be proven under the sole RSA assumption [11]. However, RSA assumption does not hold in the post-quantum setting due to Shor's quantum algorithm [13]. And we are not aware of a quantum-hard assumption for which these two security definitions (partial-domain one-wayness and one-wayness assumptions) are equivalent. So we need a quantum partial-domain one-way trapdoor permutation to use in the OAEP transform. In contrast, if the post-quantum security of OAEP+ exists, we can use a quantum one-way trapdoor permutation that is a weaker assumption.

In this paper, we fill this gap. We show that OAEP+ is secure in the quantum random oracle model. Our proof technique is to define a notion of plaintext-extractability, show that OAEP+ is plaintext-extractable and use it to prove IND-CCA security in the quantum random oracle model.

OAEP+ Transform. We informally present how OAEP+ encrypts a message m. Let G, H, H' be random oracles and f be a trapdoor permutation. To encrypt m, it chooses a random element r and computes a ciphertext c as follows:

$$s = \underbrace{(G(r) \oplus m)}_{[s]^n} \| \underbrace{H'(r, m)}_{[s]_{k_1}}, \ t = r \oplus H(s), \ c = f(s, t).$$

1.1 Our Contribution

We investigate the security of OAEP+ in the quantum random oracle model. We define a notion of plaintext-extractability in the (quantum) random oracle model. Our notion is different from PA1 since the adversary is given the possibility of eavesdropping some valid ciphertexts (through an oracle $\mathcal{E}_{pk}^{H}$) in contrast to PA1. It is not PA2 either because the extractor Ext is given access to the inputs/outputs of all queries to the random oracles (including random oracle queries used by $\mathcal{E}_{pk}^{H}$) in contrast to PA2.

We informally discuss why our plaintext-extractability notion is sufficient to prove the IND-CCA security of OAEP+ under the one-wayness of the underlying permutation. Our argument is classical but it would be extended to the quantum random oracle model in Section 4.

We start with IND-CCA game (**Game 0**) in which the adversary given access to the random oracles and decryption oracle outputs two messages m_0, m_1. The challenger chooses a random bit b and encrypts m_b and sends this challenge ciphertext c^* to the adversary. The adversary is allowed to make decryption queries, except for the challenge ciphertext, and random oracle queries. Finally, the adversary outputs a bit b' and wins if $b' = b$.

Then, we define a game (**Game 1**) in which the challenger instead of using the secret key (f^{-1}) to answer decryption queries, it uses the extractor algorithm Ext. Note that in this game, the challenger simulates the decryption queries, therefore, Ext has access to all queries to the random oracles. The indistinguishably of these two games hold due to the plaintext-extractability of OAEP+.

We define another game (**Game 2**) in which the challenger aborts and return a random bit if the adversary submits the randomness r^* (that has been used to compute c^*) as a query to either G or H'. Obviously, the probability of the abort event is negligible before the challenge query since r^* has not been used yet. We show that if r^* is queried after the challenge query, this breaks the one-wayness of the underlying permutation.

Let $\mathcal{A}$ be an adversary that distinguishes **Game 1** and **Game 2**, that is, $\mathcal{A}$ queries r^* as a post-challenge query with a non-negligible probability. Now it comes to the reduction adversary $\mathcal{B}$. Note that the input of the adversary $\mathcal{B}$ is a value c^* that is an image of f on some random values s^*, t^*. In other words, c^* is generated without any queries to the random oracles. Therefore, the adversary $\mathcal{B}$ chooses random oracles G, H and H', a random bit b, runs $\mathcal{A}$ and answers to its decryption queries using Ext. Upon receiving the challenge query m_0, m_1 from $\mathcal{A}$, the adversary sends c^* as the challenge query.

The adversary $\mathcal{B}$ guesses, randomly, the first query in which the randomness r^* (that is used to generate c^*) will be submitted to G or H'. Therefore, the adversary $\mathcal{B}$ can find r^* with a non-negligible probability. It outputs $s^* := (G(r^*) \oplus m_b) \| H'(r^*, m_b)$ and $t^* := r^* \oplus H(s^*)$ as the pre-image of f on c^*.

Back to **Game 2**, when r^* is not submitted as a query to G and H', the values of $G(r^*)$ and $H'(r^*, m_b)$ are distributed uniformly at random. Therefore, the adversary is able to guess b only with a probability of $1/2$ and this finishes the proof.

Difference with the implication PA2+IND-CPA $\implies$ IND-CCA [1, 2]. Note that in the sketch above, the extractor algorithm knows all the random oracle queries and in the reduction, the adversary $\mathcal{B}$ possess a value that is obtained by computing f on some random values s^*, t^* and without any random oracle queries. Therefore, we do not need the strong security requirement PA2 to conclude the IND-CCA security. This is of course different when we want to show IND-CCA security from IND-CPA and a plaintext-awareness notion. In this general implication, the reduction adversary $\mathcal{B}$ attacking CPA security gets its challenge ciphertext c^* through an encryption oracle (the challenger of CPA game) and the adversary $\mathcal{B}$ does not know the random oracle queries that have been used to compute c^*. For this general implication, indeed, we need PA2 notion in which the extractor Ext is not given access to the random oracle queries used to compute the challenge ciphertext c^*. However, in our case, c^* is generated without making any random oracle query. Note that both the actual decryption algorithm and the extractor return $\perp$ if c^* is submitted as a decryption query.

2 Preliminaries

Notations. The notation $x \xleftarrow{\$} X$ means that x is chosen uniformly at random from the set X. For a natural number n, $[n]$ means the set $\{1, \cdots, n\}$. $\Pr[P : G]$ is the probability that the predicate P holds true where free variables in P are assigned according to the program in G. The function $\mathsf{negl}(\eta)$ is any non-negative function that is smaller than the inverse of any non-negative polynomial $p(\eta)$ for sufficiently large η. For a function f, f_x denotes the evaluation of f on the input x, that is $f(x)$. For a bit-string x of size more-than-equal k, $[x]_k$ are the k least significant bits of x and $[x]^k$ are the k most significant bits of x. For two bits b and b', $[b = b']$ is 1 if $b = b'$ and it is 0 otherwise. QPT is a quantum polynomial-time algorithm.

2.1 Quantum Computing

We present basics of quantum computing in this subsection. The interested reader can refer to [12] for more information. For two vectors $|\Psi\rangle = (\psi_1, \psi_2, \cdots, \psi_n)$ and $|\Phi\rangle = (\phi_1, \phi_2, \cdots, \phi_n)$ in $\mathbb{C}^n$, the inner product is defined as $\langle \Psi, \Phi \rangle = \sum_i \psi_i^* \phi_i$ where ψ_i^* is the complex conjugate of ψ_i. Norm of $|\Phi\rangle$ is defined as $\| |\Phi\rangle \| = \sqrt{\langle \Phi, \Phi \rangle}$. The n-dimensional Hilbert space $\mathcal{H}$ is the complex vector space $\mathbb{C}^n$ with the inner product defined above. A quantum system is a Hilbert space $\mathcal{H}$ and a quantum state $|\psi\rangle$ is a vector $|\psi\rangle$ in $\mathcal{H}$ with norm 1. A unitary operation over $\mathcal{H}$ is a transformation $\mathbb{U}$ such that $\mathbb{U}\mathbb{U}^\dagger = \mathbb{U}^\dagger\mathbb{U} = \mathbb{I}$ where $\mathbb{U}^\dagger$ is the Hermitian transpose of $\mathbb{U}$ and $\mathbb{I}$ is the identity operator over $\mathcal{H}$. Norm of an operator $\mathbb{U}$ is $\|\mathbb{U}\| = \max_{|\psi\rangle} \|\mathbb{U}|\psi\rangle\|$. The computational basis for $\mathcal{H}$ consists of $\log n$ vectors $|b_i\rangle$ of length $\log n$ with 1 in the position i and 0 elsewhere.

An orthogonal projection $\mathbb{P}$ over $\mathcal{H}$ is a linear transformation such that $\mathbb{P}^2 = \mathbb{P} = \mathbb{P}^\dagger$. A measurement on a Hilbert space is defined with a family of projectors that are pairwise orthogonal. An example of measurement is the

computational basis measurement in which any projection is defined by a basis vector. The output of computational measurement on a state $|\Psi\rangle$ is i with probability $\|\langle b_i, \Psi\rangle\|^2$ and the post measurement state is $|b_i\rangle$. For a general measurement $\{\mathbb{P}_i\}_i$, the output of this measurement on a state $|\Psi\rangle$ is i with probability $\|\mathbb{P}_i |\Psi\rangle\|^2$ and the post measurement state is $\frac{\mathbb{P}_i|\Psi\rangle}{\|\mathbb{P}_i|\Psi\rangle\|}$.

For two operators $\mathbb{U}_1$ and $\mathbb{U}_2$, the commutator is $[\mathbb{U}_1, \mathbb{U}_2] = \mathbb{U}_1\mathbb{U}_2 - \mathbb{U}_2\mathbb{U}_1$. For two quantum systems $\mathcal{H}_1$ and $\mathcal{H}_2$, the composition of them is defined by the tensor product and it is $\mathcal{H}_1 \otimes \mathcal{H}_2$. For two unitary $\mathbb{U}_1$ and $\mathbb{U}_2$ defined over $\mathcal{H}_1$ and $\mathcal{H}_2$ respectively, $(\mathbb{U}_1 \otimes \mathbb{U}_2)(\mathcal{H}_1 \otimes \mathcal{H}_2) = \mathbb{U}_1(\mathcal{H}_1) \otimes \mathbb{U}_2(\mathcal{H}_2)$. In this paper, QFT over an n-qubits system is $\mathbb{H}^{\otimes n}$.

If a system is in the state $|\Psi_i\rangle$ with the probability p_i, we interpret this with a quantum ensemble $E = \{(|\Psi_i\rangle, p_i)\}_i$. Different outputs of a quantum algorithm can be represented as a quantum ensemble. The density operator corresponding with the ensemble E is $\rho = \sum_i p_i |\Psi_i\rangle\langle\Psi_i|$ where $|\Psi_i\rangle\langle\Psi_i|$ is the operator acting as $|\Psi_i\rangle\langle\Psi_i| : |\Phi\rangle \to \langle \Psi_i, \Phi\rangle |\Psi_i\rangle$. The trace distance of two density operators ρ_1, ρ_2 is defined as $\mathrm{TD}(\rho_1, \rho_2) := \frac{1}{2}\operatorname{tr}|\rho_1 - \rho_2|$ where tr is the trace of a square matrix (the sum of entries on the main diagonal) and $|\rho_1 - \rho_2| := \sqrt{(\rho_1 - \rho_2)^\dagger(\rho_1 - \rho_2)}$. Note that the trace distance of two pure states $|\Psi\rangle, |\Phi\rangle$ is defined as $\mathrm{TD}(|\Psi\rangle\langle\Psi|, |\Phi\rangle\langle\Phi|)$.

Any classical function $f : X \to Y$ can be implemented as a unitary operator $\mathbb{U}_f$ in a quantum computer where $\mathbb{U}_f : |x, y\rangle \to |x, y \oplus f(x)\rangle$ and it is clear that $\mathbb{U}_f^\dagger = \mathbb{U}_f$. A quantum adversary has standard oracle access to a classical function f if it can query the unitary $\mathbb{U}_f$.

2.2 Definitions

We define a public-key encryption scheme, the IND-CCA security notion in the quantum random oracle model and the quantum (partial-domain) one-wayness.

Definition 1. *A scheme $\mathcal{E}$ with three polynomial-time (in the security parameter η) algorithms* $\mathrm{Gen}, \mathrm{Enc}, \mathrm{Dec}$) *is called a public-key encryption scheme if:*

1. *The key generation algorithm* Gen *is a probabilistic algorithm which on input 1^η outputs a pair of keys, $(\mathsf{pk}, \mathsf{sk}) \leftarrow \mathrm{Gen}(1^\eta)$, called the public key and the secret key for the encryption scheme, respectively.*
2. *The encryption algorithm* Enc *is a probabilistic algorithm which takes as input a public key pk and a message m and outputs a ciphertext $c \leftarrow \mathrm{Enc}_{\mathsf{pk}}(m)$.*
3. *The decryption algorithm is a deterministic algorithm that takes as input a secret key sk and a ciphertext c and returns the message $m := \mathrm{Dec}_{\mathsf{sk}}(c)$. It is required that the decryption algorithm returns the original message, i.e., $\mathrm{Dec}_{\mathsf{sk}}(\mathrm{Enc}_{\mathsf{pk}}(m)) = m$, for every $(\mathsf{pk}, \mathsf{sk}) \leftarrow \mathrm{Gen}(1^\eta)$ and every m. The algorithm* Dec *returns $\perp$ if ciphertext c is not decryptable.*

In the following, we define the IND-CCA security notion in the quantum random oracle model. The IND-CCA security notion for a public-key encryption scheme allows the adversary to make quantum random oracle queries but the

challenge query and decryption queries are classical. We define Dec' as:

$$\mathrm{Dec}'(c) \to \begin{cases} \perp & \text{if } c^* \text{ is defined} \quad \wedge \quad c = c^* \\ \mathrm{Dec}_{sk}(c) & \text{otherwise} \end{cases},$$

where c^* is the challenge ciphertext and $\perp$ is a value outside of the output space. We say that a quantum algorithm $\mathcal{A}$ has quantum access to the random oracle H if $\mathcal{A}$ can submit queries in superposition and the oracle H answers to these queries by applying a unitary transformation that maps $|x,\ y\rangle$ to $|x,\ y \oplus H(x)\rangle$.

Definition 2 (IND-CCA in the quantum random oracle model). *A public-key encryption scheme $\Pi = (\mathrm{Gen}, \mathrm{Enc}, \mathrm{Dec})$ is IND-CCA secure if for any* QPT *adversary $\mathcal{A}$*

$$\Pr\left[b = 1 : b \leftarrow \mathsf{Exp}^{CCA,qRO}_{\mathcal{A},\mathcal{E}}(\eta)\right] \leq 1/2 + \mathit{negl}(\eta),$$

where $\mathsf{Exp}^{CCA,qRO}_{\mathcal{A},\mathcal{E}}(\eta)$ game is define as:
$\underline{\mathsf{Exp}^{CCA,qRO}_{\mathcal{A},\mathcal{E}}(\eta)}$ *game:*
Key Gen: *The challenger runs* $\mathrm{Gen}(1^\eta)$ *to obtain a pair of keys* $(\mathsf{pk}, \mathsf{sk})$ *and chooses random oracles.*

Query: *The adversary $\mathcal{A}$ given the public key* pk*, the oracle access to* Dec' *and the* ***quantum*** *access to the random oracles, chooses two* ***classical*** *messages m_0, m_1 of the same length and sends them to the challenger. The challenger chooses a random bit b and responds with $c^* \leftarrow \mathrm{Enc}_{\mathsf{pk}}(m_b)$.*

Guess: *The adversary $\mathcal{A}$ continues to query the decryption oracle and the random oracles. Finally, the adversary $\mathcal{A}$ produces a bit b'. The output of the game is $[b = b']$.*

Definition 3 (Quantum one-way function). *We say a permutation $f : \{0,1\}^{n+k_1} \times \{0,1\}^{k_0} \to \{0,1\}^m$ is quantum one-way if for any* QPT *adversary A,*

$$\Pr\left[(\tilde{s}, \tilde{t}) = (s,t) : s \xleftarrow{\$} \{0,1\}^{n+k_1},\ t \xleftarrow{\$} \{0,1\}^{k_0},\ (\tilde{s}, \tilde{t}) \leftarrow A(f(s,t))\right] \leq \mathit{negl}(\eta).$$

Definition 4 (Quantum partial-domain one-way function). *We say a permutation $f : \{0,1\}^{n+k_1} \times \{0,1\}^{k_0} \to \{0,1\}^m$ is quantum partial-domain one-way if for any* QPT *adversary A,*

$$\Pr\left[\tilde{s} = s : s \xleftarrow{\$} \{0,1\}^{n+k_1},\ t \xleftarrow{\$} \{0,1\}^{k_0},\ \tilde{s} \leftarrow A(f(s,t))\right] \leq \mathit{negl}(\eta).$$

We use the 'gentle-measurement lemma' [16] in the proof. Informally, it states that if an output of a measurement is almost certain for a quantum state, the measurement does not disturb the state much.

Lemma 1 (gentle-measurement lemma). *Let $\mathbb{M} = \{\mathbb{P}_i\}_i$ is a measurement. For any state $|\Psi\rangle$, if there exists an i such that $\|\mathbb{P}_i |\Psi\rangle\|^2 \geq 1 - \epsilon$, then* $\mathrm{TD}(|\Psi\rangle, \mathbb{M}|\Psi\rangle) \leq \sqrt{\epsilon} + \epsilon$.

2.3 Compressed Standard Oracle

Generally, it is not possible to copy a quantum state due to no-cloning theorem and destructive nature of quantum measurements. However, in a recent work, Zhandry showed that for quantum queries to a random oracle, a sort of recording is possible. Note that the conventional way to query a random oracle in superposition is to choose a uniformly at random function H and answers to the query with the unitary $\mathbb{U}_H$. However, one can consider another approach in which the oracle starts with a private state that keeps a uniform superposition of all functions and the query is answered as:

$$|x,\ y\rangle \sum_H \frac{1}{\sqrt{|\Omega_H|}} |H\rangle \to \sum_H \frac{1}{\sqrt{|\Omega_H|}} |x,\ y \oplus H(x)\rangle |H\rangle ,$$

where Ω_H is the set of all functions H. Following the perspective above, Zhandry [18] developed the CStO that its private register can be implemented efficiently, symmetrically stores the inputs/outputs of the adversary's queries in its private register and it is perfectly indistinguishable from the standard oracle (StO).

Lemma 2 (Lemma 4 in [18]). CStO *and* StO *are perfectly indistinguishable.*

We import the representation of CStO from [8]. Let $\mathfrak{D} = \otimes_{x\in X}\mathfrak{D}_x$ be the oracle register. The state space of $\mathfrak{D}_x$ is generated with vectors $|y\rangle$ for $y \in Y \cup \{\perp\}$. Let $F_{\mathfrak{D}_x}$ be a unitary acting on $\mathfrak{D}_x$ that maps $|\perp\rangle$ to $\mathsf{QFT}|0\rangle$ and vice versa. And for any vector orthogonal to $|\perp\rangle$ and $\mathsf{QFT}|0\rangle$, F is identity. We define CStO to be the following unitary acting on the input register, the output register and the $\mathfrak{D}$ register.

$$\mathsf{CStO} = \sum_x |x\rangle\langle x| \otimes F_{\mathfrak{D}_x} CNOT_{Y\mathfrak{D}_x} F_{\mathfrak{D}_x},$$

where $CNOT_{Y\mathfrak{D}_x}|y, y_x\rangle = |y \oplus y_x, y_x\rangle$ for $y, y_x \in Y$ and it is identity on $|y, \perp\rangle$. The initial state of $\mathfrak{D}$ register is $\otimes_{x\in X}|\perp\rangle$.

In the following, we present preliminaries for Theorem 3.1 in [8]. For a fixed relation $R \subset X \times Y$, Γ_R is the maximum number of y's that fulfill the relation R where the maximum is taken over all $x \in X$:

$$\Gamma_R = \max_{x\in X} |\{y \in Y | (x, y) \in R\}|.$$

We define a projector $\Pi^x_{\mathfrak{D}_x}$ that checks if the register $\mathfrak{D}_x$ contains a value $y \neq \perp$ such that $(x, y) \in R$:

$$\Pi^x_{\mathfrak{D}_x} := \sum_{y\ s.t.\ (x,y)\in R} |y\rangle\langle y|_{\mathfrak{D}_x}.$$

Let $\bar{\Pi}^x_{\mathfrak{D}_x} = \mathbb{I}_{\mathfrak{D}_x} - \Pi^x_{\mathfrak{D}_x}$. We define the measurement $\mathbb{M}$ to be the set of projectors $\{\Sigma^x\}_{x\in X\cup\{\emptyset\}}$ where

$$\Sigma^x := \bigotimes_{x'<x} \bar{\Pi}^{x'}_{\mathfrak{D}_{x'}} \otimes \Pi^x_{\mathfrak{D}_x} \text{ for } x \in X \text{ and } \Sigma^\emptyset := \mathbb{I} - \sum_x \Sigma^x. \tag{1}$$

Informally, the measurement $\mathbb{M}$ checks for the smallest x for which $\mathfrak{D}_x$ contains a value $y \neq \perp$ such that $(x, y) \in R$. If no register $\mathfrak{D}_x$ contains a value $y \neq \perp$ such that $(x, y) \in R$, the outcome of $\mathbb{M}$ is $\emptyset$. We define a purified measurement $\mathbb{M}_{\mathfrak{D}P}$ corresponding to $\mathbb{M}$ that XORs the outcome of the measurement to an ancillary register:

$$\mathbb{M}_{\mathfrak{D}P} |\phi, z\rangle_{\mathfrak{D}P} \to \sum_{x \in X \cup \{\emptyset\}} \Sigma^x |\phi\rangle_{\mathfrak{D}} |z \oplus x\rangle_P .$$

The following lemma states that CStO and $\mathbb{M}_{\mathfrak{D}P}$ almost commute if Γ_R is small proportional to the size of Y.

Lemma 3 (Theorem 3.1 in [8]). *For any relation R and Γ_R defined above, the commutator $[\mathsf{CStO}, \mathbb{M}_{\mathfrak{D}P}]$ is bounded as follows:*

$$\|[\mathsf{CStO}, \mathbb{M}_{\mathfrak{D}P}]\| \leq 8 \cdot 2^{-n/2} \sqrt{2\Gamma_R}.$$

3 Plaintext-extractability

We define "plaintext-extractable" notion below in the random oracle model and quantum random oracle model. Our notion lies between plaintext-awareness notions PA1 and PA2 [1]. Our notion is stronger than PA1 notion [4] because the adversary is allowed to eavesdrop some ciphertexts in contrast to PA1 that the adversary is not able to eavesdrop. Our notion is weaker than PA2 [1] because in our notion the extractor has access to all random oracle queries, in contrast, in PA2 notion the adversary does not know the random oracle queries that have been used to generate the eavesdropped ciphertexts.

3.1 Random Oracle Model

The random oracle model [3] is a powerful model in which the security of a cryptographic scheme is proven assuming the existence of a truly random function that is accessible by all parties including the adversary.

Informally, we say a public-key encryption scheme is plaintext-extractable if there exists an extractor algorithm Ext that given access to the list of all queries to the random oracle can simulate the decryption oracle.

Let $\mathcal{E}^H_{\mathsf{pk}}$ indicates an encryption oracle that upon receiving a query m_0, m_1 from the adversary, it chooses a random bit b, encrypts m_b and sends the resulting ciphertext to the adversary. All ciphertexts obtained from $\mathcal{E}^H_{\mathsf{pk}}$ are stored in **List** and for any $c \in$ **List**, the decryption oracle $\mathrm{Dec}^H_{\mathsf{sk}}$ returns $\perp$.

Definition 5. *Let $\mathfrak{L}_H$ be the list of inputs/outputs of all queries to the random oracle H and* **List** *be the list of ciphertexts obtained from $\mathcal{E}^H_{\mathsf{pk}}$. Let η be the security parameter. We say a public-key encryption scheme $\Pi_H = (\mathrm{Gen}, \mathrm{Enc}, \mathrm{Dec})$ is*

[1] Recently, the classical plaintext-awareness notions PA0, PA1 and PA2 are adopted to the post-quantum setting, however, in the standard model [10].

plaintext-extractable in the random oracle model if there exists an algorithm Ext *such that for any polynomial-time distinguisher* $\mathcal{D}$*, the following holds:*

$$|\Pr\Big[\mathcal{D}^{\mathrm{Dec}^H_{\mathsf{sk},\mathbf{List}},H,\mathcal{E}^H_{\mathsf{pk}}}(\mathsf{pk})=1:(\mathsf{pk},\mathsf{sk})\leftarrow \mathrm{Gen}(1^\eta), H\xleftarrow{\$}\Omega_H\Big]-$$
$$\Pr\Big[\mathcal{D}^{\mathsf{Ext}(\mathsf{pk},\mathfrak{L}_H,\mathbf{List}),H,\mathcal{E}^H_{\mathsf{pk}}}(\mathsf{pk})=1:(\mathsf{pk},\mathsf{sk})\leftarrow \mathrm{Gen}(1^\eta), H\xleftarrow{\$}\Omega_H\Big]|\leq \mathit{negl}(\eta).$$

In the definition above, the random oracle H can consist of several random oracles $\{H_i\}_i$ and is defined as $H(i,x):=H_i(x)$. (This remark has been made since OAEP+ uses three random oracles.)

Remark. The definition above can be generalized to any oracle $\mathcal{E}^H_{\mathsf{pk}}$ that upon receiving a query M from the adversary, it randomly generates a message m (note that m may depend on M), encrypts it and sends the resulting ciphertext to the adversary. In more details, we say an encryption scheme is plaintext-extractable if there exists an extractor that works for any $\mathcal{E}^H_{\mathsf{pk}}$ defined above. (This generality is not needed in our paper but it might be needed in other context. For instance this generality is crucial to prove the implication PA2+IND-CPA $\Longrightarrow$ IND-CCA in [2].)

3.2 Quantum Random Oracle Model

We define plaintext-extractibility in the quantum random oracle model [6] in which queries to the random oracles are quantum (superposition of inputs). This is necessary in the post-quantum setting since a quantum adversary attacking a scheme based on a real hash function is necessarily able to evaluate that function in superposition. Hence the random oracle model must reflect that ability if one requests post-quantum security.

In the definition below, an oracle with quantum access is differentiated with an underline (and an oracle without an underline is accessed classically).

Definition 6. *Let* $\mathfrak{D}_H$ *be a database of* CStO_H*. Let* η *be the security parameter. We say a public-key encryption scheme* Π_H *is plaintext-extractable in the quantum random oracle model if there exists an algorithm* Ext *such that for any* QPT *distinguisher* $\mathcal{D}$*, the following holds:*

$$|\Pr\Big[\mathcal{D}^{\mathrm{Dec}^{\mathsf{CStO}_H}_{\mathsf{sk},\mathbf{List}},\underline{\mathsf{CStO}_H},\mathcal{E}^{\mathsf{CStO}_H}_{\mathsf{pk}}}(\mathsf{pk})=1:(\mathsf{pk},\mathsf{sk})\leftarrow \mathrm{Gen}(1^\eta), H\xleftarrow{\$}\Omega_H\Big]-$$
$$\Pr\Big[\mathcal{D}^{\mathsf{Ext}(\mathsf{pk},\mathfrak{D}_H,\mathbf{List}),\underline{\mathsf{CStO}_H},\mathcal{E}^{\mathsf{CStO}_H}_{\mathsf{pk}}}(\mathsf{pk})=1:(\mathsf{pk},\mathsf{sk})\leftarrow \mathrm{Gen}(1^\eta), H\xleftarrow{\$}\Omega_H\Big]|\leq \mathit{negl}(\eta).$$

In the definition above, the random oracle H can consist of several random oracles $\{H_i\}_i$ and $H(i,x):=H_i(x)$. However, the first component of a quantum query (the index i) is restricted to be a classical value. In other words, the adversary is not allowed to query all oracles simultaneously by submitting $\sum_{i,x}\alpha_{i,x}|i,k\rangle$. (This restriction is not limiting since in OAEP+, the adversary is allowed to query the random oracles G,H,H' separately.)

4 Security of OAEP+

We define OAEP+ transformation below.

Definition 7 (OAEP+). *Let* $G : \{0,1\}^{k_0} \to \{0,1\}^n$, $H : \{0,1\}^{n+k_1} \to \{0,1\}^{k_0}$ *and* $H' : \{0,1\}^{n+k_0} \to \{0,1\}^{k_1}$ *be random oracles. The encryption scheme* $OAEP+ = (\mathrm{Gen}, \mathrm{Enc}, \mathrm{Dec})$ *is defined as:*

1. Gen*: Specifies an instance of the injective function* f *and its inverse* f^{-1}*. Therefore, the public key and secret key are* f *and* f^{-1} *respectively.*
2. Enc*: Given a message* $m \in \{0,1\}^n$*, the encryption algorithm computes*

$$s := (G(r) \oplus m)||H'(r,m) \quad \textit{and} \quad t := r \oplus H(s),$$

   where $r \xleftarrow{\$} \{0,1\}^{k_0}$*, and outputs the ciphertext* $c := f(s,t)$*.*
3. Dec*: Given a ciphertext* c*, the decryption algorithm does the following: Compute* $f^{-1}(c) = (s,t)$*, query the random oracle* H *on input* s*, query the random oracle* G *on input* $t \oplus H(s)$ *and compute* $m' := [s]^n \oplus G(t \oplus H(s))$*. Then, if* $H'(t \oplus H(s), m') = [s]_{k_1}$*, it returns* m'*, otherwise, it returns* $\perp$*.*

Note that k_0 *and* k_1 *depend on the security parameter* n*.*

We prove that OAEP+ is IND-CCA secure in the quantum random oracle model. First, we show that OAEP+ is plaintext-extractable and use it to show the IND-CCA security.

To show the plaintext-extractability, the overall strategy is to start with a game in which the adversary has access to the actual decryption oracle, define some indistinguishable intermediate games and reach the last game for which the challenger does not use the secret key for decryption.

In the following, the algorithm $\mathrm{Dec}_{f^{-1}}$ is the decryption algorithm of OAEP+ except for the challenge ciphertext c^* that outputs $\perp$. The number of queries to the random oracles G, H, H' is shown by $q_G, q_H, q_{H'}$, respectively, and q_D is the number of decryption queries.

Theorem 1. *OAEP+ is plaintext-extractable in the quantum random oracle model.*

Proof. **Game 0**. We start with Game 0 in which the quantum polynomial-time distinguisher $\mathcal{D}$ has classical access to the decryption oracle $\mathrm{Dec}_{f^{-1}}$, quantum access to the random oracles G, H, H' and classical access to the encryption oracle $\mathcal{E}_f^{G,H,H'}$.

Game 1. We replace the random oracles G, H, H' with the compressed standard oracles $\mathsf{CStO}_G, \mathsf{CStO}_H, \mathsf{CStO}_{H'}$, respectively. These changes are indistinguishable for the adversary by Lemma 2. Let $\mathfrak{D}_G$, $\mathfrak{D}_H$ and $\mathfrak{D}_{H'}$ denote the databases of these oracles.

Game 2. We modify the decryption oracle $\mathrm{Dec}_{f^{-1}}$ to a decryption oracle $\mathrm{Dec}^{(1)}_{f^{-1}}$ that works as follows. Let $\mathfrak{D}_{H'}$ denotes the database of $\mathsf{CStO}_{H'}$. We define the relation $R_c^{H'}$ to be the set of all $((r,m), H'_{(r,m)})$ such that

$$[[f^{-1}(c)]^{n+k_1}]_{k_1} = H'_{(r,m)}. \tag{2}$$

Given the relation $R_c^{H'}$, the projectors $\Sigma_c^{(r,m)}$ for $(r,m) \in \{0,1\}^{n+k_0}$ and $\Sigma_c^{\emptyset}$ are defined similar to Equation (1). Now the measurement

$$\mathbb{M}^{H'} = \{\Sigma_c^{(r,m)}\}_{(r,m)\in\{0,1\}^{n+k_0}\cup\{\emptyset\}}$$

checks if there exists a pair in $\mathfrak{D}_{H'}$ satisfying the relation $R_c^{H'}$ or not. If there is more than one pair satisfying the relation $R_c^{H'}$, the smallest (r,m) will be the output of $\mathbb{M}^{H'}$[2]. If there is no such a pair the output of $\mathbb{M}^{H'}$ is $\emptyset$. Let $\mathbb{M}^c_{\mathfrak{D}_{H'},P_{H'}}$ be the following purified measurement corresponding to $\mathbb{M}^{H'}$:

$$\mathbb{M}^c_{\mathfrak{D}_{H'},P_{H'}} |\phi, z\rangle_{\mathfrak{D}_{H'}P_{H'}} \to \sum_{(r,m)\in\{0,1\}^{n+k_0}\cup\{\emptyset\}} \Sigma_c^{(r,m)} |\phi\rangle_{\mathfrak{D}_{H'}} |z \oplus (r,m)\rangle_{P_{H'}}.$$

Note that $\mathbb{M}^c_{\mathfrak{D}_{H'},P_{H'}}$ is an involution, that is, $\mathbb{M}^c_{\mathfrak{D}_{H'},P_{H'}}\mathbb{M}^c_{\mathfrak{D}_{H'},P_{H'}} = \mathbb{I}$. For each decryption query on an input c, the decryption algorithm $\mathrm{Dec}^{(1)}_{f^{-1}}$ first applies the $\mathbb{M}^c_{\mathfrak{D}_{H'},P_{H'}}$ unitary with the $P_{H'}$ register initiated with 0. Then it executes $\mathrm{Dec}_{f^{-1}}$. Finally it applies the $\mathbb{M}^c_{\mathfrak{D}_{H'},P_{H'}}$ again.

$$\mathrm{Dec}^{(1)}_{f^{-1}} = \mathbb{M}^c_{\mathfrak{D}_{H'},P_{H'}}\ \mathrm{Dec}_{f^{-1}}\ \ \mathbb{M}^c_{\mathfrak{D}_{H'},P_{H'}}.$$

We show that Game 1 and Game 2 are indistinguishable. Note that we measure the database $\mathfrak{D}_{H'}$ and this measurement might be detectable to the adversary. In order to undo this measurement we apply the measurement again, however, after applying $\mathrm{Dec}_{f^{-1}}$. Since $\mathrm{Dec}_{f^{-1}}$ queries H', the measurement on $\mathfrak{D}_{H'}$ does not commute with $\mathrm{Dec}_{f^{-1}}$, trivially. Therefore, we use Lemma 3 to show that these two almost commute and therefore this measurement is not detectable to the adversary.

Recall that $\Gamma_{R_c^{H'}}$ is the maximum values of $H'_{(r,m)}$ that satisfies the relation (2) where the maximum is taken over inputs (r,m). Since $[f^{-1}(c)]^{n+k_1}$ is a single value given c, $\Gamma_{R_c^{H'}} = 1$. By Lemma 3, $\mathbb{M}^c_{\mathfrak{D}_{H'},P_{H'}}$ almost commutes with $\mathrm{Dec}_{f^{-1}}$ and the adversary can distinguish these two games with a probability at most $q_D 2^{-k_1/2+7/2}$.

Game 3. We modify the decryption oracle $\mathrm{Dec}^{(1)}_{f^{-1}}$ to a decryption oracle $\mathrm{Dec}^{(2)}_{f^{-1}}$ that works as follows. It first applies $\mathbb{M}^c_{\mathfrak{D}_{H'},P_{H'}}$, if the register $P_{H'}$ is empty, it returns $\perp$, otherwise, it executes $\mathrm{Dec}_{f^{-1}}$. Finally it applies $\mathbb{M}^c_{\mathfrak{D}_{H'},P_{H'}}$. To show

[2]Outputting the smallest (r,m) is a convention to have a correct definition of the projector. Since a random oracle is quantum collision-resistance [17], only with a negligible probability there will be more than one pair satisfying the relation (2).

that these two games are indistinguishable, we show that when the register $P_{H'}$ is empty, the decryption oracle $\mathrm{Dec}^{(1)}_{f^{-1}}$ (or $\mathrm{Dec}_{f^{-1}}$) returns $\perp$ with a high probability. Let assume the adversary submits a decryption query c for which the register $P_{H'}$ is empty, that is, there is no pair $\big((r,m), H'_{(r,m)}\big)$ in $\mathfrak{D}_{H'}$ such that the relation (2) holds. Let $f^{-1}(c) = (s_c, t_c)$. The decryption algorithm $\mathrm{Dec}_{f^{-1}}$ checks if $H'\Big(t_c \oplus H(s_c), [s_c]^n \oplus G(t_c \oplus H(s_c))\Big) = [s_c]_{k_1}$ and this equality holds with a probability at most $1/2^{k_1}$ because H' is a random oracle and $\Big(t_c \oplus H(s_c), [s_c]^n \oplus G(t_c \oplus H(s_c))\Big)$ has not been queried to H' by the adversary since $P_{H'}$ is empty. Overall, the adversary can distinguish these two games with a probability at most $q_D/2^{k_1}$.

Game 4. Let $\mathfrak{D}_G$ denotes the database of CStO_G. We modify the decryption oracle $\mathrm{Dec}^{(2)}_{f^{-1}}$ to a decryption oracle $\mathrm{Dec}^{(3)}_{f^{-1}}$ that on the input c works as follows. It first applies $\mathbb{M}^c_{\mathfrak{D}_{H'}, P_{H'}}$, if the register $P_{H'}$ is empty, it returns $\perp$. Otherwise if the register $P_{H'}$ contains a pair (r', m'), it applies a purified measurement $\mathbb{M}^G$ on the database $\mathfrak{D}_G$ that returns 1 if there exists a pair $(r', G_{r'}) \in \mathfrak{D}_G$ such that

$$[[f^{-1}(c)]^{n+k_1}]^n \oplus m' = G_{r'} \tag{3}$$

and returns 0 otherwise. The output of this measurement is stored in the register P_G that starts with $|0\rangle$. Then it applies $\mathrm{Dec}_{f^{-1}}$, $\mathbb{M}^G$ and $\mathbb{M}^c_{\mathfrak{D}_{H'}, P_{H'}}$ respectively. (Note that $\mathbb{M}^G$ is defined similar to $\mathbb{M}^c_{\mathfrak{D}_{H'}, P_{H'}}$ in Game 2.)

In order to show that these two games are indistinguishable, we show that $\mathrm{Dec}_{f^{-1}}$ and $\mathbb{M}^G$ almost commutes. (Then $\mathbb{M}^G$ will cancel out with its second application.) By Lemma 3, these two games are indistinguishable with a probability at most $q_D 2^{-n+7/2}$.

Game 5. We modify the decryption oracle $\mathrm{Dec}^{(3)}_{f^{-1}}$ to a decryption oracle $\mathrm{Dec}^{(4)}_{f^{-1}}$ that on the input c works similar to $\mathrm{Dec}^{(3)}_{f^{-1}}$ unless if the output of $\mathbb{M}^G$ is 0, it returns $\perp$.

$$\mathrm{Dec}^{(4)}_{f^{-1}}(c) = \begin{cases} \perp & \text{if } P_{H'} \text{ is empty} \\ \perp & \text{if } P_G \text{ contains } 0\ . \\ \mathrm{Dec}_{f^{-1}}(c) & \text{otherwise} \end{cases}$$

In order to show that these two games are indistinguishable, we need to show that the decryption algorithms $\mathrm{Dec}^{(3)}_{f^{-1}}$ and $\mathrm{Dec}^{(4)}_{f^{-1}}$ return the same output with a high probability. If P_G contains 1, both algorithms return $\mathrm{Dec}_{f^{-1}}(c)$. We prove that when P_G contains 0, $\mathrm{Dec}_{f^{-1}}(c)$ is $\perp$ with a high probability. Let $f^{-1}(c) = (s_c, t_c)$. Note that $\mathrm{Dec}_{f^{-1}}$ checks if

$$H'\Big(t_c \oplus H(s_c), [s_c]^n \oplus G(t_c \oplus H(s_c))\Big) = [s_c]_{k_1}$$

or not. We show that since P_G is 0, the query $\big(t_c \oplus H(s_c), [s_c]^n \oplus G(t_c \oplus H(s_c))\big)$ will be submitted to H' only with a negligible probability. Note that when P_G

is 0 the value $t_c \oplus H(s_c)$ has not been queried to the random oracle G. This means that the adversary has obtained the value $G(t_c \oplus H(s_c))$ without querying $t_c \oplus H(s_c)$ to G. This holds with a probability $1/2^n$.

When $\big(t_c \oplus H(s_c), [s_c]^n \oplus G(t_c \oplus H(s_c))\big)$ has not been queried to H', $\mathrm{Dec}_{f^{-1}}$ returns $\perp$ with a probability at least $1 - 1/2^{k_1}$ because H' is a random oracle. Overall, the adversary can distinguish these two games with a negligible probability.

Game 6. Let $\mathfrak{D}_H$ denotes the database of CStO_H. We modify the decryption oracle $\mathrm{Dec}^{(4)}_{f^{-1}}$ to a decryption oracle $\mathrm{Dec}^{(5)}_{f^{-1}}$ that on the input c works as follows. The decryption oracle $\mathrm{Dec}^{(5)}_{f^{-1}}$ is similar to $\mathrm{Dec}^{(4)}_{f^{-1}}$ except if the register $P_{H'}$ is not empty and P_G is not zero it sets $s' = (G_{r'} \oplus m')||H'_{(r',m')}$. Then it applies a purified measurement $\mathbb{M}^H$ on the database $\mathfrak{D}_H$ that returns 1 if there exists a pair $(s', H_{s'}) \in \mathfrak{D}_H$ such that

$$H_{s'} = [f^{-1}(c)]_{k_0} \oplus r'. \tag{4}$$

Otherwise it returns 0. The output of this measurement is stored in the register P_H that starts with $|0\rangle$. Note that the measurement $\mathbb{M}^H$ is applied again after $\mathrm{Dec}_{f^{-1}}$. By Lemma 3, these two games are indistinguishable with a probability at most $q_D 2^{-k_0+7/2}$. (Note that $\mathbb{M}^H$ is defined similar to $\mathbb{M}^c_{\mathfrak{D}_{H'}, P_{H'}}$ in Game 2.)

Game 7. We modify the decryption oracle $\mathrm{Dec}^{(5)}_{f^{-1}}$ to a decryption oracle $\mathrm{Dec}^{(6)}_{f^{-1}}$ that on the input c works similar to $\mathrm{Dec}^{(5)}_{f^{-1}}$ unless if the output of $\mathbb{M}^H$ is 0, it returns $\perp$.

$$\mathrm{Dec}^{(6)}_{f^{-1}}(c) = \begin{cases} \perp & \text{if } P_{H'} \text{ is empty} \\ \perp & \text{if } P_G \text{ contains } 0 \\ \perp & \text{if } P_H \text{ contains } 0 \\ \mathrm{Dec}_{f^{-1}}(c) & \text{otherwise} \end{cases}.$$

In order to show that these two games are indistinguishable, we need to show that the decryption algorithms $\mathrm{Dec}^{(4)}_{f^{-1}}$ and $\mathrm{Dec}^{(5)}_{f^{-1}}$ return the same output with a high probability. If P_H contains 1, both algorithms return $\mathrm{Dec}_{f^{-1}}(c)$. We prove that when P_H contains 0, $\mathrm{Dec}_{f^{-1}}(c)$ is $\perp$ with a high probability. Let $f^{-1}(c) = (s_c, t_c)$. By the relations (2) and (3), we can write

$$s_c = (G_{r'} \oplus m')||H'_{(r',m')} = s'.$$

Since P_H is 0, s_c has not been queried to the random oracle H and therefore $H(s_c)$ remains a uniformly random value from the adversary's perspective. This means that the equality

$$H'\Big(t_c \oplus H(s_c), [s_c]^n \oplus G(t_c \oplus H(s_c))\Big) = [s_c]_{k_1}$$

holds with a probability at most $1/2^{k_1}$. Overall, the adversary is able to distinguish these two games with a probability at most $q_D/2^{k_1}$.

Game 8. In this game, we change $\mathrm{Dec}^{(6)}_{f^{-1}}$ to a decryption oracle $\mathrm{Dec}^{(7)}_{f}$ that does not use f^{-1} to decrypt. Note that the decryption oracle $\mathrm{Dec}^{(6)}_{f^{-1}}$ uses f^{-1} in the measurements $\mathbb{M}^{c}_{\mathfrak{D}_{H'},P_{H'}}, \mathbb{M}^{G}$ and $\mathbb{M}^{H}$. So instead of applying these measurements, we search over all pairs in $\mathfrak{D}_{H'}$. Namely, for each pair $(r', m') \in \mathfrak{D}_{H'}$, the decryption oracle $\mathrm{Dec}^{(7)}_{f}$ checks if $(r', G_{r'})$ is in $\mathfrak{D}_G$. If yes, it sets $s' = (G_{r'} \oplus m')||H'_{(r',m')}$. Then it checks if $(s', H_{s'}) \in \mathfrak{D}_H$. If yes, $\mathrm{Dec}^{(7)}_{f}$ checks if $c = f(s', r' \oplus H_{s'})$. If the equality holds, it returns m' and aborts. If there is no pair $(r', m') \in \mathfrak{D}_{H'}$ that make the decryption aborts, the output of $\mathrm{Dec}^{(4)}_{f}$ will be $\perp$.

We show that these decryption algorithms $\mathrm{Dec}^{(6)}_{f^{-1}}$ and $\mathrm{Dec}^{(7)}_{f}$ are indistinguishable. It is clear that if $P_{H'}$ is empty or one of P_G or P_H registers contain 0 for a ciphertext c, both decryption algorithms return $\perp$. If for a ciphertext c, $P_{H'}$ is not empty and P_G and P_H registers contain 1, this means that the relations (2), (3) and (4) hold for $f^{-1}(c) = (s_c, t_c)$ and a pair $((r', m'), H_{(r',m')}) \in \mathfrak{D}_{H'}$. That is,

$$s' = (G_{r'} \oplus m')\|H_{(r',m')},\ [s_c]_{k_1} = H'_{(r',m')},\ [s_c]^n = m' \oplus G_{r'} \text{ and } t_c = H_{s'} \oplus r'.$$

It is clear that $s' = s_c$ and $r' = t_c \oplus H(s_c)$. In this case, $\mathrm{Dec}^{(7)}_{f}$ returns m'. On the other hand, the decryption algorithm $\mathrm{Dec}_{f^{-1}}$ checks if

$$H'\Big(t_c \oplus H(s_c), [s_c]^n \oplus G(t_c \oplus H(s_c))\Big) = [s_c]_{k_1}$$

and if this equality holds, it returns $[s_c]^n \oplus G(t_c \oplus H(s_c))$. Now it is obvious that $\mathrm{Dec}^{(6)}_{f^{-1}}$ return m' as well. This finishes the proof because $\mathrm{Dec}^{(7)}_{f}$ does not use f^{-1} to decrypt. □

We use $\mathrm{Dec}^{(7)}_{f}$ to prove the IND-CCA security of OAEP+ in the quantum random oracle model. The overall strategy is to start with the IND-CCA game, define some indistinguishable intermediate games and reach the last game for which the adversary's success probability is 1/2.

Theorem 2. *If the underlying permutation is quantum one-way, then the OAEP+ scheme is IND-CCA secure in the quantum random oracle model.*

Proof. We reduce an adversary that attacks in the IND-CCA sense to an adversary $\mathcal{B}$ that inverts the permutation f. Note that in all games below, $\mathsf{CStO}_G, \mathsf{CStO}_H, \mathsf{CStO}_{H'}$ denote the compressed oracles corresponding the random oracles G, H, H', respectively, b is a random bit chosen by the challenger, m_0, m_1 are challenge messages submitted by the adversary, r^* is a uniformly at random element, c^* is the challenge ciphertext that is computed as: $c^* = f(s^*, t^*)$ where $s^* = (G(r^*) \oplus m_b)||H'(r^*, m_b)$ and $t^* = H(s^*) \oplus r^*$.

Game 0: We start with IND-CCA game in the quantum random oracle model in which the adversary $\mathcal{A}$ wins if it guesses the challenge bit b. Note that we use

compressed oracles in this game.

Game 1: We replace the decryption algorithm $\mathrm{Dec}_{f^{-1}}$ with $\mathrm{Dec}_f^{(7)}$ constructed in Theorem 1.

Game 2: This is identical to Game 1, except the challenger measures all the queries to CStO_G and $\mathsf{CStO}_{H'}$ with the projective measurements

$$\mathbb{M}_{r^*} = \{P_1 = |r^*\rangle\langle r^*|\,,\ P_0 = \mathbb{I} - |r^*\rangle\langle r^*|\}.$$

If the output of $\mathbb{M}_{r^*}$ is 1, it aborts and returns a random bit.

Let q_{G1} and $q_{H'1}$ be the total number of queries submitted to G and H' before the challenge query. Let q_{G2} and $q_{H'2}$ be the total number of queries submitted to G and H' after the challenge query.

If there is no query to CStO_G and $\mathsf{CStO}_{H'}$ with a non-negligible weight on the state $|r^*\rangle$, we can use Lemma 1 (gentle-measurement lemma) to show that these two games are indistinguishable. In more details, let ρ_i is the state of the i-th query and let $\mathbb{M}_{r^*}(\rho_i)$ returns 1 with the probability ϵ_i. By the gentle-measurement lemma, the trace distance between $\mathbb{M}_{r^*}(\rho_i)$ and ρ_i is at most $\sqrt{\epsilon_i} + \epsilon_i$. So overall, these two games are distinguishable with the advantage of at most $2(q_G + q_{H'})\sqrt{\max_i\{\epsilon_i\}}$. Therefore, if $\max_i\{\epsilon_i\}$ is negligible, two games are indistinguishable.

Since r^* is a random value that has not been used before the challenge query $\mathbb{M}_{r^*}(\rho_i)$ returns 1 with a probability at most $1/2^{k_0}$ for any $i \in [q_{G1}+q_{H'1}]$. So the measurements before the challenge query are distinguishable with a probability at most $2(q_{G1} + q_{H'1})\sqrt{2^{-k_0}}$ that is negligible.

It is left to show that the measurements after the challenge query are indistinguishable. Let assume $\mathcal{A}$ makes a query to CStO_G or $\mathsf{CStO}_{H'}$ after the challenge query with a non-negligible weight on $|r^*\rangle$ with a probability ϵ. We can construct an adversary $\mathcal{B}$ that breaks the quantum one-wayness of f. The adversary $\mathcal{B}$ on input $c^*\big(:= f(s^*, t^*)$ for uniformly random $s^*, t^*\big)$, runs $\mathcal{A}$ and guesses randomly in which query r^* will be submitted. The adversary $\mathcal{B}$ chooses i from $[q_{G2} + q_{H'2}]$ uniformly at random and simulates the random oracle queries and decryption oracle queries right until this query. Upon receiving the challenge query from $\mathcal{A}$, the adversary $\mathcal{B}$ sends c^*. We describe $\mathcal{B}$ in more details.

H-queries. For H-queries, the adversary $\mathcal{B}$ uses CStO_H where H is a random oracle.

Let Find be an operator that on inputs $r, c^*, \mathfrak{D}_H$, checks if there exists a pair (s, H_s) in $\mathfrak{D}_H$ such that $c^* = f(s, r \oplus H_s)$. If there exists such a pair it returns $(1, s)$. Otherwise, it returns $(0, 0^{n+k_1})$. Note that since f is a permutation, the Find unitary either returns $(0, 0^{n+k_1})$ or returns $(1, s^*)$.

G-queries. Let $\tilde{G}$ be a random oracle with the same domain and co-domain as G. For each query to G, $\mathcal{B}$ first applies Find operator with an ancillary register $Q_{b'}Q_s$ of $(1 + n + k_1)$ qubits initiated with zero. Then, if the query is conducted before the challenge query or the $Q_{b'}$ is set to 0, it forwards the query to $\mathsf{CStO}_{\tilde{G}}$,

otherwise, it XORs $m_b \oplus [s^*]^n$ to the output register:

$$G : |r, y\rangle |\mathfrak{D}_H\rangle \to \begin{cases} |r, y \oplus \tilde{G}(r)\rangle & \text{if } m_b \text{ is not defined} \\ |r, y \oplus \tilde{G}(r)\rangle & \text{if } \mathsf{Find}(r, c^*, \mathfrak{D}_H) = (0, 0^{n+k_1}) \\ |r, y \oplus (m_b \oplus [s^*]^n)\rangle & \text{if } \mathsf{Find}(r, c^*, \mathfrak{D}_H) = (1, s^*) \end{cases}.$$

And finally it applies the Find operator again. Since f is a permutation, there exists only one r such that $c^* = f(s^*, r \oplus H_{s^*})$ and that is r^*. For any $r \neq r^*$ the oracle G and the random oracle $\tilde{G}$ are the same. Recall that the adversary $\mathcal{B}$ guesses that a query with a non-negligible weight on $|r^*\rangle$ occurs in the i-th query. (This holds with a non-negligible probability $\epsilon/(q_{G2} + q_{H'2})$.) Therefore, the simulation of G-queries is indistinguishable from a random oracle $\tilde{G}$ right before the i-th query.

H'-queries. Let $\tilde{H}$ be a random oracle with the same domain and co-domain as H'. For each query $|r, m\rangle$, $\mathcal{B}$ first applies Find operator with an ancillary register $Q_{b'}Q_s$ of $(1 + n + k_1)$ qubits initiated with zero. Then, if the query is conducted before the challenge query or the $Q_{b'}$ is set to 0 or $m \neq m_b$, it forwards the query to $\mathsf{CStO}_{\tilde{H}}$, otherwise, it XORs $[s^*]_{k_0}$ to the output register:

$$H' : |r, m, y\rangle |\mathfrak{D}_H\rangle \to \begin{cases} |r, m, y \oplus \tilde{H}(r, m)\rangle & \text{if } m_b \text{ is not defined} \\ |r, m, y \oplus \tilde{H}(r, m)\rangle & \text{if } \mathsf{Find}(r, c^*, \mathfrak{D}_H) = (0, 0^{n+k_1}) \\ |r, m, y \oplus \tilde{H}(r, m)\rangle & \text{if } \mathsf{Find}(r, c^*, \mathfrak{D}_H) = (1, s^*) \wedge m \neq m_b \\ |r, m, y \oplus [s^*]_{k_1}\rangle & \text{if } \mathsf{Find}(r, c^*, \mathfrak{D}_H) = (1, s^*) \wedge m = m_b \end{cases}.$$

Similar to above, the simulation of H'-queries is indistinguishable from $\tilde{H}$ for queries right before the i-th query.

The challenge query. Upon receiving m_0 and m_1 from $\mathcal{A}$, the adversary $\mathcal{B}$ returns c^* as the challenge ciphertext. Note that the way we simulate G-queries and H'-queries, $G(r^*) := m_b \oplus [s^*]^n$, $H'(r^*, m_b) = [s^*]_{k_1}$ and $c^* = f(s^*, r^* \oplus H_{S^*})$ that is a perfect simulation of the challenge query.

Decryption queries. $\mathcal{B}$ uses the oracle $\mathrm{Dec}_f^{(7)}$ on inputs $\mathfrak{D}_H$, $\mathfrak{D}_{\tilde{G}}$ and $\mathfrak{D}_{\tilde{H}}$ for the decryption queries. Note that we reprogram G and H' only on the input r^* for which $c^* = (s^*, r^* \oplus H_{s^*})$. Since $\mathrm{Dec}_f^{(4)}$ on input c^* does not use its database and returns $\perp$, the simulation of the decryption queries is perfect.

Output of $\mathcal{B}$. The adversary $\mathcal{B}$ measures the i-th random oracle query to CStO_G or $\mathsf{CStO}_{H'}$ with $\mathbb{M}_{r^*}$. Then, the adversary searches over the database $\mathfrak{D}_H$ to find a pair (s^*, H_{s^*}) such that $c^* = f(s^*, r^* \oplus H_{s^*})$. If it finds such a pair, it returns $(s^*, r^* \oplus H_{s^*})$ as the inverse of f on c^* and aborts. Otherwise, it returns $s^* = (\tilde{G}(r^*) \oplus m_b)||\tilde{H}(r^*, m_b)$ and $r^* \oplus H(s^*)$ as the inverse of f on the input c^*. Note that when there is no pair (s^*, H_{s^*}) in $\mathfrak{D}_H$ such that $c^* = f(s^*, r^* \oplus H_{s^*})$, that is $\mathsf{Find}(r^*, c^*, \mathfrak{D}_H) = (0, 0^{n+k_1})$, the G-queries and H'-queries are answered with the random oracle $\tilde{G}$ and $\tilde{H}$, respectively. Therefore, the equation $c^* = f(x, r^* \oplus H(x))$ holds for $x = (\tilde{G}(r^*) \oplus m_b)||\tilde{H}(r^*, m_b)$. Overall, the adversary $\mathcal{B}$ can break the one-wayness of f with a probability at least $\epsilon/(q_{G2} + q_{H'2})$. Since f is quantum one-way, ϵ is negligible and this means Game 1 and Game 2 are indistinguishable.

Now, it is clear that Game 2 returns 1 with the probability 1/2 because if one of the measurements returns 1, the output of the game is a random bit. If none of the measurements return 1, $G(r^*)$ and $H'(r^*, m_b)$ remain an uniformly random value for $\mathcal{A}$ and consequently $m_b \oplus G(r^*)$ is an uniformly random value for $\mathcal{A}$. So the probability that $\mathcal{A}$ guesses b is 1/2. Finally, since each two consecutive games are indistinguishable, the probability that $\mathcal{A}$ guesses b in Game 0 is $1/2 + \mathsf{negl}(\mathsf{n})$ and this finishes the proof of the theorem. □

5 Conclusion and Future Direction

In this paper, we show that a weaker notion than PA2 (our plaintext-extractability notion) is sufficient to show the IND-CCA security when the reduction adversary tries to invert an injective function. We show the IND-CCA security of OAEP+ in QROM by first showing that OAEP+ is plaintext-extractable in QROM.

We argue that OAEP+ might even satisfy a stronger notion of plaintext-extractability, namely, the post-quantum PA2 introduced in [10]. Our high-level argument is that since the random oracle H' is used to sew the randomness and the message inside of the ciphertext, an adversary to attack PA2 might fail to output a valid ciphertext for which its corresponding plaintext is unknown to the adversary due to unpredictability of H'. We leave detailed investigation of this claim as a future work.

In addition, we leave investigating the security of OAEP+ with respect to a quantum IND-CCA notion that allows quantum challenge queries [7] as a future direction. (Note that the IND-qCCA scrutiny [5] of OAEP+ will follow with small modification to our proof. The IND-qCCA is a notion with classical challenge queries and quantum decryption queries.)

Acknowledgment. We would like to thank anonymous reviewers for their useful comments and suggestions.

References

1. M. Bellare, A. Desai, D. Pointcheval, and P. Rogaway. Relations among notions of security for public-key encryption schemes. In *CRYPTO 1998*, volume 1462, pages 26–45. Springer, 1998.
2. M. Bellare and A. Palacio. Towards plaintext-aware public-key encryption without random oracles. In *International Conference on the Theory and Application of Cryptology and Information Security*, pages 48–62. Springer, 2004.
3. M. Bellare and P. Rogaway. Random oracles are practical: A paradigm for designing efficient protocols. In D. E. Denning, R. Pyle, R. Ganesan, R. S. Sandhu, and V. Ashby, editors, *CCS '93, Proceedings of the 1st ACM Conference on Computer and Communications Security, Fairfax, Virginia, USA, November 3-5, 1993.*, pages 62–73. ACM, 1993.
4. M. Bellare and P. Rogaway. Optimal asymmetric encryption. In A. D. Santis, editor, *Advances in Cryptology - EUROCRYPT '94, Workshop on the Theory and Application of Cryptographic Techniques, Perugia, Italy, May 9-12, 1994,*

Proceedings, volume 950 of *Lecture Notes in Computer Science*, pages 92–111. Springer, 1994.

5. Boneh and M. Zhandry. Secure signatures and chosen ciphertext security in a quantum computing world. In R. Canetti and J. A. Garay, editors, *Advances in Cryptology - CRYPTO 2013 - 33rd Annual Cryptology Conference, Santa Barbara, CA, USA, August 18-22, 2013. Proceedings, Part II*, volume 8043 of *Lecture Notes in Computer Science*, pages 361–379. Springer, 2013.
6. D. Boneh, Ö. Dagdelen, M. Fischlin, A. Lehmann, C. Schaffner, and M. Zhandry. Random oracles in a quantum world. In D. H. Lee and X. Wang, editors, *Advances in Cryptology - ASIACRYPT 2011 - 17th International Conference on the Theory and Application of Cryptology and Information Security, Seoul, South Korea, December 4-8, 2011. Proceedings*, volume 7073 of *Lecture Notes in Computer Science*, pages 41–69. Springer, 2011.
7. C. Chevalier, E. Ebrahimi, and Q. H. Vu. On the security notions for encryption in a quantum world. *IACR Cryptol. ePrint Arch.*, 2020:237, 2020.
8. J. Don, S. Fehr, C. Majenz, and C. Schaffner. Online-extractability in the quantum random-oracle model. In O. Dunkelman and S. Dziembowski, editors, *Advances in Cryptology - EUROCRYPT 2022 - 41st Annual International Conference on the Theory and Applications of Cryptographic Techniques, Trondheim, Norway, May 30 - June 3, 2022, Proceedings, Part III*, volume 13277 of *Lecture Notes in Computer Science*, pages 677–706. Springer, 2022.
9. E. Ebrahimi. Post-quantum security of plain OAEP transform. In *PKC 2022*, volume 13177, pages 34–51. Springer, 2022.
10. E. Ebrahimi and J. van Wier. Post-quantum plaintext-awareness. Cryptology ePrint Archive, Paper 2022/937, 2022. `https://eprint.iacr.org/2022/937`.
11. E. Fujisaki, T. Okamoto, D. Pointcheval, and J. Stern. RSA-OAEP is secure under the RSA assumption. *J. Cryptology*, 17(2):81–104, 2004.
12. M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information (10th Anniversary edition)*. Cambridge University Press, 2016.
13. P. W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM J. Comput.*, 26(5):1484–1509, 1997.
14. V. Shoup. OAEP reconsidered. In *Advances in Cryptology - CRYPTO 2001, 21st Annual International Cryptology Conference, Santa Barbara, California, USA, August 19-23, 2001, Proceedings*, pages 239–259, 2001.
15. E. E. Targhi and D. Unruh. Post-quantum security of the fujisaki-okamoto and OAEP transforms. In *TCC 2016-B*, volume 9986 of *Lecture Notes in Computer Science*, pages 192–216, 2016.
16. A. J. Winter. Coding theorem and strong converse for quantum channels. *IEEE Trans. Inf. Theory*, 45(7):2481–2485, 1999.
17. M. Zhandry. A note on the quantum collision and set equality problems. *Quantum Inf. Comput.*, 15(7&8):557–567, 2015.
18. M. Zhandry. How to record quantum queries, and applications to quantum indifferentiability. In A. Boldyreva and D. Micciancio, editors, *Advances in Cryptology - CRYPTO 2019 - 39th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 18-22, 2019, Proceedings, Part II*, volume 11693 of *Lecture Notes in Computer Science*, pages 239–268. Springer, 2019.