Security analysis sketch for ThreeBears

Mike Hamburg*
Based on collaboration with Dominique Unruh and Eike Kiltz

November 29, 2017

Abstract

This is a sketch of how to do a security proof of chosen-ciphertext security for a post-quantum key encapsulation mechanism such as ThreeBears. We intend to collaborate on a more comprehensive version in 2018.

*Rambus Security Division

Contents

1 Introduction

This technical note accompanies the ThreeBears submission. It sketches a proof that ThreeBears is secure in the quantum random oracle model (QROM) [BDF+11], assuming that the Integer Module Learning with Errors (I-MLWE) distinguishing problem is hard.

The proof is by a sequence of games, where successive games have a small Euclidean distance between their quantum states, which implies a small difference in the probability that the adversary succeeds.

2 Quantum computing

We use the model of quantum computing from [BDF+11]. An n-bit classical computer has a state $s \in \{0,1\}^n$. But an n-qubit quantum computer instead has a complex *amplitude* z_s to be in each of these states. More formally, the state of a quantum computer is a unit vector

$$|\psi\rangle := \sum_{s \in \{0,1\}^n} z_s \cdot |s\rangle \in \mathbb{C}^{2^n}$$

where each classical state s gets its own basis element $|s\rangle$ ("ket s"). As a general notation convention, we use the Greek letters $|\phi\rangle$ or $|\psi\rangle$ to refer to a fully quantum state. We use Latin $|s\rangle$ to refer to the sth canonical basis vector in the space $\mathbb{C}^{2^n}$, and likewise for numeric states like $|0\rangle$. That is, $|s\rangle$ is the quantum version of the classical state s, but $|\psi\rangle$ might be an arbitrary superposition of classical states.

The quantum computation proceeds as a sequence of unitary transformations of $\mathbb{C}^{2^n}$ and calls to quantum oracles for classical functions F. The oracles are also unitary, and are modeled as in [BDF+11] by mapping

$$|x\rangle \otimes |y\rangle \to |x\rangle \otimes |y \oplus F(x)\rangle$$

To model a classical oracle for F, we add a query log which stores the input

and output of each query to the oracle:

$$|x\rangle \otimes |y\rangle \otimes |zero\rangle \quad \rightarrow \quad |x\rangle \otimes |y \oplus F(x)\rangle \otimes |zero \oplus (x, F(x))\rangle$$

Here the value $|zero\rangle$ is set to $|0\rangle$ in the input, and a separate state is used in each query, so it is actually $|0\rangle$ at the beginning of the oracle call. The practical implementation of this is just a circuit that measures x and responds with $F(x)$, but we describe it this way so that we can talk about the distance between states after the adversary has queried a classical oracle.

A quantum process can conditionally halt by calling a classical oracle Halt(data), where data $= 0$ indicates no halt and data $\neq 0$ indicates a halt.

This model of quantum processes is deterministic, but since the final measurement is random (in the Copenhagen interpretation), it can fully describe a randomized algorithm.

2.1 Proof strategy

Our proof is by a series of games. Between games, we will change the oracles slightly. Since the adversary is deterministic until the final measurement, we will measure how far this moves the final state $|\psi\rangle$, instead of what it does to the probability of each output. We can then convert between the two under lemma 3 and corollary 6.

3 Quantum lemmas

We will need several lemmas about quantum computation in our proof. These lemmas may be of independent interest. First, let's see how far we move the adversary by changing the oracle.

Lemma 1 (Perturbation from small random changes)**.** *Let $\mathcal{O}$ and $\mathcal{P}$ be oracles drawn from some joint distribution, both taking input from a set $\mathcal{X}$. Suppose there is some $\epsilon \geq 0$ such that for each $x \in \mathcal{X}$,*

$$\Pr\left[\mathcal{P}(x) \neq \mathcal{O}(x) : \text{input}, \mathcal{O}\right] \leq \epsilon$$

Let $|\psi\rangle$ *resp* $|\psi'\rangle$ *be the final states of* $\mathcal{A}^{\mathcal{O}}(\text{input})$ *resp* $\mathcal{A}^{\mathcal{P}}(\text{input})$*. Then*

$$\text{Exp}\left[\left||\psi'\rangle - |\psi\rangle\right|\right] \leq 2q\sqrt{\epsilon}$$

If the oracle is classical, then instead

$$\text{Exp}\left[\left||\psi'\rangle - |\psi\rangle\right|\right] \leq \sqrt{2q\epsilon}$$

In both cases, the expectation is over possible oracles $\mathcal{P}$*, given a particular* $\mathcal{O}$ *and input.*

Proof. See appendix A.1. □

This lemma applies best if the oracles differ in places that are independent of $\mathcal{O}$ (or, because the lemma's bounds are symmetric, independent of $\mathcal{P}$).

But what if they differ in a place that an adversary might compute? We would like to show that the adversary still can't tell the difference without actually querying the oracles in these places. We do this by measuring *classically* whether $\mathcal{O}$ and $\mathcal{P}$ differ on the queried value, but nothing else.

We then have the following lemma, which may be regarded as a variant of Unruh's lemma [Unr14, TU16].

Lemma 2 (Punctured oracles)**.** *Let* $\mathcal{O}$ *be a quantum oracles for a functions from* $\mathcal{X}$ *to* $\mathcal{Y}$*, and* $\mathcal{P}$ *be an oracle or function from* $\mathcal{X}$ *to* $\{0,1\}$*. Let* $\mathcal{O}\backslash\mathcal{P}$ *denote the oracle*

$$(\mathcal{O}\backslash\mathcal{P})(x) := \begin{cases} \mathcal{O}(x) & \text{if } \mathcal{P}(x) = 0 \\ \mathit{Halt}(x) & \text{if } \mathcal{P}(x) = 1 \end{cases}$$

Let $|\psi\rangle$ *resp* $|\psi'\rangle$ *be the final states of* $\mathcal{A}^{\mathcal{O}}(\text{input})$ *resp* $\mathcal{A}^{\mathcal{O}\backslash\mathcal{P}}$*. Then*

$$\left||\psi'\rangle - |\psi\rangle\right| \leq \sqrt{(q+1)\cdot \Pr\left[\mathcal{A}^{\mathcal{O}\backslash\mathcal{P}}(\text{input}) \text{ halts with } x : \mathcal{P}(x) = 1\right]}$$

Proof. See appendix A.2. □

We now show some ways to apply these lemmas to probabilities instead of quantum states. The first is simplest, and is useful for showing a property such as one-way-ness, where the goal is reached with low probability.

Lemma 3 (Perturbation to small probabilities)**.** *Suppose that the adversary's goal is to output a value in some set* Goal*, and suppose that it ends up in a quantum state that's either* $|\psi\rangle$ *or* $|\psi'\rangle$*. Then*

$$\left| \sqrt{\Pr\left[\text{measure}(|\psi\rangle) \in \text{Goal}\right]} - \sqrt{\Pr\left[\text{measure}(|\psi'\rangle) \in \text{Goal}\right]} \right| \leq \left| |\psi'\rangle - |\psi\rangle \right|$$

Proof. See appendix A.3. □

Lemma 4 (Random puncturing)**.** *Let* $\mathcal{O}$ *resp* $\mathcal{P}$ *be chosen from some joint distribution of functions* $\mathcal{X} \to \mathcal{Y}$ *resp* $\mathcal{X} \to \{0,1\}$*. Suppose there is some* $\epsilon \geq 0$ *such that for all* $x \in \mathcal{X}$*,*

$$\Pr\left[\mathcal{P}(x) = 1 : \mathcal{O}, \text{input}\right] \leq \epsilon$$

Then

$$\Pr\left[\mathcal{A}^{\mathcal{O}\backslash\mathcal{P}}(\text{input}) \text{ halts with } x : \mathcal{P}(x) = 1\right] \leq 2q^2\epsilon$$

In particular, if $\mathcal{P}(x) = 1$ *for exactly one input* x *which is independent of* input *and* $\mathcal{O}$*, then*

$$\Pr\left[\mathcal{A}^{\mathcal{O}\backslash\mathcal{P}}(\text{input}) = (x, n)\right] \leq 2q^2/\text{card}\left(\mathcal{X}\right)$$

Proof. See appendix A.4. □

We could also use these results for indistinguishability arguments, but the following lemma is stronger (and also stronger than [BV93], lemma 3.2.6):

Lemma 5 (Perturbation to L^1 distance)**.** *Let* $|\psi\rangle$ *and* $|\psi'\rangle$ *be quantum states, and* $\mathcal{D}$ *and* $\mathcal{D}'$ *be the distributions produced by measuring them, and let*

$$\left|\mathcal{D}' - \mathcal{D}\right|_1 := \sum_{d \in \mathcal{D} \cup \mathcal{D}'} \left|\Pr(d \leftarrow \mathcal{D}') - \Pr(d \leftarrow \mathcal{D})\right|$$

be the L^1 *statistical distance between these distributions. Then*

$$\left|\mathcal{D}' - \mathcal{D}\right|_1 \leq 2 \cdot \left| |\psi'\rangle - |\psi\rangle \right|$$

Proof. See appendix A.5. □

Corollary 6 (Perturbation to advantage)**.** *Suppose* $|\psi\rangle$ *resp* $|\psi'\rangle$ *are the final states of two quantum algorithm* $\mathcal{A}$ *resp* $\mathcal{A}'$ *that output either 0 or 1. Then*

$$\left|\ \Pr(\mathcal{A}(\text{input})=1) - \Pr(\mathcal{A}'(\text{input})=1)\ \right| = \frac{1}{2}\left|\mathcal{D}' - \mathcal{D}\right|_1 \leq \left|\ |\psi'\rangle - |\psi\rangle\ \right|$$

4 IND-KPA security

We define the IND-KPA ("indistinguishability under known-plaintext attack") advantage of an algorithm $\mathcal{A}$ against a public-key encryption system as follows. A challenger runs:

$$\begin{aligned}
(\text{pk}, \text{sk}) &\leftarrow \text{Keygen}(\text{coins}_1) \text{ where } \text{coins}_1 \xleftarrow{R} \mathcal{C}_{\text{Keygen}} \\
(m_0, m_1, b) &\xleftarrow{R} \mathcal{M} \times \mathcal{M} \times \{0,1\} \\
\text{ct} &\leftarrow \text{Enc}(\text{pk}, m_b, \text{coins}_2) \text{ where } \text{coins}_2 \leftarrow \mathcal{C}_{\text{Enc}} \\
b' &\leftarrow \mathcal{A}(\text{pk}, m_0, m_1, \text{ct})
\end{aligned}$$

The adversary's output b' is a guess at b, i.e. at which message the challenger encrypted. Its advantage is defined as

$$\text{Adv}_{\text{KPA}}(\mathcal{A}) := \left|\Pr\left[b' = 1 : b = 1\right] - \Pr\left[b' = 1 : b = 0\right]\right|$$

We note that IND-KPA security follows trivially from IND-CPA security (indistinguishability under chosen plaintext attack).

4.1 Core IND-KPA security of ThreeBears

Let ThreeBears$_R$ be the core of ThreeBears, where the coins are chosen at random instead of using cSHAKE. Specifically, let

$$\begin{aligned}
\text{Keygen}_R &= ((M, A), a) : M \leftarrow R^{d\times d}; a \leftarrow \chi^d; \epsilon_a \leftarrow \chi^d; A \leftarrow Ma + \epsilon \\
\text{Enc}_R((M, A), m) &= (C, E) : b \leftarrow \chi^d; \epsilon_b \leftarrow \chi^d; \epsilon' \leftarrow \chi; \\
&\quad C \leftarrow b^\top M + \epsilon_b^\top;\ D \leftarrow b^\top A + \epsilon'; E \leftarrow \text{encode}(D, m) \\
\text{Dec}_R(a, (C, E)) &= \text{decode}(Ca, E)
\end{aligned}$$

where $\text{encode}(K, m)$ is within δ of uniformly random when $K \xleftarrow{R} R$, regardless of m.

Then for any IND-KPA adversary $\mathcal{A}$ against ThreeBears$_R$, there is an I-MLWE adversary $\mathcal{A}'$ with $e = d + 1$, running in about the same time as $\mathcal{A}$, such that

$$\text{Adv}_{\text{KPA}}(\mathcal{A}) \leq 2 \cdot \text{Adv}_{\text{I-MLWE}}(\mathcal{A}') + \delta$$

The proof is a straightforward sequence of four games:

- Game 0 is the real game.
- In Game 1, the public key is drawn from $\mathcal{D}_{\text{uniform}}$ instead of $\mathcal{D}_{\text{MLWE}}$. Distinguishing between this and Game 0 is as hard as $\text{MLWE}((\mathbb{Z}/N\mathbb{Z})^{d\times d}, \chi)$.
- In Game 2, the values $((M, A), D)$ are drawn from $\mathcal{D}_{\text{uniform}}$ instead of $\mathcal{D}_{\text{MLWE}}$. Distinguishing between this and Game 2 is as hard as $\text{MLWE}((\mathbb{Z}/N\mathbb{Z})^{(d+1)\times d}, \chi)$.
- Finally in Game 3, E is instead uniformly random, which is within δ of its distribution in Game 2.

The same proof applies to IND-CPA security.

4.2 Backdoored variant of ThreeBears

Since we use explicit rejection, we need a way for the simulator to decrypt an encrypted message by using the random oracle. That is, we need to define a

backdoored version of ThreeBears which differs only in the random oracle. To do this, recall that in the CCA-secure mode the coins for encryption are derived by computing

$$(b, \epsilon_b, \epsilon', \texttt{sharedSecret}) \leftarrow H(\texttt{matrixSeed}, \texttt{encSeed})$$

where with high probability the `matrixSeed` uniquely determines the public key, and thus determines both M and A. Therefore, we can modify the random oracle to work as follows:

- First choose $(b, \epsilon', \texttt{sharedSecret})$ by hashing $(\texttt{matrixSeed}, \texttt{encSeed})$.
- Look up the public key component A based on `matrixSeed`, and compute the ciphertext component

$$E \leftarrow \text{encode}(b^\top A + \epsilon', \texttt{encSeed})$$

- Choose ϵ_b by hashing E with a private random oracle G. There should be a negligible probability of collision on E, so this is very close to a uniformly random function of $(\texttt{matrixSeed}, \texttt{encSeed})$.

We further modify the oracle called for matrix seed expansion to always produce an invertible matrix; this happens anyway with overwhelming probability $> 1 - \frac{1}{N-1}$, because $\mathbb{Z}/N\mathbb{Z}$ is a field. The simulator can now decrypt with no possibility of failure (for well-formed capsules) as follows:

- Hash $\epsilon_b \leftarrow G(E)$ and compute

$$b = M^{-1}(C^\top - \epsilon_b)$$

- Compute $b^\top A$ and $\texttt{encSeed} \leftarrow \text{decode}(b^\top A, E)$. Now $b^\top A$ is close enough to $b^\top A + \epsilon'$ that decoding always produces the correct `encSeed`.
- Finally, check that re-encryption with the recovered `encSeed` produces the same ciphertext, and if not output $\perp$ just like the real decryption algorithm.

5 CCA_2 security

Theorem 1 (Informal)**.** *Let $\mathcal{A}$ be an IND-CCA_2 adversary against the CCA-secure variants of* ThreeBears*. Suppose that $\mathcal{A}$ treats cSHAKE as a random oracle, and makes at most q queries to it. Then there is an adversary $\mathcal{A}'$ using similar resources to $\mathcal{A}$ such that*

$$\begin{aligned} \mathrm{Adv}_{\mathrm{IND-CCA}_2}(\mathcal{A}) &\leq \sqrt{(q+1)\cdot(\mathrm{Adv}_{\mathrm{IND-KPA}}(\mathcal{A}') + 2q^2/2^{8\cdot\texttt{encSeedBytes}})} \\ &+ 2q\sqrt{2^{-8\cdot\texttt{privateKeyBytes}}} + 4q\sqrt{\delta} + \epsilon_0 \end{aligned}$$

where ϵ_0 is negligible and δ is the decryption failure probability. In particular, if $\mathrm{Adv}_{\mathrm{IND-CCA}_2}(\mathcal{A}) \approx 1$ then

$$q \approx \min\left(\sqrt{1/\delta},\ 2^{4\cdot\texttt{privateKeyBytes}},\ 2^{8/3\cdot\texttt{encSeedBytes}},\ 1/\mathrm{Adv}_{\mathrm{IND-KPA}}(\mathcal{A}')\right)$$

Proof. The proof is by a sequence of games. Let $|\psi_i\rangle$ be the adversary's final state before measurement in Game i.

Game 0 Game 0 is the real CCA game.

Game 1 Game 1 is the same as Game 0, except that the simulator backdoors calls to $H(\texttt{matrixSeed}$ of challenge public key$, \ldots)$ using the random oracle as described in Section 4.2. This produces a negligibly-different probability of success, because the backdoored random oracle is within some negligible statistical distance ϵ_0 from uniformly random.

Game 2 Game 2 is the same as Game 1, except that the challenge public key is created using Keygen_R instead of Keygen. Equivalently, it is the same except that the random oracle is changed at the seeds used to create the challenge public key. Then by lemma 1,

$$\mathrm{Exp}\left[||\psi_2\rangle - |\psi_1\rangle|\right] \leq 2q\sqrt{2^{-8\cdot\texttt{privateKeyBytes}}}$$

Game 3 Game 3 is the same as Game 2, except that we modify H so that decryption can't fail. The simulator knows the challenge private key, so it can test whether a given set of coins would cause a decryption failure. It then rejection-samples possible outputs of H until decryption would succeed. This changes H on a δ-fraction of inputs.

Which inputs cause failure in Game 2 is independent of Game 3. By lemma 1,

$$\mathrm{Exp}\left[\||\psi_3\rangle - |\psi_2\rangle|\right] \leq 2q\sqrt{\delta}$$

Game 3′ Game 3′ is the same as Game 3, except that on a decryption query for a ciphertext ct, the oracle recovers m using the backdoor decoder instead of with Dec. Since neither decryption algorithm can fail, this produces the same output, so

$$|\psi_{3'}\rangle = |\psi_3\rangle$$

Game 4 Game 4 is the same as Game 3′, except that H is no longer modified to prevent the ordinary decryption algorithm from failing. The decryption oracle still can't fail, since it uses the backdoor decoder. Again, the inputs on which this changes Game 4 are independent from anything that happens in Game 3′. Then

$$\mathrm{Exp}\left[\||\psi_4\rangle - |\psi_3\rangle|\right] \leq 2q\sqrt{\delta}$$

In Game 4, the simulator doesn't use the private key anymore.

Game 5 Let

$$\mathcal{P}(x) := \begin{cases} 1 & \text{if } x = (\mathrm{pk}, m) \\ 0 & \text{otherwise} \end{cases}$$

In Game 5, the simulator still creates the challenge ciphertext coins using $H(\mathrm{pk}, m)$, but then it runs $\mathcal{A}^{H\setminus\mathcal{P}}$. This takes about the same time as $\mathcal{A}^H$. By lemma 2,

$$\||\psi_5\rangle - |\psi_4\rangle| \leq \sqrt{(q+1)\cdot \Pr\left[\mathcal{A}^{H\setminus\mathcal{P}}(\text{input}) \text{ halts with } (\mathrm{pk}, m)\right]}\Bigg($$

Now $H\backslash\mathcal{P}$ is independent of the challenge coins, since it halts instead of returning when queried on (pk, m). So those coins are uniformly random and independent of everything else in Game 5. In other words, the encryption is equivalent to encrypting with Enc_R.

But now, what would happen if we ran the adversary with a different challenge ciphertext, $\mathrm{Enc}_R(\mathrm{pk}, m')$ where m' is random and unrelated to m? This would create a new input, input$'$, and the real m would be used only to puncture the oracle $H\backslash\mathcal{P}$. By lemma 4,

$$\Pr\left[\mathcal{A}^{H\backslash\mathcal{P}}(\text{input}') \text{ halts with } (\mathrm{pk}, m)\right] \leq 2q^2/\mathrm{card}\,(\mathcal{M})$$

Therefore we can treat $\mathcal{A}^{H\backslash\mathcal{P}}$ as an IND-KPA adversary against $\textsc{ThreeBears}_R$ where

$$\Pr\left[\mathcal{A}^{H\backslash\mathcal{P}}(\text{input}) \text{ halts with } (\mathrm{pk}, \mathrm{seed})\right] \begin{pmatrix} \leq & \mathrm{Adv}_{\text{IND-KPA}}(\mathcal{A}^{H\backslash\mathcal{P}}) \\ + & 2q^2/\mathrm{card}\,(\mathcal{M}) \end{pmatrix}$$

Therefore,[1]

$$\mathrm{Exp}\left[||\psi_5\rangle - |\psi_4\rangle|\right] \leq \sqrt{(q+1)\cdot(\mathrm{Adv}_{\text{IND-KPA}}(\mathcal{A}^{H\backslash\mathcal{P}}) + q^2/2^{8\cdot\texttt{encSeedBytes}})}$$

Summing up Summing the perturbations from Game 1 through Game 5, we have

$$\begin{aligned} \mathrm{Exp}\left[||\psi_5\rangle - |\psi_1\rangle|\right] &\leq 4q\sqrt{\delta} + 2q\sqrt{2^{-8\cdot\texttt{privateKeyBytes}}} + \\ &+ \sqrt{(q+1)\cdot(\mathrm{Adv}_{\text{IND-KPA}}(\mathcal{A}^{H\backslash\mathcal{P}}) + 2q^2/2^{8\cdot\texttt{encSeedBytes}})} \end{aligned}$$

By corollary 6,

$$\mathrm{Adv}_{\text{IND-CCA}_2}(\mathcal{A}) \leq ||\psi_5\rangle - |\psi_1\rangle| + \epsilon_0$$

This completes the proof. Note that we have not assumed that the adversary has only classical access to the decryption oracle. □

[1] We are using AM-QM here, since we have a probability inside the square root and need an expectation outside the square root.

Not also that by lemma 3,

$$\sqrt{\mathrm{Adv}_{\text{OW-CCA}_2}(\mathcal{A})} - \sqrt{2^{8\cdot\texttt{sharedSecretBytes}}} \leq ||\psi_5\rangle - |\psi_1\rangle| + \epsilon_0$$

Let's compare the terms of $||\psi_5\rangle - |\psi_1\rangle|$ to known attacks.

- $2q\sqrt{2^{-8\cdot\texttt{privateKeyBytes}}}$ represents an attack on the key generation seed using Grover's algorithm.
- $4q\sqrt{\delta}$ represents a failure attack, like [HGNP+03]. Here the adversary uses Grover's algorithm [Gro96] to find a ciphertext that causes a decryption failure. In order to use Grover's algorithm, the adversary needs either quantum access to a decryption oracle or a way to predict whether a given decryption will fail. Realistically, the adversary will probably have neither, so this attack should be less powerful than $4q\sqrt{\delta}$.
- ϵ_0 is an artifact of the backdoor technique. It roughly captures our certainty that the adversary cannot successfully encrypt without knowing the message he's encrypting.
- $(q+1)\cdot\mathrm{Adv}_{\text{IND-KPA}}(\mathcal{A})$ represents an attack on the underlying encryption scheme, but it is loose by a factor of $(q+1)$.
- $(q+1)\cdot 2q^2/2^{8\cdot\texttt{encSeedBytes}}$ is an attack on the message used in the challenge encryption. This should work with probability on the order of $q^2/2^{8\cdot\texttt{encSeedBytes}}$, so this term is also loose by a factor of about $q+1$.

6 Acknowledgements

Special thanks to Dominique Unruh and Eike Kiltz for collaboration on the proof outline, and for refining and finding holes in the arguments.

Special thanks to Daniel Kane for his help in proving 5.

Special thanks to Fernando Virdia and Amit Deo for checking our math.

References

[BBBV97] Charles H Bennett, Ethan Bernstein, Gilles Brassard, and Umesh Vazirani. Strengths and weaknesses of quantum computing. *SIAM journal on Computing*, 26(5):1510–1523, 1997.

[BDF+11] Dan Boneh, Özgür Dagdelen, Marc Fischlin, Anja Lehmann, Christian Schaffner, and Mark Zhandry. Random oracles in a quantum world. In Dong Hoon Lee and Xiaoyun Wang, editors, *ASIACRYPT 2011*, volume 7073 of *LNCS*, pages 41–69. Springer, Heidelberg, December 2011.

[BDK+17] Joppe Bos, Léo Ducas, Eike Kiltz, Tancrède Lepoint, Vadim Lyubashevsky, John M. Schanck, Peter Schwabe, and Damien Stehlé. CRYSTALS – kyber: a CCA-secure module-lattice-based KEM. Cryptology ePrint Archive, Report 2017/634, 2017. `http://eprint.iacr.org/2017/634`.

[BV93] Ethan Bernstein and Umesh V. Vazirani. Quantum complexity theory. In *25th ACM STOC*, pages 11–20. ACM Press, May 1993.

[Gro96] Lov K. Grover. A fast quantum mechanical algorithm for database search. In *28th ACM STOC*, pages 212–219. ACM Press, May 1996.

[HGNP+03] Nick Howgrave-Graham, Phong Q. Nguyen, David Pointcheval, John Proos, Joseph H. Silverman, Ari Singer, and William Whyte. The impact of decryption failures on the security of NTRU encryption. In Dan Boneh, editor, *CRYPTO 2003*, volume 2729 of *LNCS*, pages 226–246. Springer, Heidelberg, August 2003.

[SXY17] Tsunekazu Saito, Keita Xagawa, and Takashi Yamakawa. Tightly-secure key-encapsulation mechanism in the quantum

random oracle model. Cryptology ePrint Archive, Report 2017/1005, 2017. http://eprint.iacr.org/2017/1005.

[TU16] Ehsan Ebrahimi Targhi and Dominique Unruh. Post-quantum security of the fujisaki-okamoto and OAEP transforms. In Martin Hirt and Adam D. Smith, editors, *TCC 2016-B, Part II*, volume 9986 of *LNCS*, pages 192–216. Springer, Heidelberg, October / November 2016. doi:10.1007/978-3-662-53644-5_8.

[Unr14] Dominique Unruh. Revocable quantum timed-release encryption. In Phong Q. Nguyen and Elisabeth Oswald, editors, *EUROCRYPT 2014*, volume 8441 of *LNCS*, pages 129–146. Springer, Heidelberg, May 2014. doi:10.1007/978-3-642-55220-5_8.

A Proofs of lemmas

A.1 Proof of lemma 1

Lemma 1 (Perturbation from small random changes)**.** *Let $\mathcal{O}$ and $\mathcal{P}$ be oracles drawn from some joint distribution, both taking input from a set $\mathcal{X}$. Suppose there is some $\epsilon \geq 0$ such that for each $x \in \mathcal{X}$,*

$$\Pr\left[\mathcal{P}(x) \neq \mathcal{O}(x) : \text{input}, \mathcal{O}\right] \leq \epsilon$$

Let $|\psi\rangle$ resp $|\psi'\rangle$ be the final states of $\mathcal{A}^{\mathcal{O}}(\text{input})$ resp $\mathcal{A}^{\mathcal{P}}(\text{input})$. Then

$$\text{Exp}\left[\left||\psi'\rangle - |\psi\rangle\right|\right] \leq 2q\sqrt{\epsilon}$$

If the oracle is classical, then instead

$$\text{Exp}\left[\left||\psi'\rangle - |\psi\rangle\right|\right] \leq \sqrt{2q\epsilon}$$

In both cases, the expectation is over possible oracles $\mathcal{P}$, given a particular $\mathcal{O}$ and input.

Proof. We'll deal with the quantum case first. Let Δ be the set where the oracles differ. Let $|\psi_i\rangle$ be the final state of the $\mathcal{A}^{\mathcal{O}_i}(\text{input})$, where $\mathcal{O}_i$ answers the first i queries from $\mathcal{O}$ and the rest from $\mathcal{P}$. Let

$$\phi_i := \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} z_{xy} \cdot |xy\rangle$$

be its state at the beginning of the ith oracle query. Now

$$|\psi_i\rangle = U(\mathcal{O}(|\phi_i\rangle)) \quad \text{and} \quad |\psi_{i-1}\rangle = U(\mathcal{P}(|\phi_i\rangle))$$

for some unitary transformation U, so that

$$\begin{aligned}
\left||\psi_i\rangle - |\psi_{i-1}\rangle\right|^2 &= \left|\mathcal{O}(|\phi_i\rangle) - \mathcal{P}(|\phi_i\rangle)\right|^2 \\
&= \left|\sum_{x\in\Delta,\ y\in\mathcal{Y}} z_{xy}\cdot|x\rangle\otimes\left(|y\oplus\mathcal{O}(x)\rangle - |y\oplus\mathcal{P}(x)\rangle\right)\right|^2 \\
&= \left|\sum_{x\in\Delta,\ y\in\mathcal{Y}} \left(z_{xy\oplus\mathcal{O}(x)} - z_{xy\oplus\mathcal{P}(x)}\right)\cdot|x\rangle\otimes|y\rangle\right|^2 \\
&= \sum_{x\in\Delta,\ y\in\mathcal{Y}} \left|z_{xy\oplus\mathcal{O}(x)} - z_{xy\oplus\mathcal{P}(x)}\right|^2 \\
&\leq \sum_{x\in\Delta,\ y\in\mathcal{Y}} 4\left|z_{xy}\right|^2
\end{aligned}$$

Then in expectation,

$$\begin{aligned}
\mathrm{Exp}\left[\left|\mathcal{O}(\phi_i) - \mathcal{P}(\phi_i)\right|\right] &\leq \sqrt{\mathrm{Exp}\left[\left|\mathcal{O}(\phi_i) - \mathcal{P}(\phi_i)\right|^2\right]} \\
&\leq \sqrt{\mathrm{Exp}\left[\sum_{x\in\Delta,\ y\in\mathcal{Y}} 4\left|z_{xy}\right|^2\right]} \\
&\leq 2\sqrt{\epsilon}
\end{aligned}$$

The quantum case of the lemma then follows by summing over all i and applying the triangle inequality.

We note that the leading 4 may not be tight here. For example [BBBV97], Theorem 3.3 seems to imply a bound of $q\sqrt{\epsilon}$ instead.

The classical case is easier. We'll use the variable Q to refer to a query log, so that

$$\psi = \sum_Q |\psi_Q\rangle\otimes|Q\rangle \quad \text{and} \quad \psi' = \sum_Q |\psi'_Q\rangle\otimes|Q\rangle$$

Then

$$|\psi'\rangle - |\psi\rangle = \sum_{(x,\mathcal{P}(x))\in Q, x\in\Delta} |\psi_Q\rangle\otimes|Q\rangle - \sum_{(x,\mathcal{O}(x))\in Q, x\in\Delta} |\psi'_Q\rangle\otimes|Q\rangle$$

Because the query logs in the two sums are different – the ones on the left contain $(x, \mathcal{P}(x))$ and the ones on the right $(x, \mathcal{O}(x))$ where $\mathcal{O}(x) \neq \mathcal{P}(x)$ – all the terms in the sums are mutually orthogonal. Therefore

$$\begin{aligned} \mathrm{Exp}\left[\left||\psi'\rangle - |\psi\rangle\right|\right] &\leq \sqrt{\mathrm{Exp}\left[\left||\psi'\rangle - |\psi\rangle\right|^2\right]} \\ &= \sqrt{2\Pr\left[\mathcal{A} \text{ queried anything in } \Delta\right]} \\ &\leq \sqrt{2q\epsilon} \end{aligned}$$

as claimed. □

A.2 Proof of lemma 2

Lemma 2 (Punctured oracles)**.** *Let $\mathcal{O}$ be a quantum oracles for a functions from $\mathcal{X}$ to $\mathcal{Y}$, and $\mathcal{P}$ be an oracle or function from $\mathcal{X}$ to $\{0,1\}$. Let $\mathcal{O}\backslash\mathcal{P}$ denote the oracle*

$$(\mathcal{O}\backslash\mathcal{P})(x) := \begin{cases} \mathcal{O}(x) & \text{if } \mathcal{P}(x) = 0 \\ \mathit{Halt}(x) & \text{if } \mathcal{P}(x) = 1 \end{cases}$$

Let $|\psi\rangle$ resp $|\psi'\rangle$ be the final states of $\mathcal{A}^{\mathcal{O}}(\text{input})$ resp $\mathcal{A}^{\mathcal{O}\backslash\mathcal{P}}$. Then

$$\left||\psi'\rangle - |\psi\rangle\right| \leq \sqrt{(q+1)\cdot\Pr\left[\mathcal{A}^{\mathcal{O}\backslash\mathcal{P}}(\text{input}) \textit{ halts with } x : \mathcal{P}(x) = 1\right]}$$

Proof. Let's start by defining an algorithm $\mathcal{A}_{\mathrm{c}}^{\mathcal{O},\mathcal{P}}$ that counts how many times $\mathcal{A}$ queries values x such that $\mathcal{P}(x) = 1$. It allocates $\lceil \log_2 q \rceil$ extra qubits, which are all 0 in the input and are unused by $\mathcal{A}$.[2] After each query x to $\mathcal{O}$, $\mathcal{A}_{\mathrm{c}}^{\mathcal{O},\mathcal{P}}$ applies a unitary Count transform to x and the counter:

$$\mathrm{Count}(|x\rangle \otimes |i\rangle) = \begin{cases} |x\rangle \otimes |i+1\rangle & \text{for each } x \text{ where } \mathcal{P}(x) = 0 \\ |x\rangle \otimes |i\rangle & \text{for each } x \text{ where } \mathcal{P}(x) = 1 \end{cases}$$

Consider the state $|\psi''\rangle$ at the end of $\mathcal{A}_{\mathrm{c}}^{\mathcal{O},\mathcal{P}}(\text{input})$). This may be written

$$|\psi''\rangle = \sum_{i=0}^{q} |\psi_i''\rangle \otimes |i\rangle$$

[2] Technically this means $|\psi\rangle = \mathcal{A}^{\mathcal{O}}(\text{input}) \otimes |0\rangle$ instead of just $\mathcal{A}^{\mathcal{O}}(\text{input})$.

Because Count only changes the counter bits, we must have

$$|\psi\rangle = \sum_{i=0}^{q} |\psi_i''\rangle \otimes |0\rangle$$

Therefore,

$$\begin{aligned}
\big|\, |\psi''\rangle - |\psi\rangle \,\big|^2 &= \left|\sum_{i=1}^{q} |\psi_i''\rangle\right|^2 + \sum_{i=1}^{q} \big||\psi_i''\rangle\big|^2 \\
&\leq \sum_{i=1}^{q} \Big(\big|\,|\psi_i''\rangle\big|\Big)^2 + \sum_{i=1}^{q} \big|\psi_i''\rangle\big|^2 \quad \text{(triangle inequality)} \\
&\leq q \cdot \sum_{i=1}^{q} \big|\psi_i''\rangle\big|^2 + \sum_{i=1}^{q} \big||\psi_i''\rangle\big|^2 \quad \text{(AM-QM inequality)} \\
&= (q+1) \sum_{i=1}^{q} \big|\psi_i''\rangle\big|^2
\end{aligned}$$

For the searching algorithm $\mathcal{A}^{\mathcal{O}\backslash\mathcal{P}}$, instead the oracle halts when $\mathcal{P}(x) = 1$. Let's do the same calculations for its final state $|\psi'\rangle$. Let $\mathcal{H}_1$ be the set of halt-oracle transcripts where the system halted during a call to $\mathcal{O}\backslash\mathcal{P}$, and $\mathcal{H}_0$ the set where it did not. Expand

$$|\psi'\rangle = \sum_{h \in \mathcal{H}_0 \cup \mathcal{H}_1} |\psi_h'\rangle \otimes |i\rangle$$

where $\sum_{h \in \mathcal{H}_0} |\psi_h'\rangle = |\psi_0''\rangle$ by construction. Since

$$\sum_{i=0}^{q} \big|\psi_i''\rangle\big|^2 = \sum_{h \in \mathcal{H}_0 \cup \mathcal{H}_1} \big||\psi_h'\rangle\big|^2 = 1$$

we must also have

$$\sum_{i=1}^{q} \big|\psi_i''\rangle\big|^2 = \sum_{h \in \mathcal{H}_1} \big||\psi_h'\rangle\big|^2$$

Therefore

$$\big|\, |\psi'\rangle - |\psi\rangle \,\big|^2 = \left|\sum_{i=1}^{q} |\psi_i''\rangle\right|^2 + \sum_{h \in \mathcal{H}_1} \big||\psi_h'\rangle\big|^2$$

where

$$\left|\sum_{i=1}^{q} |\psi_i''\rangle\right|^2 \leq q \cdot \sum_{i=1}^{q} \left||\psi_i''\rangle\right|^2 = q \cdot \sum_{h \in \mathcal{H}_1} \left||\psi_h'\rangle\right|^2$$

so that

$$\left|\, |\psi'\rangle - |\psi\rangle \,\right|^2 \leq (q+1) \sum_{h \in \mathcal{H}_1} \left||\psi_h'\rangle\right|^2$$

But $\sum_{h \in \mathcal{H}_1} ||\psi_h'\rangle|^2$ is exactly the probability that $\mathcal{O}\backslash\mathcal{P}$ halted. And if does halt, then by construction it halts with some value x such that $\mathcal{P}(x) = 1$. Taking the square root of both sides gives

$$\left||\psi'\rangle - |\psi\rangle\right| \leq \sqrt{(q+1) \cdot \Pr\left[\mathcal{B}^{\mathcal{O}\backslash\mathcal{P}}(\text{input}) \text{ halts with } x : \mathcal{P}(x) = 1\right]}$$

as claimed. □

A.3 Proof of lemma 3

Lemma 3 (Perturbation to small probabilities)**.** *Suppose that the adversary's goal is to output a value in some set* Goal*, and suppose that it ends up in a quantum state that's either $|\psi\rangle$ or $|\psi'\rangle$. Then*

$$\left| \sqrt{\Pr\left[\text{measure}(|\psi\rangle) \in \text{Goal}\right]} - \sqrt{\Pr\left[\text{measure}(|\psi'\rangle) \in \text{Goal}\right]} \right| \leq \left||\psi'\rangle - |\psi\rangle\right|$$

Proof. Let P be the projection map to $\text{span}(|G\rangle : G \in \text{Goal})$. Then

$$\sqrt{\Pr[\mathcal{A}^{\mathcal{O}}(\text{input}) \in \text{Goal}]} = |P(|\psi\rangle)|$$

and likewise for $|\psi'\rangle$. Since P is a contraction map,

$$\left||P(|\psi'\rangle) - |P(|\psi\rangle)|\right| \leq \left|P(|\psi'\rangle) - P(|\psi\rangle)\right| \leq \left||\psi'\rangle - |\psi\rangle\right|$$

as claimed. □

A.4 Proof of lemma 4

Lemma 4 (Random puncturing)**.** *Let $\mathcal{O}$ resp $\mathcal{P}$ be chosen from some joint distribution of functions $\mathcal{X} \to \mathcal{Y}$ resp $\mathcal{X} \to \{0,1\}$. Suppose there is some $\epsilon \geq 0$ such that for all $x \in \mathcal{X}$,*

$$\Pr\left[\mathcal{P}(x) = 1 : \mathcal{O}, \text{input}\right] \leq \epsilon$$

Then

$$\Pr\left[\mathcal{A}^{\mathcal{O}\backslash\mathcal{P}}(\text{input}) \text{ halts with } x : \mathcal{P}(x) = 1\right] \leq 2q^2\epsilon$$

In particular, if $\mathcal{P}(x) = 1$ for exactly one input x which is independent of input *and $\mathcal{O}$, then*

$$\Pr\left[\mathcal{A}^{\mathcal{O}\backslash\mathcal{P}}(\text{input}) = (x, n)\right] \leq 2q^2/\text{card}\left(\mathcal{X}\right)$$

Proof. We follow the exact same proof as for lemma 1, except that in the step

$$\left|\sum_{x\notin\Delta,\ y\in\mathcal{Y}} z_{xy} \cdot |x\rangle \otimes \left(|y \oplus \mathcal{O}(x)\rangle - |y \oplus (\mathcal{O}\backslash\mathcal{P})(x)\rangle\right)\right|^2 \leq \sum_{x\in\Delta,\ y\in\mathcal{Y}} 4\left|z_{xy}\right|^2$$

we instead have

$$\left|\sum_{x\notin\Delta,\ y\in\mathcal{Y}} z_{xy} \cdot |x\rangle \otimes \left(|y \oplus \mathcal{O}(x)\rangle \otimes |0\rangle - |y\rangle \otimes |1\rangle\right)\right|^2 \leq \sum_{x\in\Delta,\ y\in\mathcal{Y}} 2\left|z_{xy}\right|^2$$

which is a factor of 2 tighter. Now $\mathcal{A}^{\mathcal{O}}$ doesn't halt during $\mathcal{O}$ queries, so by lemma 3 the probability that $\mathcal{B}^{\mathcal{O}\backslash\mathcal{P}}$ halts in this way is at most $||\psi'\rangle - |\psi\rangle|^2 \leq 2q^2\epsilon$ as claimed. □

A.5 Proof of lemma 5

Lemma 5 (Perturbation to L^1 distance)**.** *Let $|\psi\rangle$ and $|\psi'\rangle$ be quantum states, and $\mathcal{D}$ and $\mathcal{D}'$ be the distributions produced by measuring them, and let*

$$\left|\mathcal{D}' - \mathcal{D}\right|_1 := \sum_{d\in\mathcal{D}\cup\mathcal{D}'} \left|\Pr(d \leftarrow \mathcal{D}') - \Pr(d \leftarrow \mathcal{D})\right|$$

be the L^1 statistical distance between these distributions. Then

$$\left|\mathcal{D}' - \mathcal{D}\right|_1 \leq 2 \cdot \left|\left|\psi'\right\rangle - \left|\psi\right\rangle\right|$$

Proof. To rephrase this, let $|\psi_1\rangle$ and $|\psi_2\rangle$ be unit vectors, and $\{e_i\}$ be an orthonormal basis. We wish to show that

$$\sum_i \left|\langle e_i, |\psi_1\rangle\rangle^2 - \langle e_i, |\psi_2\rangle\rangle^2\right| \leq 2 \cdot ||\psi_1\rangle - |\psi_2\rangle|$$

Let α and β be any two unit vectors; then since for all real x, y we have $x \cdot y \leq \frac{1}{2}(x^2 + y^2)$, we also have

$$\begin{aligned}
\sum_i |\langle \alpha, e_i\rangle \cdot \langle \beta, e_i\rangle| &\leq \frac{1}{2}\sum_i \left|\langle \alpha, e_i\rangle\right|^2 + \left|\langle \beta, e_i\rangle\right|^2 \\
&= \frac{1}{2}(|\alpha|^2 + |\beta|^2) \\
&= 1
\end{aligned}$$

Plugging in

$$\alpha, \beta := \frac{|\psi_1\rangle + |\psi_2\rangle}{||\psi_1\rangle + |\psi_2\rangle|}, \frac{|\psi_1\rangle - |\psi_2\rangle}{||\psi_1\rangle - |\psi_2\rangle|}$$

we get

$$\sum_i \frac{\left|\langle e_i, \psi_1\rangle^2 - \langle e_i, \psi_2\rangle^2\right|}{||\psi_1\rangle + |\psi_2\rangle| \cdot ||\psi_1\rangle - |\psi_2\rangle|} = \sum_i \left|\langle \alpha, e_i\rangle \cdot \langle \beta, e_i\rangle| \leq 1\right.$$

We complete the proof by multiplying both sides by $||\psi_1\rangle + |\psi_2\rangle| \cdot ||\psi_1\rangle - |\psi_2\rangle|$, and noting that $||\psi_1\rangle + |\psi_2\rangle| \leq 2$.

Special thanks to Daniel Kane for the key ingredients of this proof. □