Quantum Collision-Resistance of Non-Uniformly Distributed Functions

Ehsan Ebrahimi Targhi, Gelo Noel Tabia, and Dominique Unruh

University of Tartu, Estonia

Abstract. We study the quantum query complexity of finding a collision for a function f whose outputs are chosen according to a distribution with min-entropy k. We prove that $\Omega(2^{k/9})$ quantum queries are necessary to find a collision for function f. This is needed in some security proofs in the quantum random oracle model (e.g. Fujisaki-Okamoto transform).

Keywords: Quantum, Collision, Non-uniform Distribution, Query Complexity.

1 Introduction

Let D be a distribution with min-entropy k over set Y and f be a function whose outputs are drawn according to the distribution D. In this paper, we study the difficulty of finding a collision for unknown function f in the quantum query model. Recall that a collision for function f consists of two distinct inputs x_1 and x_2 such that $f(x_1) = f(x_2)$. Classically, by application of the birthday attack it is easy to observe that $\Theta(2^{k/2})$ queries are necessary and sufficient to find a collision with constant probability. However, in quantum query model this number of queries may be high for the reason that one quantum query may contain the whole input-output values of the function.

Zhandry [Zha15] shows that $\Theta(2^{k/3})$ quantum queries are necessary and sufficient to find a collision for the function f when D is a uniform distribution. However, he leaves the non-uniform case as an open problem. One motivation for studying the quantum collision problem for a non-uniform distribution is the interest in proving the security of classical cryptographic schemes against quantum adversaries. Hash functions are crucial cryptographic primitives that are used to construct many encryption schemes and cryptographic schemes. They are usually modeled as random functions and they are used inside other functions. Therefore the output of combination of a function f and a random function H may not be distributed uniformly and finding a collision for this non-uniformly distributed $f \circ H$ may break the security of the scheme. For example the well-known Fujisaki-Okamoto construction [FO99] uses a random function H to produce the randomness for an encryption scheme f. The security relies on the fact that the adversary can not find two inputs of the random function that lead to the same ciphertext. This is roughly equivalent to saying that $f \circ H$ is collision-resistant. In

fact, our result is a crucial ingredient for analyzing a variant of Fujisaki-Okamoto construction in the quantum setting [ETU15].

We prove an $\Omega(2^{k/9})$ lower bound for the quantum query complexity of the function f and leave as an open problem to verify whether or not Zhandry's bound applies to the function f. The proof procedure is as follows. We apply the Leftover Hash Lemma [HILL93] to the function f to extract the number of bits that are indistinguishable from uniformly random bits. After applying the Leftover Hash Lemma, the output distribution of $h \circ f$, where h is a universal hash function, is indistinguishable from the uniform distribution over a set. Note that a collision for function f is a collision for $h \circ f$. Let A be a quantum adversary that has quantum access to f and finds a collision for $h \circ f$. Using the existence of A, we show that there exists a quantum algorithm B that has quantum access to $h \circ f$ and finds a collision for $h \circ f$ with the same probability and the same number of queries as algorithm A. Theorem 1.1 by Zhandry [Zha12] shows that two distribution are indistinguishable if and only if they are oracle-indistinguishable. Therefore, $h \circ f$ is indistinguishable from a random function (recall that the output of $h \circ f$ is indistinguishable from the uniform distribution by Leftover Hash Lemma) and as a result any quantum algorithm B is unable to differentiate between $h \circ f$ and a random function. By using an existing result for finding a collision for a random function presented by Zhandry [Zha15, Theorem 7], we obtain an upper bound for the probability of finding a collision for function $h \circ f$. Therefore, we get an upper bound for the probability of success for the quantum collision problem applied to the function f.

The quantum collision problem has been studied in various previous works. In the following, we mention the existing results on the number of queries that are necessary to find a collision. An $\Omega(N^{1/3})$ lower bound for function f is given by Aaronson and Shi [AS04] and Ambainis [Amb05] where f is a two-to-one function with the same domain and co-domain and N is the domain size. Yuen [Yue14] proves an $\Omega(N^{1/5}/\text{polylog}N)$ lower bound for the quantum collision problem for a random function f with same domain and co-domain. He reduces the distinguishing between a random function and a random permutation problem to the distinguishing between a function with r-to-one part and a function without r-to-one part. His proof is a merger of using the r-to-one lower bound from [AS04] and using the quantum adversary method [Amb00]. Zhandry [Zha15] improves Yuen's bound to the $\Omega(N^{1/3})$ and also removes the same size domain and co-domain constraint. He uses the existing result from [Zha12] to prove his bound.

The sufficient number of quantum queries to find a collision is given in the following works. A quantum algorithm that requires $O(N^{1/3})$ quantum queries and finds a collision for any two-to-one function f with overwhelming probability is given by Brassard, Hϕyer and Tapp [BHT97]. Ambainis [Amb07] gives a quantum algorithm that requires $O(N^{2/3})$ queries to find two equal elements among N given elements and therefore it is an algorithm for finding a collision in an arbitrary function f given the promise that f has at least one collision.

Yuen [Yue14] shows that the collision-finding algorithm from [BHT97] is able to produce a collision for a random function with same domain and co-domain using $O(N^{1/3})$ queries. Zhandry shows that $O(M^{1/3})$ queries are adequate to find a collision for a random function $f : [N] \to [M]$ where $N = \Omega(M^{1/2})$. He uses Ambainis's element distinctness algorithm [Amb07] as a black box in his proof. Zhandry's bound also implies that we can not expect a lower bound for the query complexity of finding a collision for a non-uniform function better than $O(2^{k/3})$.

2 Preliminaries

In this section, we present some definitions and existing results that are needed in this paper. Notation $x \xleftarrow{\$} X$ shows that x is chosen uniformly at random from set X. If D is a distribution over X, then notation $x \leftarrow D$ shows that x is chosen at random according to the distribution D. $\Pr[P : G]$ is the probability that the predicate P holds true where free variables in P are assigned according to the program in G. We say that the quantum algorithm A has quantum access to the oracle $O : \{0,1\}^{n_0} \to \{0,1\}^{n_1}$, denoted by A^O, where A can submit queries in superposition and the oracle O answers to the queries by a unitary transformation that maps $|x,\ y\rangle$ to $|x,\ y \oplus O(x)\rangle$.

Definition 1. *Let D_1 and D_2 be distributions on a set X. The statistical distance between D_1 and D_2 is*

$$\mathrm{SD}(D_1, D_2) = \frac{1}{2} \sum_{x \in X} \big| \Pr[D_1(x)] - \Pr[D_2(x)] \big|.$$

Definition 2. *Let D be a distribution on a set X. The min-entropy of this distribution is defined as*

$$H_\infty(D) = -\log \max_{x \in X} \Pr[D(x)].$$

Definition 3. *We say that function $f : \{0,1\}^{n_1} \to \{0,1\}^{n_2}$ has min-entropy k if,*

$$-\log \max_{y \in \{0,1\}^{n_2}} \Pr[y = f(x) : x \xleftarrow{\$} \{0,1\}^{n_1}] = k.$$

Definition 4 (Universal Hash Function [CW79]). *A family of functions $H = \{h : \{0,1\}^n \to \{0,1\}^m\}$ is called a universal family if for all distinct $x, y \in \{0,1\}^n$:*

$$\Pr[h(x) = h(y) : h \xleftarrow{\$} H] \leq 1/2^m.$$

Lemma 1 (Leftover Hash Lemma [HILL93]). *Let D be a distribution with min-entropy k and e be a positive integer. Let $h : \{0,1\}^m \times \{0,1\}^n \to \{0,1\}^{k-2e}$ be a universal hash function. Then,*

$$\mathrm{SD}\Big(\big(h(y,x), y\big), \big(z, y\big)\Big) \leq 2^{-e-1}$$

where $x \xleftarrow{D} \{0,1\}^n, y \xleftarrow{\$} \{0,1\}^m$ *and* $z \xleftarrow{\$} \{0,1\}^{k-2e}$.

Lemma 2 ([Zha12]). *Let* D_1 *and* D_2 *be efficiently sampleable distributions over some set* Y*, and let* X *be some other set. For* $i = 1, 2$*, let* D_i^X *be the distributions of functions* F_i *from* X *to* Y *where for each* $x \in X$*,* $F_i(x)$ *is chosen at random according to the distribution* D_i*. Then if* A *be a quantum algorithm that makes* q *queries and distinguish* D_1^X *from* D_2^X *with non-negligible probability* ϵ*, we can construct a quantum algorithm* B *that distinguishes samples from* D_1 *and* D_2 *with probability at least* $\frac{3\epsilon^2}{64\pi^2 q^3}$.

Lemma 3 (Theorem 7 [Zha15]). *Let* $h : \{0,1\}^n \to \{0,1\}^m$ *be a random function. Then any quantum algorithm making* q *number of queries to* h *outputs a collision for* h *with probability at most* $\frac{C(q+2)^3}{2^m}$ *where* C *is a universal constant.*

3 Main Result

Let $\Pr[\mathsf{Coll}(O; A^O) : O \leftarrow D]$ be the probability of finding a collision in function O that is drawn according to the distribution D using a quantum algorithm A with quantum access to the function O.

Lemma 4. *Let* D *be a distribution over* $\{0,1\}^{n_1}$*. Let* $f : \{0,1\}^{n_1} \to \{0,1\}^{n_2}$ *be a public function and* $X = \{0,1\}^{n_0}$*. If* A *is a quantum algorithm that makes* q *queries to function* O *drawn from distribution* D^X *and finds a collision for* $f \circ O$ *with some probability, then there exists a quantum algorithm* B *that makes* q *queries to* $f \circ O$ *and outputs a collision for* $f \circ O$ *with the same probability.*

Proof. Let $S_y = f^{-1}(\{y\})$ for $y \in \operatorname{Im} f$. We define distribution D_y over S_y as

$$\Pr[D_y(z)] := \frac{\Pr[D(z)]}{\sum_{z \in S_y} \Pr[D(z)]}.$$

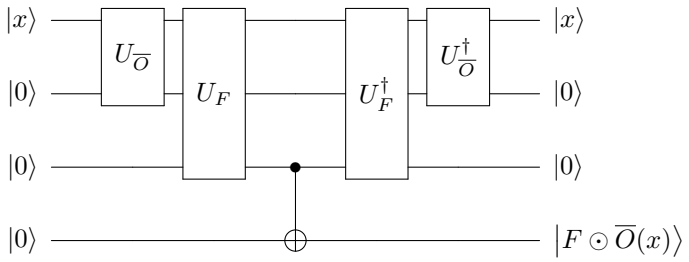
Let D' be the distribution of functions F from $\{0,1\}^{n_0} \times \operatorname{Im} f$ to $\{0,1\}^{n_1}$ where for each $x \in \{0,1\}^{n_0}$ and $y \in \operatorname{Im} f$, $F(x,y)$ is chosen at random in S_y according to the distribution D_y. Let $(F \odot g)(x) := F\big(x, g(x)\big)$. We show that output of O and output of $F \odot (f \circ O)$ have the same distribution when F is chosen according to distribution D'. For every $x \in \{0,1\}^{n_0}$ and $z \in \{0,1\}^{n_1}$:

$$\begin{aligned}
&\Pr[\big(F \odot (f \circ O)\big)(x) = z : \ O \leftarrow D^X, \ F \leftarrow D'] \\
&\quad = \Pr[F\Big(x, \ f\big(O(x)\big)\Big) = z : \ O \leftarrow D^X, \ F \leftarrow D'] \\
&\quad = \Pr[F\big(x, \ f(z')\big) = z : \ z' \leftarrow D, \ F \leftarrow D'] \\
&\quad = \Pr[z'' = z : \ z' \leftarrow D, \ z'' \leftarrow D_{f(z')}] \\
&\quad \overset{(*)}{=} \Pr[z'' = z \wedge z' \in S_{f(z)} : \ z' \leftarrow D, \ z'' \leftarrow D_{f(z')}] \\
&\quad \overset{(**)}{=} \Pr[z' \in S_{f(z)} : \ z' \leftarrow D] \Pr[z'' = z : \ z'' \leftarrow D_{f(z)}] \\
&\quad = \Big(\sum_{z' \in S_{f(z)}} \Pr[D(z')]\Big) \cdot \frac{\Pr[D(z)]}{\sum_{z' \in S_{f(z)}} \Pr[D(z')]} = \Pr[D(z)],
\end{aligned}$$

where $(*)$ holds for the reason that if $z'' = z$ be true, then z' will be in the set $S_{f(z)}$ and $(**)$ uses the conditional probability. As a result:

$$\Pr[\mathsf{Coll}(f \circ O; A^O) : O \leftarrow D^X] = \Pr[\mathsf{Coll}(f \circ O; A^{F \odot f \circ O}) : O \leftarrow D^X, F \leftarrow D'].$$

Now, we construct quantum algorithm B. Algorithm B runs A and answers to its query as follows: (i) query $(f \circ O)(x) := y$, (ii) pick $z \leftarrow D_y$, and (iii) set $O(x) := z$. That is, B runs $A^{F \odot f \circ O}$ with $F \leftarrow D'$. Let $\overline{O} = f \circ O$. The way that quantum algorithm B handles quantum queries is shown in the following circuit.



Algorithm B returns the output of A after q queries. Therefore, we prove the existence of quantum algorithm B stated in the lemma.

Theorem 1. *Let D be a distribution with $H_\infty(D) \geq k$ over set $\{0,1\}^{n_1}$. Let O be a function drawn from distribution D^X. Then any quantum algorithm A making q queries to O returns a collision for O with probability at most $\frac{C'(q+2)^{9/5}}{2^{k/5}}$ where C' is a universal constant. That is,*

$$\Pr[\mathsf{Coll}(O; A^O) : O \leftarrow D^X] \leq \frac{C'(q+2)^{9/5}}{2^{k/5}}.$$

Let $h : \{0,1\}^m \times \{0,1\}^{n_1} \to \{0,1\}^{k-2e}$ be a universal hash function. Lemma 1 implies that:

$$\mathrm{SD}(h_y(x), z) \leq 2^{-e-1} \tag{1}$$

where $h_y(x) := h(y,x)$, $x \leftarrow D$, $y \xleftarrow{\$} \{0,1\}^m$ and $z \xleftarrow{\$} \{0,1\}^{k-2e}$.
The upper bound can be concluded by following steps:

$$\begin{aligned}
&\Pr[\mathsf{Coll}(O; A^O) : O \leftarrow D^X] \\
&\overset{(i)}{\leq} \Pr[\mathsf{Coll}(h_y \circ O; A^O) : O \leftarrow D^X] \\
&\overset{(ii)}{=} \Pr[\mathsf{Coll}(h_y \circ O; B^{h_y \circ O}) : O \leftarrow D^X] \\
&\overset{(iii)}{\leq} \Pr[\mathsf{Coll}(O^*; B^{O^*}) : O^* \xleftarrow{\$} (\{0,1\}^{n_1} \to \{0,1\}^{k-2e})] + \sqrt{64\pi^2 q^3 2^{-e-1}/3} \\
&\overset{(iv)}{\leq} \frac{C(q+2)^3}{(2^{k-2e})} + \sqrt{\frac{64\pi^2 q^3}{3(2^{e+1})}}
\end{aligned}$$

where

(i) follows from the fact that collisions for O will also be collisions for $h_y \circ O$, and that $h_y \circ O$ can have other collisions;
(ii) follows from Lemma 4 that implies the existence of quantum algorithm B;
(iii) can be seen as follows: Let D_1 be output distribution of $h_y \circ O$ and D_2 be uniform distribution over $\{0,1\}^{k-2e}$. Equation 1 implies that for every adversary A,

$$|\Pr[A(y) = 1 : y \leftarrow D_1] - \Pr[A(y) = 1 : y \leftarrow D_2]| \leq 2^{-e-1}.$$

Using Lemma 2, we can conclude that

$$\Big| \Pr[\mathsf{Coll}(h_y \circ O; B^{h_y \circ O}) : O \leftarrow D^X] - \Pr[\mathsf{Coll}(O^*; B^{O^*}) : O^* \xleftarrow{\$} (\{0,1\}^{n_1} \to \{0,1\}^{k-2e})] \Big| \leq \sqrt{64\pi^2 q^3 2^{-e-1}/3};$$

and finally
(iv) follows from applying Lemma 3 to the random function O^*.

So far, we have the upper bound

$$\eta_e := \frac{2^{2e}\mu}{2^k} + \frac{\nu}{2^{e/2}}, \qquad \text{where } \mu := C(q+2)^3 \text{ and } \nu := \frac{8\pi q^{3/2}}{\sqrt{6}}.$$

It is minimized by choosing

$$e = \frac{2}{5}k + \frac{2}{5}\log\frac{\nu}{4\mu}.$$

Substituting this value of e gives us

$$\Pr[\mathsf{Coll}(O; A^O) : O \leftarrow D^X] \leq \frac{2^{2/5}\mu^{1/5}\nu^{4/5}}{2^{k/5}} \leq \frac{C'(q+2)^{9/5}}{2^{k/5}}.$$

.

Corollary 1. *Let $f : \{0,1\}^{n_1} \to \{0,1\}^{n_2}$ be a function with min-entropy k. Let $O : \{0,1\}^* \to \{0,1\}^{n_1}$ be a random function. Then any quantum algorithm A making q queries to O returns a collision for $f \circ O$ with probability at most $O\left(\frac{q^{9/5}}{2^{k/5}}\right)$.*

We apply Lemma 4 to obtain the quantum algorithm B that has access to $f \circ O$ and finds a collision for $f \circ O$ with the same number of queries and the same probability as the quantum algorithm A. Then the result follows by Theorem 1 for the reason that the output distribution of $f \circ O$ has min-entropy k.

4 Acknowledgments

We would like to thank the anonymous reviewers for their comments. This work was supported by the Estonian ICT program 2011-2015 (3.2.1201.13-0022), the

European Union through the European Regional Development Fund through the sub-measure "Supporting the development of R&D of info and communication technology", by the European Social Fund's Doctoral Studies and Internationalisation Programme DoRa, by the Estonian Centre of Excellence in Computer Science, EXCS.

References

Amb00. Andris Ambainis. Quantum lower bounds by quantum arguments. In F. Frances Yao and Eugene M. Luks, editors, *Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing, May 21-23, 2000, Portland, OR, USA*, pages 636–643. ACM, 2000.

Amb05. Andris Ambainis. Polynomial degree and lower bounds in quantum complexity: Collision and element distinctness with small range. *Theory of Computing*, 1(1):37–46, 2005.

Amb07. Andris Ambainis. Quantum walk algorithm for element distinctness. *SIAM J. Comput.*, 37(1):210–239, 2007.

AS04. Scott Aaronson and Yaoyun Shi. Quantum lower bounds for the collision and the element distinctness problems. *J. ACM*, 51(4):595–605, July 2004.

BHT97. Gilles Brassard, Peter Høyer, and Alain Tapp. Quantum algorithm for the collision problem. *ACM SIGACT News (Cryptology Column)*, 28:14–19, 1997.

CW79. Larry Carter and Mark N. Wegman. Universal classes of hash functions. *J. Comput. Syst. Sci.*, 18(2):143–154, 1979.

ETU15. Ehsan Ebrahimi Targhi and Dominique Unruh. Quantum security of the Fujisaki-Okamoto transform. `http://2015.qcrypt.net/wp-content/uploads/2015/09/Poster10_Ehsan-Ebrahimi.pdf`, 2015.

FO99. Eiichiro Fujisaki and Tatsuaki Okamoto. Secure integration of asymmetric and symmetric encryption schemes. In *Proceedings of the 19th Annual International Cryptology Conference on Advances in Cryptology*, CRYPTO '99, pages 537–554, London, UK, UK, 1999. Springer-Verlag.

HILL93. Johan Håstad, Russell Impagliazzo, Leonid A. Levin, and Michael Luby. Construction of a pseudo-random generator from any one-way function. *SIAM Journal on Computing*, 28:12–24, 1993.

Yue14. Henry Yuen. A quantum lower bound for distinguishing random functions from random permutations. *Quantum Information & Computation*, 14(13-14):1089–1097, 2014.

Zha12. Mark Zhandry. How to construct quantum random functions. In *53rd Annual IEEE Symposium on Foundations of Computer Science, FOCS 2012, New Brunswick, NJ, USA, October 20-23, 2012*, pages 679–687. IEEE Computer Society, 2012.

Zha15. Mark Zhandry. A note on the quantum collision and set equality problems. *Quantum Information & Computation*, 15(7&8):557–567, 2015.