Hash function in quantum world

Hash function은 quantum computing으로부터 안전한가?

본 보고서는 hash function이 quantum computing으로 부터 안전하다는 것을 증명하는 보고서입니다. 여기서는 구체적인 증명보다는 아이디어 위주로 서술하고, 구체적인 증명 과정은 해당 논문을 참고하시길 바랍니다

1. 안전한 hash function 이란?

Hash function은 임의의 길이의 데이터를 고정된 길이의 output으로 반환하는 함수이다. 이 함수는 암호학에서 데이터의 integrity를 증명하는데 사용된다. 데이터의 integrity란 데이터가 전송 과정 중에 손실되지 않는 것을 의미한다. Hash function이 integrity를 보장하기 위해서는 다음 3가지 속성을 만족해야 한다.

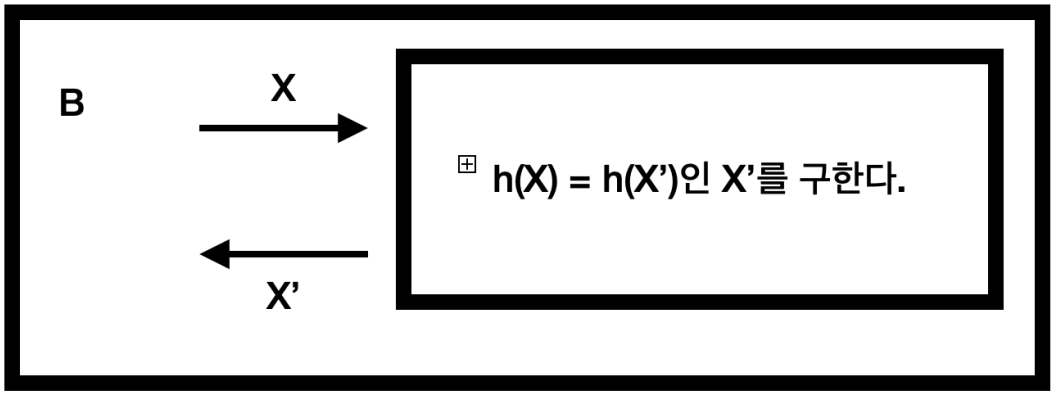
Hash function을 h, input data가 x일 때, 그 결과를 y라고 하자. (h(x) = y)

1) pre-image resistance: 임의의 y가 주어졌을 때, h(x) = y 를 찾을 확률은 negligible해야 한다.
2) 2nd-pre-image resistance: 임의의 x가 주어졌을 때, x != x'인 h(x) = h(x')을 찾을 확률은 negligible해야 한다.
3) Collision resistance: h(x) = h(x')인 서로 다른 x, x'를 찾을 확률은 negligible해야 한다.

(negligible: 모든 양의 다항식 p에 대해, 충분히 큰 n에서는 항상 f(n) < 1/p(n) 을 만족할 정도로 0으로 너무 빨리 작아져서, 어떤 다항식으로 곱해도 0에 수렴하는 함수, 즉, 다항시간 만큼 시도한다면 성공할 확률이 거의 0이라는 것을 의미한다. 예를 들어, $1/2^n$ 등이 있다.)

위의 세가지 속성 중 우리가 증명하고자 하는 속성은 collision resisitance이다. 이유는 어떤 함수가 collision resistance를 만족하면, pre-image resistance와 2nd-pre-image resistance를 만족하기 때문이다. 이제 이를 증명하겠다.

1. 임의의 function이 collision resistance하면, 2nd-pre-image resistance하다.
증명: 대우법을 사용해 증명하겠다. 즉, 2nd-pre-image resistance하지 않다면, collision resistance하지 않는다를 증명하겠다. 2nd-pre-image 문제를 해결하는 공격자 A가 있다고 할 때, B를 다음과 같이 모델링하자.



1) B는 임의의 message X를 선택한다. 2) X를 A에게 보내고 A는 h(X) = h(X')인 X'를 구한다. (2nd-pre-image) 3) B는 (X, X')를 반환한다. (collision)

즉, 2nd-pre-image를 non-negligible 할 확률로 찾을 수 있으면, collision도 non-negligible 할 확률로 찾을 수 있으므로 증명되었다.

2. 함수의 range size가 domain size에 비해 충분히 작다고 하자. 그러면 collision-resistance하면 pre-image resistance하다.

Appendix B.1에 증명되어있다.[1]

2. preliminaries

Reduction이란 우리가 이해하기 어려운 문제를 이해하기 쉬운 문제로 바꾸는 것을 의미한다. 정의는 다음과 같다.

$f : X \to Y$ 라고 하자. 모든 $x \in X$ 에 대해 $f(x) = g(p(x))$ 인 다항시간 알고리즘 p(x)가 존재하면, f를 g로 reduce할 수 있다고 한다. 기호로는 $f \leq_p g$ 라고 한다. 이것은 g를 풀 수 있다면, f도 풀 수 있다는 것을 의미한다.

reduction은 암호학에서 자주 사용된다. 예를 들어, RSA 문제를 푸는 것을 factoring으로 변환해, 양자 컴퓨터에서 RSA는 안전하지 않음을 증명한다.

Symmetric function이란 입력의 순서는 상관없고, 1의 개수에 따라서 함수값이 결정되는 boolean function이다. 정의는 다음과 같다.

$f : \{0,1\}^n \to \{0,1\}$이고, $x = (x_1, x_2, \ldots, x_n) \in \{0,1\}^n$일 때, x의 1의 개수를 i 라고 하자. 모든 x에 대해 f(x)가 오직 i에 의해 결정되면, f를 Symmetric function이라고 한다.

임의의 function을 symmetric function으로 바꾸는 과정은 input의 가능한 모든 순열에 대해 합을 구한 뒤 평균을 내는 것이다. 식으로 쓰면 다음과 같다.
input이 주어졌을 때, 가능한 permutation의 집합을 S라고 하자. 그러면
$f_{\text{sym}}(x) = \frac{1}{|S|} \sum_{\pi \in S} f(\pi(x))$ 이다.
예를 들어, x = (0,1,1)이고 symmetric function을 구해보자. 그러면
$f_{\text{sym}}((1,1,0)) = \frac{1}{3}(f(0,1,1) + f(1,0,1) + f(1,1,0))$ 이 성립한다. 또한,
$f_{\text{sym}}((1,1,0)) = f_{\text{sym}}((1,0,1)) = f_{\text{sym}}((0,1,1))$ 이 성립하는 것을 알 수 있다.

Lower bound는 임의의 다항식이 특정 조건을 만족할 때, 다항식의 차수의 하한에 대한 내용이다. 이것이 중요한 이유는 hash function의 collision을 구하기 위해 필요한 query 수의 하한을 다항식의 차수로 나타내기 때문이다.

다음은 paturi가 증명한 lower bound다.[2]
$q(\alpha) \in \mathbb{R}[\alpha]$를 차수가 d인 다항식이라 하고, a,b 를 a < b인 정수, $\xi \in [a, b]$를 실수라고 하자. 만약 다음 조건이 만족된다면:
1. 모든 정수 $i \in [a, b]$ 에 대하여 $|q(i)| \leq 1$이고,
2. 어떤 상수 c에 대하여 $|q(\lfloor \xi \rfloor) - q(\xi)| \geq c$ 라면, 다음이 성립한다.

$d = \Omega\left(\sqrt{(\xi - a + 1)(b - \xi + 1)}\right)$. 특히 다음의 lower bound를 가진다. $d = \Omega(\sqrt{b-a})$

quantum blackbox model이란 quantum oracle이 하나 있고, superposition된 qubit을 query하고 superposition된 qubit을 답으로 주는 모델이다. blackbox라고 불리는 이유는 quantum oracle의 내용은 모르고, 질문에 대한 답만 받을 수 있기 때문이다. 구체적인 내용은 다음과 같다.
정의역 $[n]$과 치역 $[N]$을 가지는 함수 f가 오라클로 주어졌다고 가정하자. 양자 알고리즘은 쿼리 전후에 큐비트 상태를 조작하는 유니터리 연산(U)과 오라클 쿼리(O_f)를 교대로 총 T번 수행한 후, 마지막으로 투영 연산(P)을 통해 수락 확률(acceptance probability)을 구한다. 즉, 연산 과정은 다음과 같은 순서로 진행된다.$U_0 \to O_f \to U_1 \to \cdots \to U_{T-1} \to O_f \to U_T \to P$ 의 연산에 대한 초기 상태를 $|0\rangle$이라 할 때, 최종적인 수락 확률 $P(f)$는 다음과 같이 정의된다 .$P(f) := \|PU_T O_f U_{T-1} \cdots O_f U_0 |0\rangle\|^2$ 이때, $f(i) = j$인지 여부를 나타내는 Boolean 변수 $\delta_{i,j}$에 대하여, 수락 확률 함수 $P(f)$는 차수가 $2T$이하인 다항식으로 표현된다는 중요한 성질을 가진다. [3]

3. Hash function의 lower bound
이제 우리가 구하고자 하는 문제와 증명 과정에 대해 설명하겠다.

$D_{2\to1}$(n, N): 함수 D가 1-1 대응함수인지, 2-1 대응함수인지 구분하는 문제이다.
$D_{2\to1}^{1/2}$(n, N): 입력의 절반은 치역의 상위 부분(N/2,…,N)으로 반드시 2-1 대응될 때, 나머지 절반은 1,…, N/2으로 2-1 인지 1-1 대응함수인지 구분하는 문제이다.

증명하는 큰 개요는 다음과 같다.[4]

1) reduction: $D_{2\to1}^{1/2}$(n, n)은 $D_{2\to1}$(n, 3n/2) 으로 reduce 될 수 있다. 즉, $D_{2\to1}^{1/2}(n,n) \le_Q D_{2\to1}(n,3n/2)$ 이 성립한다.
2) Lower bound: 어떠한 quantum algorithms이라도 $D_{2\to1}^{1/2}$(n, n) 을 풀기 위해서는 최소한 $\Omega(n^{1/3})$의 query가 필요하다.
3) 따라서 $D_{2\to1}$을 풀기 위해서는 최소한 $n^{1/3}$의 query가 필요하고, n을 기존의 hash 값보다 3배 늘리면 같은 수준의 보안을 달성할 수 있다. (여기서 n은 가능한 input의 수이므로, bit수를 3배 늘리면 된다.)

reduction을 하는 이유는 다항식의 급격한 변화를 주기 위함이다. $D_{2\to1}$은 전체적으로 균일한 구조를 가져 이것을 나타내는 다항식은 완만하게 변하는 경향이 있다. 따라서 높은 차수를 증명하기 어렵다. 반면 $D_{2\to1}^{1/2}$은 입력의 절반(n/2)은 고정하고 나머지 절반(m)은 변수로 둔다. 이렇게 하면 m = n/2인 지점에서 함수가 급격하게 변화가 생기며, 이는 다항식의 차수가 높아야만 가능하기 때문에 lower bound를 증명할 수 있다.

4. Conclusion
본 연구의 가장 중요한 점은 hash function은 어떠한 quantum 알고리즘에도 안전하다는 것이다. 따라서 blockchain에서 rsa 기반 사용하는 알고리즘들을 hash로 바꾼다면 양자 세계에서도 안전한 blockchain을 만들 수 있다.

References
[1]: Phillip Rogaway and Thomas Shrimpton, “Cryptographic Hash-Function Basics: Definitions, Implications, and Separations for Preimage Resistance, Second-Preimage Resistance, and Collision Resistance,” FSE 2004, LNCS 3017, Springer, full version 2009

[2]: R. Paturi. On the degree of polynomials that approximate symmetric Boolean functions (preliminary version). In Proceedings of the Twenty-Fourth Annual ACM Symposium on the Theory of Computing, pages 468–474, Victoria, British Columbia, Canada, May 1992.
[3]: R. Beals, H. Buhrman, R. Cleve, M. Mosca, and R. de Wolf. Quantum lower bounds by polynomials. In 39th Annual Symposium on Foundations of Computer Science, pages 352–361, Los Alamitos, CA, Nov. 1998. IEEE.
[4]: Y. Shi, "Quantum lower bounds for the collision and the element distinctness problems," in *Proceedings of the 43rd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2002, pp. 513-519.