Post-Quantum Cryptography

1 Introduction

In the last few lectures, we saw that a certain cryptographic task, namely key exchange, can be achieved with unconditional security (assuming authenticated channels) using quantum information. Yet, other common tasks, including commitments, coin-flipping and oblivious transfer, provably cannot.

In the next few lectures, we will switch gears and move to a world where the honest parties are classical but the adversaries are quantum. This is referred to as *post-quantum cryptography*. We will revisit classical cryptographic primitives and tasks and see which ones can be achieved in such a world, and the tools it takes to get there. In particular, we will see post-quantum secure commitment, zero knowledge and proof of knowledge protocols. But first, let's start with a primer on classical cryptography.

2 Classical Cryptography: The Easy Case

One-way Functions. A function family $\mathcal{F} = \{F_n\}_{n\in\mathbb{N}}$ where each F_n maps n bits to $m = m(n)$ bits is one-way if for every probabilistic polynomial-time adversary $\mathcal{A}$,

$$\Pr[F_n(x') = F_n(x) : x \leftarrow \{0,1\}^n; \mathcal{A}(F_n(x)) = x'] = \mathrm{negl}(n)$$

where $\mathrm{negl}(\cdot)$ is a negligible function. That is, given $F_n(x)$ for a random x in the domain, no adversary can produce *a pre-image* for it.

If f is hard to invert for quantum algorithms, that is, changing the quantifier in the above definition from probabilistic polynomial-time to quantum polynomial-time, we immediately have quantum-secure one-way functions (we will call them QOWF). The subset sum problem as well as the learning with errors problem, as we will define later, give us candidates for QOWFs. **Anand**: There are functions that we believe are classical OWFs but know are not QOWFs, e.g. ones based on factoring.

If f is a one-to-one, onto function, and it is one-way, then f is a one-way permutation.

Anand: Concretely, the candidate function based on LWE is $F(A, s, e) = (A, As + e \bmod q)$, where A is a matrix and s, e are vectors in $\mathbb{F}_q$, with e consisting of "small" numbers. We will learn a lot more about this later.

Anand: Audience question: is there a candidate for post-quantum OWPs? Answer: no candidate yet!

Pseudorandom Generators. The same holds for pseudorandom generators (PRG) which are function families $\mathcal{G} = \{G_n\}_{n\in\mathbb{N}}$ where each G_n expands n bits to $m(n) > n$ bits. The security requirement says that for any probabilistic polynomial-time adversary $\mathcal{A}$,

$$\mathsf{Adv}(\mathcal{A}) := \Big|\Pr[\mathcal{A}(y) = 1 : y \leftarrow \{0,1\}^m] - \Pr[\mathcal{A}(y) = 1 : x \leftarrow \{0,1\}^n; y = G_n(x)]\Big| = \mathrm{negl}(n)$$

where $\mathrm{negl}(\cdot)$ is a negligible function. That is, no p.p.t. adversary can distinguish between a uniformly random m-bit string and the output of the pseudorandom generator applied to a uniformly random n-bit string.

The "HILL" construction by Håstad, Impagliazzo, Levin and Luby (HILL) of PRGs from OWFs as well as their proof of security are entirely black-box, and go through even if the adversary is quantum. That is, if there is a QOWF, then there is a QPRG.

Anand: However, the HILL security reduction is *not* automatic if the adversary has quantum auxiliary information! This case might still be open!

3 Pseudorandom Functions

The situation changes with pseudorandom functions. You should think of pseudorandom functions as pseudorandom generators that stretch by a large, possibly exponential, amount. For example, from n bits to 2^n bits. However, in that case, it doesn't make sense to ask the function to write down the whole output in one go as it would take exponential time. Instead, we ask for random access: i.e. given the key/seed $\sigma \in \{0,1\}^n$ and an index $i \in [2^n]$, produce the i^{th} bit of the output in $\mathsf{poly}(n)$ time.

Formally, a pseudorandom function family $\mathcal{F} = \{F_K : \{0,1\}^\ell \to \{0,1\}^m : K \in \{0,1\}^n\}$ is called pseudorandom if for every probabilistic polynomial-time oracle adversary $\mathcal{A}$,

$$\Pr[\mathcal{A}^{F_K}(1^n) = 1] - \Pr[\mathcal{A}^{R}(1^n) = 1]$$

where R is a uniformly random function from $\{0,1\}^\ell$ to $\{0,1\}^m$.[1] In the quantum world, one could consider two cases.

Classical Queries. One could change the definition above to let the adversary be a quantum polynomial-time algorithm, but restrict her to make classical queries. In this case, as with PRGs, post-quantum security follows immediately, assuming that the underlying assumption that the PRF is built on is post-quantum secure.

Quantum Queries. The more challenging situation is where the adversary can query on a superposition of inputs. For example, they could query the oracle $\mathcal{O}$ (which is either F_K or R) on the state

$$|\psi\rangle = \sum_{x \in \{0,1\}^n} \alpha_x |x\rangle |0^n\rangle$$

and get as output

$$\sum_{x \in \{0,1\}^n} \alpha_x |x\rangle |0^n\rangle \to \sum_{x \in \{0,1\}^n} \alpha_x |x\rangle |\mathcal{O}(x)\rangle$$

Thus, in one shot, the adversary can get a superposition over exponentially many outputs of the oracle. If a PRF is post-quantum secure, is it secure against adversaries that make quantum queries? **Anand**: Why even consider quantum queries? For one, security in a stronger model is always better! For another, we will possibly be in Quantumania in a few decades, where this will be relevant. Alternatively, if the adversary has an obfuscated circuit description of the PRF (which is kind of the sitaution in real life), then the adversary can just implement that circuit coherently to make quantum queries.

Theorem 1. *There exists a post-quantum secure PRF that is* not *secure against quantum-query adversaries.*

Proof. The key idea is to embed a period-finding problem into the PRF. Given a PRF $F_K : \{0,1\}^n \to \{0,1\}^m$, construct $F'_{K,r}$ such that

$$F'_{K,r}(x) = F'_{K,r}(x \oplus r)$$

for every x in the domain. (For example, a constructive way of doing this would be to let $r = (1,s)$, i.e. the first bit of r is 1, and let $F_K(1,y) = F_K(0, y \oplus r)$ for every $y \in \{0,1\}^{n-1}$).

[1] Here, we will for simplicity restrict our attention to functions that use n-bit seeds. In reality, one has to define such a family for each n.

Now, r can be recovered using Simon's algorithm with quantum queries. Indeed, the adversary queries the oracle on

$$\frac{1}{2^{n/2}} \sum_{x \in \{0,1\}^n} |x\rangle |0^m\rangle$$

and measures the second register to obtain

$$\frac{1}{\sqrt{2}}(|x\rangle + |x \oplus r\rangle)$$

for some x. Do a Hadamard transform and measure to get a vector d which must have the property that $\langle d, r\rangle = 0 \pmod 2$. Once you collect sufficiently many such vectors, you can do a Gaussian elimination to figure out r.

It is also not hard to show that $F'_{K,r}$ is secure with classical queries. Given a q-query algorithm, calculate the probability that $x_i \oplus x_j = r$ for some two queries x_i and x_j. This is at most $q^2/2^m$ which is negligible as long as m is large enough. □

For positive results, let's look at the Goldreich-Goldwasser-Micali (GGM) "tree" construction. On the one hand, if you are willing to lose an exponential factor in the input length, security against quantum queries follows immediately. However, the traditional proof that handles polynomial-time classical adversaries that, in particular, make at most a polynomial number of queries to $F_K(\cdot)$, breaks down. On the other hand, Zhandry showed that the GGM construction (as well as several other classical constructions), are secure against adversaries making quantum queries.

4 The Right Definitions?

Secret-key Encryption. *Weak* pseudorandom functions are a relaxed notion of PRFs where the adversary obtains polynomially many pairs $(r_i, F_K(r_i))$ for random inputs $r_i \leftarrow \{0,1\}^n$. That is, the adversary does not have the power to query the PRF oracle on chosen inputs. Post-quantum secure weak PRFs are sufficient to construct IND-CPA secure secret-key encryption schemes, as long as the adversary does not query the challenge oracle on a superposition of messages.

On the other hand, if the adversary can make a challenge query on superpositions, say

$$|0\rangle \otimes \sum_{m \in \mathcal{M}} |m\rangle \otimes |0\rangle$$

where $\mathcal{M}$ is the message space, and get

$$|0\rangle \otimes \sum_{m \in \mathcal{M}} |m\rangle \otimes |\mathsf{Enc}_K(0)\rangle$$

if $b = 0$ and

$$|0\rangle \otimes \sum_{m \in \mathcal{M}} |m\rangle |\mathsf{Enc}_K(m)\rangle$$

for $b = 1$. The adversary can now easily tell what b is. For example, measure the third register. In the $b = 0$ case, measuring does not disturb the first and second register; while, in the $b = 1$ case, the second register collapses to some m. Doing a Hadamard transform on the second register now finishes the distinguisher's job. For more discussion on alternate definitions, see Gagliardoni, Hülsing and Schaffner's work.

Message Authentication Codes (MAC). In the case of MACs, after making q superposition queries, the adversary cannot produce a MAC on a "new" message, but how do you define what's new? Boneh and Zhandry came up with a $q+1$ definition where they asked the adversary to produce MACs on $q+1$ distinct messages after making q quantum queries. However, this leaves open the possibility of the adversary querying for MACs of messages starting with 0 and producing the MAC of a message starting with 1. For more refined definitions, see Garg, Yuen and Zhandry.

Anand: Finding a satisfactory definition here is still open. If you satisfy Vinod, he will buy you a cookie.

5 Commitments

Another big difference between classical and post-quantum security comes up in the case of commitment schemes. This happens even if the interaction between the adversary and the honest parties is classical, so this is a real problem. Our exposition here closely follows Ma and Vazirani.

Recall the notion of (classical) cryptographic commitments (to classical messages) which asks for a pair of algorithms $(\mathsf{Com}, \mathsf{Ver})$ where

- The key-generation algorithm outputs a commitment key k.
- The commitment algorithm (non-interactive for simplicity) $\mathsf{Com}_k(m, r)$ outputs a commitment c.
- The verification algorithm, given the commitment key k, a commitment c, and a message-randomness pair (m, r), outputs 1 if and only if $\mathsf{Com}_k(m, r) = c$.

We will usually ignore the commitment key and assume that all algorithms get access to k.

- *Hiding:* For any two messages m_0, m_1, the distributions of $\mathsf{Com}(m_0; \cdot)$ and $\mathsf{Com}(m_1; \cdot)$ are indistinguishable. The indistinguishability can be statistical or computational.
- *Classical Binding:* It is hard to find a string c and message-randomness pairs (m_0, r_0) and (m_1, r_1) such that
$$\mathsf{Com}(m_0; r_0) = \mathsf{Com}(m_1; r_1) = c$$
Binding can be statistical or computational. Note that if the commitment scheme is statistically (or perfectly) hiding, it has to be computationally binding. Computationally binding schemes can produce succinct commitments whereas statistically binding ones cannot, making them particularly attractive.

Anand: Added: Why is this a good definition? Let's try it out for an example setting: suppose we want to run an auction, where parties commit to their bids in advance, and then open them. How does this definition prevent a party from, say, maliciously opening to the lowest bid plus one dollar? The idea is that if a party in the auction could cheat, this means that should be possible for it to open to two different values m_0, m_1, depending on the other bids that it observes. Now, since the committer is classical, it can be *rewound*: run it once with one history of observed bids to get an opening to m_0, and then reset the machine to its previous state and run it with the other history to get an opening to m_1. This yields an adversary that breaks the definition of classical binding.

Constructions. We will describe two canonical constructions of commitment schemes: one of a perfectly binding, computationally hiding commitment scheme from one-way permutations; and another of a computationally binding, statistically hiding commitment scheme from collision-resistant hash functions.

Construction 1. Let $f : \{0,1\}^n \to \{0,1\}^n$ be a one-way permutation. To commit to a bit b, pick a random $r \leftarrow \{0,1\}^n$ and output $c = (f(r), h(r) \oplus m)$ where h is a hardcore predicate for f (e.g. the Goldreich-Levin predicate). This construction is perfectly binding and therefore binding against even quantum attackers. It is computationally hiding, and therefore hiding against QPT adversaries as long as f is a QOWP.

Construction 2. Let $\mathcal{F}$ be a family of collision-resistant hash functions, and $\mathcal{H}$ be a pairwise independent family of functions from $\{0,1\}^n$ to $\{0,1\}^m$. The commitment key is $f \leftarrow \mathcal{F}$ and $h \leftarrow \mathcal{H}$. To commit to a message b (not necessarily a bit), draw an $r \leftarrow \{0,1\}^n$ and output $(f(r), h(r) \oplus b)$. Statistical hiding follows from the fact that r has min-entropy given $f(r)$ (since f is compressing) and h extracts randomness (via the leftover hash lemma). Computational binding follows directly from the collision-resistance of the family $\mathcal{F}$.

An alternate construction of a succinct, computationally binding, but not necessarily hiding, commitment scheme simply outputs $f(b)$. This is clearly computationally binding and succinct.

Anand: Statistical binding is not great for two reasons. First, stat. binding implies computational hiding, which means you don't have *everlasting* hiding. Secondly, it's impossible for a statistically bindng commitment to be succinct: the commitment must be at least as long as the message just by the pigeonhole principle. So we really want post-quantum security of something like Construction 2.

Post-Quantum Security. Imagine an attacker $\widetilde{S}$ (against the binding property) producing a commitment string c (say, commitment to a bid in an auction) and later, opening c to any given message m (in this example, determining the bid after the fact). This is clearly very bad and violates what we intuitively think of as binding.

Indeed, the classical binding definition above prohibits this type of behavior from classical adversaries. Indeed, if classical adversaries can open to any message at will, they can definitely produce two message-randomness pairs, violating the classical binding requirement.

How about a quantum adversary? A quantum adversary could create a superposition over all messages and random strings, compute the commitment algorithm in superposition and create

$$\sum_{m,r} |m\rangle_{\mathcal{M}} |r\rangle_{\mathcal{R}} |\mathsf{Com}(m;r)\rangle_{C}$$

Measure the C register to get a commitment c, and send it to the receiver. Now, the adversary is left with a state

$$|\psi_c\rangle := \sum_{m,r : \mathsf{Com}(m;r)=c} |m\rangle_{\mathcal{M}} |r\rangle_{\mathcal{R}}$$

Later, *potentially*, the adversary could apply a unitary U_0 to $|\psi_c\rangle$ and measure to get $(0, r_0)$ and another unitary U_1 to $|\psi_c\rangle$ and measure to get $(1, r_1)$. However, once the adversary produced one of these openings, they destroyed $|\psi_c\rangle$, and will not be able to produce the other opening. This leads us to the following strange behavior which is classically impossible: the adversary can open the commitment c to either 0 or to 1, but we cannot use him to extract both an opening to 0 and to 1.

Anand: (TODO: check this) Moreover, if we're statistically hiding, we *know* from Uhlmann's theorem that a *possibly inefficient* U_0 and U_1 satisfying these properties necessarily always exist!

Do such weird commitment schemes exist? This turns out to be a far more interesting question that it might sound at first. [ARU14] show the existence of such a commitment scheme in a quantum oracle model. [Zha19, DS23] show a "win-win" result: either every post-quantum binding commitment scheme is also collapse-binding or one can construct very interesting objects called quantum lightning and one-shot signatures that are not known to exist from standard cryptographic assumptions.

An Attempt: Sum-Binding. Note that if $|\psi_c\rangle$ contains an equal superposition of 0 and to 1, the adversary can trivially open to 0 with probability p_0 and to 1 with probability p_1 where $p_0 + p_1 = 1$. This is not an attack. However, if $p_0 + p_1 \geq 1 + \frac{1}{\mathsf{poly(n)}}$, where n is the security parameter, we call it a successful attack. This discussion naturally leads to the definition of sum-binding which asks that for every QPT adversary $\mathcal{A}$, there is a negligible function $\mathrm{negl}(\cdot)$ such that $p_0 + p_1 \leq 1 + \mathrm{negl}(n)$.

This definition, while appealing at first, is problematic in several respects: (1) it is a priori unclear how to generalize this definition to more than a single bit; (2) it is unclear how to use this as a component in larger protocols; and (3) it is unclear how to build a multi-bit commitment with any reasonable guarantee from a sum-binding commitment scheme.

Collapse-Binding [Unr16]. The (non-obvious) definition of collapse-binding was proposed by Unruh in 2016, and has quickly become the de facto standard of a binding definition for commitments.

Definition 2. *A commitment scheme* Com *is termed collapse-binding (or collapsing) if every QPT adversary* $\mathcal{A}$ *has negligible advantage in distinguishing between the following games:*

In both Game 0 *and Game* 1*,* $\mathcal{A}$ *comes up with a classical commitment string* c *and a quantum opening*

$$|\psi\rangle = \sum_{m,r} \alpha_{m,r} |m\rangle_{\mathcal{M}} |r\rangle_{\mathcal{R}}$$

Verify the opening: compute

$$\sum_{m,r} \alpha_{m,r} |m\rangle_{\mathcal{M}} |r\rangle_{\mathcal{R}} |\mathsf{Com}(m;r)\rangle_{C} \ ,$$

measure the third register, and if it is different from c*, abort. In Game* 0*, return the* $\mathcal{M}$ *and* $\mathcal{R}$ *registers to* $\mathcal{A}$*. In Game* 1*, measure the* $\mathcal{M}$ *register, and return both the* $\mathcal{M}$ *and* $\mathcal{R}$ *registers to* $\mathcal{A}$*. The adversary returns a bit* b'*, and she succeeds if* $b' = b$*.*

Remark An alternate definition (in the case of *bit* commitment schemes) would have the adversary apply a phase flip in case $b = 1$. That is, they would return

$$(Z \otimes I) \sum_{m,r} \alpha_{m,r} |m\rangle_{\mathcal{M}} |r\rangle_{\mathcal{R}} = |0\rangle_{\mathcal{M}} \otimes \sum_{r:\mathsf{Com}(0,r)=c} \alpha_{0,r} |r\rangle_{\mathcal{R}} - |1\rangle_{\mathcal{M}} \otimes \sum_{r:\mathsf{Com}(1,r)=c} \alpha_{1,r} |r\rangle_{\mathcal{R}} \ ,$$

to $\mathcal{A}$. This definition is equivalent to the one above: one way to intuitively see this is to note that in the collapse-binding definition, one takes a qubit with the density matrix

$$\rho = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$$

and measure it in the standard basis to get

$$\rho' = \begin{pmatrix} \alpha & 0 \\ 0 & \delta \end{pmatrix}$$

In the second case, one gets (by a random choice of b)

$$\rho'' = \frac{1}{2}\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} + \frac{1}{2}Z\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}Z = \begin{pmatrix} \alpha & 0 \\ 0 & \delta \end{pmatrix} = \rho'$$

Properties of collapse binding.

Lemma 3. *Collapse-binding implies sum-binding.*

Note: very recent, highly nontrivial work showed that for single-bit commitments the reverse implication holds!

We will prove the lemma in two steps.

1. Collapse-binding implies that you "can't change your mind after the fact." (We'll formalize this).
2. This implies that you're sum-binding.

Anand: Ended here.

Swap Test. To test the overlap between two states $|\psi_0\rangle$ and $|\psi_1\rangle$, prepare

$$|+\rangle \otimes |\psi_0\rangle \otimes |\psi_1\rangle$$

and do a controlled-swap where the first register is the control bit. This results in

$$\frac{1}{\sqrt{2}} \cdot (|0\rangle |\psi_0\rangle |\psi_1\rangle + |1\rangle |\psi_1\rangle |\psi_0\rangle) = |+\rangle \otimes \frac{|\psi_0\rangle |\psi_1\rangle + |\psi_1\rangle |\psi_0\rangle}{2} + |-\rangle \otimes \frac{|\psi_0\rangle |\psi_1\rangle - |\psi_1\rangle |\psi_0\rangle}{2}$$

Measure the first register in the X basis, the probability of getting a $|+\rangle$ is

$$\left\|\frac{|\psi_0\rangle |\psi_1\rangle + |\psi_1\rangle |\psi_0\rangle}{2}\right\|^2 = \frac{1}{4} \cdot (2 + 2\langle\psi_0|\psi_1\rangle\langle\psi_1|\psi_0\rangle) = \frac{1}{4} \cdot (2 + 2|\langle\psi_0|\psi_1\rangle|^2) = \frac{1}{2} \cdot (1 + |\langle\psi_0|\psi_1\rangle|^2)$$

A Different Overlap Test. Assume that you can prepare the state

$$\frac{1}{\sqrt{2}}(|0\rangle |\psi_0\rangle + |1\rangle |\psi_1\rangle)$$

(Note that if you are handed $|\psi_0\rangle$ and $|\psi_1\rangle$, you may not be able to prepare this state; however, in our use case below, we will show how to do this.) Measure the first qubit in the X basis, and output the bit. The intuition is that if $\psi_0 = \psi_1 = \psi$ are identical, the state is $|+\rangle |\psi\rangle$, and measuring the first qubit gives us deterministically $|+\rangle$. On the other hand, if ψ_0 and ψ_1 are orthogonal, then $|\psi_0\rangle + |\psi_1\rangle$ and $|\psi_0\rangle - |\psi_1\rangle$ are orthogonal and have the same length, so measuring the first qubit gives us $|+\rangle$ with probability $\frac{1}{2}$. In general,

$$\frac{1}{\sqrt{2}}(|0\rangle |\psi_0\rangle + |1\rangle |\psi_1\rangle) = \frac{1}{2}\Big(|+\rangle \otimes (|\psi_0\rangle + |\psi_1\rangle) + |-\rangle \otimes (|\psi_0\rangle - |\psi_1\rangle)\Big)$$

The probability that measuring the first qubit gives a $|+\rangle$ is

$$\frac{1}{4} \cdot \| |\psi_0\rangle + |\psi_1\rangle \|^2 = \frac{1}{2}(1 + \mathsf{Re}(\langle\psi_0|\psi_1\rangle))$$

Collapse-Binding Implications. Define a projector onto valid openings of 0:

$$\Pi_0 \ := |0\rangle\langle 0| \otimes \sum_{r_0 : \mathsf{Com}(0;r_0)=c} |r_0\rangle\langle r_0|$$

and similarly

$$\Pi_1 \ := |1\rangle\langle 1| \otimes \sum_{r_1 : \mathsf{Com}(0;r_1)=c} |r_1\rangle\langle r_1|$$

Consider an adversary who comes up with a state $|\phi\rangle$ which can be opened to 0 or 1. That is, measuring the first qubit of $U_0 |\phi\rangle$ gives 0 with probability p_0 and measuring the first qubit of $U_1 |\phi_1\rangle$ gives 1 with probability p_1 where $p_0 + p_1 \gg 1$. Equivalently, setting $|\psi\rangle = U_0 |\phi\rangle$ and $U = U_1 U_0^\dagger$, we have

$$||\Pi_0 |\psi\rangle||^2 + ||\Pi_1 U |\psi\rangle||^2 = p_0 + p_1 \overset{want}{=} 1 + \mathrm{negl}(n)$$

This formalizes the condition that no adversary can open in two ways.

What does it mean that the adversary can't change their mind? $\mathcal{A}$ cannot come up with some state $|\psi\rangle$, project to Π_0 to get $\Pi_0 |\psi\rangle$ which consists of 0 openings, and then apply some unitary map to get $U\Pi_0 |\psi\rangle$ which has a significant overlap with the Π_1 subspace consisting of openings of 1. That is,

$$||\Pi_1 U \Pi_0 |\psi\rangle||^2 = \mathrm{negl}(n)$$

Lemma 4. *If* $||\Pi_1 U \Pi_0 |\psi\rangle||^2 = \varepsilon$, *then* $||\Pi_0 |\psi\rangle||^2 + ||\Pi_1 U |\psi\rangle||^2 \leq 1 + 2\varepsilon + \varepsilon^2$.

That is, if $\mathcal{A}$ cannot change their mind, they cannot open in two ways, and therefore satisfy sum-binding.

Proof.

$$\begin{aligned}
||\Pi_0 |\psi\rangle||^2 + ||\Pi_1 U |\psi\rangle||^2 &= ||\Pi_0 |\psi\rangle||^2 + ||\Pi_1 U \Pi_0 |\psi\rangle + \Pi_1 U (I - \Pi_0) |\psi\rangle||^2 \\
&\leq ||\Pi_0 |\psi\rangle||^2 + \Big(||\Pi_1 U \Pi_0 |\psi\rangle|| + ||\Pi_1 U (I - \Pi_0) |\psi\rangle|| \Big)^2 \\
&= ||\Pi_0 |\psi\rangle||^2 + \varepsilon^2 + 2\varepsilon ||\Pi_1 U (I - \Pi_0) |\psi\rangle|| + ||\Pi_1 U (I - \Pi_0) |\psi\rangle||^2 \\
&\leq ||\Pi_0 |\psi\rangle||^2 + \varepsilon^2 + 2\varepsilon + ||\Pi_1 U (I - \Pi_0) |\psi\rangle||^2 \\
&\leq ||\Pi_0 |\psi\rangle||^2 + ||(I - \Pi_0) |\psi\rangle||^2 + \varepsilon^2 + 2\varepsilon \\
&= 1 + 2\varepsilon + \varepsilon^2
\end{aligned}$$

where the first inequality is by triangle inequality, and the second since

$$||\Pi_1 U (I - \Pi_0) |\psi\rangle|| \leq ||U (I - \Pi_0) |\psi\rangle|| = ||(I - \Pi_0) |\psi\rangle|| \leq || |\psi\rangle || = 1$$

using the fact that Π_0 and Π_1 are projectors and U is a unitary, and the third since Π_1 is a projector and U a unitary. □

We will now show that if the commitment is collapse-binding,

$$||\Pi_1 U \Pi_0 |\psi\rangle||^2 = \langle\psi| \Pi_0 U^\dagger \Pi_1 U \Pi_0 |\psi\rangle = \mathrm{negl}(n)$$

Lemma 5. *If the commitment is collapse-binding, for any state $|\psi\rangle$ and polynomial-time computable unitary U,*

$$\|\Pi_1 U \Pi_0 |\psi\rangle\|^2 = \mathrm{negl}(n) .$$

Proof. For contradiction, suppose there is a state $|\psi\rangle$ and a polynomial-time computable unitary U such that

$$\|\Pi_1 U \Pi_0 |\psi\rangle\|^2 = \varepsilon(n) > 1/p(n) .$$

for some polynomial function $p(\cdot)$. This must mean that $\|\Pi_0 |\psi\rangle\|^2 > 1/p'(n)$ for some polynomial function $p'(\cdot)$ as applying U does not charge the norm and projecting to Π_1 can only decrease it.

Define

$$|\phi_0\rangle = \Pi_0 |\psi\rangle \quad \text{and} \quad |\phi_1\rangle = \Pi_1 U \Pi_0 |\psi\rangle$$

The collapse-binding adversary does the following:

1. Prepare the state (proportional to) $|0\rangle|\phi_0\rangle + |1\rangle|\phi_1\rangle$: to do this,

   (a) measure $|\psi\rangle$ with the projective measurement $\{\Pi_0, I - \Pi_0\}$;

   (b) if the outcome is $I - \Pi_0$, give up and output a random bit.

   (c) Otherwise, do a controlled-U on $|+\rangle \otimes \Pi_0 |\psi\rangle$ to get

   $$|0\rangle \otimes \Pi_0 |\psi\rangle + |1\rangle \otimes U \Pi_0 |\psi\rangle$$

   (d) Measure with $\Pi' = |0\rangle\langle 0| \otimes I + |1\rangle\langle 1| \otimes \Pi_1$ to get

   $$|0\rangle \otimes \Pi_0 |\psi\rangle + |1\rangle \otimes \Pi_1 U \Pi_0 |\psi\rangle = |0\rangle \otimes |\phi_0\rangle + |1\rangle \otimes |\phi_1\rangle$$

   Again, if the outcome corresponds to $I - \Pi'$, discard and output a random bit.

2. Send the M, R registers of the state to the collapse-binding challenger to get

   $$|0\rangle|\phi_0\rangle + (-1)^b |1\rangle|\phi_1\rangle$$

3. Do a controlled-$U^\dagger$.

4. Measure the first qubit in the X basis. Output $b' = 0$ if the outcome is $|+\rangle$ and $b' = 1$ if the outcome is $|-\rangle$.

The probability that an abort event does not occur in step 1(b) is at least

$$\|\Pi_0 |\psi\rangle\|^2 > 1/p'(n)$$

as observed above, and that it does not occur in step 1(d) is at least

$$\|\Pi_1 U \Pi_0 |\psi\rangle\|^2 > 1/p(n)$$

If neither abort event occurs, the state prepared in step 3 is $|0\rangle|\tilde{\phi_0}\rangle + (-1)^b |1\rangle|\tilde{\phi_1}\rangle$ where

$$|\tilde{\phi_0}\rangle = \Pi_0 |\psi\rangle \quad \text{and} \quad |\tilde{\phi_1}\rangle = U^\dagger \Pi_1 U \Pi_0 |\psi\rangle$$

Assume $b = 0$. (The same analysis goes through when $b = 1$, and the probability of measuring a $|-\rangle$.) The probability that measuring the first qubit in the X basis results in a $|+\rangle$ is

$$\begin{aligned}
\Pr[+] &= \left\| \frac{|\tilde{\phi}_0\rangle + |\tilde{\phi}_1\rangle}{\sqrt{2}} \right\|^2 \\
&= \frac{1}{2} \cdot \left(\langle \tilde{\phi}_0 | \tilde{\phi}_0 \rangle + \langle \tilde{\phi}_1 | \tilde{\phi}_1 \rangle + \langle \tilde{\phi}_0 | \tilde{\phi}_1 \rangle + \langle \tilde{\phi}_1 | \tilde{\phi}_0 \rangle \right) \\
&= \frac{1}{2} \cdot \left(1 + 2\mathsf{Re}(\langle \tilde{\phi}_1 | \tilde{\phi}_0 \rangle) \right) \\
&= \frac{1}{2} + \mathsf{Re}(\langle \tilde{\phi}_1 | \tilde{\phi}_0 \rangle) \\
&= \frac{1}{2} + \langle \psi | \Pi_0 U^\dagger \Pi_1 U \Pi_0 | \psi \rangle \\
&> \frac{1}{2} + \frac{1}{p(n)}
\end{aligned}$$

(Note that $\langle \tilde{\phi}_1 | \tilde{\phi}_0 \rangle$ is real.) Including the abort event, we get an advantage of $\frac{1}{p(n)^2}$. □

References

- Zhandry'12: https://eprint.iacr.org/2012/182.pdf
- d'Agnoll-Spooner'23: https://eprint.iacr.org/2022/786.pdf
- Gagliardoni-Hülsing-Schaffner'16: https://arxiv.org/pdf/1504.05255.pdf
- Aaronson-Atia-Susskind'20: https://arxiv.org/pdf/2009.07450.pdf

6 Zero-Knowledge Proofs of Knowledge

- Describe Blum’s Hamiltonicity protocol.
- Collapse-binding implies the soundness of Blum.
- Collapse-binding as the right definition: composable; works with quantum rewinding; compatible with statistical hiding, short commitments.
- There are explicit constructions of collapse-binding commitment schemes from, e.g. lossy trapdoor functions.

References

[ARU14] Andris Ambainis, Ansis Rosmanis, and Dominique Unruh. Quantum attacks on classical proof systems: The hardness of quantum rewinding. In *55th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2014, Philadelphia, PA, USA, October 18-21, 2014*, pages 474–483. IEEE Computer Society, 2014.

[DS23] Marcel Dall'Agnol and Nicholas Spooner. On the necessity of collapsing for post-quantum and quantum commitments. In Omar Fawzi and Michael Walter, editors, *18th Conference on the Theory of Quantum Computation, Communication and Cryptography, TQC 2023, July 24-28, 2023, Aveiro, Portugal*, volume 266 of *LIPIcs*, pages 2:1–2:23. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2023.

[Unr16] Dominique Unruh. Computationally binding quantum commitments. In Marc Fischlin and Jean-Sébastien Coron, editors, *Advances in Cryptology - EUROCRYPT 2016 - 35th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Vienna, Austria, May 8-12, 2016, Proceedings, Part II*, volume 9666 of *Lecture Notes in Computer Science*, pages 497–527. Springer, 2016.

[Zha19] Mark Zhandry. Quantum lightning never strikes the same state twice. In Yuval Ishai and Vincent Rijmen, editors, *Advances in Cryptology - EUROCRYPT 2019 - 38th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Darmstadt, Germany, May 19-23, 2019, Proceedings, Part III*, volume 11478 of *Lecture Notes in Computer Science*, pages 408–438. Springer, 2019.