1 Structure of the protocol

1.1 Overview

In this lab, we'll look at a commitment protocol which uses a hash function to commit to a message using some randomness. The security of these schemes depend on the function being used, and so we will look at two different choices of functions f for the protocol. The overall structure is the same in both cases. The protocol using a function f is as follows:

To commit a fixed 1-bit message m, Alice does the following:

- Uniformly picks some $n-$bit randomness $r \overset{\$}{\leftarrow} \{0,1\}^n$.
- Computes commitment value $c = f(m||r)$ where $||$ is bitstring concatenation.
- Sends Bob the value c.

Later, to show Bob the value that she committed to, Alice then sends over m, r and Bob verifies that the initial value c that was sent to him was computed using m and r. So he checks if $f(m||r) = c$.

However, since we're using quantum computers, things are slightly more complicated as Alice does not send a classical values but instead sends a series of qubits.

1.2 Making Alice Unitary

On Figure 1 Alice has also been made unitary - choosing the random value r is being done by applying $H^{\otimes n}$ and then performing a partial trace. Remember that the density matrix for a uniformly chosen classical value is

$$\rho_u 2^{-n} \sum_{x \in \{0,1\}^n} |x\rangle\langle x|$$

And this is not the same as just taking $|0\ldots0\rangle$ and applying $H^{\otimes n}$ to it.

$$H^{\otimes n}|0\ldots0\rangle = \sum_{x \in \{0,1\}^n} 2^{-\frac{n}{2}}|x\rangle$$

$$\rho_h = 2^{-n} \sum_{x \in \{0,1\}^n} \sum_{\tilde{x} \in \{0,1\}^n} |x\rangle\langle\tilde{x}|$$

However, we can add a buffer system, and CNOT the values over.

$$(\sum_{x \in \{0,1\}^n} 2^{\frac{-n}{2}}|x\rangle) \otimes |0\ldots0\rangle \rightarrow^{CNOT} \sum_{x \in \{0,1\}^n} 2^{-\frac{n}{2}}|x\rangle \otimes |x\rangle$$

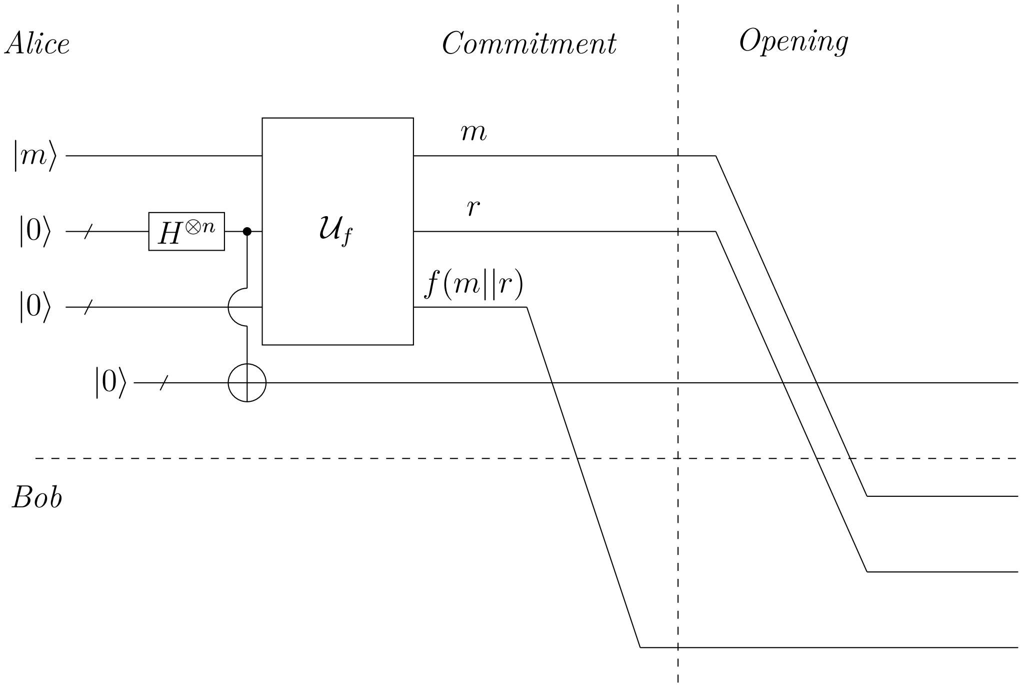

Figure 1: The circuit for the commitment protocol using function f.

$$\begin{aligned}\rho_b &= 2^{-n}(\sum_{x\in\{0,1\}^n} |x\rangle \otimes |x\rangle)(\sum_{\tilde{x}\in\{0,1\}^n} |\tilde{x}\rangle \otimes |\tilde{x}\rangle)^\dagger \\ &= 2^{-n} \sum_{x\in\{0,1\}^n} \sum_{\tilde{x}\in\{0,1\}^n} (|x\rangle \otimes |x\rangle)(\langle\tilde{x}| \otimes \langle\tilde{x}|) \\ &= 2^{-n} \sum_{x\in\{0,1\}^n} \sum_{\tilde{x}\in\{0,1\}^n} |x\rangle\langle\tilde{x}| \otimes |x\rangle\langle\tilde{x}|\end{aligned}$$

And now tracing away the second buffer register (which is also done by restricting our view to the first system) leaves only elements where $x = \tilde{x}$ as those are the only ones with non-zero trace.

$$tr_b\ \rho_b = 2^{-n} \sum_{x\in\{0,1\}^n} |x\rangle\langle x|$$

Which is exactly the density matrix of picking a uniformly random classical bitstring.

1.3 Description of the state

We now describe the state of the system in the commitment phase of the protocol. In the beginning, Alice starts out in the state

$$|\Psi_0^m\rangle = |m, 0, 0, 0\rangle$$

where the 0's stand for $0 \ldots 0$ bitstrings of all zeroes, of n, l, n length respectively. The first n-bit zero is where the randomness will go, the l-bit zero is where the function of the output will go, and the third zero is for the copy of the n-bit randomness that gets traced out.

Then the Hadamard gate is applied to the register for the randomness.

$$|\Psi_1^m\rangle = (I \otimes H^n \otimes I \otimes I)|\Psi_0^m\rangle = |m\rangle \otimes (\sum_{r\in\{0,1\}^n} 2^{-\frac{n}{2}}|r\rangle) \otimes |0\rangle \otimes |0\rangle = \sum_{r\in\{0,1\}^n} 2^{-\frac{n}{2}}|m, r, 0, 0\rangle$$

And then the randomness is CNOT'ed over to the buffer register.

$$|\Psi_2^m\rangle = \sum_{r\in\{0,1\}^n} 2^{-\frac{n}{2}}|m, r, 0, r\rangle$$

Then the unitary $\mathcal{U}_f$ is applied to $|\Psi_2^m\rangle$ to calculate the commitment value.

$$|\Psi_3^m\rangle = (\mathcal{U}_f \otimes I)|\Psi_2^m\rangle = \sum_{r\in\{0,1\}^n} 2^{-\frac{n}{2}}|m, r, f(m||r), r\rangle$$

And then the commitment part is sent to Bob.

$$|\Psi^m\rangle = \sum_{r\in\{0,1\}^n} 2^{-\frac{n}{2}}|m, r, r\rangle_A \otimes |f(m||r)\rangle_B$$

And when we trace out the extra buffer register that contains the randomness copy, we have a nice density matrix representation.

$$\rho_{AB}^m = \sum_{r\in\{0,1\}^n} 2^{-n}|m, r\rangle\langle m, r|_A \otimes |f(m||r)\rangle\langle f(m||r)|_B$$

2 Security definitions

Since we are going to show that the schemes are not perfectly secure, we need to first define what security actually means. Commitment schemes are described using two properties that go against one another.

The binding property states that the commitment value that Alice produces needs to "lock in" her choice of message. That once she has committed to some secret message and "put it in a box" using the commitment scheme, she can't then open the commitment in a way that fools us into thinking that she committed using another message instead.

The hiding property states that the commitment value should not reveal what the message is - that committing to a message m_1 looks the same as committing to a message m_2.

These goals go against one another because the binding property requires there to be a strong link between the commitment value and the secret message, while the hiding property wants to link to be weak, in order to allow multiple messages to give similar looking commitment values.

However, we can still achieve useful and secure commitment schemes by not requiring perfection. This is similar to encryption, where we can only achieve perfect secrecy if the key size is at least as large as the message size, which would be impractical. And yet we still have strong, though imperfect, systems.

2.1 0-Binding property

For the ε-binding property, we play a game with Alice. We let her commit to some value. Then we pick a message b and tell her to provide a valid opening such that the verification passes.

In the definition, we say that Alice has a fixed algorithm A for creating the commitment, and then algorithms A_0 for when we tell her to open to 0, and A_1 for when we tell her to open to 1.

We say that a protocol is ε-binding if the probability of Alice succeeding in both of the cases is $1 + \varepsilon$. More specifically, we say that

$$P_0 + P_1 = 1 + \varepsilon$$

where P_b is the probability that Alice produces a valid opening after running A and then A_b.

2.2 0-Hiding property

For the 0-hiding property, we want the protocol to be perfectly hiding - meaning the commitment has to perfectly hide which $m \in \{0, 1\}$ was committed. This means that Bob can't distinguish from the commitment message two runs of the protocol - ρ^0_{AB} where Alice committed 0, and ρ^1_{AB} where Alice committed 1.

Formally, we can express this as

$$TD(tr_A \; \rho^0_{AB}, tr_A \; \rho^1_{AB}) = 0$$

Or alternatively as

$$tr_A \; \rho^0_{AB} = tr_A \; \rho^1_{AB}$$

We can begin by tracing out Alice's part of the system from the above state to find $tr_A\ \rho^m_{AB}$.

$$tr_A\ \rho^m_{AB} = \sum_{r\in\{0,1\}^n} 2^{-n}\ tr_A\ (|m,r\rangle\langle m,r|_A \otimes |f(m||r)\rangle\langle f(m||r)|_B)$$
$$\sum_{r\in\{0,1\}^n} 2^{-n}\ (tr\ |m,r\rangle\langle m,r|_A)\cdot |f(m||r)\rangle\langle f(m||r)|_B$$

Notice that since each $|m,r\rangle$ is a basis state, $|r,m\rangle\langle r,m|$ consists of just one 1 on the diagonal, therefore it has trace 1. Alternatively, we can also see that from the fact that $tr\ |\Psi\rangle\langle\Psi| = |||\Psi\rangle||^2$.

Therefore we can easily trace Alice's part out.

$$\begin{aligned} tr_A\ \rho^m_{AB} &= \sum_{r\in\{0,1\}^n} 2^{-n}\ (tr\ |m,r\rangle\langle m,r|_A)\otimes |f(m||r)\rangle\langle f(m||r)|_B \\ &= \sum_{r\in\{0,1\}^n} 2^{-n}\ 1\cdot |f(m||r)\rangle\langle f(m||r)|_B \\ &= \sum_{r\in\{0,1\}^n} 2^{-n}\ |f(m||r)\rangle\langle f(m||r)|_B \end{aligned}$$

This is as descriptive as we can be about the state without any knowledge of the function f. To analyze this further, we must specify what kind of function we are using.

3 Example 1: Injective function

The first function we're examining is an injective one, that is for every pair of inputs that are distinct, $\forall x, x' \in \{0,1\}^{n+1} x \neq x' \Rightarrow f(x) \neq f(x')$. In simple terms, it means that if the outputs to the function are different from one another, then the outputs will be as well.

Our functions have the type $f : \{0,1\}^{n+1} \to \{0,1\}^l$. That is, they take a message bit and n bits of randomness, and give us an l-bit string as the output. Notice that by fixing the message bit, we have two distinct functions, $f_1(x) = f(1||r)$ and $f_0(x) = f(0||x)$. These functions are both of the type $f_m : \{0,1\}^n \to \{0,1\}^l$, so they have the same domain and codomain.

We can then visualize the functions as in Figure 3. Note that there are no inputs r, r' such that $f_0(r) = f_1(r')$ as this would mean $f(0||r) = f(1||r')$ which contradicts the injectivity. So the ranges of f_0 and f_1 are completely disjoint.

3.1 Breaking the 0-Hiding property

We now examine if the above commitment scheme is 0-hiding when we use the above function to commit. Recall that we wanted to show that $TD(tr_A\ \rho^0_{AB}, tr_A\ \rho^1_{AB}) = 0$

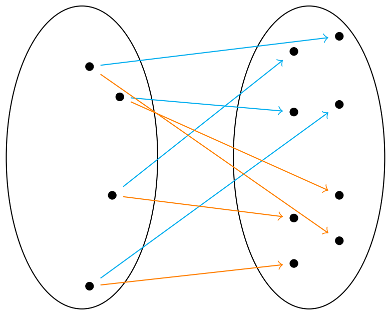

Figure 2: Visualization of functions f_0 (blue) and f_1 (orange) in case 1.

and that

$$tr_A\ \rho^m_{AB} = \sum_{r\in\{0,1\}^n} 2^{-n}\ |f(m||r)\rangle\langle f(m||r)|_B$$

Now, let's see what the actual matrices of $tr_A\ \rho^0_{AB}$ and $tr_A\ \rho^1_{AB}$ would be.

First notice that they are purely diagonal matrices, as they're of the form $\sum_y |y\rangle\langle y|$ where the y denote the possible outputs.

Second, the diagonal elements are completely disjoint in both matrices. Meaning that if there is a non-zero element on index i in $tr_A\ \rho^0_{AB}$, then in $tr_A\ \rho^0_{AB}$ the element at index i **must** be zero. This is because otherwise the output $|i\rangle\langle i|$ would appear in both cases with some non-zero probability, which means that for some r, r' we have that $f(0||r) = f(1||r') = i$. But this clearly can't be, because different inputs get mapped to different outputs, and the first bit is guaranteed to be different for the two inputs.

When computing the trace distance explicitly, we can show that this scheme is not hiding at all:

$$\begin{aligned}
&TD(tr_A\ \rho^0_{AB}, tr_A\ \rho^1_{AB}) = \\
&= \frac{1}{2} tr\ |\sum_{r\in\{0,1\}^n} 2^{-n}\ |f(0||r)\rangle\langle f(0||r)|_B - \sum_{r\in\{0,1\}^n} 2^{-n}\ |f(1||r)\rangle\langle f(1||r)|_B| \\
&= \frac{1}{2} tr\ |\sum_{r\in\{0,1\}^n} 2^{-n}\ |f(0||r)\rangle\langle f(0||r)|_B - |f(1||r)\rangle\langle f(1||r)|_B| \\
&= \frac{1}{2} tr\ |\sum_{r\in\{0,1\}^n, m\in\{0,1\}} 2^{-n}\ |f(m||r)\rangle\langle f(m||r)|_B \cdot (-1)^m|
\end{aligned}$$

Now because of the injectivity, all elements of this sum are distinct, and this is a sum over $2 \cdot 2^n$ elements. Since this is also a diagonal matrix, we can take the trace

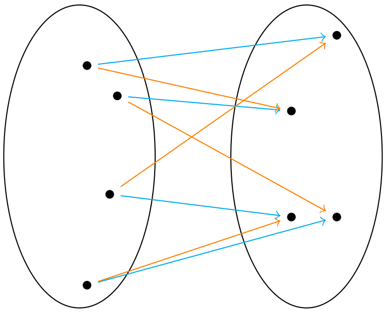

Figure 3: Visualization of functions f_0 (blue) and f_1 (orange) in case 2.

of the absolute value by taking the absolute value of the diagonal elements.

$$= \frac{1}{2} tr \mid \sum_{r \in \{0,1\}^n, m \in \{0,1\}} 2^{-n} \; |f(m||r)\rangle\langle f(m||r)|_B \cdot (-1)^m|$$

$$= \frac{1}{2} tr \sum_{r \in \{0,1\}^n, m \in \{0,1\}} 2^{-n} \; 1 = \frac{1}{2} 2^{-n} \cdot 2 \cdot 2^n = 1$$

And since the trace distance is 1 instead of 0, the scheme is 1-hiding, which is to say not at all. The cases between $m = 0$ and $m = 1$ are perfectly distinguishable.

4 Example 2: Two bijections

For the second function, we have a function of the type $f : \{0,1\}^{n+1} \to \{0,1\}^n$. In addition, we have the following property:

$$\forall y, m : \exists_1 r : f(m||r) = y$$

Which states that for every output y and message m, you can find a unique randomness r to get that output, $f(m||r) = y$. This leads to some interesting and useful effects. For one, notice that f_0 and f_1 are now bijections.

To see that f_0 is surjective, see that for every output element y we can find an r such that $f(0||r) = y$. This follows directly by the definition of the property.

To see that f_0 is injective, we have to pay attention to the dimensions. Because f_0 maps randomness from $\{0,1\}^n$ to output elements in $\{0,1\}^n$, we can't have that $f(0||r) = f(0||r')$ for $r \neq r'$. If there were two r's which did map to the same output element, then the number of all possible output elements would be $2^n - 1$. This would mean that one output element would not have a valid preimage.

4.1 Proving 0-Hiding property

Again, we want to show that $TD(tr_A\ \rho^0_{AB}, tr_A\ \rho^1_{AB}) = 0$ knowing that

$$tr_A\ \rho^m_{AB} = \sum_{r\in\{0,1\}^n} 2^{-n}\ |f(m||r)\rangle\langle f(m||r)|_B$$

In this case, it is easier to show that $tr_A\ \rho^0_{AB} = tr_A\ \rho^1_{AB}$ directly. Consider the density matrix of $tr_A\ \rho^0_{AB}$. We have that

$$tr_A\ \rho^0_{AB} = \sum_{r\in\{0,1\}^n} 2^{-n}\ |f(0||r)\rangle\langle f(0||r)|_B = \sum_{y\in\{0,1\}^n} |y\rangle\langle y|$$

Since the sum is over all 2^n possible values for r, and the output of $f_0(r)$ is distinct for every r, the set of all images is also all possible 2^n bitstrings. Notice that the same argumentation applies for f_1.

$$tr_A\ \rho^1_{AB} = \sum_{r\in\{0,1\}^n} 2^{-n}\ |f(1||r)\rangle\langle f(1||r)|_B = \sum_{y\in\{0,1\}^n} |y\rangle\langle y|$$

And so the two density matrices are the same, giving us 0-hiding. To get some intuition as to why this is, notice that for every output value y, I can find r such that $f(0||r) = y$. But I can also find an r' such that $f(1||r') = y$. And thus seeing only the images gives no information about the message.

However we need to be careful when applying reasoning like this as quantum cryptography can be anything but intuitive, and so we always need to have strict formal results.

4.2 Breaking the 0-Binding property

To show that the scheme is not perfectly binding, we will show that Alice can commit using message m and then provide an opening that is valid for the other message $\overline{m}$.

Recall that after running the protocol in the commitment phase for message m, the state of the system is

$$|\Psi^m\rangle = \sum_{r\in\{0,1\}^n} 2^{-\frac{n}{2}} |m, r, r\rangle_A \otimes |f(m||r)\rangle_B$$

We want to turn $|\Psi^0\rangle$ to $|\Psi^1\rangle$. To do this we will use the simulataneous Schmidt decomposition (Lemma 26 in lecture notes). We have already shown that $tr_A\ |\Psi^0\rangle\langle\Psi^0| = tr_A\ |\Psi^1\rangle\langle\Psi^1|$. That was exactly what we did in the 0-hiding section - that when Alice's part is traced away, then the view is the same.

The simultaneous Schmidt decomposition states that for a choice of orthonormal sets $\{|\alpha\rangle_i\}$ and $\{|\beta\rangle_i\}$ there are reals $\lambda_i \geq 0$ with $\sum_i \lambda_i^2 = 1$ such that

$$|\Psi^0\rangle = \sum_i \lambda_i |\alpha_i\rangle \otimes |\beta_i\rangle \ and \ |\Psi^1\rangle = \sum_i \lambda_i |\tilde{\alpha_i}\rangle \otimes |\beta_i\rangle$$

Now, let's choose our basis sets. We will take $|\beta_y\rangle = |y\rangle$, the set of all possible outputs of the functions. For the α sets, we will pick the input sets of both f_0 and f_1. Meaning that $|\alpha_y\rangle$ will correspond to the state Alice must have in order for f_0 to output y.

$$|\alpha_y\rangle = |0, f_0^{-1}(y), f_0^{-1}(y)\rangle$$
$$|\tilde{\alpha_y}\rangle = |1, f_1^{-1}(y), f_1^{-1}(y)\rangle$$

And all $\lambda_y = 2^{-\frac{n}{2}}$. $f_0^{-1}(y)$ is the inversion of f_0 which gives us the randomness r needed such that $f(0||r) = y$. Now all we need is a conversion from $|\alpha_y\rangle$ to the corresponding $|\tilde{\alpha}_y\rangle$. We want to go from the state that produces y when committing 0 to the state that produces y when committing 1.

More specifically, we want to find U such that $U|\alpha_y\rangle = |\tilde{\alpha_y}\rangle$. If I know f_0 and f_1^{-1} then this is easy. Suppose I am given some randomness r and I am tasked with finding r' such that $f(0||r) = f(1||r')$. Then I can apply f_0 to find $y = f(0||r)$. And then I can apply f_1^{-1} to find the unique randomness that gives me $f(1||r') = y$. Thus I can find the r' from r using $r' = f_1^{-1}(f_0(r))$. This is what the unitary U will do to both registers which contain the randomness.

$$U|a, b, c\rangle \rightarrow |\overline{a}, f_1^{-1}(f_0(b)), f_1^{-1}(f_0(c))\rangle$$

The reason why it is not expressed as $U|0, r, r\rangle \rightarrow \ldots$ is because the unitary needs to be a valid function for all basis states, which includes states aren't of the form $|0, r, r\rangle$ (so states where the last two registers differ).

Now that we have the U, Alice can turn a state that is a run of the protocol with the message 0 to a run of the protocol with message 1.

$$(U \otimes I)|\Psi^0\rangle = |\Psi^1\rangle$$

And thus Alice has an attack against the binding property of the commitment scheme. Her commitment algorithm A will just be to commit to the message 0. Then if she is asked to open to 0, she doesn't run U but just sends over the valid opening that she has. If she is asked to open to 1, she runs U and thus has a valid opening for 1 that she sends over.

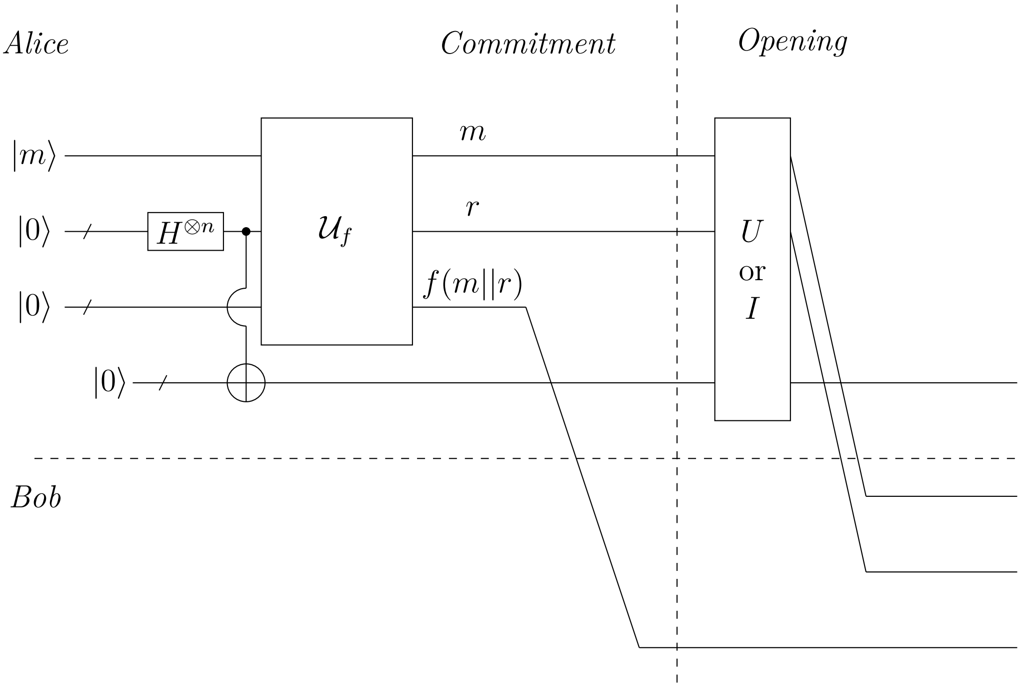

Figure 4: The circuit for the commitment protocol using function f.