Lecture 19: Impossibility of Quantum Bit Commitment

March 30, 2006

In the previous lecture we discussed the BB84 quantum key-distribution protocol, which allows two physically separated parties to construct a secure private key, and therefore communicate privately by means of the one-time pad. Although the problem of communicating privately is of paramount importance in cryptography, there are other cryptographic tasks that one may consider that key distribution protocols do not address. In this lecture we will discuss one of them: *bit commitment*. This is a particularly interesting problem from the point of view of quantum information because it turns out to be impossible, and this impossibility is due to a fundamental fact about bipartite quantum states (which we have already discussed).

Bit commitment

First let us define what bit commitment is. Alice has a bit b that she wishes to *commit* to Bob, but she doesn't want Bob to know what it is until she chooses to *reveal* it. Although Bob should not be able to determine b before Alice reveals, he should be sure that Alice cannot change the bit after it is committed. The two key properties of any bit commitment protocol are therefore that it must be:

- **Binding**. Alice should not be able to change the bit she committed.
- **Concealing**. Bob should not be able to identify the bit that Alice committed until she reveals it.

One can imagine a "mechanical" implementation of bit commitment as follows. Alice writes b on a piece of paper, locks it in a safe, and sends the safe to Bob. Bob receives the safe, but he doesn't have the key. He cannot open the safe without the key, so he cannot determine b, and therefore the concealing property holds. Because the safe is in Bob's possession, Alice cannot open it and change the bit, so the binding property also holds. When Alice wishes to reveal the bit, she sends the key to Bob.

Of course this mechanical interpretation is not satisfactory with respect to information processing purposes—we would like an implementation based on information. (It could also be argued that the binding and concealing properties, and in particular the concealing property, are based on strong physical assumptions. It is probably impossible to build a safe that can only be opened with a unique key. This is beside the point, however, because our real interest is with an information-based implementation.)

Similar to key distribution, it is impossible to implement bit commitment using classical information without using assumptions about computational intractability. It is a natural question to ask whether quantum information allows one to implement bit commitment.

Before addressing this question, it may be helpful to briefly motivate bit commitment. Why would we want to implement bit commitment? The answer is that it is a very interesting cryptographic primitive from which several interesting protocols can be built. For example, bit commitment allows for secure multi-party computations (such as voting), zero-knowledge proofs for NP-complete problems, and coin-flipping (as well as more complicated variants, such as playing poker).

Sketch of impossibility proof

The impossibility of quantum bit commitment relies on the following fact, which we proved in Lecture 15.

Fact. Suppose $|\phi\rangle, |\psi\rangle \in \mathcal{A} \otimes \mathcal{B}$ satisfy $\mathrm{Tr}_{\mathcal{A}} |\phi\rangle\langle\phi| = \mathrm{Tr}_{\mathcal{A}} |\psi\rangle\langle\psi|$. Then there exists a unitary operator $U \in \mathrm{L}(\mathcal{A})$ such that $(U \otimes I)|\phi\rangle = |\psi\rangle$.

It happens to be the case that there are approximate versions of the above fact, but because we have not discussed meaningful distance measures for quantum states it will not be possible to go into greater detail about this. Expressed informally, the approximate versions are of this form: if

$$\mathrm{Tr}_{\mathcal{A}} |\phi\rangle\langle\phi| \approx \mathrm{Tr}_{\mathcal{A}} |\psi\rangle\langle\psi|$$

then there exists a unitary operator $U \in \mathrm{L}(\mathcal{A})$ such that $(U \otimes I)|\phi\rangle \approx |\psi\rangle$.

Now let us apply the above fact to the problem of bit commitment. To begin, suppose that Alice and Bob use a "purely quantum" protocol to supposedly implement bit commitment. What we mean by this is that Alice and Bob apply only unitary operations and send quantum information back and forth—so assuming Alice and Bob are both being honest there are no measurements made until the end of the protocol, and there is no noise or other non-unitary transformations during the protocol.

There are necessarily two phases of any bit commitment protocol: the commit phase and the reveal phase. During the commit phase, Alice and Bob may perform some sequence of unitary operations and send qubits back and forth to one another any number of times. Assume that Alice and Bob's quantum systems at the end of the commit phase have corresponding vector spaces $\mathcal{A}$ and $\mathcal{B}$. There are two possible pure states of the entire system at this point: $|\psi_0\rangle$ or $|\psi_1\rangle$, depending on whether Alice intended to commit 0 or 1, respectively.

Under the assumption that the protocol is perfectly concealing, it is the case that

$$\mathrm{Tr}_{\mathcal{A}} |\psi_0\rangle\langle\psi_0| = \mathrm{Tr}_{\mathcal{A}} |\psi_1\rangle\langle\psi_1|;$$

if this were not so, Bob would be able to perform some measurement of his portion of $|\psi_0\rangle$ or $|\psi_1\rangle$ and gain at least partial information about which bit Alice committed. This implies the existence of a unitary operation $U \in \mathrm{L}(\mathcal{A})$ that Alice can perform on her qubits alone that would transform $|\psi_0\rangle$ to $|\psi_1\rangle$:

$$(U \otimes I)|\psi_0\rangle = |\psi_1\rangle.$$

Therefore, the binding property must completely fail to hold—Alice can switch back and forth between $|\psi_0\rangle$ and $|\psi_1\rangle$ without Bob's help or knowledge. For example, Alice may simply run the

original protocol and "commit" to 0, and later right before the reveal phase, she may either apply U (to switch her commitment to 1) or do nothing (leaving the commitment as 0).

Now, you might ask what happens when a protocol specifies that Alice and Bob must perform certain measurements or exchange classical rather than quantum information. It turns out that the situation is essentially the same. This is because the cheating party can decide to simulate all measurements and other non-unitary operations by using only unitary operations, which is always possible using auxiliary qubits as we discussed a few lectures ago. The fact that the non-cheating party may perform measurements or other non-unitary operations will not affect the cheater's ability to cheat—we may view that the non-cheater uses only unitary operations and simply chooses not to interact with the auxiliary qubits that would be used to do this.

Finally, you might ask what happens if perfect security is not required, but in instead only approximate security is permitted. This is indeed the more interesting situation, and it is handled by the approximate versions of the fact that was used above. Specifically, if it is the case that after the commit phase that Bob can learn a little bit, but not too much, about Alice's commitment, then we must have

$$\mathrm{Tr}_{\mathcal{A}} |\psi_0\rangle \langle\psi_0| \approx \mathrm{Tr}_{\mathcal{A}} |\psi_1\rangle \langle\psi_1| ,$$

where "$\approx$" has some technical meaning that we have not discussed. (Usually one uses either the notion of *fidelity* or *trace distance* to quantify the notion of approximate equality in such a situation.) This will not necessarily allow Alice complete freedom to change her commitment, but she will have almost complete freedom. There will exist U such that $(U \otimes I) |\psi_0\rangle \approx |\psi_1\rangle$, meaning that Alice will be able to change her commitment in a way that Bob will probably not be able to notice.

Example of an incorrect protocol

In order to illustrate the above discussion, let us consider a protocol that may initially appear to implement bit commitment (in a way that is perfectly concealing and approximately binding). This example, including a cheating strategy, was given in the same paper that proposed the BB84 key exchange protocol (so it was never really believed to be a correct protocol).

Example 1. Consider the following protocol, where we assume Alice wishes to commit to the bit $b \in \{0, 1\}$.

Commit phase. Let $S_0 = \{|0\rangle , |1\rangle\}$ and $S_1 = \{|+\rangle , |-\rangle\}$. Alice prepares a qubit X in a uniformly chosen state $|\phi\rangle \in S_b$, and sends X to Bob. (Bob does not need to do anything in the commit phase other than store the qubit sent by Alice.)

Reveal phase. When Alice wishes to reveal her commitment to Bob, she reveals a classical specification of $|\phi\rangle$, for instance:

$$\begin{aligned} 00 &\leftrightarrow |\phi\rangle = |0\rangle \\ 01 &\leftrightarrow |\phi\rangle = |+\rangle \\ 10 &\leftrightarrow |\phi\rangle = |1\rangle \\ 11 &\leftrightarrow |\phi\rangle = |-\rangle. \end{aligned}$$

In other words, the second bit is b, while the first bit specifies which element of S_b was selected. To check that Alice was being truthful, Bob measures the qubit X sent by Alice in the basis S_b. If the measurement result does not match with the state $|\phi\rangle$ sent by Alice, then Bob has caught Alice cheating. (Otherwise he has not.)

Let us examine the protocol to see what is wrong. (There has to be something wrong if we believe the previous argument that bit commitment is impossible.) First we check to see if it is concealing. If Alice wishes to commit 0, she sends Bob the qubit X in state $|0\rangle$ or state $|1\rangle$, each with equal probability. As Bob does not know which one Alice chooses, his description of the state of X is given by the density matrix

$$\frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1| = \frac{1}{2}I,$$

which is the totally mixed state. On the other hand, if Alice wishes to commit a 1, she sends either the state $|+\rangle$ or the state $|-\rangle$, again each with probability 1/2. In this case, the density matrix describing Bob's knowledge of X is

$$\frac{1}{2}|+\rangle\langle +| + \frac{1}{2}|-\rangle\langle -| = \frac{1}{2}I,$$

which is identical to the first case. So, the protocol is indeed perfectly concealing—Bob cannot determine any information about b at any time before the reveal phase.

This means the binding property must not hold. It is not difficult to come up with a faulty argument for why the binding property should hold (where Bob catches Alice cheating with some nonzero probability when she does so), under the assumption that Alice prepares X in some state $|\phi\rangle$. This is a bad assumption, though.

One way that Alice can cheat is as follows. She starts by preparing two qubits W and X in the state $\frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|11\rangle$, and sends X to Bob. The reduced state of Bob's qubit X at this point happens to be the totally mixed state, but this doesn't really matter—Bob is being honest, so he would not measure his qubit until the reveal phase. Alice has not really committed to anything. Time passes and the reveal phase comes. At this point, Alice may effectively decide to reveal that she "committed" $b = 0$ or $b = 1$, whichever she wants.

If she wishes to reveal $b = 0$, she measures W in the standard basis. If she gets the result 0, she sends 00 to Bob, and if she gets 1, she sends 10 to Bob. Bob measures and gets precisely the same outcome as Alice, leading him to believe that she was being honest and had committed $b = 0$ all along.

If Alice instead wishes to reveal $b = 1$, she measures W in the $\{|+\rangle, |-\rangle\}$ basis, interpreting the result as 0 or 1 respectively. (Alternately, she performs a Hadamard transform on W and measures.) She then sends the measurement outcome followed by $b = 1$ to Bob. Suppose Alice's measurement outcome was 0. Then the state of X becomes $|+\rangle$. Similarly, if Alice measures 1, the state of X becomes $|-\rangle$. This is because

$$\begin{aligned}(H \otimes I)\left(\frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|11\rangle\right) &= \frac{1}{2}|00\rangle + \frac{1}{2}|10\rangle + \frac{1}{2}|01\rangle - \frac{1}{2}|11\rangle \\ &= \frac{1}{\sqrt{2}}|0\rangle|+\rangle + \frac{1}{\sqrt{2}}|1\rangle|-\rangle.\end{aligned}$$

When Bob measures to "check" Alice's honesty, he gets always gets the outcome that leads him to believe Alice was honest.

The binding property has failed completely as we suspected.

There have been many attempts to bypass the proof that bit commitment is impossible using quantum information, but they never succeed. Indeed, there are some people who continue to write papers arguing that the proof of the impossibility of quantum bit commitment is possible to circumvent. It isn't—their arguments are always based on a misunderstanding of quantum information, cryptography, or mathematics more generally.