The Trouble with Quantum Bit Commitment

Dominic Mayers *

Université de Montréal †

8 March 1996

Abstract

In a recent paper, Lo and Chau explain how to break a family of quantum bit commitment schemes, and they claim that their attack applies to the 1993 protocol of Brassard, Crépeau, Jozsa and Langlois (BCJL). The intuition behind their attack is correct, and indeed they expose a weakness common to all proposals of a certain kind, but the BCJL protocol does not fall in this category. Nevertheless, it is true that the BCJL protocol is insecure, but the required attack is more subtle. Here we provide the first complete proof that the BCJL protocol is insecure.

* Supported in part by Canada's NSERC and Québec's FCAR.

† Département IRO, Université de Montréal, C.P. 6128, succursale centre-ville, Montréal (Québec), Canada H3C 3J7. e-mail: mayersd.umontreal.ca.

1 Introduction

Recently, Lo and Chau have made available on the `quant-ph` archives [7] a preprint that explains how to break a family of quantum bit commitment schemes, and they claim that their attack applies to the protocol of Brassard, Crépeau, Jozsa and Langlois [3], hereafter called the BCJL protocol. Furthermore, they express the concern that secure quantum bit commitment may not be possible at all. The intuition behind their attack against the BCJL protocol is correct, and indeed they expose a weakness common to all proposals of a certain kind (including [2, 1]), but the BCJL protocol does *not* fall in this category (see the opening paragraph in Section 4). Therefore, the attack of Lo and Chau does not apply to the BCJL protocol. Nevertheless, it is true that the BCJL protocol is insecure, but the required attack, which we have known for quite some time [9], is more subtle. Here we provide the first complete proof that the BCJL protocol is insecure.

We have also considered several variations on the BCJL theme. Neither Lo and Chau's attack nor the correct attack on the BCJL protocol explained below apply to these variations. One of these variations consists in having the photons travel in the reverse direction compared with the original BCJL protocol. This is natural for many cryptographic applications. Nevertheless, all these variations fail as well for different reasons related to subtle points in quantum information theory [5, 6] that only began to be understood at the time the BCJL paper was written. A proof that none of these variations work will be the subject of a forthcoming paper: the current paper focuses on the correct attack against the original BCJL protocol.

Lo and Chau ask the question: "Is quantum bit commitment really possible?". Despite our success at breaking the original BCJL protocol and several variations on it, the answer is YES. In a further paper, soon to be released, we shall offer a different quantum bit commitment protocol and give a complete proof of security against all attacks consistent with the laws of quantum physics. Of course, our new protocol does not fall in the family of protocols investigated by Lo and Chau.

Lo and Chau wrote: "The security of other quantum cryptographic protocols say for oblivious transfer [...] remains to be examined." This would indeed have been a serious concern had it been impossible to provide a secure quantum bit commitment scheme because quantum oblivious transfer depends on the security of bit commitment [13]. On the other hand, we disagree with the following sentence from Lo and Chau: "One might wonder if all of quantum cryptography may stumble under closer scrutinies" because our earlier proof of security for quantum key distribution [8] would hold even if secure quantum bit commitment had not been possible despite the fact that it is based on an earlier "proof" of security for quantum oblivious transfer that would fail in the absence of a secure bit commitment scheme. The reason is that the proof of security for quantum key distribution does not depend on the security of quantum oblivious transfer, but rather on the (correct) proof that quantum oblivious transfer would be secure if implemented on top of a secure bit commitment scheme.

2 Bit Commitment

Any cryptographic task defines the relationship between inputs and outputs respectively entered and received by the task's participants. In bit commitment, Alice enters a bit b. At a later time, Bob may request this bit and, whenever he does, he receives this bit, otherwise he learns nothing about b.

In a naive but concrete realization of bit commitment, Alice puts the bit into a strong-box of which she keeps the key and then gives this strong-box to Bob. At a later time, if Bob requests the bit, Alice gives the key to Bob. The main point is that Alice cannot change her mind about the bit b and Bob learns nothing about it unless he obtains the key.

3 The protocol of BCJL

Now, let us consider how BCJL attempts to realize bit commitment.

COMMIT(b)

1. Bob chooses a linear code C (with some required properties) and announces it to Alice.
2. Alice chooses a perfectly random string $r \in \{0,1\}^n$ and announces it to Bob.

3. Alice chooses a perfectly random string $\theta = \theta_1 \ldots \theta_n \in \{+, \times\}^n$ and a string $c = c_1 \ldots c_n$ uniformly distributed over $\{c \in \{0,1\}^n \mid c \odot r = b\}$.

4. Alice sends n photons to Bob in a product state $|c\rangle_\theta = |c_1\rangle_{\theta_1} \otimes \ldots \otimes |c_n\rangle_{\theta_n}$.

5. Bob measures the n photons in a perfectly random basis $\hat{\theta} = \hat{\theta}_1 \ldots \hat{\theta}_n \in \{+, \times\}^n$ and obtains the classical outcome $\hat{c} = \hat{c}_1 \ldots \hat{c}_n$.

To unveil the bit b, Alice announces θ, c and b to Bob. Bob computes a function $T_b(\theta, c, \hat{\theta}, \hat{c}, C, r)$ to test whether or not he should accept Alice's claim. The exact description of the function T_b is irrelevant for our analysis.

The pair (C, r) corresponds to the information that is shared between Alice and Bob just before Alice sends the photons. Most of our analysis is done for (C, r) fixed, so (C, r) is suppressed in most of the notations. For instance, from here on, (C, r) is suppressed in the input of the functions $T_b(\theta, c, \hat{\theta}, \hat{c})$.

We denote $p(c, \theta | b)$ the probability of the state $|c\rangle_\theta$ given the bit b that Alice has in mind. Such a random distribution of pure states is called a mixture. The BCJL authors [3] explains that, given b, the results of any physical measurement whatever on the mixture prepared by Alice depend only on the density operator $\rho_b = \sum_{c,\theta} p(c, \theta | b) \; |c\rangle_{\theta\,\theta}\langle c|$. This means that a dishonest Alice could send any other mixture $\{(|\psi\rangle, p(\psi))\}$ such that $\sum_\psi p(\psi) |\psi\rangle\langle\psi| = \rho_b$, for some $b \in \{0,1\}$, without being detected. It is correctly shown in [3] that such a strategy does not work. Their Theorem 3.7 implies that, for all practical purposes, any pure state $|\psi\rangle$ commits Alice to a single bit b. More precisely, once Bob has received the n photons in a pure state $|\psi\rangle$, there exists one value b such that, except with a negligible probability, Alice cannot convince Bob that she had the opposite bit $\bar{b}$ in mind, that is, $(\forall\ \theta, c)$, $\Pr(T_{\bar{b}}(\theta, c, \hat{\Theta}, \hat{C}) = ok \mid \psi)$ is exponentially small. The conclusion in [3] is that the protocol is secure against Alice.

However, as explained by these authors, preparing a mixture is not the only way to prepare a density matrix. Alice may prepare the density matrix ρ_b of the n photons by introducing another system A kept on her side and preparing the new incremented system in a pure state $|\phi\rangle$. Note that a pure state $|\phi\rangle$ of the incremented system is not in general the product of a pure state of A with a pure state of the n photons.

This possibility was considered in [3]. In the appendix of their paper, they mention the true fact that as far as the results of Bob's measurement is concerned, this alternative approach is equivalent to a preparation of a mixture by Alice. It is true that no matter what Alice does on her side, at the best, she will be found in a situation where everything behaves as if such a mixture had been sent to Bob. For any such a mixture, it is true that Alice may only open a single bit b. However, this is not sufficient to show the security against Alice. The problem, which we explain in this paper, is that by delaying her measurement on the system A that she kept on her side Alice may choose the mixture and thus the bit b after the commit.

To explain this problem, we consider another simple realization of bit commitment. This realization makes use of the $\binom{2}{1}$-string-OT primitive in which Bob enters two strings $\psi^{(0)}$ and $\psi^{(1)}$, and Alice enters a bit b and receives $\psi^{(b)}$. Alice learns nothing about $\psi^{(\bar{b})}$ and Bob learns nothing about b.

SIMPLE_COMMIT(b)

1. Bob chooses two strings $\psi^{(0)}$, $\psi^{(1)}$.
2. Alice and Bob execute $\binom{2}{1}$-string-OT$((\psi^{(0)}, \psi^{(1)}), b)$ so that Alice obtains $\psi = \psi^{(b)}$.

To unveil the bit b, Alice announces ψ and b to Bob. Bob computes the function $T_b(\psi, \psi^{(0)}, \psi^{(1)})$ which returns *ok* if and only if $\psi = \psi^{(b)}$.

Note that the function T_b does not depend only upon the information that Bob has when SIMPLE_COMMIT ends, but upon $(\psi, \psi^{(0)}, \psi^{(1)})$ which includes the information ψ that Alice has. In the protocol SIMPLE_COMMIT, Alice may choose to obtain $\psi = \psi^{(0)}$ or $\psi = \psi^{(1)}$ at her choice and, in this way, she chooses the bit b. For this protocol to be secure it is not sufficient that Alice cannot have both $\psi^{(0)}$ and $\psi^{(1)}$ in a given execution. A problem would arise if Alice could decide after the end of the protocol SIMPLE_COMMIT which of $\psi^{(0)}$ or $\psi^{(1)}$ she obtains. More precisely, the execution of the subtask $\binom{2}{1}$-string-OT should be finished when SIMPLE_COMMIT ends.

The protocol COMMIT has a problem which is similar to the potential problem that we just described in the context of SIMPLE_COMMIT. Let

W be the input $(c, \theta, \hat{c}, \hat{\theta})$ of the functions T_0 and T_1. We consider W as a random variable. We denote $W^{(0)}$ the random variable W conditioned by $b = 0$ and $W^{(1)}$ the same random variable conditioned by $b = 1$. Both $W^{(0)}$ and $W^{(1)}$ refer to the honest protocol. If the protocol is correct, except with negligible probability, we should have $T_b(W^{(b)}) = ok$, that is, if Alice has been honest and has chosen the bit b, then Bob should accept it. The problem with the protocol COMMIT is that Alice, by delaying her measurement, can choose after the commit to have W behave either as $W^{(0)}$ or $W^{(1)}$.

Let us consider what happens when Alice executes a measurement on the system A. It is enough to consider the cases where the incremented system is in a pure state. Let $\mathcal{M}$ be a complete measurement executed by Alice on the system A and $\psi = \psi^{(\mathcal{M})}$ be the classical outcome of this measurement. It is explained in [5] that, after that Alice has obtained the classical outcome ψ, the final state of the incremented system is a product $|\psi\rangle_{\mathcal{M},A} \otimes |\psi\rangle_{\mathcal{M},B}$ where $|c\rangle_{\mathcal{M},A}$ and $|c\rangle_{\mathcal{M},B}$ are respectively pure states of A and the n photons. In what follows, we suppress $\mathcal{M}$ in some notations. As far as Alice is concerned, after that she has received the information ψ, the n photons are in the pure state $|\psi\rangle_B$. A dishonest Alice chooses $\mathcal{M}$ after the commit as a function of b, then, she executes $\mathcal{M}$ and computes (c, θ) in view of ψ, and announces (b, c, θ) to Bob. Without loss of generality, we may assume that (c, θ) is the classical outcome of $\mathcal{M}$ because the additional computation may be considered as a part of $\mathcal{M}$. Now, we describe a condition which is sufficient to show that the protocol fails.

Condition 1 *There exists a state $|\phi\rangle$ of the incremented system such that, for every b, there exists a measurement $\mathcal{M}_b$ on the system A such that*

- *(c, θ) has the same probability distribution as in the honest case in which Alice chooses b,*
- *whenever Alice obtains (c, θ), the n photons on Bob's side collapse in the state $|c\rangle_\theta$ as in the honest protocol.*

Alice executes $\mathcal{M}_b$ after Bob has executed his measurement on the n photons. However, these two measurements commute, that is, the random variable W is the same whether Alice measures before or after Bob. The only thing that matters is the distribution of W. Therefore, let us assume that Alice measures before Bob. In such a case, after Alice's measurement, the situation is exactly as in the honest protocol. We conclude that when the measurement $\mathcal{M}_b$ is chosen by Alice, we have $W = W^{(b)}$, that is, W behaves as it would in the honest protocol when Alice chooses the bit b. In such a case, as mentioned before, T_b is expected to return *ok*. So Alice can cheat. In the remainder of this paper, we show that *for all practical purposes* condition 1 holds.

4 A simpler case

This section considers the security of any bit commitment protocol in which Alice commits herself to b by sending photons to Bob and where the density matrices for $b = 0$ and $b = 1$ are identical. This is precisely the case that was independently considered by Lo and Chau [7]. This is sufficient to break the old protocol in [2] (which is not surprising since a simple EPR-type attack was already included in the same paper [2]) as well as a more recent protocol proposed by Ardehali [1]. However, this analysis is insufficient to break the BCJL protocol since the density matrices ρ_b prepared by Alice, when she has respectively $b = 0$ and $b = 1$ in mind, are not identical. In fact, the main thrust of the BCJL paper was to prove that Bob could not cheat *despite* the fact that he was sent slightly different density matrices by Alice depending on which bit she wanted to commit to. (Clearly, the protocol could not be secure if the density matrices had differed too much, because then Bob would be able to distinguish between $b = 0$ and $b = 1$ without any help from Alice.) In this section, we show that the condition 1 holds under the simplifying assumption $\rho_0 = \rho_1$. In the next section, we shall consider the situation that really applies to the BCJL protocol.

In the commit part, Alice prepares a pure state of the incremented system such that the density matrix for the n photons is $\rho = \rho_0 = \rho_1$, that is, the same density matrix that would have been honestly prepared by Alice no matter whether she had $b = 0$ or $b = 1$ in mind. Now, let us assume that, after the commit, Alice wants to convince Bob that she had some bit b of her choice in mind. It is shown in [5] that by choosing the appropriate measurement on A, Alice may choose any mixture $\{(|\psi\rangle, p(\psi)\}$ such that $\sum_c p(\psi)|\psi\rangle\langle\psi| = \rho$. We recall that when Alice chooses the mixture $\{(|\psi\rangle, p(\psi)\}$, she receives the classical outcome ψ with probability $p(\psi)$ and, furthermore, the classical information ψ received by Alice uniquely determines the collapsed state $|\psi\rangle$ of the n photons on Bob's side. In particular, Alice may choose $\{(|\psi\rangle, p(\psi)\}$ to be the mixture $\{(|c\rangle_\theta, p(c,\theta|b))\}$. We have that the classical information ψ received by Alice, which uniquely determines the collapsed state $|\psi\rangle = |c\rangle_\theta$ on Bob's side, is (c,θ). This shows that condition 1 holds.

5 The real situation

Now, we consider the real situation in the BCJL protocol, where the density matrices ρ_0 and ρ_1 are not identical. If the protocol is to be secure against Bob, the density matrices ρ_0 and ρ_1 must respect some constraint. We express this constraint in terms of measurements on the n photons that return a binary classical outcome $X \in \{0,1\}$. We recall that Alice prepares the density matrix ρ_b when she has b in mind, that is, when $B = b$. Without loss of generality, we take the convention that $\Pr(X=0|B=0) \geq \Pr(X=0|B=1)$. We denote X_b the random variable X conditioned by $B = b$ so that $\Pr(X = x|B = b) = \Pr(X_b = x)$. The constraint is

$$PE - \frac{1}{2} = [\sum_{b=0}^{1} \Pr(B=b) \Pr(X_b = \bar{b})] - \frac{1}{2} \leq 2^{-\alpha n}.$$

This constraint says that no matter which measurement Bob uses to decide between $B = 0$ and $B = 1$, the probability of error is exponentially close to 1/2.

From this constraint, with some algebra, one may obtain that the Kolmogorov distance defined by

$$K = \sum_{x=0}^{1} |\Pr(X_0 = x) - \Pr(X_1 = x)|$$

is bounded by $4 \times 2^{-\alpha n}$. Note that the above constraint has been defined only in terms of measurements that return a binary classical outcome X. On the other hand, the Kolmogorov distance is defined for any measurement, that is, a measurement with an arbitrary classical outcome J. Let us prove that we did not loose any generality, that is, let us consider the Kolmogorov distance for an arbitrary measurement. The sum is over j in this case. Let A

be the set of possible values for J. Again, we denote J_b the random variable J conditioned by $B = b$. We define $A_0 = \{j \in A \mid \Pr(J_0 = j) \geq \Pr(J_1 = j)\}$ and $A_1 = A - A_0 = \{j \in A \mid \Pr(J_1 = j) > \Pr(J_0 = j)\}$. We have that

$$\begin{aligned} K &= \sum_{j \in A} |\Pr(J_0 = j) - \Pr(J_1 = j)| \\ &= (\Pr(J_0 \in A_0) - \Pr(J_1 \in A_0)) + (\Pr(J_1 \in A_1) - \Pr(J_0 \in A_1)) \\ &= |(\Pr(J_0 \in A_0) - \Pr(J_1 \in A_0))| + |(\Pr(J_0 \in A_1) - \Pr(J_1 \in A_1))| \end{aligned}$$

Let M_j be the operation associated with the result j. We define $M'_0 = \sum_{j \in A_0} M_j$ and $M'_1 = \sum_{j \in A_1} M_j$. One may easily check that M'_0 and M'_1 define an incomplete measurement with a binary classical outcome X such that $\Pr(X_b = x) = \Pr(J_b \in A_x)$. So we have obtained the desired results.

Now, let us consider the Bhattacharyya-Wootters distance

$$BW = \sum_{j \in A} \Pr(J_0 = j)^{\frac{1}{2}} \Pr(J_1 = j)^{\frac{1}{2}}$$

It is explained in [10] that $(1 - BW) \leq K/2$. Therefore, we have $BW \geq (1 - 4 \times 2^{-\alpha n})$. Furthermore, in [4, 12] it is shown that the minimum of BW over all possible measurement is the fidelity F between ρ_0 and ρ_1. So, we have $F \geq (1 - 4 \times 2^{-\alpha n})$.

A theorem due to Uhlmann [6, 11] says that the fidelity between two mixed states ρ_0 and ρ_1 on a system H is given by

$$F = \max |\langle \phi_0 | \phi_1 \rangle|^2$$

where the maximum is taken over the purifications ϕ_0 and ϕ_1 of ρ_0 and ρ_1 respectively. Let ϕ_0 and ϕ_1 be two purifications such that

$$|\langle \phi_0 | \phi_1 \rangle|^2 = F \geq (1 - 4 \times 2^{-\alpha n}).$$

We have that ϕ_0 and ϕ_1 are almost the same state.

Now, we are ready to describe Alice's strategy. Alice prepares the incremented system in the state $|\phi_0\rangle$. She could also use $|\phi_1\rangle$ or any normalized linear combination of $|\phi_0\rangle$ and $|\phi_1\rangle$. Clearly, using the same argument that works under the hypothesis $\rho_0 = \rho_1$, Alice can choose a measurement $\mathcal{M}_0$ that returns $\psi = (\theta, c)$ which she can use to convince Bob that she had $b = 0$

in mind. Now, assume that Alice wants to convince Bob that she had $b = 1$ in mind. We know that she could do it, if she had prepared $|\phi_1\rangle$ rather than $|\phi_0\rangle$. Let us consider the measurement $\mathcal{M}_1$ that works when Alice prepares $|\phi_1\rangle$. This measurement and Bob's measurement in the random basis $\hat{\theta}$, both together, determine an overall measurement $\mathcal{M}_1^{(\hat{\theta})}$ on the incremented system. This measurement returns the classical outcome $(c, \theta, \hat{c})$, that is, the classical outcomes of both Alice and Bob.

Let $M^{(\hat{\theta})}_{c,\theta,\hat{c}} = |c, \theta, \hat{c}\rangle\langle u_{c,\theta,\hat{c}}|$ be the operation associated with the outcome $(c, \theta, \hat{c})$ of the measurement $\mathcal{M}_1^{(\hat{\theta})}$. We have $\Pr(c, \theta, \hat{c}, \hat{\theta}) = (1/2^n)\langle u_{c,\theta,\hat{c}}|\phi\rangle$ where $|\phi\rangle$ is the pure state prepared by Alice. If b = 1 and $|\phi\rangle = |\phi_1\rangle$, by construction, $T_b(c, \theta, \hat{c}, \hat{\theta})$ returns *ok*, except with a negligible probability. Now, Alice has prepared $|\phi_0\rangle$, not $|\phi_1\rangle$. But, the two states are almost identical. Therefore, even if Alice has prepared $|\phi_0\rangle$, if she uses the measurement $\mathcal{M}_1$, except with a negligible probability, the function T_b will accept the input $(c, \theta, \hat{c}, \hat{\theta})$ where $b = 1$. This concludes the proof that the BCJL protocol is broken.

Note that as a consequence, Yao's proof of security for Quantum Oblivious Transfer [13] fails because it is built on insecure foundations (through no fault of Yao). Ironically, as we stated in the Introduction, the proof of security for Quantum Key Distribution shown in [8] stands despite the fact that it draws on Yao's work because it does not depend on the security of Bit Commitment.

References

[1] M. Ardehali, "A perfectly secure quantum bit commitment protocol", Los Alamos preprint archive `quant-ph/9505019`, May 1995.

[2] C.H. Bennett and G. Brassard, "Quantum cryptography: Public key distribution and coin tossing", *Proceedings of IEEE International Conference on Computers, Systems and Signal Processing*, Bangalore, India, December 1984, pp. 175–179.

[3] G. Brassard, C. Crépeau, R. Jozsa, D. Langlois, "A quantum bit commitment scheme provably unbreakable by both parties", in *Proceeding of the 34th annual IEEE Symposium on foundations of Computer Science*, November 1993, pp. 362–371.

[4] C. A. Fuchs and C. M. Caves, "Mathematical techniques for quantum communication theory", *Open Systems and Information Dynamics*, vol. 3(3), pp. 1 – 12, to be published.

[5] L. P. Hughston, R. Jozsa, and W. K. Wootters, "A complete classification of quantum ensembles having a given density matrix", *Physics Letters A*, vol. 183, pp. 14 – 18, 1993.

[6] R. Jozsa, "Fidelity for mixed quantum states", *Journal of Modern Optics*, vol. 41(12), pp. 2315 – 2323, 1994.

[7] H.-K. Lo, H. F. Chau, "Is quantum bit commitment really possible?", Los Alamos preprint archive `quant-ph/9603004`, March 1996.

[8] D. Mayers, On the security of the quantum oblivious transfer and key distribution protocols, *Advances in Cryptology: Proceeding of Crypto '95*, Lecture Notes in Computer Science.

[9] D. Mayers explained the details of his attack against the BCJL protocol at the 4th workshop on quantum information theory organized by G. Brassard in Montréal, October 1995.

[10] G. T. Toussaint, "Some functional lower bounds on the expected divergence for multihypothesis pattern recognition, communication, and radar systems", *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-1, pp. 384 – 385, 1971.

[11] A. Uhlmann, "The 'transition probability' in the state space of a *–algebra", *Reports on Mathematical Physics*, vol. 9, pp. 273 – 279, 1976.

[12] W. K. Wootters, "Statistical distance and Hilbert space", *Physical Review D*, vol. 23, pp. 357 – 362, 1981.

[13] A. Yao, Security of Quantum Protocols Against Coherent Measurements, in *Proceedings of the 26th Symposium on the Theory of Computing*, June 1995.