# Lecture 2: Quantum bit commitment and authentication

This lecture is the second and final lecture in the CryptoWorks21 Novice program in quantum cryptography tools. All future lectures are not part of the the CryptoWorks 21 Novice program.

## 1 Quantum bit commitment is impossible

A protocol for *bit commitment* solves the following problem. Alice has a bit $b$ that she wishes to *commit* to Bob, but she doesn't want Bob to learn the value of $b$ until she chooses to *reveal* it at a later time.

[*** picture of a safe and key ***]

In other words, a protocol for bit commitment must meet the following conditions:

**Binding.** Alice should not be able to alter the value of the bit once she has committed.

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

**Applications:** Bit commitment $\implies$ multi-party computation, zero-knowledge proofs for NP-complete problems, coin tossing. (That is, secure protocols for these tasks can be constructed from a secure bit commitment protocol.)

**Example 1** (Bit commitment $\implies$ coin tossing)**.** The following protocol uses a bit commitment primitive to implement coin tossing.

1. Alice guesses a random bit $b$ and commits this guess to Bob.
2. Bob flips a coin (selects a random bit $c$) and reports the result to Alice.
3. Alice reveals her guess. The output is $c \oplus b$.

It is clear that Bob cannot cheat: if Alice is honest then the result of the coin flip is always uniformly random. In order for Alice to cheat she must be able to change the value of her bit $b$ after she has commited to it, contradicting the binding property of bit commitment. □

### 1.1 Bit commitment in a classical universe

Unconditionally secure bit commitment is impossible. Any protocol for bit commitment requires computational assumptions such as the existence of one-way functions. There exist protocols that are unconditionally binding and computationally concealing and *vice versa*.

**Example 2** (Bit commitment from a pseudo-random generator [Nao91])**.** Let $R$ be a PRG from $n$ bits to $3n$ bits and suppose Alice wishes to commit a bit $b$.

1. Bob selects a random $3n$-bit string $s$ and sends it to Alice.
2. Alice selects a random $n$-bit string $y$ and computes the $3n$-bit string $R(y)$.
3. **Commit.** If $b = 1$ then Alice sends $R(y)$ to Bob. Otherwise, she sends $R(y) \oplus s$ to Bob.

4. **Reveal.** Alice sends $y$ to Bob, who can then check whether he received $R(y)$ or $R(y) \oplus s$.

**Statistically binding.** Alice cannot cheat with probability greater than $2^{-n}$ even if she has unlimited computational power at her disposal. Why? Because if she wants to cheat then she needs to find $y'$ with $R(y') = R(y) \oplus s$. However, $R(y) \oplus s$ is uniformly distributed among $2^{3n}$ possible strings, whereas $R(y')$ is drawn from a pool of only $2^n$ strings. Thus, the probability that there exists some $y'$ with the desired property is vanishingly small.

**Computationally concealing.** In order to distinguish $R(y)$ from $R(y) \oplus s$ Bob must distinguish true randomness ($R(y) \oplus s$) from pseudo-randomness ($R(y)$), implying that he could break the PRG. □

### 1.2 Unconditionally secure quantum bit commitment is impossible

Can unconditionally secure bit commitment be achieved with quantum information? No. Here's a sketch of the proof.

By the Church of a Larger Hilbert Space, we can assume without loss of generality that Alice and Bob begin with a pure state and perform only local unitary operations, with all measurements deferred until the end of the protocol. Let $\mathcal{A}, \mathcal{B}$ denote the local registers of Alice, Bob.

Any protocol for bit commitment has two *phases*: a commit phase and a reveal phase. Suppose Alice wishes to commit bit $b$ and let $|\psi_b\rangle \in \mathcal{AB}$ denote the pure state the entire system at the end of the commit phase. In a perfectly concealing protocol Bob cannot infer any information about $b$ from his local portion $\mathcal{B}$. Thus, it must be that

$$\mathrm{Tr}_{\mathcal{A}}(|\psi_0\rangle\langle\psi_0|) = \mathrm{Tr}_{\mathcal{A}}(|\psi_1\rangle\langle\psi_1|).$$

By the unitary equivalence of purifications, it follows that there exists a unitary $U$ acting only on $\mathcal{A}$ with

$$(U \otimes I_{\mathcal{B}})|\psi_0\rangle = |\psi_1\rangle.$$

In particular, Alice can switch back and forth between $|\psi_0\rangle, |\psi_1\rangle$ after the commit phase without Bob's knowledge or consent; the binding property does not hold.

The previous argument showed that a *perfectly* concealing protocol cannot be binding. But what about a *statistically* concealing protocol? The impossibility proof sketch is the same except now we need to carry around epislons and convert between the trace norm and fidelity. As a bonus, we can compute a lower bound on the probability with which one party can cheat in any bit commitment protocol.

**Theorem 3.** *Suppose that Bob can deduce the bit $b$ before the reveal phase with bias no larger than $\varepsilon$. Then Alice can convince Bob to accept the opposite bit value $\overline{b}$ after the commit phase with probability at least $1 - 2\sqrt{\varepsilon}$. It follows that one party can always cheat with probability at least $17.1\%$.*

**Legend.** In 1993 a group of prominent researchers published a paper in the proceedings of a top-tier computer science conference [BCJL93] in which they claimed to exhibit an unconditionally secure quantum bit commitment scheme. A flaw was subsequently found in their scheme, and quantum bit commitment was proven impossible. There is some dispute over who should be credited with the discovery of this flaw and the impossibility proof. These days, credit is typically given to both Mayers [May97] and Lo and Chau [LC97] as independent discoverers.

## 2 Authentication of quantum messages

### 2.1 Review of classical authentication

Let $K, M, C$ denote finite sets of keys, messages, and ciphertexts, respectively. (Think of these as sets of bit strings.) An *authentication scheme* (also called a *message authentication code (MAC)*) consists of encoding

and decoding functions

$$E : K \times M \to C$$
$$D : K \times C \to M \times \{\text{valid}, \text{invalid}\}$$

Such a scheme is said to be $\varepsilon$-*secure* if it has the following conditions:

**Completeness.** Untampered messages are accepted with certainty: $D_k(E_k(m)) = (m, \text{valid})$ for all messages $m$ and keys $k$.

**Soundness.** Tampered messages are rejected with high probability over the choice of key $k$. That is, for any adversary $A : C \to C$ and any distinct messages $m, m'$ we have

$$\Pr_{k \in K}\left[D_k(A(E_k(m))) = (m', \text{valid})\right] < \varepsilon$$

where the security parameter $\varepsilon$ should drop exponentially in the bit length of ciphertexts.

**Example 4** (Wegman-Carter)**.** Let $H$ be a set of hash functions on $M$ and for each $h \in H$ define

$$E_h : m \mapsto (m, h(m))$$
$$D_h : (x, y) \mapsto \begin{cases} (x, \text{valid}) & \text{if } h(x) = y \\ (x, \text{invalid}) & \text{otherwise} \end{cases}$$

If $H$ is a *universal class of hash functions* then $E_h, D_h$ are a secure authentication scheme with key set $K = H$. □

**Observe:** This scheme authenticates but does not encrypt—the plaintext message is transmitted in the clear!

(In case you're wondering, a set $H$ of hash functions is called a *universal class* if for each pair $m, m' \in M$ of distinct messages it holds that

$$\Pr_{h \in H}\left[h(m) = h(m')\right] \leq \frac{1}{|H|}.$$

In other words, sampling uniformly from $H$ has the same effect as sampling uniformly from *all possible* hash functions.)

## 2.2 Definition of quantum authentication

Getting the definition right is tricky, owing to the fact that quantum information has a continuum of pure states. For discussion of definitional issues see the original work of Barnum *et al.* [BCG$^+$02].

Let's cut to the chase with the correct definition. Let $K$ denote a finite set of (classical) keys. Let $\mathcal{M}, \mathcal{C}$ denote spaces associated with quantum systems for the message and ciphertext, respectively, and let $\mathcal{Q}$ be a two-dimensional space with basis states $\{|\text{valid}\rangle, |\text{invalid}\rangle\}$. A *quantum authentication scheme* consists of encoding and decoding channels for each key $k \in K$:

$$E_k : \mathrm{L}(\mathcal{M}) \to \mathrm{L}(\mathcal{C})$$
$$D_k : \mathrm{L}(\mathcal{C}) \to \mathrm{L}(\mathcal{M} \otimes \mathcal{Q})$$

For each pure state $|\psi\rangle \in \mathcal{M}$ of the message space let

$$\Pi_{|\psi\rangle} = (I_{\mathcal{M}} - |\psi\rangle\langle\psi|) \otimes |\text{valid}\rangle\langle\text{valid}|$$

denote the projection onto the subspace of $\mathcal{M} \otimes \mathcal{Q}$ spanned by message states orthogonal to $|\psi\rangle$ that are nonetheless accepted by the scheme.

A quantum authentication scheme is said to be $\varepsilon$-*secure* if it meets the following conditions for every $|\psi\rangle \in \mathcal{M}$:

**Completeness.** Untampered messages are accepted with certainty:

$$D_k(E_k(|\psi\rangle\langle\psi|)) = |\psi\rangle\langle\psi| \otimes |\text{valid}\rangle\langle\text{valid}|.$$

**Soundness.** Tampered messages are rejected with high probability over the choice of key $k$. That is, for any adversary $A : \mathrm{L}(\mathcal{C}) \to \mathrm{L}(\mathcal{C})$ let

$$\rho_{A,|\psi\rangle} = \frac{1}{|K|} \sum_{k \in K} D_k(A(E_k(|\psi\rangle\langle\psi|)))$$

denote the mixed state of $\mathcal{M} \otimes \mathcal{Q}$ at the end of the protocol, averaged over all choices of key $k$. The soundness requirement is that for all $|\psi\rangle, A$ it holds that

$$\left\langle \Pi_{A,|\psi\rangle}, \rho \right\rangle \stackrel{\text{def}}{=} \mathrm{Tr}\left(\Pi_{A,|\psi\rangle}\rho\right) \leq \varepsilon$$

where the security parameter $\varepsilon$ should drop exponentially in the number of qubits of $\mathcal{C}$.

There are several quantum authentication schemes out there. Here are two of the simplest:

**Example 5** (Clifford scheme [ABOE10])**.** $m$ message qubits are authenticated into $m + d$ cipher qubits. Keys are $(m + d)$-qubit Clifford circuits $C$.

$E_C$ Append $d$ ancilla qubits in state $|0\rangle^{\otimes d}$, then apply $C$.

$D_C$ Apply the inverse $C^*$, then measure the $d$ ancilla qubits to ensure they're in the $|0\rangle$ state.

This scheme achieves security $2^{-d}$. □

**Example 6** (Trap scheme [BGS12])**.** Based on any $[[n, 1, d]]$ quantum error-correcting code $C$. One message qubit is authenticated into $3n$ cipher qubits. Keys are pairs $(\sigma, P)$ where $\sigma$ is a permutation on $3n$ elements and $P$ is a $3n$-qubit Pauli operator.

$E_{(\sigma,P)}$ Encode the message qubit according to $C$. Append $2n$ *trap qubits* in state $|0\rangle^{\otimes n}|+\rangle^{\otimes n}$, then permute all $3n$ qubits according to $\sigma$, then encrypt all $3n$ qubits by applying $P$.

$D_{(\sigma,P)}$ Decrypt by applying $P^*$, de-permute by applying $\sigma^{-1}$, and decode by applying $C^*$. Measure the trap qubits and the syndrome qubits of $C$ and reject if any errors are detected.

This scheme achieves security at least $(2/3)^{d/2}$. □

## 2.3 Authentication requires encryption

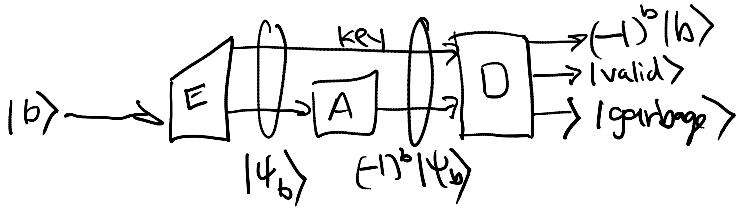
Recall that classical authentication schemes need not encrypt the message. By contrast, notice that both of our examples of quantum authentication schemes also encrypt the message. This is no coincidence: it is impossible to authenticate a quantum state without also encrypting it.

Why? Suppose toward a contradiction that we have a quantum authentication scheme that does not also encrypt. For a given message state $|\psi\rangle$ let $\rho_{|\psi\rangle}$ denote the authentication of $|\psi\rangle$ under this scheme.

Let us begin with an extreme example: suppose that there are message states $|0\rangle, |1\rangle$ whose ciphertexts $\rho_{|0\rangle}, \rho_{|1\rangle}$ are perfectly distinguishable. Then $\rho_{|0\rangle}, \rho_{|1\rangle}$ must have orthogonal support.

For $b \in \{0, 1\}$ let $\Pi_b$ denote the projection onto the support of $\rho_{|b\rangle}$ and consider an adversary who applies the unitary $\Pi_0 - \Pi_1$. That is, the adversary applies a $-1$ phase to the support of $\rho_{|1\rangle}$.

May-06-13
4:45 PM




New Section 1 Page 1

Figure 1: An adversary who can distinguish $\rho_{|0\rangle}$ from $\rho_{|1\rangle}$ can can apply a conditional phase to map $\rho_{|0\rangle+|1\rangle}$ to $\rho_{|0\rangle-|1\rangle}$, thus breaking the authentication scheme. In this figure, $|\psi_b\rangle$ denotes the purification (including the secret key) of the authenticated state $\rho_{|b\rangle}$ seen by the adversary.

When we try to authenticate $|0\rangle + |1\rangle$ the adversary maps $\rho_{|0\rangle+|1\rangle}$ to $\rho_{|0\rangle-|1\rangle}$. Thus, the decoder can be made to accept $|0\rangle - |1\rangle$ with certainty even though the original message was $|0\rangle + |1\rangle$. Our hypothesized scheme is insecure.

What if $\rho_{|0\rangle}, \rho_{|1\rangle}$ are not *perfectly* distinguishable, but only *mostly* distinguishable? One can still prove a bound on the security of the scheme.

**Proposition 7** ([BCG$^+$02])**.** *If $|0\rangle, |1\rangle$ are message states with $\|\rho_{|0\rangle} - \rho_{|1\rangle}\|_{\mathrm{Tr}} > 2 - \varepsilon$ then an adversary can successfully tamper with $\rho_{|0\rangle+|1\rangle}$ with probability at least $1 - \varepsilon$.*

What if $\rho_{|0\rangle}, \rho_{|1\rangle}$ are only *slightly* distinguishable? One can still prove a bound on the security of the scheme. We use the well known fact that distinguishability of $t$ copies of two states increases exponentially in $t$. Specifically,

$$\|\rho_{|0\rangle} - \rho_{|1\rangle}\|_{\mathrm{Tr}} \geq \delta \implies \|\rho_{|0\rangle^{\otimes t}} - \rho_{|1\rangle^{\otimes t}}\|_{\mathrm{Tr}} \geq 2 - \exp(-t\delta^2/4).$$

Then by Proposition 7 an adversary can change the message state $|0^t\rangle + |1^t\rangle$ to $|0^t\rangle - |1^t\rangle$ with high probability. All this can be formalized to prove:

**Theorem 8** ([BCG$^+$02])**.** *Any $\varepsilon$-secure quantum authentication scheme has the property that for any message states $|\phi\rangle, |\psi\rangle$ it holds that*

$$\|\rho_{|\phi\rangle} - \rho_{|\psi\rangle}\|_{\mathrm{Tr}} \leq O(\varepsilon^{1/6}).$$

*In other words, such a scheme must also encrypt message states so as to achieve distinguishability bias $O(\varepsilon^{1/6})$.*

**Why doesn't classical authentication also require encryption?**

The original authors said it best [BCG$^+$02]:

The difference can be understood as a complementarity feature of quantum mechanics: authenticating a message in one basis requires encrypting it in the complementary Fourier-transformed basis. This is essentially another realization of the principle that measuring data in one basis disturbs it in any complementary basis. For classical messages, therefore, encryption is not required: only one basis is relevant. In contrast, for quantum messages, we require authentication in all bases and therefore we must also require encryption in all bases.

## References


[ABOE10] Dorit Aharonov, Michael Ben-Or, and Elad Eban. Interactive proofs for quantum computations. In *Proceedings of Innovations in Computer Science*, pages 453–469, 2010. arXiv:0810.5375v2 [quant-ph]. 4

[BCG+02] Howard Barnum, Claude Crépeau, Daniel Gottesman, Adam Smith, and Alain Tapp. Authentication of quantum messages. In *Proceedings of the 43rd IEEE Symposium on Foundations of Computer Science*, FOCS 2002, pages 449–458, 2002. arXiv:quant-ph/0205128. 3, 5

[BCJL93] Gilles Brassard, Claude Crépeau, Richard Jozsa, and Denis Langlois. A quantum bit commitment scheme provably unbreakable by both parties. In *Proceedings of the 34th IEEE Symposium on Foundations of Computer Science*, FOCS 1993, pages 42–52, 1993. 2

[BGS12] Anne Broadbent, Gus Gutoski, and Douglas Stebila. Quantum one-time programs. To appear in CRYPTO 2013. arXiv:1211.1080 [quant-ph], 2012. 4

[LC97] Hoi-Kwong Lo and H. F. Chau. Is quantum bit commitment really possible? *Physical Review Letters*, 78:3410–3413, 1997. arXiv:quant-ph/9603004. 2

[May97] Dominic Mayers. Unconditionally secure quantum bit commitment is impossible. *Physical Review Letters*, 78:3414–3417, 1997. arXiv:quant-ph/9605044. 2

[Nao91] Moni Naor. Bit commitment using pseudorandomness. *Journal of Cryptology*, 4(2):151–158, 1991. 1