Lecture Notes

Quantum Cryptography Week 7:

Quantum cryptography using untrusted devices

Contents

This week we introduce a new variant of the BB'84 quantum key distribution protocol we studied last week. This variant is due to Ekert [Eke91] and is often referred to as the E'91 protocol for quantum key distribution (since our protocol won't exactly follow Ekert's original proposal we will simply call it the "DIQKD protocol". Although it looks rather similar to the BB'84 protocol, or more specifically its purified version, the key difference is that we use a different test for step 7. in the protocol (see the description of BB'84 from last week's notes).

Instead of the "matching outputs" test considered in the BB'84 protocol, Ekert's protocol uses a test based on the CHSH game (recall the game from week 2!). The advantage of using this test is that it allows us to prove that the protocol is secure without relying on Alice and Bob performing trusted measurements on their qubit in each round — in fact, without even relying on the assumption that the system they measure is a qubit! This stronger notion of security is called *device-independent security*, and we'll define it in more detail later in these notes.

Before introducing Ekert's protocol and its analysis, we first return to the CHSH game and give a more detailed analysis of its properties than we did in week 2. This game turns out to have a striking property, which forms the key to its use in the DIQKD protocol. This is the property of *rigidity*, which states that optimal strategies for the players in the game are unique in a very strong sense: any strategy that achieves the optimum success probability $p^*_{\text{CHSH}} = \cos^2 \pi/8$, or even close to the optimum, must be equivalent (in a sense to be made precise later) to the strategy we saw in week 2. There is no alternative! As an immediate consequence we get that any strategy with close to optimal success probability must involve a shared entangled state between Alice and Bob that is equivalent to an EPR pair, just as the optimal strategy we described does. This fact does not need us to assume any a priori knowledge about the state or the measurements used in the strategy.

7.1 Testing EPR pairs

Recall that in the CHSH game the referee sends each of the two players, Alice and Bob, a uniformly random bit $x, y \in \{0,1\}$ respectively. The players have to return outcomes $a, b \in \{0,1\}$ such that the CHSH condition $a \oplus b = x \wedge y$ is satisfied. We saw that the maximum success probability of classical non-communicating players in this game is $p_{\text{CHSH}} = 3/4$, while if Alice and Bob are quantum there is a strategy that allows them to succeed with probability $p^*_{\text{CHSH}} = \cos^2 \pi/8 \approx 0.85$.

In the strategy we described, Alice and Bob share an EPR pair $|\phi^+\rangle_{AB}$ and make the following measurements. When $x = 0$, Alice measures her qubit in the standard basis $\{|0\rangle, |1\rangle\}$, and when $x = 1$ she measures in the Hadamard basis $\{|+\rangle, |-\rangle\}$. When $y = 0$, Bob measures his qubit in the basis $\{\cos(\pi/8)|0\rangle + \sin(\pi/8)|1\rangle, -\sin(\pi/8)|0\rangle + \cos(\pi/8)|1\rangle\}$ and when $y = 1$, he measures in the basis $\{\cos(\pi/8)|0\rangle - \sin(\pi/8)|1\rangle, \sin(\pi/8)|0\rangle + \cos(\pi/8)|1\rangle\}$. Since these measurements are binary projective measurements, with POVM elements of the form $\{\Pi, \mathbb{I} - \Pi\}$, we can equivalently describe them using the associated *observables* $O = \mathbb{I} - 2\Pi$. Note that O is a Hermitian operator which squares to identity. For Alice's measurements the observables are

$$A_0 = |0\rangle\langle 0| - |1\rangle\langle 1| = Z \ (x = 0) \quad \text{and} \quad A_1 = |+\rangle\langle +| - |-\rangle\langle -| = X \ (x = 1).$$

For Bob we have

$$B_0 = H \ (y = 0) \quad \text{and} \quad B_1 = \tilde{H} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ -1 & -1 \end{pmatrix} \ (y = 1).$$

We introduced this as a "good" strategy for the players: it certainly beats the classical bound $p_{\text{CHSH}} = 3/4$, and achieves $p^*_{\text{CHSH}} = \cos^2 \pi/8$. But could there be better strategies, achieving an even larger value? Or, even if they are not better, different strategies, based on using a different type of entangled state, for achieving the same success probability?

We're going to show that this is not the case: the maximum success probability of any quantum strategy in the CHSH game, as complicated as it may be, is p^*_{CHSH}. Moreover, any strategy achieving this value must be "equivalent" to the strategy described above. What do we mean by equivalent? We couldn't possibly hope to claim that the strategy is strictly unique. For example, if Alice and Bob were to rotate their basis choices by the same angle, then since the EPR pair is itself rotation invariant their success probability would remain unchanged. The theorem shows that this local degree of freedom is essentially the only flexibility that the players have in designing an optimal strategy.

Theorem 7.1.1 — CHSH rigidity. Suppose given an entangled state $|\psi\rangle_{AB} \in \mathbb{C}^{d_A} \otimes \mathbb{C}^{d_B}$ and observables A_0, A_1 for Alice and B_0, B_1 for Bob such that the corresponding strategy has a success probability $p^*_{\text{CHSH}} = \cos^2 \pi/8$ in the CHSH game. Then there exist local isometries $U_A : \mathbb{C}^{d_A} \to \mathbb{C}^2 \otimes \mathbb{C}^{d_{A'}}$ and $V_B : \mathbb{C}^{d_B} \to \mathbb{C}^2 \otimes \mathbb{C}^{d_{B'}}$ such that

$$U_A \otimes V_B |\psi\rangle_{AB} = |\phi^+\rangle \otimes |\text{junk}\rangle_{\text{A}'\text{B}'}$$

and

$$\begin{aligned}
(U_A \otimes V_B)(A_0 \otimes \mathbb{I}_B)|\psi\rangle &= ((Z \otimes \mathbb{I})|\phi^+\rangle) \otimes |\text{junk}\rangle, \\
(U_A \otimes V_B)(A_1 \otimes \mathbb{I}_B)|\psi\rangle &= ((X \otimes \mathbb{I})|\phi^+\rangle) \otimes |\text{junk}\rangle, \\
(U_A \otimes V_B)(\mathbb{I}_A \otimes B_0)|\psi\rangle &= ((\mathbb{I} \otimes H)|\phi^+\rangle) \otimes |\text{junk}\rangle, \\
(U_A \otimes V_B)(\mathbb{I}_A \otimes B_1)|\psi\rangle &= ((\mathbb{I} \otimes \tilde{H})|\phi^+\rangle) \otimes |\text{junk}\rangle.
\end{aligned}$$

In words, the theorem says that if a strategy achieves the optimal value in CHSH then up to some local rotations on Alice and Bob's spaces it looks exactly as the strategy described above. We called the rotation "isometries" because their range might not be the whole space; in particular it is not necessarily the case that d_A or d_B are even. The state $|\text{junk}\rangle$ is an arbitrary state that does not matter for the purposes of analyzing the strategy. This state is unavoidable, as any strategy can always be made to appear more complicated by extending the entangled state arbitrarily, and making the players' measurements act as identity on the extended space.

Note also the theorem presupposes that the players' strategy can be described by observables, or equivalently binary projective measurements. More generally we may consider players that apply a non-projective POVM. However, a POVM can always be simulated with a projective measurement acting on a larger space, so the assumption is without loss of generality.

R In practice it will never be possible to certify that a given device implements a strategy with optimal success probability in the CHSH game: at best, by repeated testing it will be possible to verify that it achieves a success probability at least $p^*_{\text{CHSH}} - \delta$, where $\delta > 0$ is a quantity depending on the quality of the device and on the accuracy of the testing performed (i.e. the number of repetitions of the game). To handle this it is important to obtain "robust" analogues of Theorem 7.1.1. Such a result is known, where the exact equalities in Theorem 7.1.1 are replaced by approximations in trace distance with an error scaling as $O(\sqrt{\delta})$ [**mckague2012robust**].

Before we get to the proof of the theorem we make a small detour and explore the notion of angle between a pair of projection operators.

7.1.1 Principal angles and Jordan's lemma

Consider two lines through the origin in the complex plane $\mathbb{C}^2$. Each line is described by a unit vector $|u\rangle$, $|v\rangle$, and (ignoring any orientation) the angle between the two lines is the unique $\theta \in [0, \pi/2)$ such that $\cos^2 \theta = |\langle u|v\rangle|^2$. Up to a change of basis we can always consider that $|u\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and (up to an irrelevant phase) $|v\rangle = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}$. A more pedantic way to describe the angle between the two lines is through the associated rank-1 projections $P = |u\rangle\langle u|$ and $Q = |v\rangle\langle v|$: there will always exist a choice of basis for $\mathbb{C}^2$ in which

$$P = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \qquad \text{and} \qquad Q = \begin{pmatrix} \cos^2\theta & \cos\theta\sin\theta \\ \cos\theta\sin\theta & \sin^2\theta \end{pmatrix},$$

for some $\theta \in [0, \pi/2)$.

How do we generalize the notion of angle to higher dimensional subspaces? The notion of principal angle provides an inductive definition. Suppose P and Q are two orthogonal projections in $\mathbb{C}^d$. (We identify the projections with the space on which they project.) The smallest principal angle between P and Q is defined as $\theta_1 \in [0, \pi/2)$ such that

$$\cos^2\theta_1 = \sup_{|u\rangle\in P, |v\rangle\in Q} |\langle u|v\rangle|^2,$$

where by $|u\rangle \in P$ we mean any unit vector in the range of P, i.e. such that $P|u\rangle = |u\rangle$. This is a natural definition: we are finding the lines lying in P and Q that form the smallest possible angle . If P and Q intersect, then they share a vector and $\theta_1 = 0$.

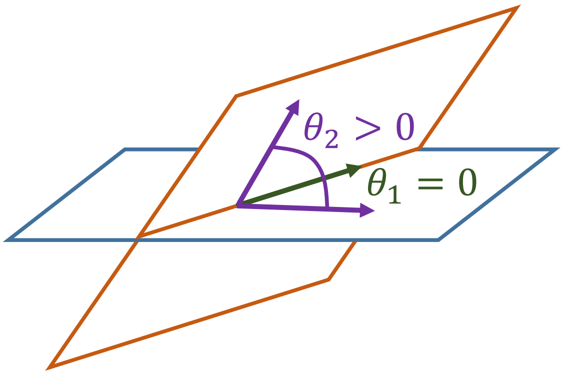

Figure 7.1: Principal angles between two 2-dimensional subspaces in 3 dimensions. The subspaces intersect, and the smallest angle is $\theta_1 = 0$. The second principal angle is $\theta_2 > 0$.

We define principal angles $\theta_2, \dots, \theta_d$, where $d = \min(\operatorname{rank} P, \operatorname{rank} Q)$, inductively via

$$\cos^2\theta_i = \sup_{\substack{|u_i\rangle\in P,\, |u_i\rangle\perp \mathrm{Span}\{|u_1\rangle,\dots,|u_{i-1}\rangle\} \\ |v_i\rangle\in Q,\, |v_i\rangle\perp \mathrm{Span}\{|v_1\rangle,\dots,|v_{i-1}\rangle\}}} |\langle u_i|v_i\rangle|^2,$$

where $|u_1\rangle, \dots, |u_{i-1}\rangle$ are unit vectors in P that achieve the optimum in the definition of $\theta_1, \dots, \theta_{i-1}$ respectively, and similarly for the $|v_j\rangle$ and Q.

Jordan's lemma states that associated with the principal angles comes a very convenient simultaneous block decomposition of P and Q.

Lemma 1 — Jordan's lemma. Let P and Q be two projection operators in $\mathbb{C}^d$. Then there exists a basis of $\mathbb{C}^d$ in which P and Q are simultaneously block diagonal, with blocks of size one or two such that either (for one-dimensional blocks)

$$P, Q \in \{(0), (1)\},$$

or (for two-dimensional blocks)

$$P = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad Q = \begin{pmatrix} \cos\theta_i^2 & \cos\theta_i\sin\theta_i \\ \cos\theta_i\sin\theta_i & \sin\theta_i^2 \end{pmatrix},$$

with $\theta_1, \dots, \theta_d \in (0, \pi/2]$, $d = \min(\operatorname{rank} P, \operatorname{rank} Q)$, the principal angles between P and Q.

The proof of the lemma is not very hard, and considers an alternate definition of the principal angles via the singular values of the operator PQ (see e.g. Exercise VII.1.10 of [Bha13]).

7.1.2 Proof of the rigidity theorem

The proof of Theorem 7.1.1 proceeds in two steps. In the first step we use Jordan's lemma to reduce the case of general strategies to the case of "qubit strategies", for which the shared state is a two-qubit entangled states and the players' observables single-qubit observables. In the second step we analyze qubit strategies in detail and show that they must take the form of Pauli measurements on an EPR pair.

1. Reduction to qubit strategies.

Consider an arbitrary strategy $|\psi\rangle_{AB}$, A_0, A_1, B_0, B_1. Apply Jordan's lemma to the projections $P = \frac{1}{2}(\mathbb{I}+A_0)$ and $Q = \frac{1}{2}(\mathbb{I}+A_1)$. The lemma gives a basis for Alice's space $\mathbb{C}^{d_A}$ such that both P and Q are block-diagonal in that basis, with blocks of size at most 2×2. Then $A_0 = 2P - \mathbb{I}$ and $A_1 = 2Q - \mathbb{I}$ are block-diagonal in the same basis.

This block-diagonal decomposition lets us reformulate Alice's strategy as follows: each of her two-outcome projective measurements is equivalent to a measurement which (i) applies a multiple-outcome projective measurement that projects on the individual blocks of the decomposition, and (ii) depending on the block obtained as outcome performs the basis measurement associated with the restriction of A_0 (or A_1) to that block.

Exercise 7.1.1 Suppose that after application of Jordan's lemma we discover a basis

$$\{|u_1\rangle, |u_2\rangle, |u_3\rangle, |u_4\rangle, |u_5\rangle\} \tag{7.1}$$

of $\mathbb{C}^5$ in which

$$A_0 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad A_1 = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} & 0 & 0 & 0 \\ -\frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 \end{pmatrix}.$$

Consider the two-outcome projective measurements associated with A_0 and A_1. Give an equivalent description of these measurements as the combination of a projective measurement $\{\Pi_0, \Pi_1, \Pi_2\}$ followed by a basis measurement involving at most 2 basis elements. The projective measurement should be independent of Alice's input x, while the basis measurement should depend both on the outcome of the projective measurement and Alice's input. ■

The same argument can be applied to Bob's observables. Now the key point is that, since the block decomposition is the same for A_0 and A_1 (resp. B_0 and B_1), step (i) associated with projection on the blocks does not depend on the player's question. Thus the step could be performed even before the game even starts, without affecting their success probability! But then the players are really playing the game with a qubit strategy — whichever qubit strategy corresponds to the outcomes they obtained when applying the projective measurement from step (i).

This reformulation of an arbitrary strategy shows that it can always be reduced to a convex combination of qubit strategies, and it will be sufficient to analyze the latter.

2. Optimal qubit strategies.

To prove the theorem we first express the success probability p^*_{win} of a given quantum strategy in terms of the observables A_x and B_y.

Exercise 7.1.2 Using the definition of the winning criterion $a \oplus b = x \wedge y$ and the relation between observables and binary measurements, show that

$$p^*_{\text{win}} = \frac{1}{2} + \frac{1}{8}\langle \psi | A_0 \otimes B_0 + A_0 \otimes B_1 + A_1 \otimes B_0 - A_1 \otimes B_1 | \psi \rangle. \tag{7.2}$$

■

Let's call the operator appearing inside the bra-ket in (7.2) the CHSH operator,

$$\text{CHSH} = \mathrm{A}_0 \otimes \mathrm{B}_0 + \mathrm{A}_0 \otimes \mathrm{B}_1 + \mathrm{A}_1 \otimes \mathrm{B}_0 - \mathrm{A}_1 \otimes \mathrm{B}_1 .$$

The main trick in the proof is to consider the square of this operator. Using $A_0^2 = A_1^2 = B_0^2 = B_1^2 = \mathbb{I}$, we get

$$\begin{aligned}\text{CHSH}^2 &= \big((A_0 + A_1) \otimes B_0 + (A_0 - A_1) \otimes B_1\big)^2 \\ &= (A_0 + A_1)^2 \otimes \mathbb{I} + (A_0 - A_1)^2 \otimes \mathbb{I} + (A_0 + A_1)(A_0 - A_1) \otimes B_0 B_1 \\ &\qquad + (A_0 - A_1)(A_0 + A_1) \otimes B_1 B_0 \\ &= 4\mathbb{I} + [A_0, A_1] \otimes [B_1, B_0],\end{aligned} \tag{7.3}$$

where $[A_0,A_1] = A_0A_1 - A_1A_0$ and $[B_1,B_0] = B_1B_0 - B_0B_1$ are the commutators. Since the operator norm (the largest singular value) of $[A_0,A_1]$ and $[B_0,B_1]$ is each at most 2, the norm of CHSH^2 (which is simply the largest overlap of CHSH^2 with a unit vector) is at most 8. Plugging back into (7.2), even an optimal choice of $|\psi\rangle$ (i.e. an eigenvector of CHSH associated to its largest singular value) will give a value at most $p^*_{\text{win}} \leq 1/2 + \sqrt{8}/8 = \cos^2 \pi/8 = p^*_{\text{CHSH}}$. Thus $\cos^2 \pi/8$ is indeed the maximum probability of success in the CHSH game.

Note that so far we have not used the reduction to qubit strategies discussed in the previous section, and the preceding argument is completely general. Let's now assume we are working with a qubit strategy which achieves the optimal $p^*_{\text{win}} = p^*_{\text{CHSH}}$. Then all inequalities discussed above must be tight. In particular, $|\psi\rangle$ must be an eigenvector of CHSH with eigenvalue $2\sqrt{2}$, and as a consequence of (7.3) $|\psi\rangle$ must also be an eigenvector of $[A_0,A_1] \otimes [B_0,B_1]$ with associated eigenvalue 4. Squaring this operator,

$$\big([A_0,A_1]^2 \otimes [B_0,B_1]^2\big)|\psi\rangle = 16|\psi\rangle.$$

Using further that $[A_0,A_1]^2 \leq 4\mathbb{I}$ and $[B_0,B_1]^2 \leq 4\mathbb{I}$ we get that necessarily

$$\big([A_0,A_1]^2 \otimes \mathbb{I}\big)|\psi\rangle = \big(\mathbb{I} \otimes [B_0,B_1]^2\big)|\psi\rangle = 4|\psi\rangle, \tag{7.4}$$

as neither operator can reduce the norm of $|\psi\rangle$. Assume $|\psi\rangle$ is not trivial, in the sense that its reduced density matrices on A and B have rank 2 (if this is not the case then it is easy to see that the strategy boils down to a classical strategy, which cannot achieve a success probability larger than $p_{\text{CHSH}} = 3/4$). Tracing out the A or B qubits in (7.4) and inverting the reduced density matrix of $|\psi\rangle$ on the remaining qubit gives us the operator equalities $A_0A_1 = -A_1A_0$ and $B_1B_0 = -B_0B_1$: Alice's and Bob's observables pairwise anti-commute. It turns out that anti-commutation is a surprisingly strong constraint, as shown in the following exercise.

Exercise 7.1.3 Suppose that R and S are two observables on $\mathbb{C}^2$ such that $RS = -SR$. Then there exists a basis of $\mathbb{C}^2$ in which $R = Z$ and $S = X$. [Hint: first show that we cannot have $R = \mathbb{I}$ or $R = -\mathbb{I}$, and deduce the eigenvalues of R. Use this to write R in a convenient form, and then use the anti-commutation relation to express S.] ■

Applying the results of the exercise to A_0 and A_1 we obtain a rotation U_A on Alice's qubit such that $U_A A_0 U_A^\dagger = Z$ and $U_A A_1 U_A^\dagger = X$. Similarly, for Bob's observables we may find a unitary U_B such that $U_B B_0 U_B^\dagger = H$ and $U_B B_1 U_B^\dagger = \tilde{H}$. Note that for Bob we are using H and $\tilde{H}$ in lieu of X and Z, but any pair of single-qubit observables will do. To conclude it remains to show the following.

Exercise 7.1.4 Show that the operator

$$Z \otimes H + X \otimes H + X \otimes \tilde{H} - Z \otimes \tilde{H}$$

has largest eigenvalue $2\sqrt{2}$, with a unique associated eigenvector equal to $|\phi^+\rangle$. ■

3. Putting everything together.

We are almost done with the proof of Theorem 7.1.1. To summarize, we start with an arbitrary strategy $|\psi\rangle_{AB}$, A_0, A_1, B_0, B_1 with success probability $p^*_{\mathrm{win}} = p^*_{\mathrm{CHSH}}$ in the CHSH game. Using part 1. this strategy can be decomposed in a convex combination of qubit strategies. More formally, there are projective measurements $\Pi^A = \{\Pi^A_1, \ldots, \Pi^A_{k_A}\}$ and $\Pi^B = \{\Pi^B_1, \ldots, \Pi^N_{k_B}\}$ for Alice and Bob, made of projectors with rank at most 2 each, such that $A_x = \sum_j \Pi^A_j A_x \Pi^A_j$ and $B_y = \sum_j \Pi^B_j B_y \Pi^B_j$. The associated block decomposition can be specified by a unitary changes of basis U'_A and U'_B on Alice and Bob's systems respectively.

Using the first steps of part 2., we know that any strategy can have success probability at most p^*_{CHSH}, therefore all the qubit strategies, given by $(\Pi^A_j \otimes \Pi^B_\ell |\psi\rangle, \Pi^A_j A_x \Pi^A_j, \Pi^B_\ell B_y \Pi^B_\ell)$ for any $j \in \{1, \ldots, k_A\}$ and $\ell \in \{1, \ldots, k_B\}$, must have success probability p^*_{CHSH} (otherwise the overall strategy wouldn't achieve the optimal success probability).

By the remainder of part 2., for of these qubit strategies there exists a local change of basis U^A_j and U^B_ℓ in which it is equivalent to the canonical optimal strategy. By combining the unitaries U_A (resp. U_B), which specify the blocks, with the unitaries U^A_j (resp. U^B_ℓ), which identify a basis for each block in which $\Pi^A_j A_0 \Pi^A_j = Z$, $\Pi^A_j A_1 \Pi^A_j = X$, and similarly for Bob and $H, \tilde{H}$, we obtain the isometries claimed in the theorem: the proof is complete!

7.2 A device independent QKD protocol

In the previous section we gave mathematical justification for our intuition that the CHSH game can serve as a good test for entanglement. We're now going to see how the game can be embedded as a test in a key distribution protocol to make the protocol *device-independent*. Let's first explore more precisely what this notion of security covers — and does not cover.

7.2.1 Device-independent security

Device independence is a notion of security for cryptography that is motivated by the practical difficulty of characterizing the quantum mechanical devices, such as photon emitters or receptors, used in protocols such as the one for BB'84. The protocol calls for Alice to "prepare a qubit in the Hadamard basis", and for Bob to "measure his qubit in the $\pi/8$-rotated basis". When Alice prepares her qubit, and when Bob measures it, can they really trust their equipment to implement the prescribed task? What if Bob's measurement apparatus fails some percentage of the time: should he treat these failures as noise, or could they be adversarial (for example, the failure rate could vary depending on his basis choice or on the state of the qubit)? What if Alice's preparation device sometimes created two qubits, instead of a single one, without her noticing; could the additional qubit be intercepted by Eve and provide her with additional information, without Alice or Bob noticing? The following exercise shows that these such misbehavior of Alice and Bob's equipment can lead to serious security issues.

■ **Example 7.2.1 — Taken from (Pir+09).** Consider the purified variant of the BB'84 protocol. Suppose that Eve prepares a state ρ_{ABE} of the following form:

$$\rho_{ABE} = \sum_{x,z=0}^{1} |x,z\rangle\langle x,z|_A \otimes |x,z\rangle\langle x,z|_B \otimes |x,z\rangle\langle x,z|_E\,. \tag{7.5}$$

Now suppose Alice and Bob's measurement devices, instead of measuring a single qubit in the standard or Hadamard bases, as they think the device does, in fact performs the following:

- When the device is told to measure in the standard basis, it measures the first qubit of the two-qubit system associated with the device in (7.5) in the standard basis;
- When the device is told to measure in the Hadamard basis, it measures the second qubit of the two-qubit system associated with the device in (7.5) in the standard basis.

If the devices perform as described they perfectly pass all tests performed in the protocol: indeed, when the basis choice is the same the outcome is the same, whereas when the bases are different the outcomes are perfectly uncorrelated. But any key extracted from ρ_{ABE} in (7.5) is completely insecure! (Exercise: Give an explicit attack for Eve.) ■

The difficulty is not only theoretical. In fact, one of the first "attacks" on the BB'84 protocol is that the photon receptor used in an early experiment made a different clicking noise when it measured in one of Bob's bases, thereby "leaking" Bob's basis choice to any eavesdropper within earshot! Many such side-channel attacks have been demonstrated, and implemented, in practice. Some of the most effective are called "detector blinding" attacks, in which the eavesdropper can take complete control of Bob's receptor by shining a very bright laser right into it (without Bob noticing!).

Device-independence aims to guarantee security even in the context of such seemingly dramatic failures of Alice and Bob's equipment. But we have to be careful what we promise exactly. For example, at the extreme we could imagine that Bob's device contains radio equipment that automatically transmits all its measurement results to Eve: in this case security is compromised, but there is no way for Bob to detect the radio transmitter unless he opens the device. In a similar vein, if the random number generator used by Alice to make her basis choices is biased, or controlled by Eve, then security cannot hold. The specific kinds of failures that are allowed by a device-independent proof of security thus have to be specified on a case-by-case basis. For quantum key distribution we will make the following assumptions:

1. Alice and Bob's labs are perfectly isolated: once the protocol starts no information enters or exits their respective labs that is not specified in the protocol.
2. Alice and Bob's random number generators are perfect.
3. The devices used by Alice and Bob to perform measurements are arbitrary. These devices are initialized in a state ρ_{ABE} that may be chosen by the adversary. At each step of the protocol, each of Alice and Bob's devices makes a measurement when instructed, and always produces an outcome $x \in \{0,1\}$. The measurement that is performed is arbitrary. In particular the device may have memory and behave differently in each round.
4. At the end of the protocol the devices are discarded and will never be re-used. It is assumed that they will never fall in Eve's hands.

Device-independence refers to the freedom in assumption 3., which allows the devices to perform any kind of measurement, on any state; both may have been decided on by Eve as part of her "attack".

The last assumption is important: as will be apparent from the protocol, the devices themselves know what Alice and Bob's raw key is, and could potentially store it in memory. It is important that this memory is never allowed to leak to any adversary.

7.2.2 The protocol

The security of the DIQKD protocol we are about to give, a variant of Ekert's original proposal for quantum key distribution [Eke91], is based on the rigidity properties of the CHSH game that we explored earlier on: a high success probability in the game can be used to certify an EPR pair, even when the measurements being performed could a priori be arbitrary.

Before we proceed, there is a small technicality we have to deal with. In the honest optimal strategy for the CHSH game it is never the case that Alice and Bob use the same basis, and thus they never produce perfectly correlated outcomes. In order to produce a key it will be convenient for them to be able to rely on (almost) perfectly correlated outcomes for at least one choice of a pair of inputs. Therefore in the protocol we think of Bob's device as having 3, instead of 2, possible inputs: the inputs $\tilde{\theta} \in \{0,1\}$ correspond to the usual CHSH inputs (for which the ideal device would measure using observables H and $\tilde{H}$ respectively), and the additional input $\tilde{\theta} = 2$ instructs the device to measure in the standard basis, so that on inputs $(\theta, \tilde{\theta}) = (0,2)$ the devices are expected to produce matching outcomes (of course, in practice the device may implement any measurement it likes).

Protocol 1 Device independent QKD. Outputs $k \in \{0,1\}^\ell$ to both Alice and Bob.

1. Alice chooses a uniformly random basis string $\theta = \theta_1, \dots, \theta_n \in \{0,1\}^n$ and sequentially instructs her measurement device to measure in the bases θ. The device returns a string of outcomes $x = x_1, \dots, x_n$.
2. Bob chooses a uniformly random basis string $\tilde{\theta} = \tilde{\theta}_1, \dots, \tilde{\theta}_n \in \{0,1,2\}^n$ and sequentially instructs his measurement device to measure in the bases $\tilde{\theta}$. The device returns a string of outcomes $\tilde{x} = \tilde{x}_1, \dots, \tilde{x}_n$.
3. Alice and Bob tell each other their basis strings θ and $\tilde{\theta}$ respectively over the CAC.
4. Alice selects a random subset $T \subseteq [n]$ of size $n/2$ and announces T to Bob. They set $T' = \{j \in T, \tilde{\theta}_j \in \{0,1\}\}$, $T'' = \{j \in T, \theta_j = 0 \wedge \tilde{\theta}_j = 2\}$, and $R = \{j \notin T, \theta_j = 0 \wedge \tilde{\theta}_j = 2\}$.
5. Alice and Bob announce x_T and $\tilde{x}_T$ to each other over the CAC. They compute the success probabilities $p_{\text{win}} = |\{j \in T', x_j \oplus \tilde{x}_j = \theta_j \wedge \tilde{\theta}_j\}|/|T'|$ and $p_{\text{match}} = |\{j \in T'', x_j = \tilde{x}_j\}|/|T''|$. If $p_{\text{win}} < \cos^2 \pi/8 - \delta$ or $p_{\text{match}} < 1 - \delta$ they abort.
6. Alice and Bob perform information reconciliation and privacy amplification on their respective outcomes $x_R, \tilde{x}_R$.

As already mentioned security of the protocol is based on the rigidity of the CHSH correlations. However, as we already saw in last week's analysis, the kind of strong guarantees provided by Theorem 7.1.1 are very difficult to expand to the analysis of a full protocol, where not just one but many CHSH games are played sequentially. Luckily, these guarantees are also more than we really need: ultimately, what need to show is security of the classical key — however it is obtained at the quantum mechanical level. In fact, due to the last steps of information reconciliation and privacy amplification (which are unchanged from the BB'84 protocol) the only thing we really need to establish is uncertainty in Alice's outputs x_R, given the side information E. To show this, we use yet another variant of the guessing game, this time based on the CHSH correlations.

7.2.3 A CHSH-based guessing game

Consider the following guessing game. There are three players, Alice, Bob and Eve. Alice receives an input $\theta \in \{0,1\}$, Bob receives a $\tilde{\theta} \in \{0,1,2\}$, and Eve receives no input (equivalently, her input is always the same). The players produce outcomes $x, \tilde{x}, z \in \{0,1\}$ respectively. They win the game if and only if the following conditions hold:

- If $\tilde{\theta} \in \{0,1\}$ then $x \oplus \tilde{x} = \theta \wedge \tilde{\theta}$.
- If $\theta = 0$ and $\tilde{\theta} = 2$ then $x = z$.

Lemma 2 — CHSH guessing lemma. Consider an arbitrary strategy for the players in the CHSH guessing game. Let p_{win} be the probability that the first test passes (conditioned on $\tilde{\theta} \in \{0,1\}$) and p_{id} the probability that the second test passes (conditioned on $\theta = 0$ and $\tilde{\theta} = 2$). Suppose that $p_{\text{win}} \geq \cos^2 \pi/8 - \delta$. Then $p_{\text{id}} \leq 1/2 + 2\delta^{1/2}$.

We will not give a proof of the lemma here. There are many ways it can be shown, yielding bounds of varying quality. The simplest analysis would consider a relaxation of the problem where the three players are allowed any kind of *non-signaling strategy*: in this case a bound can be obtained via linear programming. The bound can then be strengthened by considering the fact that the players must be quantum, using a semidefinite relaxation of the problem. But the optimal bound can be obtained by a direct analytic calculation, using the fact that Alice only has two possible inputs to reduce to the two-dimensional case via an application of Jordan's lemma. This is done in [Pir+10], from which the bound given here, which is due to [VV14], can be derived.

7.3 Security of device-independent quantum key distribution

Let's analyze the security of our DIQKD protocol. Our goal is to show that there is an $\varepsilon > 0$ (the error) and a $\kappa > 0$ (the key rate), depending on the parameters of the protocol, such that the following holds:

> For any strategy of the eavesdropper Eve, specified by an initial state ρ_{ABE} of the devices and a choice of measurements to be made at every step in the protocol, either Alice and Bob abort in step 5. of the protocol with probability larger than ε, or Alice's outcomes x_R at step 6. satisfy $\mathrm{H}^{\varepsilon}_{\min}(X_R|EK) \geq \kappa n$, where K denotes all the communication exchanged on the CAC during the protocol.

A few comments regarding this statement. First, by focusing on establishing a sufficiently large rate of min-entropy in Alice's raw key bits we are putting the steps of information reconciliation and privacy amplification behind us. We studied these steps in detail in previous weeks and understand them well, but they have to be performed, and as a result the length of key produced will be slightly reduced.

Second, ε enters in the statement twice. First, we are assuming that the probability of an abort in step 5. is not too large, not larger than $1 - \varepsilon$. The reason this is needed is that conditioning on very low probability events can have drastic consequences. There is always the chance that Eve prepares states that have a very high failure probability, but such that conditioned on passing all the tests (which might still happen with low probability — for instance in the extremely unlikely event that the sets T' and T'' are both empty!) the protocol becomes completely broken. Second, ε also appears in the bound $\mathrm{H}^{\varepsilon}_{\min}(X_R|EK) \geq \kappa n$. This bound is evaluated on the joint state of Alice and Eve in step 6., conditioned on not aborting in step 5. It is unrealistic to hope to prove a bound directly on the min-entropy of that state. For instance, even though Alice and Bob did not abort there is always a small chance that Eve still attacked a large number of rounds of the protocol (by preparing a malicious entangled state) but got extremely lucky in the tests. Thus we will only be able to show that the state at step 6. is close, in trace distance, to a state whose min-entropy is large; this is the meaning of the ε in the smooth min-entropy condition $\mathrm{H}^{\varepsilon}_{\min}(X_R|EK)$.

Now that we understand precisely our target — let's prove security! There are two main steps. The first is to use the testing condition from step 5. to infer a lower bound on the conditional min-entropy $\mathrm{H}_{\min}(X_j|EK)$ in individual rounds of the protocol, for $j \in R$. The second is to combine these bounds into a bound on the whole string X_R.

We will show how both steps can be performed under the restriction that the eavesdropper is limited to so-called *collective attacks*. A collective attack is one in which the initial state of the devices takes the form $\rho_{ABE}^{\otimes n}$, and moreover the measurement performed by the device in each round is the same (on the same inputs), i.e. the device is memoryless. (The name "collective" comes from the fact that at the end of the protocol we still allow Eve to perform a joint measurement simultaneously on all the E systems, as well as all the classical information she has acquired, when making her best guess for the key.)

The most general attacks, without these two assumptions, are called *coherent attacks*. These allow Eve to introduce significant complications by choosing an initial state that is entangled across all rounds; in fact the state may not have n pre-specified qubits and the device could measure the same, or partially overlapping, high-dimensional systems in different rounds. This makes the analysis much more involved, and we will only outline an important tool that can be used to adapt our security proof against collective attacks to a full proof of security against coherent attacks.

7.3.1 Collective attacks

The assumption of collective attacks allows us to model the behavior of the device in each round as independent from its actions in previous (or subsequent) rounds. In particular, the device has a well-defined success probability in the CHSH game: if it is given inputs $\theta, \tilde{\theta} \in \{0,1\}$ in any particular round, how well does it perform in the game?

This is precisely the quantity that is estimated at step 5. of the protocol. Let $Z_1, \dots, Z_k$, where $k = |T'|$, be binary random variables such that Z_j equals 1 if the CHSH condition in round j is satisfied. Then $p_{\text{win}} = |T'|^{-1} \sum_{j \in T'} Z_j$. Note that this is an "observed" quantity; let $\hat{p}_{\text{win}}$ be the "true value", i.e. the probability of success of the device in the CHSH game. How different can p_{CHSH} and $\hat{p}_{\text{CHSH}}$ be?

We can think of the inputs for the rounds T' as being selected after the set of rounds T' itself is chosen by Alice: for instance, we could imagine Bob choosing rounds in which $\tilde{\theta}_J = 2$ at random, and Alice choosing a random set T; this defines the set T' but the players still have the freedom to choose specific inputs for those rounds. Since the probability of any given round lying in T is $1/2$, and independently the probability that Bob chooses $\tilde{\theta}_j = 2$ is $1/3$, the expected size of $|T'|$ is $n/6$. To show that the chance that the actual size differs from the expected size by too much is small we need a simple concentration inequality.

Theorem 7.3.1 — Chernoff bound (Che81). Let $X_1, \dots, X_n$ be i.i.d. random variables taking values in $\{0,1\}$, and $\mu = \mathrm{E}[X_i]$. Then for all $0 < \alpha < 1$,

$$\Pr\left(\left| \frac{1}{n} \sum_{i=1}^{n} X_i - \mu \right| > \alpha\mu \right) \le 2e^{-\frac{\alpha^2 \mu n}{3}}.$$

If we apply the proposition with $\mu = 1/6$ and $\alpha = 1/4$ we obtain that the probability that $|T'| < n/8$ is at most $e^{-n/(3\cdot 6\cdot 16)}$. Let's assume this is not the case. Then we can apply the same bound once more to obtain

$$\Pr\left(\sum_{j \in T'} Z_j > (1+\alpha)|T'|\hat{p}_{\text{win}} \right) \le 2e^{-\frac{\alpha^2 \hat{p}_{\text{win}} |T'|}{3}}.$$

Hence, using our lower bound on the size of $|T'|$ as well as $\hat{p}_{\text{win}} \ge 1/2 - \sqrt{2}/4$ (exercise: why?),

$$\Pr\left(\hat{p}_{\text{win}} < \frac{1}{1+\alpha} p_{\text{win}} \right) \le 2e^{-\frac{\alpha^2 n}{C}}$$

for some large constant C.

So far we have managed to show that, except with probability exponentially small in n, provided the protocol does not abort in step 5. of the protocol it must be the case that $\hat{p}_{\text{win}} \ge p_{\text{win}}/(1+\alpha) \ge \cos^2 \pi/8 - 2\delta$ (if we choose $\alpha = \delta$). Now is time to apply the CHSH guessing lemma, Lemma 2.

The condition $p_{\mathrm{id}} \leq 1/2 + 2(2\delta)^{1/2}$ that results gives a direct bound on the guessing probability of the device,

$$\mathrm{H}_{\min}(X_j|E) \geq -\log_2\left(\frac{1}{2} + 2(2\delta)^{1/2}\right) \geq 1 - C\sqrt{\delta} \tag{7.6}$$

for some small constant C.

Using the assumption that the device behaves identically and independently in each round of the protocol, the bound (7.6) does not only apply in the tested rounds $j \in T'$, but also in the rounds $j \in R$ used for the raw key. Thus as a final step we use (7.6) for $j \in R$, together with the fact that the devices are in tensor product form, to add up the entropies and conclude that $\mathrm{H}_{\min}(X_R|E) \geq |R|(1 - C\sqrt{\delta})$ — exactly what we set out to show!

A final subtlety is that this bound on the min-entropy holds under the conditions $|T'| \geq n/8$ and $\hat{p}_{\mathrm{win}} \geq p_{\mathrm{win}}(1-\delta)$. As we showed, conditioned on not aborting both conditions hold except with probability ε that is exponentially small. Taking this into account we obtain a lower bound on the smooth min-entropy of the raw key, $\mathrm{H}^{\varepsilon}_{\min}(X_R|E) \geq |R|(1 - C\sqrt{\delta})$. This bound is sufficient for privacy amplification to apply (the smoothing parameter ε will simply have to be added to the error of the extractor used for privacy amplification). Thus, provided that privacy amplification and information reconciliation are implemented correctly, Alice and Bob can generate a secure key.

7.3.2 Coherent attacks

The two-step approach we followed to analyze security against collective attacks no longer works against coherent attacks. First, since the devices may now have memory we cannot directly infer properties of the devices in the rounds used for the raw key from its behavior in the testing rounds. Second, since the global state prepared by Eve is no longer assumed to have a tensor product form we can no longer claim that the min-entropy adds up across rounds.

The first difficulty can be handled by using a variant of the concentration bound in Theorem 7.3.1 that applies to processes which may have memory, but still have a sequential nature and satisfy certain regularity properties. Such bounds are called martingale inequalities; one of the most useful is due to Azuma. By applying that inequality it is possible to obtain a similar bound as in (7.6) on the min-entropy per round for the raw key rounds from success of the CHSH test in the test rounds.

The second difficulty is more thorny. Given a lower bound $\mathrm{H}_{\min}(X_j|E) \geq h$ for some $h > 0$ for all $j \in R$, can we conclude a meaningful lower bound on $\mathrm{H}_{\min}(X_R|E)$? Unfortunately in general the answer is no: the quantum conditional min-entropy (in contrast to conditional von Neumann entropy) doesn't satisfy a nice form of the chain rule. To make progress we again need to use the sequential nature of our process. At this point there are different approaches to finishing the proof, and we mention just one, based on a technical result called "entropy accumulation theorem" (EAT) [DFR16]. The EAT gives conditions under which min-entropy "accumulates", and these conditions are satisfied by our setup. (The most important conditions are that the outputs are generated sequentially in each round, and are only a function of the state of the devices in that round; moreover the test, when it is performed, should be a deterministic function of the inputs and outputs in the round.)

Once it applies, the EAT is rather powerful, and it provides essentially the same consequences are we were able to derive in the case of collective attacks (except with a small loss in the parameter ε). Let's state the final result as a theorem.

Theorem 7.3.2 The DIQKD protocol, Protocol 1, satisfies the following properties. There is a $0 < \kappa \leq 1$ and $C \geq 1$ (depending on the tolerance parameter δ) such that the following hold for $\ell = \kappa n$ and $\varepsilon \leq 2^{-Cn}$.

First, there is an implementation of the devices such that the protocol does not abort with probability at least $1-\varepsilon$.

Second, for any implementation of the devices, either the protocol aborts with probability larger than $1-\varepsilon$, or conditioned on not aborting Alice and Bob each produces a key of length ℓ such that $\Pr(K_A \neq K_B) \leq \varepsilon$ and

$$(1-\Pr(abort))D(\rho_{K_AE}, U_\ell \otimes \rho_E) \leq \varepsilon,$$

where E denotes all the side information available to the eavesdropper at the end of the protocol.

In our analysis we considered the min-entropy per round, and argued that it could be added up to obtain a bound on the min-entropy of the string x_R corresponding to Alice's raw key. A stronger bound can be obtained by using the fact that, when considering a large number of samples of a random variable X, the min-entropy converges to the von-Neumann entropy:

$$\frac{1}{n}\mathrm{H}^{\varepsilon}_{\min}(X_1\cdots X_n) \approx_{n\to\infty} H(X)$$

for i.i.d. X, provided the smoothing parameter ε is chosen sufficiently large. This is called the "asymptotic equipartition property". Using this property it is possible to show that a lower bound on the von Neumann entropy in each round is enough to conclude a lower bound on the min-entropy of the whole string. Since the von Neumann entropy can in general be larger than the min-entropy this leads to better bounds on the key rate.

Acknowledgments

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International Licence. The lecture notes are written by Nelly Ng, Thomas Vidick and Stephanie Wehner. We thank David Elkouss, Kenneth Goodenough, Jonas Helsen, Jérémy Ribeiro, and Jalex Stark for proofreading.

Bibliography

[Bha13] Rajendra Bhatia. *Matrix analysis*. Volume 169. Springer Science & Business Media, 2013 (cited on page 6).

[Che81] Herman Chernoff. “A note on an inequality involving the normal distribution”. In: *The Annals of Probability* (1981), pages 533–535 (cited on page 12).

[DFR16] Frederic Dupuis, Omar Fawzi, and Renato Renner. “Entropy accumulation”. In: *arXiv preprint arXiv:1607.01796* (2016) (cited on page 13).

[Eke91] Artur K Ekert. “Quantum cryptography based on Bell’s theorem”. In: *Physical review letters* 67.6 (1991), page 661 (cited on pages 3, 10).

[Pir+09] Stefano Pironio et al. “Device-independent quantum key distribution secure against collective attacks”. In: *New Journal of Physics* 11.4 (2009), page 045021 (cited on page 9).

[Pir+10] Stefano Pironio et al. “Random numbers certified by Bell’s theorem”. In: *Nature* 464.7291 (2010), pages 1021–1024 (cited on page 11).

[VV14] Umesh Vazirani and Thomas Vidick. “Fully device-independent quantum key distribution”. In: *Physical review letters* 113.14 (2014), page 140501 (cited on page 11).