Quantum Cryptography – Short Notes

Sofia Qvarfort

March 31, 2016

Contents

1 Introduction

Condition for information-theoretic security achieved if

$$P(y|x) = P(y) \tag{1}$$

where x is the message and y is the codeword.

Interpretation: A probability distribution for y conditioned on x must only contain information about y and no information about x.

Shannon's Theorem states that for a secret key $\mathcal{K}$, a plaintext $\mathcal{X}$ and the ciphertext $\mathcal{Y}$, it follows that

$$|\mathcal{K}| \geq |\mathcal{Y}| \geq |\mathcal{X}| \tag{2}$$

Interpretation: To have perfect security, we must share a key as long as the message.

Consequences: Very inconvenient to have perfect security by this method since the key must be very long.

Public-key cryptography uses one public key and one private key. The public key is distributed and used for encryption, while the private key is secret and is used for decoding only.

RSA is a form of public-key cryptography. The private key is two integers (a, b) and the public key is their product $a \cdot b = c$. It is conjectured to be computationally hard to obtain a and b, hence RSA is secure.

However, it relies on two assumptions:

- Computational power limitations of the adversary
- Unproven mathematical conjectures

2 Quantum Key Distribution

Idea to have secure transfer of a secret key that can then be used for communication.

Alice randomly prepares photons in either the Z or the X basis with probability 1/2. Bob chooses a measurement basis with probability 1/2 and measures the particle. Thus, about 1/2 of the total photons will be used for the key.

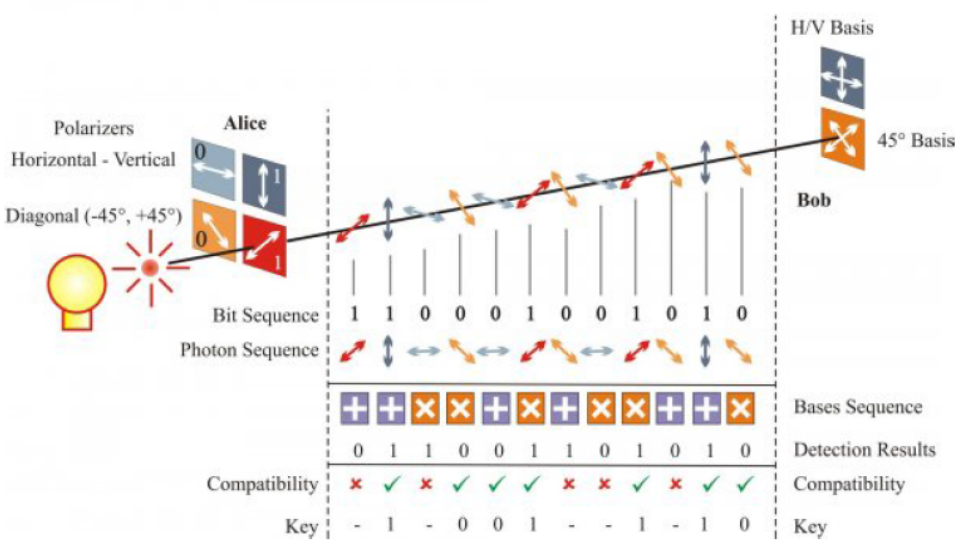

Figure 1: Quantum Key Distribution setup.

Vertical polarisation corresponds to the states $|0\rangle$ and $|1\rangle$.

Horizontal polarisation corresponds to the states $|+\rangle$ and $|-\rangle$.

Knowledge of adversary Eve similarly chooses a measurement basis with probability 1/2. Thus, with probability 1/4 she learns something about the key. This is not very good.

Privacy amplification can be used to improve the security of the resulting key. It transforms the N-bit string into an $(N - S)$-bit string, whose distance from the ideal, completely secure string is e^{-S}.

2.1 State discrimination

Here we derive the error of discriminating between two states ρ_0 and ρ_1. We are given each state with probability p_0 or p_1 such that $p_0 + p_1 = 1$.

Consider choosing a POVM with elements A_0 and A_1, where $A_0 + A_1 = 1$ that would optimally distinguish between these two states. The probability of successfully distinguishing the state is

$$p_{success} = p_0 P(A_0|\rho_0) + p_1 P(A_1|\rho_1) \tag{3}$$

However, the probability of failure (that is, using the wrong POVM element) is

$$\begin{aligned} p_{error} &= p_0 P(A_1|\rho_0) + p_1 P(A_0|\rho_1) && (4)\\ &= p_0 \mathrm{Tr}\left[A_1\rho_0\right] + p_1 \mathrm{Tr}\left[A_0\rho_1\right] && (5)\\ &= p_0 \mathrm{Tr}\left[A_1\rho_0\right] + p_1 \mathrm{Tr}\left[\mathbb{I} - A_1\rho_1\right] && (6)\\ &= p_1 + \mathrm{Tr}\left[A_1 M\right] && (7) \end{aligned}$$

where $M = p_0\rho_0 - p_1\rho_1$. Let

$$M = \sum_k \lambda_k \left|\phi_k\right\rangle \left\langle\phi_k\right| \tag{8}$$

The A_1 that minimises p_{error} is that which projects onto the negative eigenvalues of M. So then,

$$p_{error} = p_1 + \sum_{k:\lambda_k<0} \lambda_k \tag{9}$$

We now want to replace the expression for the negative eigenvalues with something better. We know that

$$\|M\|_1 = \sum_k |\lambda_k| = \sum_{k:\lambda_k\geq 0} \lambda_k - \sum_{k:\lambda_k<0} \lambda_k \tag{10}$$

We also know that

$$\mathrm{Tr}\left[M\right] = p_0\mathrm{Tr}\left[\rho_0\right] - p_1\mathrm{Tr}\left[\rho_1\right] = p_0 - p_1 = \sum_k \lambda_k \tag{11}$$

Thus, we find

$$p_0 -_1 -\|M\|_1 = 2\sum_{k:\lambda_k<0} \lambda_k \tag{12}$$

And finally, we can write

$$p_{error} = \frac{1}{2} - \frac{1}{2}\|M\|_1 = \frac{1}{2} - \frac{1}{2}\|p_0\rho_0 - p_1\rho_1\|_1 \tag{13}$$

This is the final failure probability. Note that orthogonal states have zero failure probability, because their norm (their distance) is maximised.

2.2 Optimal individual attacks

Individual attack means that Eve interacts with and measures each photon separately.

Coherent attack another name of individual attack.

Result comparison Alice and Bob choose a subset of their results with matching basis and compare them publicly. Any discrepancy is assumed to be due to an adversary listening in on the channel.

Disturbance usually denoted D is the disturbance caused by Eve and detected by Alice and Bob.

Symmetries Eve will use symmetries in her attack that match those in Alice's and Bob's protocol. That is, it is symmetric under $0 \leftrightarrow 1$ and $Z \leftrightarrow X$.

Question: I am not entirely sure what the symmetries mean. The $Z - X$ symmetry means that the disturbance should the same regardless of the axis that Eve chooses.

The optimal attack is a symmetric attack where Eve lets her quantum system interact with the states that Alice and Bob send through the channel. Eve will wait until Alice and Bob publish the measurement bases that they used, so that we then will have to distinguish the states that she has. For example, if Alice prepares in the Z basis, Eve will have to distinguish the states

$$\rho_0 = F\left|E_{00}\right\rangle\left\langle E_{00}\right| + D\left|E_{01}\right\rangle\left\langle E_{01}\right| \tag{14}$$
$$\rho_1 = F\left|E_{11}\right\rangle\left\langle E_{11}\right| + D\left|E_{10}\right\rangle\left\langle E_{10}\right| \tag{15}$$

These are mixed states, because Eve does not know whether she has E_{00} or E_{01}. Thus, the best situation arises when

$$\langle E_{00}|E_{10}\rangle = \langle E_{01}|E_{11}\rangle = 0 \tag{16}$$

This will however not make the probability of failure zero.

I guess that this optimal attack requires Eve to have a pretty good quantum memory.

Failure probability is the average of the F event and the D event.

$$p_{error} = \frac{1}{2}\left(1 - F\sqrt{1-\alpha^2} - D\sqrt{1-\beta^2}\right) \tag{17}$$

Using Lagrange multipliers, we find

$$\alpha = \beta = 1 - 2D \tag{18}$$

as the optimal solutions, which gives

$$p_{error} = \frac{1}{2} - \sqrt{D(1-D)} \tag{19}$$

Note that $p_{error} = 0$ for $D = 1/2$.

Strategy $D = 1/2$ also causes the key that Alice and Bob create to be completely useless. They will know that Eve is listening to everything. Thus, the best strategy for Eve is to pretend to be noise. Alice and Bob can never distinguish between actual noise and the interferences caused by Eve.

3 Shannon Information Theory

Relative frequency the number of times a letter x in a sequence with N letters from the alphabet $1 \dots d$ appears is $N(i)$. The relative frequency is therefore

$$\nu(x) = \frac{N(x)}{N} \tag{20}$$

That is, it answers the question: How many times does this letter appear in a sequence of N letters?

It is bounded from above by

$$|\{\nu(x)\}| \leq (N+1)^d \tag{21}$$

for an alphabet of size d.

Question: I understand how N^d would be the highest possible frequency if I had say $xxxx \dots x$ N times. But why $N + 1$?

Answer: We count from 0.

Number of sequences called Ω with relative frequency $\nu(x)$ is

$$\Omega(\nu(x)) = \binom{N}{N(1) \dots N(d)} = \frac{N!}{\prod_x N(x)!} \tag{22}$$

which by Stirling's approximation,

$$n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \tag{23}$$

becomes

$$\Omega(\nu(x)) = \sqrt{\frac{2\pi N}{\prod_x 2\pi N\nu(x)}} 2^{NH(\nu)} \tag{24}$$

Shannon entropy defined as

$$H(\nu) = H(X) = -\sum_x \nu(x)\log\nu(x) \tag{25}$$

Question: Why are we considering the relative frequency here, instead of the probability distribution? After all, isn't frequency dimension-full? Well, not in this case because we have defined everything in terms of numbers.

Probability of sampling the probability that a sequence with frequency $\nu(x)$ is sampled from a source with distribution $q(x)$ is

$$P(\nu(x)|q(x)) = \Omega(\nu(x))\prod_x q(x)^{N\nu(x)} = 2^{-ND(\nu|z)} \tag{26}$$

Interpretation: We multiply the number of possible sequences by the product of all probabilities to the power of $N\nu(x)$. Recall that

$$N\nu(x) = N(x) \tag{27}$$

So the combined probability to the power of the number of letters that we get?

Question: This is not clear to me.

Relative entropy defined as

$$D(\nu|q) = \sum_x \nu(x)\log\frac{\nu(x)}{q(x)} \tag{28}$$

Infinitesimal variations if there is a small difference between the entropies such that

$$\nu(x) = q(x) + \delta(x) \tag{29}$$

we can show that the conditional probability distribution is Gaussian,

$$P(q+\delta|q) = 2^{-\frac{N}{2}\sum_x \frac{\delta(x)^2}{q(x)}} \tag{30}$$

So if we sampled $\nu(x)$, it would be distributed by average $1/\sqrt{N}$ around $q(x)$.

Typical sequences These are the sequences most likely to be chosen from a random distribution.

Probability of typical sequence is $2^{-NH(\nu)}$.

3.1 Randomness distillation

Randomness distillation can also be thought of as compression. We compress all the information that is not random. We cannot compress a truly random distribution – it has maximum entropy.

i.i.d means independent and identically distributed. It is given by

$$P(x_1, x_2 \dots x_N) = P(x_1)P(x_2)\dots P(x_N) \tag{31}$$

Number of typical sequences is $2^{NH(X)}$. If they all occur with the same probability, this probability is $P(X) = e^{-NH(X)}$.

Compression we can index every typical sequence by one of the sequences of an M-bit string. Then

$$2^M = 2^{NH(X)} \tag{32}$$

So we only need $M = NH(X)$ bits.

Entropy interpretation is: The number of bits of memory required to store each letter X is $H(X)$. It is also the number of bits of perfect randomness distillable from each letter X.

3.2 Privacy Amplification

Key idea Consider the key created in the QKD protocol. Eve might have gained some information about the key, but we want to post-process the key so that Alice and Bob ensure that the key is safe.

Information about initial data that Eve has can be written

$$P(x_1, z_1 \dots x_N, z_N) = P(x_1, z_1) \dots P(x_N, z_N) \tag{33}$$

That is, each z_i will tell us a little about Alice's x_i.

Conditional entropy derivation.

How much does Eve know about $x_1, x_2 \dots x_N$ if she has observed the sequence $\bar{z}_1, \bar{z}_2 \dots \bar{z}_N$? How many sequences $(x_1, \dots x_N)$ are compatible with $(\bar{z}_1 \dots \bar{z}_N$?

Question: exactly what do we mean by 'compatible'?

The number of 1s in $\bar{z}_1 \dots \bar{z}_N$ is roughly $NP(Z = 1)$.

The number of zs in $\bar{z}_1 \dots \bar{z}_N$ is roughly $NP(Z = z)$.

Consider now only the subsequences, chosen from $(x_1, z_1) \dots (x_N, z_N)$. Take only the sequences where $z_i = 1$ and combine them. They have length $NP(Z = 1)$. The probability distribution for x is

$$P(x|Z = 1) \tag{34}$$

Then, with high probability, the number of subspaces with $Z_i = 1$ is

$$2^{NP(Z=1)H(X|Z=1)} \tag{35}$$

and the same goes for any other subspace with $Z_i = z$. Then the total number of the full sequence is the product of the above. We are multiplying together all the subsequences for certain Z.

$$\prod_x 2^{NP(Z=z)H(Z|Z=z)} = 2^{NH(X|Z)} \tag{36}$$

Conditional entropy is defined as

$$H(X|Z) = \sum_z P(Z = z)H(x|Z = z) \tag{37}$$

This describes the ignorance that Eve has about X, given that she has information about Z.

Recall that Z is the entire distribution of z, it is the set of possible values for z.

Note that we say that Eve's sequence is different from the ideal sequence because of real world Gaussian sampling etc. (I think this is the reason).

Compression to restrict information Let Alice compress her information $(x_1, \dots x_N) \to (k_1 \dots k_M)$ so that Eve's distribution $(\bar{z}_1 \dots \bar{z}_N$ using some function f.

Let f be uniform. Then, the probability of it mapping two specific sequences $(x_1 \dots x_N)$ into the same $(k_1 \dots k_M)$ is 2^{-M}. That is because there are 2^M typical sequences. So, given that I have already mapped one sequence to the sequence $k_1 \dots k_M$, the probability that I choose another one to map to the same sequence is 2^{-M}.

Then, since Eve has $2^{NH(X|Z)}$ strings that are compatible with Alice's string, the probability that two strings have the same compression image is

$$2^{NH(X|Z)-M} \tag{38}$$

I want this probability to go to zero, because then the number of compatible strings that Eve has goes to zero.

Question: I think that the above is the number of compatible strings, is that right?

If we choose

$$M = NH(X|Z) - \sqrt{N} \tag{39}$$

the rate goes to zero as $N \to \infty$.

Summary of privacy amplification We only need a function that uniformly randomly compresses our information into certain bit strings.

3.3 Error correction

Key idea Let us say that Alice and Bob are trying to create a secret key. But let's also say that there are some errors, so that knowing $(y_1 \dots y_N)$ does not give us perfect information about $(x_1 \dots x_N)$. So, there are a number of $2^{NH(X|Y)}$ compatible sequences.

Let Bob ignore which of these sequences Alice has. Question: Why?

Let Alice then send Bob some partial information about her string. It is obtained from her string through some function g such that $g(s_1 \dots x_N) = (c_1 \dots c_L)$, where we choose

$$L = NH(X|Y) + \sqrt{N} \tag{40}$$

Then, what is the probability that two of the $2^{NH(X|Y)}$ sequences is mapped to the same $(c_1 \dots c_L)$? Basically, if two separate sequences were mapped to the same one, it would not be possible for Bob to find out which sequence Alice has – there wouldn't be a one-to-one relationship.

The probability for two strings being mapped onto the same C is

$$2^{NH(X|Y)-L} = 2^{-\sqrt{N}} \tag{41}$$

which tends to zero as $N \to \infty$.

Question: Is this basically Schumacher compression?

Question: How would this correct for errors?

Answer: Say that the errors caused some misinformation on Bob's side. The additional information will complement it so that the key is perfectly shared.

3.4 Post-processing in secret key distribution

Key idea : We now combine error correction and privacy amplification. The goal is for Alice and Bob to end up with the same string $(x_1 \dots x_N)$ with high probability.

Error correction Alice wants to correct Bob's errors. She sends

$$L = NH(X|Y) + \sqrt{N} \tag{42}$$

bits of information $(c_1 \dots c_L)$ to Bob.

This ensures that Bob also has string $(x_1 \dots x_N)$.

Eve also gains this information. She now knows that the correct string is among $2^{NH(X|Z)-L}$ strings that she may consider. That is, performing error correction publicly means that Eve can discard 2^{-L} of her strings... I think.

Privacy amplification Both Alice and Bob compress their information using f into $(k_1 \dots k_M)$. We now want to remove the information that Eve could have gained form the privacy amplification as well.

So, we must choose to compress to

$$M = NH(X|Z) - L - \sqrt{N} = -2\sqrt{N} \tag{43}$$

bits. Doing this ensures that the key is secure.

Final asymptotic efficiency rate is given by

$$H(X|E) - H(X|Y) \tag{44}$$

This s the number of perfect secret bits per raw bit sent sent in QKD.

3.5 Secret key rate of BB84

This is a long section that I don't think we went through.

The secret key rate is

$$r = H(X|Z,T) - H(X|Y) \tag{45}$$

where Z and T belong to Eve and are part of the post-processing mechanics.

4 More Quantum Key Distribution

The optimal cloning machine corresponds to the optimal individual attack on a protocol that uses the X, Y and Z bases. It copies $|\psi\rangle|0\rangle|0\rangle \to |\psi\rangle|\psi\rangle|\psi\rangle$ with 5/6 fidelity.

State purification Given a mixed state

$$\rho_A = \sum_j \lambda_j |a_j\rangle\langle a_j| \tag{46}$$

a purification is a bipartite state of the form

$$|\psi\rangle_{AE} = \sum_j \sqrt{\lambda_j}\,|a_j\rangle_A \otimes |b_j\rangle_E \tag{47}$$

where $|b\rangle_j$ is some orthonormal basis on E. These are unique up to a unitary. Then,

$$\rho_A = \mathrm{Tr}_E\left[|\psi\rangle_{AE}\langle\psi|_{AE}\right] \tag{48}$$

4.1 Entanglement based QKD

Key idea Instead of Alice sending a state to Bob, there is a source between Alice and Bob that sends them each a subsystem of an entangled pair.

Equivalence to BB84 Because we can view this as steering through measurement, since Alice effectively prepares the state by measuring it 'first', it is equivalent to he protocol we considered above.

Differences Since Alice and Bob always obtain the same density matrix ρ_{AB} from the source, they can perform state tomography until they learn what the state is.

Now, instead of measuring in a random basis, once state tomography has been performed, Alice and Bob measure in a basis that minimises the correlations with Eve.

Eve has the purification , that is, Eve has the state ρ_E such that global state is $|\psi\rangle_{ABE}$. This is the best attack that Eve can perform.

Eve's conditional states Eve's state conditioned on Alice's classical information is

$$\rho_{E|x} = \langle x| \rho_{AE} |x\rangle \tag{49}$$

Then, Eve performs state discrimination to learn x.

Joint distribution which determines which values Alice and Bob obtain is

$$P(x, y) = \langle xy| \rho_{AB} |xy\rangle \tag{50}$$

for some classical value for x and y.

Rate for entanglement based QKD is

$$r = H(X|E) - H(X|B) \tag{51}$$

4.2 Collective Attacks

Key idea The global attack is the most general attack, provided that $\rho_{AB}^{\otimes N}$ is i.i.d. The final key generally depends on correlations between bits in the raw key. If Eve only measures single bits, this information gets diluted. Hence, the global attack is the best attack.

Global state instead of considering every single ρ_{AB} state, we consider all the states sent to Alice and Bob, $\rho_{AB}^{\otimes N}$. This scheme relies on that the same state ρ_{AB} is sent every time.

Global purification if the system is i.i.d. we can choose the global purification as $|\psi\rangle_{ABE}^{\otimes N}$.

Rate bound given by

$$r \geq I(X:Y) - I(X:E) \tag{52}$$

Classical-Quantum State that Alice and Eve share (and that exists after Alice has performed a measurement) is given by

$$\rho_{AE} = \sum_x P(x) \left|x\right\rangle \left\langle x\right| \otimes \rho_{E|x} \tag{53}$$

That is, Eve is holding a quantum state conditioned on the classical outcome of Alice's measurement x.

Mutual information of global attack is given by

$$I(X:E) = S(\rho_E) - \sum_x P(x) S(\rho_{E|x}) \tag{54}$$

4.3 General attacks and the de-Finetti Theorem

Assumption for general attacks The only assumption we make is that Alice and Bob observe statistics compatible with the global state $\rho_{ABE}^{(N)}$. This can be a completely arbitrary state, as can the measurements that Alice and Bob perform.

Notation for global states Note that the state $\rho_{AB}^{\otimes N}$ is different from $\rho_{AB}^{(N)}$. The first state is a collection of identical states, whereas the second state exists in a Hilbert spaces as large as $\rho_{AB}^{\otimes N}$ but it is not necessarily i.i.d. distributed over it.

Protocol symmetries every of the N states treated in the protocol is treated on an equal footing.

The effect of averaging If Alice and Bob take the states and perform a random unitary, and then forget the outcome, it doesn't change the protocol. Applying a random unitary and forgetting the outcome essentially means that we are mixing the state. Thus,

we can treat every state in the protocol as the maximally mixed state.

Incidentally, this means that the state $\rho_{AB}^{(N)}$ is symmetric under the exchange of pairs (that is, we could swap any of the ρ_{AB}^{i} states in the protocol and still be fine).

The de-Finetti Theorem (simplified) if $\rho^{(N)}$ is a symmetric state of N systems and $\rho^{(M)}$ is the reduced state for $M < N$ of the N systems, then

$$\| \int \mathrm{d}\sigma P(\sigma)\sigma^{\otimes M} - \rho^{(M)}\|_1 \leq 2^{-(N-M)} \tag{55}$$

where $P(\sigma)$ is a probability distribution over single-system density matrices.

Interpretation: After discarding a small fraction of the N states, the remaining state is approximately i.i.d. It massively simplifies our analysis, because Alice and Bob could simply randomly throw away states until they are sure to have a key created by an i.i.d. setting. Then, they can gain the general rates showed above.

5 Post-Shannon Information Theory

Quantum memories can be used by the adversary to store her quantum system until she is ready to measure it. The advantage is that Alice and Bob might in the future use their secret key in ways that will reveal information about it.

State before Eve measures is given by

$$\rho^{real} = \sum_{k,k',s} P(k,k',x) \left|k\right\rangle \left\langle k\right|_A \otimes \left|k'\right\rangle \left\langle k'\right|_B \otimes \left|s\right\rangle \left\langle s\right|_E \otimes \rho_{E|k,k',s} \tag{56}$$

Here, k and k' are the two keys, and $|s\rangle$ is any public message that Alice and Bob sent.

The ideal key holds no entanglement with Eve. That is,

$$\rho^{ideal} = \left(\sum_k \frac{1}{|\mathcal{K}|} \left|k\right\rangle \left\langle k\right|_A \otimes \left|k\right\rangle \left\langle k\right|_B \right) \otimes \left(\sum_s P(s) \left|s\right\rangle \left\langle s\right|_E \otimes \rho_{E|s} \right) \tag{57}$$

Highest security conditions can be written

$$\|\rho^{real} - \rho^{ideal}\|_1 \leq \epsilon \tag{58}$$

Renner-Koenig Theorem states that

$$\|\rho^{real} - \rho^{ideal}\|_1 \leq \sqrt{2}^{S-N[I(X:Y)-I(X:E)]} \tag{59}$$

Consequences: It is indeed possible to come very close to the ideal state. It is really only limited by S, so if Alice and Bob keep their public communication to a minimum, it is possible to create a secret key. This means creating a smaller secret key.

Question: Do you run into a trade-off with Shannon's theorem here? Shannon's Theorem states that the key must be as long as the message for good security, but if we make a short key and want to use it for a long message, then we risk being detected.

I guess this tells us that we can't both have the cake and eat it - the more you communicate in public, the more likely it is that you will have the key compromised.

6 Non-Local Correlations

6.1 Classical, Quantum and Beyond

No-signalling condition for a probability distribution $P(a, b|x, y)$

$$P(b|y) = \sum_a P(a, b|x, y) = \sum_a P(a, b|x', y) \tag{60}$$

and

$$P(a|x) = \sum_b (a, b|x, y) = \sum_b P(a, b|x, y') \tag{61}$$

for all x, x', b. That is, since the distributions are conditioned on x, y the marginals must be independent on the output value of x and y. That is, a or b should depend only on their partner input not on any output that happens far away.

Put differently, we require

$$\sum_a P(a, b|x, y) = P(b|y) \tag{62}$$

Tracing out the input a also ensures that the outcome x cannot influence the probability distribution.

Set of non-signalling correlations is convex we can show that if $P(a,b|x,y)$ is non-signalling, then so is

$$qP_1(a,b|x,y) + (1-q)P_2(a,b|x,y) \tag{63}$$

Question: Should we not make sure that the variables a, b, x, y are different?

Answer: If we show different variables, then the comparison is useless.

Non-signalling finite points are finite. That is, the polytope that forms the non-signalling set has a finite number of extreme points.

Polytope $\mathcal{P}$ can be defined in terms of generators $\{e_1 \dots e_n\}$ or in terms of linear inequalities, $\{\vec{c}_1 \dots \vec{c}_m\}$, such that

$$\mathcal{P} = \text{conv}\{\vec{\rceil}_\infty \dots \vec{\rceil}_\backslash\} = \left\{ \sum_\rangle \surd^\rangle \vec{\rceil}_\rangle : \surd^\rangle \geq \prime, \sum_\rangle \surd^\rangle = \infty \right\} \tag{64}$$

That is, we can use the generators, the extremal points and add them in various ways using the set $\{p_i\}$, so that they form any point in $\mathcal{P}$. $\mathcal{P}$ is then the set of all points (I think).

For inequalities, we write

$$\mathcal{P} = \{\vec{\S} : \vec{\rfloor}_\infty \cdot \vec{\S} \leq \infty, \dots \vec{\rfloor}_\Updownarrow \cdot \vec{\S} \leq \infty\} \tag{65}$$

That is, we describe the bounds of the polytope (its edges) in terms of the inequalities.

Note that there are not as many generators as inequalities.

Local correlations can be written

$$P(a,b|x,y) = \sum_\lambda P(\lambda)P(a|x,\lambda)P(b|y,\lambda) \tag{66}$$

That is, along with one 'cause', x, λ also determines the value of a or b. λ is an example of a hidden variable.

Generators for local correlations the set of local correlations is generated by the extremal points,

$$P_{fg}(a,b|x,y) = \delta^a_{f(x)}\delta^b_{g(y)} \tag{67}$$

where the functions f and g map input to output. That is, $f : \mathcal{X} \to \mathcal{A}$ and $g : \mathcal{Y} \to \mathcal{B}$. Choosing a function will denote which point (or boundary) that we are dealing with.

The defining feature of an extreme point is that it cannot be written as a mixture of other points. It has to be one single element of weight 1.

Local extremal points There are 16 local extremal points (generators) for the bipartite case. They are given by

$$P(a, b|x, y) = \delta^a_{f(x)}\delta^b_{g(y)} \tag{68}$$

for all pairs of binary functions $f, g : \{0, 1\} \to \{0, 1\}$. These functions can be constant, even or odd.

Decomposing conditional distributions we find that

$$P(a|x, \lambda) = \sum_f P(f|\lambda)\delta^1_{f(x)} \tag{69}$$

That is, the conditional probability $P(f|\lambda)$ acts as a weight for the extremal points, specified entirely by $\delta^a_{f(x)}$. Note that everything will always be conditioned on λ, as this is our local variable.

Question: Should $P(f|\lambda)$ here be $P(f(x)|\lambda$?

Local distribution in terms of generators we find

$$\begin{aligned} P(a, b|x, y) &= \sum_\lambda P(\lambda)P(a|x, \lambda)P(b|y, \lambda) \text{ using Baye's Theorem} \\ &= \sum_{\lambda,f,g} P(\lambda)P(f|\lambda)\delta^a_{f(x)}P(g|\lambda)\delta^b_{g(y)} \text{ using generators} \\ &= \sum_{f,g} P(f, g)\delta^a_{f(x)}\delta^b_{g(y)} \text{ defining new } \bar{\lambda} = (f, g) \end{aligned} \tag{70}$$

Question: I could not prove that indeed

$$P(f, g) = \sum_\lambda P(\lambda)P(f|\lambda)P(g|\lambda) \tag{71}$$

Quantum correlations definition: There exists a state ρ_{AB} and measurement operators $A^a_x B^b_y$ such that

$$P(a, b|x, y) = \mathrm{Tr}\left[A^a_x \otimes B^b_y \rho_{AB}\right] \tag{72}$$

Set of quantum correlations has an infinite number of extremal points. The only fully known and characterised set of quantum points arises for the bipartite case.

Correlation function defined by

$$C_{x,y} = \sum_{a,b} (-1)^{a+b} P(a,b|x,y) \tag{73}$$

Binary correlation function gives

$$C_{x,y} = P(a=b|x,y) - P(a \neq b|x,y) \tag{74}$$

It takes the values $C_{x,y} \in [-1,1]$.

Non-signalling condition for correlation functions given by

$$P(a,b|x,y) = \begin{cases} \frac{1+C_{xy}}{4} \text{ if } a=b \\ \frac{1-C_{xy}}{4} \text{ if } a \neq b \end{cases} \tag{75}$$

which has uniform marginals for Alice and Bob

$$P(a|x) = P(b|y) = \frac{1}{2} \tag{76}$$

for all values of a, b, x, y.

How to check non-signalling Make sure that the marginals only depend on one single output value!

CHSH inequalities given by the 8 expressions:

$$-2 \leq C_{00} \pm C_{01} \pm C_{10} \pm C_{11} \leq 2 \tag{77}$$

where we shift one single minus sign around.

T'sirelson's bound given by

$$-2\sqrt{2} \leq C_{00} \pm C_{01} \pm C_{10} \pm C_{11} \leq 2\sqrt{2} \tag{78}$$

If this condition is fulfilled, the correlations are quantum and not just non-local.

PR-boxes generate maximally non-local correlations. They obey the distribution

$$P_{PR}(a,b|x,y) = \begin{cases} \frac{1}{2} \text{ if } a \oplus b = xy \\ 0 \text{ otherwise} \end{cases} \tag{79}$$

There are 8 extremal points for the PR boxes. They form tetrahedrons on the local polytope.

6.2 Monogamy of non-local correlations

Proof by contradiction Start with a 3-party distribution $P(a, b, e|x, y, z)$ such that its marginals are the maximally non-local PR-box

$$\sum_e P(a, b, e|x, y, z) = P_{RP}(a, b|x, y) \tag{80}$$

That is, Alice and Bob are holding a PR box between them. The question is: Is Eve also part of the non-local correlations?

Using Bayes rule, write

$$\sum_e P(e|z)P(a, b|x, y, e, z) = P_{PR}(a, b|x, y) \tag{81}$$

The fact that

$$P(a, b, e|x, y, z) = P(e|z)P(a, b|x, y, e, z) \tag{82}$$

by Bayes Rule should be understood from the fact that e is conditioned on z always, and thus any marginal $P(e)$ has to be $P(e|z)$. Alternatively, the situation would be the same if we ignored the entire conditioning.

However, the above is a contradiction. Since $P_{PR}(a, b|x, y)$ is an extremal points, it cannot be made up of several distributions. Thus, we find

$$P(a, b|x, y, e, z) = P(a, b|x, y) \tag{83}$$

for all values of e and z. This implies that

$$P(a, b, e|x, y, z) = P(a, b|x, y)P(e|z) \tag{84}$$

And so if Alice and Bob share correlations, Eve cannot be correlated to either Alice or Bob.

Question: So if Alice and Bob share a purely entangled pair, Eve cannot have any information. Thus, I guess it is advisable for Eve, when she distributes the entangled states, to always give them a slightly faulty state. I believe that this is what has been shown in the noisy channel above.

Monogamy of pure state entanglement There exists a simple argument for the entanglement sharing of pure states. Say that ρ_{AB} is an entangled pure state. Let is be part of a larger system. Now, if the larger system is to be correlated to ρ_{AB} needs to be a mixed state, but it is by assumption pure. Thus, there can be no correlations between ρ_{AB} and a larger system.

2-shareable distribution a distribution is 2-shareable (with respect to Bob) if exists a 3-party non-signalling distribution such that

$$\sum_{b_1} = P(a, b_1, b|x, y_1, y) = P(a, b|x, y) \tag{85}$$

$$\sum_{b_2} P(a, b, b_2|x, y, y_2) = P(a, b|x, y) \tag{86}$$

That is, the resulting distribution is the marginal distribution for two different distributions when Bob has more than 1 measurement input and output.

2-Shareability Theorem : If Bob has a 2-shareable function, he can prove that it satisfies all Bell inequalities with just two measurements.

Proof. Assume that Bob has two measurement input b_1 and b_2, along with two measurement outcomes y_1 and y_2. Then, we can write

$$P8a, b|x, y) = \sum_{b_1,b_2} P(b_1, b_2|0, 1)P(a|x, b_1, b_2, 0, 1)\delta^b_{b_1\delta^0_y + b_2\delta^1_y} \tag{87}$$

Question: I don't see where this comes from. Also, I don't see how it relates to 2-shareability.

Then, use $\lambda = (b_1, b_2)$ to get

$$P(a, b|x, y) = \sum_{\lambda} P(\lambda)P(a|x, \lambda)P(b|y, \lambda) \tag{88}$$

which shows that the distribution is local.

k-shareable distributions satisfies all Bell measurements if Bob has k measurements on his side.

Question: is it that the distribution is non-local if Bob has fewer measurements? Or is it still local, but we cannot determine it?

Symmetrising correlations can be done without losing non-locality.

Noisy PR-boxes satisfy the following distribution:

$$P_\nu(a, b|x, y) = \begin{cases} \frac{1-\nu}{2} \text{ if } a \oplus b = xy \\ \frac{\nu}{2} \text{ otherwise} \end{cases} \tag{89}$$

where δ is a noise parameter. For certain values of ν, the distribution leaves the non-local domain and enters the local polytope.

Since this is not an extremal point any more, we can write it in terms of a PR box.

$$P_\nu(a,b|x,y) = (1-2\nu)P_{PR}(a,b|x,y) + 2\nu\frac{1}{4} \tag{90}$$

Symmetry of noisy PR box can be seen by the fact that

$$C_{00} + C_{01} + C_{10} - C_{11} = 1 - 2\nu \tag{91}$$

This cover the full classical-quantum-beyond range for different values of ν.

Binary monogamy consider a distribution $P(a,b,e|x,y,z)$ with binary variables. Then, at most one of the marginals $P(a,b|x,y)$ or $P(a,e|x,z)$ is non-local.

That is, if Alice and Bob share a non-local correlation, Eve cannot be correlated with Alice as well.

No-cloning theorem for a general non-local distribution. Imagine that we start with a non-local distribution $P(a,b|x,y)$. If Alice measures on her side, using some setting a, she prepares the following distribution for Bob

$$P(b|y,a,x) \tag{92}$$

Then, let Bob attempt to clone the distribution. In some frame, Bob will do so before Alice performs the measurement. So, in some frame, Bob gets the distribution,

$$P'(b_1,b_2,a|y_1,y_2,x) = P''(b_1,b_2|y_1,y_2,a,x)P(a|x) \tag{93}$$

Note that marginalising out either b_1, y_1 or b_2, y_2 would yield two separate distributions that otherwise are the same (depend on the same a and x). That is, the distribution satisfies

$$P'(b_1,a|y_1,x) = P'(b_2,a|y_2,x) = P(b|y,a,x) \tag{94}$$

However, $P(b|y,a,x)$ is non-local. By the monogamy of correlations, we see that all three cannot be non-local. This means that either the original distribution $P(b|y,a,x)$ is local, or the cloning is forbidden.

7 Device-Independent QKD

Assumption The parties are restricted by the no-signalling requirement. Otherwise, they can be as non-local as they like.

Key idea Before, we worked out this protocol in terms of states and operators. However, since any state produced by a device might be corrupted, we wish to describe the entire formalism using probability distribution. That is, the formalism is device independent if we have security without the requirement that the quantum devices are trustworthy.

Rewriting the distribution in terms of one PR box and several local points. Start with the symmetric distribution.

$$P_\nu(a,b|x,y) = \begin{cases} \frac{1-\nu}{2} \text{ if } a \oplus b = xy \\ \frac{\nu}{2} \text{ otherwise} \end{cases} \tag{95}$$

Then, use $C = 1 - 2\nu$, to write

$$P_\nu(a,b|x,y) = (1-4\nu)P_{\nu=0}(a,b|x,y) + 4\nu P_{\nu=1/4}(a,b|x,y) \tag{96}$$

Since $\nu = 1/4$ is the fully classical value, we can write that in terms of the 8 local points. So,

$$P_\nu(a,b|x,y) = (1-4\nu)P_{\nu=0}(a,b|x,y) + 4\nu\frac{1}{8}\sum_{t=1}^{8}\delta^a_{f_t(x)}\delta^b_{g_t(y)} \tag{97}$$

Optimal individual attack Given that Alice and Bob share a noisy probability distribution $P_\nu(a,b|x,y)$, the optimal attack has global distribution $P(a,b,e|x,y,z)$. This has extremal points at

$$\{P(z,b|x,ye,z)\}\forall e \tag{98}$$

That is, the points that Eve gives Alice and Bob are, to her, extremal. Alice and Bob can't really tell. They don't know Eve's component.

Question: Is this similar to when Eve holds the purification to Alice's and Bob's state?

Proof: Assume the opposite, that $P(a,b|x,y,e,z)$ is made up of a mixture of other extremal points $P_i(a,b|x,y,e,z)$

$$P(a,b|x,y,e,z) = \sum_i P(i|e,z)P_i(a,b|x,y,e,z) \tag{99}$$

Here, $P(i|e,z)$ is the weight. Then, it follows that we can write

$$P(a,b,e,i|x,y,z) = P_i(a,b|x,y,e,z)P(i|e,z)P(e|z) \tag{100}$$

is non-signalling.

Question: I didn't really manage to come up with a good proof of this.

If Eve has this distribution and knows both i and e, she can ignore i, because the distribution has extremal conditional distributions $P(a,b|x,y,z,e,i)$. So it doesn't matter which extremal point we have, just the knowledge that it is an extremal point suffices (I think). This enables us to write

$$P8a,b,e|x,y) = (1-4\nu)P_{PR}(a,b|x,y)\delta^0_e + 4\nu\frac{1}{8}\sum_{t=1}^{8}\delta^a_{f_t(x)}\delta^b_{g_t(y)}\delta^t_e \tag{101}$$

That is, we treat the 0 point as the PR box point, and every other point is distributed.

If Eve knows x, then she also knows a with probability 4ν, and knows nothing with probability $1-4\nu$. That is, Eve can know the local correlations perfectly, but can learn nothing from the PR box.

Question: I still do not see how this shows that the extremal points make up the optimal attack.

Thus, the best attack would be for Eve to simulate, say, quantum mechanics, while still having access to the local correlations and the PR box.

Secret key rate for device independent QKD is

$$r = 1 - 4D - h(D) \tag{102}$$

7.1 Characterising the set of quantum correlations

Key idea We want to know what sets quantum correlations apart from beyond-quantum correlations

Ingredients of QM In quantum mechanics, we define a correlation function as

$$C_{xy} = \mathrm{Tr}\left[\rho_{AB}A_x \otimes B_y\right] \tag{103}$$

where A_x and B_y are operators with eigenvalues ± 1.

Deriving T'sirelson's Bound we define two matrices,

$$(M_1, M_2, M_3, M_4) = (A_0 \otimes I, A_1 \otimes I, I \otimes B_0, I \otimes B_1) \tag{104}$$

and

$$Q_{ij} = \mathrm{Tr}\left[\rho_{AB} M_i M_j\right] \tag{105}$$

Both M_i and Q are Hermitian.

Since M_i has only ± 1 eigenvalues, $M_i^2 = I$, since it is the only matrix with only $+1$ eigenvalues. Thus, we can write

$$Q = \begin{pmatrix} 1 & \alpha & C_{00} & C_{01} \\ \bar{\alpha} & 1 & C_{10} & C_{11} \\ C_{00} & C_{01} & 1 & \beta \\ C_{10} & C_{11} & \bar{\beta} & 1 \end{pmatrix} \tag{106}$$

where

$$\alpha = \mathrm{Tr}\left[\rho_{AB} A_0 A_1 \otimes I\right] \tag{107}$$

$$\beta = \mathrm{Tr}\left[\rho_{AB} \mathbb{I} \otimes B_0 B_1\right] \tag{108}$$

Theorem for distinguishing QM If the correlations C_{xy} are quantum, then there are two real numbers α, β such that Q is positive semi-definite.

Proof: Choosing a symmetric scenario where

$$C_{00} + C_{01} + C_{10} - C_{11} = C \tag{109}$$

we can solve for the eigenvalues of the matrix. Requiring that they are all zero or above yields

$$4(\alpha^2 \beta^2 + 4C^4) \leq 4 \tag{110}$$

Minimising with respect to α and β gives

$$C \leq \frac{1}{\sqrt{2}} \tag{111}$$

which is T'sirelson's bound.