

Security of Quantum Key Distribution Against All Collective Attacks

Eli Biham⁽¹⁾, Michel Boyer⁽²⁾, Gilles Brassard⁽²⁾,
Jeroen van de Graaf⁽³⁾ and Tal Mor^(4,5)

(1) Computer Science Department, Technion, Haifa 32000, Israel;

(2) DIRO, Université de Montréal, Montréal, Canada;

(3) LCC/Cenapad, UFMG, Belo Horizonte, Brazil;

(4) Electrical Engineering, UCLA, Los Angeles, Cal., USA;

(5) Electrical Engineering, College of Judea and Samaria, Ariel, Israel.

December 5, 2001

Abstract

Security of quantum key distribution against sophisticated attacks is among the most important issues in quantum information theory. In this work we prove security against a very important class of attacks called *collective attacks* (under a compatible noise model) which use quantum memories and gates, and which are directed against the final key. This work was crucial for a full proof of security (against the joint attack) recently obtained by Biham, Boyer, Boykin, Mor, and Roychowdhury [1].

1 Introduction

Processing information using quantum two-level systems (qubits), instead of classical bits, has led to many surprising results such as exponentially fast quantum algorithms, teleportation of unknown states, and quantum cryptography which was originated by Wiesner, Bennett, and Brassard (see for instance [2, 3, 4, 5, 6]). Quantum key distribution was invented in 1984 [3], to

provide a new type of solution to one of the most important cryptographic problems: the transmission of secret messages. A key distributed via quantum cryptography techniques can be secure even against an eavesdropper with unlimited computing power, while the most advanced “public key” or “secret key” schemes do not have, and never will have this type of security.

The conventional setting (which we adopt here as well) is as follows: Alice and Bob have labs that are perfectly secure, they use qubits for their quantum communication, and they have access to a classical communication channel which can be heard, but cannot be jammed (i.e. cannot be tampered with) by the eavesdropper. The last assumption can easily be justified if Alice and Bob can broadcast messages, or if they already share some small number of secret bits in advance, to authenticate the classical channel. The other two assumptions (the perfectly secure labs, and the creation of qubits) are discussed in some papers regarding practical cryptography (see for instance [7]), but more work is required to justify them.

Using this setting, many physicists and computer scientists have tried to prove the security of various quantum key distribution schemes in the last decade. The conventional measure of security is the information Eve can obtain on the final key, and a security proof usually calculates (or puts bounds on) this information. Many particular attacks have been analyzed, such as the *intercept-resend attacks* and the *individual particle attacks*, for which there is a clear intuition that classical privacy amplification provides the desired security. However, these attacks are special cases which simplify the security analysis considerably. Quantum mechanics allows much stronger attacks which use quantum gates and quantum memory and are directed against the *final key*. The strongest (most general) attacks allowed by quantum mechanics are called *joint attacks*. Such attacks are beyond current technology, but obviously, a cryptosystem is not absolutely secure if some future technology could break it. Thus, proving security against any attack allowed by the rules of physics is a vital step.

In this paper we complete the work started in [8, 9, 10] to conclude that, under a compatible error model, the four-state scheme [3] for quantum key distribution is secure against any *collective attack*, an important subclass of the joint attacks. The first version of our paper appeared in the public domain¹ already in 1998 but has not yet been published. Our result is an

¹In the Los-Alamos Archive xxx.lanl.gov/archive/quant-ph; Quant-Physics number 9801022.

important step on the route to proving the security of QKD against joint attacks. It can lead to such a proof if one generalizes this work, or if one succeeds to prove an older conjecture [9], namely, that collective attacks are the strongest subclass of the joint attacks (meaning that Eve can obtain the largest amount of mutual information if she uses a collective attack, and not if she uses any other joint attack).

Recently, by exploiting and generalizing the techniques presented here, and by expanding the analysis further, Biham, Boyer, Boykin, Mor, and Roychowdhury [1] (BBBMR) succeeded to prove the security of quantum key distribution against the most general attacks on the channel, the joint attacks: the general attack is described in [1] as a generalization of the attack on a bit (Eq. 7) to an attack of a string, and the proof of security is based on a generalization of the purification of Bob’s bit (Eq. 10) to a purification of Bob’s string, on the use of the Theorems proven here, and many additional steps required to deal with the much more complex attack. Note that the conjecture of [9] cited above is still an open problem.

Other independent proofs of the security against joint attacks were also derived prior to the proof of BBBMR ([11, 12]) and shortly after ([13]) based on various different methods², hence this old-standing important problem of the security is now considered solved. However, it is important to have several approaches since *a.*—practical quantum key distribution is not yet proven secure [7], *b.*—it is rather common in cryptography that proofs of security are found wrong or incomplete once better understanding is obtained (though this statement probably does not apply to any of the proposed proofs in the case of QKD), and *c.*—different proofs are sometimes based upon different assumptions hence often lead to different results (especially when connected to various practical considerations, such as multi-photon states and high channel losses).

2 Description of the protocol

In the four-state scheme of Bennett and Brassard (BB84) [3] the sender (Alice) sends to the receiver (Bob) a classical string x'' of length n'' using a quantum channel, by sending qubits. She sends either $|0\rangle_z$ or $|0\rangle_x = (|0\rangle_z + |1\rangle_z)/\sqrt{2}$ to encode a bit value 0, or she sends either $|1\rangle_z$ or $|1\rangle_x = (|0\rangle_z -$

² The proof in [11] assumes quantum computers, and the other proofs [12, 1, 13] apply to more realistic scenarios.

$|1\rangle_z)/\sqrt{2}$ to encode a bit value 1. Each classical bit is chosen randomly, and the basis (x or z) for encoding each classical bit using a qubit is also chosen randomly. Alice and Bob are also connected by a classical channel which is insecure against eavesdropping (it is assumed public) but unjammable (cannot be tampered with). At a later stage, after Bob has received the particles, Alice tells Bob through the classical, public channel which basis (x or z) she used for each qubit. If Bob has used the same basis for his measurement they keep the bit, otherwise they discard it. The resulting n' -bit string x' is known as the *sifted* key. If no errors and no eavesdropping occurred, then Alice's string x'_A is identical to Bob's string x'_B .

But due to imperfections in the creation, transmission and measurement of qubits transmission errors will occur. Therefore Alice and Bob agree in advance on some error rate threshold p_{allowed} . Once Alice and Bob share the sifted key (i.e. two almost identical string x'_A resp. x'_B) they estimate the error rate using a random subset of n_{test} bits. If the error rate p_{test} on the test-bits is less than the pre-agreed threshold p_{allowed} , then the test succeeds and Alice and Bob choose n bits from the remaining bits (hence, $n \leq n' - n_{\text{test}}$) to be the “information bits”. these n -bit information strings in Alice's and Bob's hands are denoted respectively x_A and x_B . The error-rate on the information bits (the value $|x_A \oplus x_B|/n$ where x_A and x_B are the n -bit information strings in Alice's and Bob's hands respectively) is promised to be similar to the error-rate on the test bits due to some law of large numbers which is applicable due to the random sampling of the test bits.

Once the test is passed, the last step is to obtain a final key of length m from these n bits by performing error correction and privacy amplification. Let us describe a particular way to perform this (alternative ways exist, but are harder to analyze): Alice and Bob choose parities of r substrings for error correction and parities of m substrings for privacy amplification. The parity of each of the r error-correction substrings is announced in order to correct x_B , and the parities of the m privacy amplification substrings are kept secret, and used as the final key.

In the most general (“joint”) attack, the eavesdropper, Eve, can do whatever she likes (the most general unitary transformation using an ancilla) to the qubits, and delay her measurement of the (extremely big) ancilla until receiving all classical data. In the collective attack, each of Alice's qubits is attacked via a separate ancilla and all ancillas are finally measured collectively. The first hints that such collective measurements will not destroy the security obtained by privacy amplification were provided in [8]. The first

complete examples (which contain privacy amplification and error correction) were provided in [14, 9] (Mayers work was based on an earlier work of Yao).

In this paper we restrict ourself to “collective” attacks [9, 10], where each qubit is attached to a separate probe, unentangled to any other probe. The measurement is delayed until after all the classical data is obtained, and is performed collectively on all probes. There are good reasons [9, 10] to believe that collective attacks are the strongest joint attacks (when n is large): Eve has no knowledge regarding which are the relevant qubits at the time she performs the unitary transformation, and the probability of performing the transformation on relevant qubits alone is exponentially small. A transformation done on relevant and irrelevant qubits together seems to cause noise without providing useful information to Eve. Furthermore, no particular joint attack has been shown to be stronger than collective attacks (in terms of Eve’s information on the final key). It may well be that the symmetric collective attack of [10], which uses 2-dimensional probes, is the strongest joint attack in terms of the amount of information it gives to the eavesdropper.

In a collective attack, when Alice sends x'' to Bob, each qubit gets entangled with a separate probe of Eve. So the global state of the Eve-Bob system is $\rho_1 \otimes \dots \otimes \rho_{n''}$ where each ρ_i is a density operator on the space $\mathcal{H}^{E_i} \otimes \mathcal{H}^{B_i}$ and where the space \mathcal{H}^{E_i} and the 2-dimensional space \mathcal{H}^{B_i} belong respectively to Eve and Bob.

To simplify the derivation of the proof of security, and to make this paper clear and a bit less technical, we leave the proofs of the four theorems provided here to the appendix. We consider $m = 1$ in the first sections and leave the general case of m bits to the last sections of this paper.

3 Bounds on information

We present some notations from information theory. Let B and X be random variables (describing the input and output of a channel). When the context is clear we write $p(b)$ for $p(B=b)$ and $p(x)$ for $p(X=x)$. The joint probability $p(x, b)$ satisfies $p(x) = \sum_{b \in B} p(x, b)$ and $p(b) = \sum_{x \in X} p(x, b)$. The conditional probability is denoted by $p_b(x) \equiv p(X=x|B=b)$ and $p_x(b) \equiv p(B=b|X=x)$. It satisfies the Bayes formula $p_b(x)p(b) = p(x, b) = p_x(b)p(x)$. The mutual information between the input and the output probability distributions, $I(X; B) = -\sum_{b \in B} p(b) \log_2 p(b) + \sum_{x \in X} p(x) \sum_{b \in B} p_x(b) \log_2 p_x(b)$, tells us the

average increase of knowledge about the input when the output is known.

Let the entropy function $h_2(p)$ be defined as $h_2(p) = -p \log_2 p - (1 - p) \log_2 (1 - p)$. For a binary input B with equal input probabilities

$$I(B; X) = \sum_{x \in X} p(x) I_2(p_x(0)) , \quad (1)$$

where $I_2(p) = 1 - h_2(p) = 1 + p \log_2 p + (1 - p) \log_2 (1 - p)$. Distinguishing the input when the output is given, is then equivalent to distinguishing the two probability distributions $p_0(x)$ and $p_1(x)$. All the probability distributions in the expression of the mutual information can be calculated from $p_0(x)$, $p_1(x)$ (since the input probabilities $p(0) = p(1) = 1/2$ are also known):

$$p(x) = \sum_b p(x, b) = \sum_b p(b) p_b(x) ;$$

$$p_x(0) = p(x, 0)/p(x) = p(B = 0)p_0(x)/p(x) .$$

Therefore, we can define another function SD , Shannon Distinguishability,

$$SD(p_0(x), p_1(x)) \equiv I(B; X) \quad (2)$$

(restricted to binary input with equal probabilities), which has the same values as the mutual information, but is a function of the two probability distributions $p_0(x)$ and $p_1(x)$, and is a measure of their distinguishability.

Suppose we are given a state (density matrix) ρ . The most general measurement giving a result x in some set X of possible outputs is given by a POVM [15] \mathcal{E} indexed by X , i.e. a family $\mathcal{E} = (E_x)_{x \in X}$ of Hermitian operators E_x with non negative eigenvalues such that $\sum_{x \in X} E_x = I$. The probability of occurrence of x given the state ρ is then equal to

$$p^{\mathcal{E}}(x) = Tr(\rho E_x) .$$

Given two equally likely states ρ_0 and ρ_1 , and a measurement procedure \mathcal{E} , let $p_0^{\mathcal{E}}(x) = Tr(\rho_0 E_x)$ and $p_1^{\mathcal{E}}(x) = Tr(\rho_1 E_x)$ be the resulting two probability distributions, and let

$$SD^{\mathcal{E}}(\rho_0, \rho_1) \equiv SD(p_0^{\mathcal{E}}(x), p_1^{\mathcal{E}}(x)) , \quad (3)$$

the Shannon Distinguishability between the density matrices once a particular POVM is used. The maximum information we can get regarding the state

we are facing is given by the optimal Shannon Distinguishability (known also as the *accessible information*)

$$SD(\rho_0, \rho_1) \equiv \sup_{\mathcal{E}} [SD^{\mathcal{E}}(\rho_0, \rho_1)] \quad (4)$$

where the supremum is taken over all POVM's on all possible sets X.

Unfortunately, there is no known analytic formula giving such an optimal mutual information. In what follows, we present two important bounds; the proofs are given in the appendix.

Theorem 1. — If $\tilde{\rho}_0$ and $\tilde{\rho}_1$ are two density matrices defined on some space $\mathcal{H}_1 \otimes \mathcal{H}_2$ and $\rho_i = Tr_2(\tilde{\rho}_i)$ are the density matrices on \mathcal{H}_1 obtained by tracing-out \mathcal{H}_2 , then,

$$SD(\rho_0, \rho_1) = SD(Tr_2(\tilde{\rho}_0), Tr_2(\tilde{\rho}_1)) \leq SD(\tilde{\rho}_0, \tilde{\rho}_1) . \quad (5)$$

The $\tilde{\rho}_i$ will be called a *lift-up* of ρ_p , and it is known as *purification* if it is a pure state. This theorem (proven independently in [16]) actually states that tracing out cannot increase information. It provides a useful upper bound on the mutual information that can be obtained about mixed states ρ_i , if we can find appropriate states $\tilde{\rho}_i$. A similar idea which says that mixing cannot increase information was used in [10] to obtain a more limited security result.

For an Hermitian matrix A , the Trace-Norm of A , denoted by $Tr|A|$, is the sum of the absolute values of the eigenvalues of A . The following upper bound is very important.

Theorem 2. [17] — For any two density matrices $\tilde{\rho}_0$ and $\tilde{\rho}_1$, defined on some space \mathcal{H} ,

$$SD(\tilde{\rho}_0, \tilde{\rho}_1) \leq \frac{1}{2} Tr|\tilde{\rho}_0 - \tilde{\rho}_1| . \quad (6)$$

The importance of this bound is in avoiding the necessity of optimizing over all possible measurements (since the Trace-Norm is independent of the measurement process), an optimization process which is in general unknown. The optimal measurement (as far as mutual information is concerned) still yields an optimal mutual information which is bounded by the expression of the Trace-Norm.

4 Error versus information

Let us assume that Eve is powerful enough to control the natural noise. Without loss of generality, we assume that Eve's probes are in some arbitrary but fixed initial (tensor product) pure state, and that each probe is in a state $|E\rangle$. In the *collective attack*, the state $|E\rangle \otimes |b\rangle$ is subjected to Eve's unitary transformation \mathcal{U} that changes the state $|b\rangle$ sent by Alice to the final global state

$$\begin{aligned} |0\rangle_z &\mapsto |E_{0,0}^z\rangle|0\rangle_z + |E_{0,1}^z\rangle|1\rangle_z \equiv |\phi_0^z\rangle \\ |1\rangle_z &\mapsto |E_{1,0}^z\rangle|0\rangle_z + |E_{1,1}^z\rangle|1\rangle_z \equiv |\phi_1^z\rangle \end{aligned} \quad (7)$$

where the $|E_{i,j}^z\rangle$ are Eve's non normalized states. Implicitly, this description corresponds to restricting natural noise to follow the spirit of the collective attacks. It is reasonable to suspect that more general noise models would not be to Eve's advantage, but being unable to prove it, we adopted the above restriction.

Bob's average error probability in the z basis [measuring $|0\rangle_z$ if $|1\rangle_z$ was sent etc.] is $p_e^z = (1/2)[\langle E_{0,1}^z|E_{0,1}^z\rangle + \langle E_{1,0}^z|E_{1,0}^z\rangle]$. Alice can also use the alternate basis x , and then the transformation \mathcal{U} can also be expressed in the x basis (replacing everywhere z by x) to yield Bob's average probability of error in the x basis

$$p_e^x = (1/2)[\langle E_{0,1}^x|E_{0,1}^x\rangle + \langle E_{1,0}^x|E_{1,0}^x\rangle] .$$

For any choice of attack we now show a connection between the error rate induced if Alice and Bob used the x basis, and Eve's benefit from the attack if the z basis was used. This is an "information vs. disturbance" argument. Due to linearity of the transformation \mathcal{U} we obtain $|E_{0,1}^x\rangle = \frac{1}{2}[(|E_{0,0}^z\rangle - |E_{1,1}^z\rangle) + (|E_{1,0}^z\rangle - |E_{0,1}^z\rangle)]$ and $|E_{1,0}^x\rangle = \frac{1}{2}[(|E_{0,0}^z\rangle - |E_{1,1}^z\rangle) - (|E_{1,0}^z\rangle - |E_{0,1}^z\rangle)]$. If we expand p_e^x in terms of the vectors in the z basis we get $p_e^x = (1/4)[\langle E_{0,0}^z - E_{1,1}^z|E_{0,0}^z - E_{1,1}^z\rangle + \langle E_{1,0}^z - E_{0,1}^z|E_{1,0}^z - E_{0,1}^z\rangle]$. Since \mathcal{U} preserves inner products, the states $|\phi_0^z\rangle$ and $|\phi_1^z\rangle$ have norm 1. Therefore, $\langle E_{0,0}^z|E_{0,0}^z\rangle + \langle E_{0,1}^z|E_{0,1}^z\rangle = 1$ and $\langle E_{1,0}^z|E_{1,0}^z\rangle + \langle E_{1,1}^z|E_{1,1}^z\rangle = 1$, yielding

$$p_e^x = \frac{1}{2}[1 - \text{Re}\{\langle E_{0,0}^z|E_{1,1}^z\rangle + \langle E_{1,0}^z|E_{0,1}^z\rangle\}] .$$

Using $\text{Re}\{\alpha\} \leq |\alpha|$ for any complex number α , we finally get

$$1 - 2p_e^x \leq |\langle E_{0,0}^z|E_{1,1}^z\rangle + \langle E_{1,0}^z|E_{0,1}^z\rangle| . \quad (8)$$

Eve's view, if the z basis was used, is obtained by tracing-out Bob from the states ϕ_b^z :

$$\begin{aligned}\rho_0^z(E) &= |E_{0,0}^z\rangle\langle E_{0,0}^z| + |E_{0,1}^z\rangle\langle E_{0,1}^z| \\ \rho_1^z(E) &= |E_{1,0}^z\rangle\langle E_{1,0}^z| + |E_{1,1}^z\rangle\langle E_{1,1}^z| .\end{aligned}\tag{9}$$

For a collective attack, these density matrices contain all the information available to Eve on a particular qubit, since bob's measurements on the test bits (which are revealed) provide no information regarding the information bits.

Note that the states ϕ_b^z are purifications of Eve's states. These purifications are orthogonal to each other, hence can be perfectly distinguished, and cannot be useful in Theorem 1.

Many other pure states also yield the same reduced density matrices for Eve (hence are purifications of Eves states). In particular

$$\begin{aligned}|\psi_0^z\rangle &= |E_{0,0}^z\rangle|0\rangle_z + |E_{0,1}^z\rangle|1\rangle_z \\ |\psi_1^z\rangle &= |E_{1,1}^z\rangle|0\rangle_z + |E_{1,0}^z\rangle|1\rangle_z\end{aligned}\tag{10}$$

are found to be useful *purifications* since the angle between them is zero if there is no disturbance. While these states have only virtual existence, they can be used in Theorem 1 to yield the desired bound, since Eve's states are the trace-out of these pure states. They span a 2-dimensional subspace \mathcal{H} of the Bob-Eve space. They are normalized, and their overlap yields $|\langle\psi_0^z|\psi_1^z\rangle| = \cos(2\alpha_z)$ for some angle $0 \leq \alpha_z \leq \pi/4$. Note also that $|\langle\psi_0^z|\psi_1^z\rangle| = |\langle E_{0,0}^z|E_{1,1}^z\rangle + \langle E_{1,0}^z|E_{0,1}^z\rangle|$ so that $1 - 2\sin^2(\alpha_z) = |\langle E_{0,0}^z|E_{1,1}^z\rangle + \langle E_{1,0}^z|E_{0,1}^z\rangle|$. Using (8) we deduce that $1 - 2p_e^x \leq 1 - 2\sin^2(\alpha_z)$ leading to

$$\sin(\alpha_z) \leq (p_e^x)^{1/2} .\tag{11}$$

The overlap is, in general, a complex number $\langle\psi_0^z|\psi_1^z\rangle = e^{i\theta}\cos(2\alpha_z)$. Let us define the two vectors

$$\begin{aligned}|0_{\mathcal{H}}^z\rangle &= [|\psi_0^z\rangle + e^{-i\theta}|\psi_1^z\rangle]/(2\cos\alpha_z) ; \\ |1_{\mathcal{H}}^z\rangle &= [|\psi_0^z\rangle - e^{-i\theta}|\psi_1^z\rangle]/(2\sin\alpha_z) .\end{aligned}\tag{12}$$

They form an orthonormal basis of the 2-dimensional subspace \mathcal{H} spanned by $|\psi_0^z\rangle$ and $|\psi_1^z\rangle$. In this new basis $|\psi_0^z\rangle = \cos(\alpha_z)|0_{\mathcal{H}}^z\rangle + \sin(\alpha_z)|1_{\mathcal{H}}^z\rangle$ and $|\psi_1^z\rangle = e^{i\theta}[\cos(\alpha_z)|0_{\mathcal{H}}^z\rangle - \sin(\alpha_z)|1_{\mathcal{H}}^z\rangle]$.

More appropriate purifications of Eve's states can be defined as follows: Let $|\Psi_0^z\rangle = |\psi_0^z\rangle$ and $|\Psi_1^z\rangle = e^{-i\theta}|\psi_1^z\rangle$. Then

$$\begin{aligned} |\Psi_0^z\rangle &= |E_{0,0}^z\rangle|0\rangle_z + |E_{0,1}^z\rangle|1\rangle_z \\ |\Psi_1^z\rangle &= e^{-i\theta}[|E_{1,1}^z\rangle|0\rangle_z + |E_{1,0}^z\rangle|1\rangle_z] \end{aligned} \quad (13)$$

and in the new basis (12) they are written in a very simple way

$$\begin{aligned} |\Psi_0^z\rangle &= \cos(\alpha_z)|0_{\mathcal{H}}^z\rangle + \sin(\alpha_z)|1_{\mathcal{H}}^z\rangle \\ |\Psi_1^z\rangle &= \cos(\alpha_z)|0_{\mathcal{H}}^z\rangle - \sin(\alpha_z)|1_{\mathcal{H}}^z\rangle . \end{aligned} \quad (14)$$

Everything that has been said about $|\psi_b^z\rangle$ and p_e^x holds by symmetry for replacing the bases, once other new basis vectors $|0_{\mathcal{H}}^x\rangle$ and $|1_{\mathcal{H}}^x\rangle$ are defined by replacing z by x in (12,10), yielding $\sin(\alpha_x) \leq (p_e^z)^{1/2}$.

Alice uses both bases with the same probability, hence Bob's overall probability of error is $p_e = \frac{1}{2}(p_e^x + p_e^z)$. Using $p_e^x \leq 2p_e$ and $p_e^z \leq 2p_e$, we obtain $\sin(\alpha_z) \leq (2p_e)^{1/2}$ and $\sin(\alpha_x) \leq (2p_e)^{1/2}$. When Eve performs an attack on a particular qubit the purifications in both bases are restricted as above. In the sequel we will simply drop the indices x and z , taking as a convention that we are dealing with the actual basis that Alice and Bob agreed upon (and which become known to Eve only after she retransmitted the particle towards Bob). Thus, for any qubit $|b\rangle_x$ or $|b\rangle_z$ transmitted from Alice to Bob, we have found that the connection between Eve's angle (in the basis actually chosen by Alice and Bob) and Bob's average error-rate is

$$\sin(\alpha) \leq (2p_e)^{1/2} . \quad (15)$$

Also we have found that the purification of Eve's state (known to her only once she learn the basis) is

$$|\Psi_b\rangle = \cos(\alpha) |0_{\mathcal{H}}\rangle + (-1)^b \sin(\alpha) |1_{\mathcal{H}}\rangle . \quad (16)$$

written using the relevant basis vectors $\{|0_{\mathcal{H}}^z\rangle ; |1_{\mathcal{H}}^z\rangle\}$ or $\{|0_{\mathcal{H}}^x\rangle ; |1_{\mathcal{H}}^x\rangle\}$ depending on the basis, z or x , chosen by Alice.

5 The state in Eve's hands

We now look at the n' remaining qubits after Alice and Bob discard those bits where the bases did not agree. Some bits are used to verify that $p_{\text{test}} \leq$

p_{allowed} , to be left with n -bit string \mathbf{x} . After retransmitting the i -th bit (denoted here by x_i) to Bob, Eve's state is $\rho_{x_i}(E)$, where x_i is either 0 or 1 ($1 \leq i \leq n$) according to the bit which Alice sent to Bob. [Note that x_i replaces b of the previous section; b is not used in this section, and shall have a different meaning in later sections.] The purification of Eve's state is $|\Psi_{x_i}\rangle = \cos(\alpha_i) |0\rangle_i + (-1)^{x_i} \sin(\alpha_i) |1\rangle_i$, where $\{|0\rangle_i; |1\rangle_i\}$ replaces $\{|0_{\mathcal{H}_i}\rangle; |1_{\mathcal{H}_i}\rangle\}$ of the notations of the previous paragraph and in the appropriate basis. Moreover $\sin(\alpha_i) \leq (2p_i)^{1/2}$, where p_i is Bob's probability of error on the i -th bit (averaged over the four possible input states), which is completely determined by Eve's transformation. Thanks to the properties of the trace, the global state of Eve's probes, $\rho_{\mathbf{x}}(E)$, is obtained by performing a partial trace of $|\Psi_{\mathbf{x}}\rangle$, the tensor product of the $|\Psi_{x_i}\rangle$.

To expand $|\Psi_{\mathbf{x}}\rangle$ we first need some notations. Boldface letters like \mathbf{j} , \mathbf{x} are used to denote strings in $\{0,1\}^n$ that are interpreted as n -vectors on the binary field. The i 'th bit in \mathbf{j} is denoted j_i [\mathbf{j}_i stands for a different n -bit vector and not for the i 'th bit of \mathbf{j}]. Boldface letters are also used in kets, with the following meaning: if $\mathbf{j} = j_1 \dots j_n$ is concatenation of n bits then $|\mathbf{j}\rangle = |j_1\rangle_1 \otimes \dots \otimes |j_n\rangle_n$ where $|0\rangle_i$ and $|1\rangle_i$ are the basis vectors of the purifications of Eve's i 'th qubit. The state

$$|\Psi_{\mathbf{x}}\rangle = \bigotimes_{i=1}^n [(\cos(\alpha_i) |0\rangle_i + (-1)^{x_i} \sin(\alpha_i) |1\rangle_i)] \quad (17)$$

can be written as

$$|\Psi_{\mathbf{x}}\rangle = \sum_{\mathbf{j} \in \{0,1\}^n} d_{\mathbf{j}} (-1)^{\mathbf{x} \cdot \mathbf{j}} |\mathbf{j}\rangle, \quad (18)$$

where $d_{\mathbf{j}} = d_{j_1} \dots d_{j_n}$ with $d_{j_i} = \cos \alpha_i$ if $j_i = 0$ and $d_{j_i} = \sin \alpha_i$ if $j_i = 1$, and where $\mathbf{x} \cdot \mathbf{j}$ is by definition $\mathbf{x} \cdot \mathbf{j} = \sum_{i=1}^n x_i j_i \bmod 2$. For instance, $|\Psi_{01}\rangle = \cos \alpha_1 \cos \alpha_2 |00\rangle - \cos \alpha_1 \sin \alpha_2 |01\rangle + \sin \alpha_1 \cos \alpha_2 |10\rangle - \sin \alpha_1 \sin \alpha_2 |11\rangle$. Then, for $|\mathbf{j}\rangle = |10\rangle$ we see that $j_1 = 1$, $j_2 = 0$, $d_{j_1} = \sin \alpha_1$, $d_{j_2} = \cos \alpha_2$, and $d_{\mathbf{j}} = d_{10} = \sin \alpha_1 \cos \alpha_2$.

Let $\mathbf{j} \oplus \mathbf{k}$ be the bitwise XOR of the strings \mathbf{j} and \mathbf{k} . Using the equality $(-1)^{\mathbf{x} \cdot \mathbf{j}} (-1)^{\mathbf{x} \cdot \mathbf{k}} = (-1)^{\mathbf{x} \cdot (\mathbf{j} \oplus \mathbf{k})}$, the density matrix corresponding to the lift-up (here purification) $|\Psi_{\mathbf{x}}\rangle \langle \Psi_{\mathbf{x}}|$ of Eve's state can be written

$$\widetilde{\rho}_{\mathbf{x}} = |\Psi_{\mathbf{x}}\rangle \langle \Psi_{\mathbf{x}}| = \sum_{\mathbf{j}, \mathbf{k} \in \{0,1\}^n} d_{\mathbf{j}} d_{\mathbf{k}} (-1)^{\mathbf{x} \cdot (\mathbf{j} \oplus \mathbf{k})} |\mathbf{j}\rangle \langle \mathbf{k}|, \quad (19)$$

for any string \mathbf{x} sent by Alice.

Let $\hat{\mathbf{x}}$ be the number of ones in \mathbf{x} , called the *Hamming weight* of \mathbf{x} . Then, $\widehat{\mathbf{x} \oplus \mathbf{j}}$ is the *Hamming distance* between \mathbf{x} and \mathbf{j} .

6 The parity bit

In order to encode one key-bit b (0 or 1) using a substring of the n bits she sent, Alice proceeds as follows: she chooses some string $\mathbf{w} \in \{0, 1\}^n$ to define the *relevant* bits in \mathbf{x} whose exclusive-OR forms the final 1-bit secret key (in other words — the ones in the string \mathbf{w} define the substring of \mathbf{x} which will be used for privacy amplification); Alice announces \mathbf{w} to Bob; Bob understands that the key-bit sent is $b = \mathbf{x} \cdot \mathbf{w}$, and can calculate the final bit b . Eve now knows \mathbf{w} (but not \mathbf{x}) and has to guess $b = \mathbf{x} \cdot \mathbf{w}$. Only strings \mathbf{x} such that $\mathbf{x} \cdot \mathbf{w} = b$ contribute to $\widetilde{\rho}_b^{\mathbf{w}}$, and all these strings are equiprobable, hence

$$\begin{aligned}\widetilde{\rho}_0^{\mathbf{w}} &\equiv \frac{1}{2^{n-1}} \sum_{\mathbf{x}: \mathbf{x} \cdot \mathbf{w} = 0} |\Psi_{\mathbf{x}}\rangle \langle \Psi_{\mathbf{x}}| ; \\ \widetilde{\rho}_1^{\mathbf{w}} &\equiv \frac{1}{2^{n-1}} \sum_{\mathbf{x}: \mathbf{x} \cdot \mathbf{w} = 1} |\Psi_{\mathbf{x}}\rangle \langle \Psi_{\mathbf{x}}| .\end{aligned}\tag{20}$$

The following Theorem applies to the states given by Eqs. (20,19), and deals with the quality of the privacy amplification.

Theorem 3. —

$$\widetilde{\rho}_0^{\mathbf{w}} - \widetilde{\rho}_1^{\mathbf{w}} = 2 \sum_{\mathbf{j} \in \{0,1\}^n} d_{\mathbf{j}} d_{\mathbf{j} \oplus \mathbf{w}} |\mathbf{j}\rangle \langle \mathbf{j} \oplus \mathbf{w}| ;\tag{21}$$

$$\text{Tr}|\widetilde{\rho}_0^{\mathbf{w}} - \widetilde{\rho}_1^{\mathbf{w}}| \leq 2 \sum_{\mathbf{j}} d_{\mathbf{j}} d_{\mathbf{j} \oplus \mathbf{w}} .\tag{22}$$

To learn the secret bit b Eve needs to distinguish between the two density matrices (in her hands) $\rho_b^{\mathbf{w}}(E) = \frac{1}{2^{n-1}} \sum_{\mathbf{x}: \mathbf{x} \cdot \mathbf{w} = b} \rho_{\mathbf{x}}(E)$ for which $\widetilde{\rho}_b^{\mathbf{w}}$ are lift-ups. Due to Theorem 1 (Eq. 5), Theorem 2 (Eq. 6) and Theorem 3, we get $SD(\rho_0^{\mathbf{w}}, \rho_1^{\mathbf{w}}) \leq SD(\widetilde{\rho}_0^{\mathbf{w}}, \widetilde{\rho}_1^{\mathbf{w}}) \leq (1/2) \text{Tr}|\widetilde{\rho}_0^{\mathbf{w}} - \widetilde{\rho}_1^{\mathbf{w}}| \leq \sum_{\mathbf{j}} d_{\mathbf{j}} d_{\mathbf{j} \oplus \mathbf{w}}$. Thus, if the error correction data is unknown to Eve, her information regarding the final key bit is

$$SD(\rho_0^{\mathbf{w}}, \rho_1^{\mathbf{w}}) \leq \sum_{\mathbf{j}} d_{\mathbf{j}} d_{\mathbf{j} \oplus \mathbf{w}} .\tag{23}$$

In our case where $d_{\mathbf{j}}$ equals $d_{j_1} \dots d_{j_n}$ with $d_{j_i} = \cos \alpha_i$ if $j_i = 0$ and $d_{j_i} = \sin \alpha_i$ if $j_i = 1$, we can calculate the above Trace-Norm further. If w_i ,

the i^{th} bit of \mathbf{w} equals 1, then the product of the i^{th} factor of $d_{\mathbf{j}}$ by the i^{th} bit of $d_{\mathbf{w} \oplus \mathbf{j}}$ is $\cos \alpha_i \sin \alpha_i$, since either $[d_{j_i} = \cos \alpha_i \text{ and } d_{j_i \oplus w_i} = \sin \alpha_i]$ or alternatively $[d_{j_i} = \sin \alpha_i \text{ and } d_{j_i \oplus w_i} = \cos \alpha_i]$, since $j_i \oplus 1 = \text{not}(j_i)$. The contribution of such terms is $(\sin 2\alpha_i)$ since the sum is over all \mathbf{j} so the term $d_{\mathbf{j}}d_{\mathbf{j} \oplus \mathbf{w}}$ contributes twice. If w_i , the i^{th} bit of \mathbf{w} equals 0, then the product of the i^{th} factor of $d_{\mathbf{j}}$ by the i^{th} bit of $d_{\mathbf{w} \oplus \mathbf{j}}$ is either $\cos^2 \alpha_i$ or $\sin^2 \alpha_i$. When summing over all \mathbf{j} , such terms sum up to yield 1. As a result the sum reduces to

$$\begin{aligned} \sum_{\mathbf{j}} d_{\mathbf{j}}d_{\mathbf{j} \oplus \mathbf{w}} &= \prod_{i:w_i=1} \sin(2\alpha_i) \prod_{i:w_i=0} 1 \\ &= \prod_{i:w_i=1} \sin(2\alpha_i) . \end{aligned} \quad (24)$$

For instance, with $\mathbf{w} = 10$ we get $\sum_{\mathbf{j}} d_{\mathbf{j}}d_{\mathbf{j} \oplus \mathbf{10}} = d_{00}d_{10} + d_{01}d_{11} + d_{10}d_{00} + d_{11}d_{01} = 2[d_{00}d_{10} + d_{01}d_{11}]$ and therefore $\sum_{\mathbf{j}} d_{\mathbf{j}}d_{\mathbf{j} \oplus \mathbf{10}} = 2[\cos \alpha_1 \cos \alpha_2 \sin \alpha_1 \cos \alpha_2 + \cos \alpha_1 \sin \alpha_2 \sin \alpha_1 \sin \alpha_2] = 2 \cos \alpha_1 \sin \alpha_1 [\cos^2 \alpha_2 + \sin^2 \alpha_2] = \sin 2\alpha_1$.

If we look at \mathbf{w} as the characteristic function of a set also denoted \mathbf{w} , one can write $i \in \mathbf{w}$ instead of $w_i = 1$ (that is, “set notations”), and thus Theorem 3 gives

$$\text{Tr}|\widetilde{\rho}_0^{\mathbf{w}} - \widetilde{\rho}_1^{\mathbf{w}}| \leq 2 \prod_{i \in \mathbf{w}} \sin(2\alpha_i) , \quad (25)$$

so that finally

$$SD(\rho_0^{\mathbf{w}}, \rho_1^{\mathbf{w}}) \leq \prod_{i \in \mathbf{w}} \sin(2\alpha_i) , \quad (26)$$

if the error correction data is unknown to Eve.

The special case where $\mathbf{w} = 11 \dots 1$ and all angles are the same (α) was solved to leading order in [8] using different methods giving $(2\alpha)^n / \sqrt{\pi n/2}$. Using our method for that case, we get $(\sin 2\alpha)^n$ as a bound.

7 Error correction

For error correction, Alice chooses r independent strings $\{\mathbf{v}_1, \mathbf{v}_2 \dots \mathbf{v}_r\}$, and calculates $\mathbf{x} \cdot \mathbf{v}_i = b_i$ for each such string. Doing so, she actually calculates the parity of various substrings of \mathbf{x} , obtaining r independent parity bits. The strings \mathbf{v}_1 to \mathbf{v}_r and the outcomes b_1 to b_r are sent to Bob and hence also to Eve. As a result, Eve knows that \mathbf{x} is a solution of the set \mathbf{E} of equations $\{\mathbf{x} \cdot \mathbf{v}_1 = b_1, \dots, \mathbf{x} \cdot \mathbf{v}_r = b_r\}$. If we let \mathbf{b} be the string $(b_i)_{1 \leq i \leq r}$ and

if \mathbf{V} is the *parity matrix* whose r rows are the strings $\mathbf{v}_1, \dots, \mathbf{v}_r$, then the set \mathbf{E} of equations can be written more succinctly as

$$\mathbf{x}\mathbf{V}^\top = \mathbf{b}$$

where \mathbf{V}^\top is the transpose of \mathbf{V} . The matrix \mathbf{V} defines a linear code \mathcal{C} with 2^{n-r} code words of length n , namely the linear space defined by $\mathbf{y}\mathbf{V}^\top = \mathbf{0}$, and spanned by all these strings \mathbf{y} . If its minimal distance (the smallest Hamming weight of a non zero code word) is $d = 2t + 1$, then the code can be used to correct up to t errors. The same holds for other sets of strings \mathbf{x} provided we are given the ‘‘syndrome’’ $\mathbf{b} = \mathbf{x}\mathbf{V}^\top$; The set of strings \mathbf{x} (for a particular nonzero syndrome) then forms a nonlinear code defined by the same matrix \mathbf{V} , with the same minimal distance as the linear code \mathcal{C} . That is, the new code consists in the coset of \mathcal{C} containing \mathbf{b} (it is an *affine* code instead of a linear code).

The strings $\mathbf{v}_1, \dots, \mathbf{v}_r$ can be chosen randomly, to yield a *random linear code*, or can be chosen according to some other procedure. The results we obtain are independent of the way the code is generated.

Let \mathbf{V} be the linear space spanned by $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$. Since the strings \mathbf{v}_i are assumed to be linearly independent, the space \mathbf{V} is of dimension 2^r over the field $\mathbf{F}_2 = \{0, 1\}$. For any $\mathbf{s} \in \mathbf{F}_2^r = \{0, 1\}^r$, let $\mathbf{v}_\mathbf{s} = \sum_{i=1}^r s_i \mathbf{v}_i$. Clearly the map $\mathbf{s} \mapsto \mathbf{v}_\mathbf{s}$ is linear and injective (from linear independence); as a result, it is an isomorphism (i.e. 1 – 1 and onto) from \mathbf{F}_2^r onto \mathbf{V} . Moreover, for any \mathbf{x} that is a solution of the system \mathbf{E} of equations (i.e. such that $\mathbf{x}\mathbf{V}^\top = \mathbf{b}$), we get

$$\mathbf{x} \cdot \mathbf{v}_\mathbf{s} = \mathbf{x} \cdot \left(\sum_{i=1}^r s_i \mathbf{v}_i \right) = \sum_{i=1}^r s_i \mathbf{x} \cdot \mathbf{v}_i = \sum_{i=1}^r s_i b_i = \mathbf{s} \cdot \mathbf{b}$$

The notations $\mathbf{v}_\mathbf{s}$ will be used extensively in the sequel. Notice that for any $\mathbf{v} \in \mathbf{V}$ Eve is able to find \mathbf{s} such that $\mathbf{v}_\mathbf{s} = \mathbf{v}$ and, as a result, knows the parity $\mathbf{s} \cdot \mathbf{b}$ of the substring of \mathbf{x} determined by \mathbf{v} .

The privacy amplification string \mathbf{w} must also be linearly independent from the set of the error correction strings i.e. $\mathbf{w} \notin \mathbf{V}$. We shall show that the Hamming distances between the privacy amplification string and each string in the linear space \mathbf{V} spanned by the error correction strings $\{\mathbf{v}_1 \dots \mathbf{v}_r\}$, will provide the security parameters of the final key.

Recall that Eve holds a state for which the pure state of (19) is a purification, but she does not know \mathbf{x} . Eve learns the set of linear constraints which are imposed on the bits of \mathbf{x} . More precisely, the additional data provides the

system $\mathbf{E} = \{\mathbf{v}_1 \cdot \mathbf{x} = b_1, \mathbf{v}_2 \cdot \mathbf{x} = b_2, \dots, \mathbf{v}_r \cdot \mathbf{x} = b_r\}$ of r linear equations such that the $r+1$ strings $\mathbf{w}, \mathbf{v}_1, \dots, \mathbf{v}_r$ are linearly independent and each b_i is either 0 or 1. For the bit b (0 or 1) obtained from the privacy amplification string, we now define

$$\widetilde{\rho}_b^{\mathbf{E}, \mathbf{w}} \equiv \frac{1}{2^{n-r-1}} \sum_{\mathbf{x}: \mathbf{x} \cdot \mathbf{w} = b \wedge \mathbf{x} \mathbf{V}^T = \mathbf{b}} |\Psi_{\mathbf{x}}\rangle \langle \Psi_{\mathbf{x}}|, \quad (27)$$

an equal mixture of the states $|\Psi_{\mathbf{x}}\rangle \langle \Psi_{\mathbf{x}}|$ such that \mathbf{x} is a ‘‘code word’’, and the secret parity bit is b .

Theorem 4. —

$$\widetilde{\rho}_0^{\mathbf{E}, \mathbf{w}} - \widetilde{\rho}_1^{\mathbf{E}, \mathbf{w}} = 2 \sum_{\mathbf{s} \in \{0,1\}^r} \sum_{\mathbf{j} \in \{0,1\}^n} d_{\mathbf{j}} d_{\mathbf{j} \oplus \mathbf{w} \oplus \mathbf{v}_{\mathbf{s}}} (-1)^{\mathbf{s} \cdot \mathbf{b}} |\mathbf{j}\rangle \langle \mathbf{j} \oplus \mathbf{w} \oplus \mathbf{v}_{\mathbf{s}}|; \quad (28)$$

$$Tr|\widetilde{\rho}_0^{\mathbf{E}, \mathbf{w}} - \widetilde{\rho}_1^{\mathbf{E}, \mathbf{w}}| \leq 2 \sum_{\mathbf{s} \in \{0,1\}^r} \sum_{\mathbf{j}} d_{\mathbf{j}} d_{\mathbf{j} \oplus \mathbf{w} \oplus \mathbf{v}_{\mathbf{s}}}. \quad (29)$$

Note that the first result generalizes Eq. (21), and the second result generalizes Eq. (22), to contain the error correction data.

Using Eq. (24) and the set notation of Eq. (25) we finally get

$$Tr|\widetilde{\rho}_0^{\mathbf{E}, \mathbf{w}} - \widetilde{\rho}_1^{\mathbf{E}, \mathbf{w}}| \leq 2 \sum_{\mathbf{s} \in \{0,1\}^r} \prod_{i \in (\mathbf{w} \oplus \mathbf{v}_{\mathbf{s}})} \sin(2\alpha_i). \quad (30)$$

Therefore, we can use Theorems (1) and (2) to get

$$SD(\rho_0^{\mathbf{E}, \mathbf{w}}, \rho_1^{\mathbf{E}, \mathbf{w}}) \leq \sum_{\mathbf{s} \in \{0,1\}^r} \prod_{i \in (\mathbf{w} \oplus \mathbf{v}_{\mathbf{s}})} \sin(2\alpha_i), \quad (31)$$

when the error correction data is known to Eve.

8 A bound on Eve’s information

The above result bounds Eve’s information in terms of parameters chosen by Eve, that is, α_i . We now show that, if the test is passed, then Eve’s information is bounded by parameters chosen by Alice and Bob.

Using $\sin(2\alpha_i) \leq 2 \sin \alpha_i \leq (8p_i)^{1/2}$ we get

$$SD(\rho_0^{\mathbf{E}, \mathbf{w}}, \rho_1^{\mathbf{E}, \mathbf{w}}) \leq \sum_{\mathbf{s} \in \{0,1\}^r} \left[\prod_{i \in (\mathbf{w} \oplus \mathbf{v}_{\mathbf{s}})} (8p_i) \right]^{1/2}. \quad (32)$$

For each $\mathbf{s} \in \{0, 1\}^r$, let $\hat{n}_{\mathbf{s}} = \widehat{\mathbf{w} \oplus \mathbf{v}_{\mathbf{s}}}$, be the Hamming distance between \mathbf{w} and $\mathbf{v}_{\mathbf{s}}$. Since $n_{\mathbf{s}}$ is the number of elements in the set of indices corresponding to $\mathbf{w} \oplus \mathbf{v}_{\mathbf{s}}$, it is also the number of factors in the product $\prod_{i \in (\mathbf{w} \oplus \mathbf{v}_{\mathbf{s}})}$. Also let

$$p_{\mathbf{s}} = \left[\sum_{i \in (\mathbf{w} \oplus \mathbf{v}_{\mathbf{s}})} p_i \right] / \hat{n}_{\mathbf{s}} \quad (33)$$

be the arithmetic mean (average) of the probabilities p_i for $i \in \mathbf{w} \oplus \mathbf{v}_{\mathbf{s}}$. The geometrical mean of the elements $8p_i$ such that $i \in \mathbf{w} \oplus \mathbf{v}_{\mathbf{s}}$ is bounded above by the arithmetic mean of the same elements, i.e. $[\prod_{i \in (\mathbf{w} \oplus \mathbf{v}_{\mathbf{s}})} (8p_i)]^{1/\hat{n}_{\mathbf{s}}} \leq 8p_{\mathbf{s}}$, and consequently

$$SD(\rho_0^{\mathbf{E}, \mathbf{w}}, \rho_1^{\mathbf{E}, \mathbf{w}}) \leq \sum_{\mathbf{s} \in \{0, 1\}^r} [8p_{\mathbf{s}}]^{\hat{n}_{\mathbf{s}}/2} . \quad (34)$$

Given that the test is passed, $p_{\text{test}} \leq p_{\text{allowed}}$, statistical analysis promise us that each of the $p_{\mathbf{s}}$ is bounded; indeed, combining two laws of large numbers of Hoeffding [18], Theorem 2 in [18] (sums of independent random variables) and its extension in section 6 in [18] (sampling from a finite population), we are promised that $p_{n'}$, the average p_i of all n' relevant bits ($n' = n + n_{\text{test}}$) satisfies $\text{Prob}[p_{n'} > p_{\text{test}} + 2\delta] \leq 2e^{-2n_{\text{test}}\delta^2}$ (since the tested bits are picked at random). Moreover, $p_{\mathbf{s}} \leq (n'/\hat{n}_{\mathbf{s}})p_{n'}$. As a result we get for all $\mathbf{s} \in \{0, 1\}^r$

$$p_{\mathbf{s}} \leq (n'/n_{\mathbf{s}})(p_{\text{test}} + 2\delta) \quad (35)$$

except with probability $p_{\text{luck}} \leq 2e^{-2n_{\text{test}}\delta^2}$. Therefore, Eve's information is generously bounded by

$$SD(\rho_0^{\mathbf{E}, \mathbf{w}}, \rho_1^{\mathbf{E}, \mathbf{w}}) \leq \sum_{\mathbf{s} \in \{0, 1\}^r} [(8n'/\hat{n}_{\mathbf{s}})(p_{\text{test}} + 2\delta)]^{(\hat{n}_{\mathbf{s}}/2)} , \quad (36)$$

except with a probability of $p_{\text{luck}} = 2e^{-2n_{\text{test}}\delta^2}$.

Now, let $\hat{n} = \min_{\mathbf{s}} \hat{n}_{\mathbf{s}}$ and let n' be even [throw one bit away if needed (before choosing the bits for the test)], and choose $n = n_{\text{test}} = n'/2$. Moreover, let $\delta = p_{\text{test}}/2$ (those choices are clearly not intended to be optimal). Then

$$8p_{\mathbf{s}} \leq 32n p_{\text{test}} / \hat{n} , \quad (37)$$

except with a probability of $2e^{-np_{\text{test}}^2/2}$. Therefore, Eve's information is bounded by

$$SD(\rho_0^{\mathbf{E}, \mathbf{w}}, \rho_1^{\mathbf{E}, \mathbf{w}}) \leq \sum_{\mathbf{s} \in \{0, 1\}^r} [32n p_{\text{test}} / \hat{n}]^{(\hat{n}_{\mathbf{s}}/2)} , \quad (38)$$

except with a probability of $p_{\text{luck}} = 2e^{-np_{\text{test}}^2/2}$.

For sufficiently small error rates, there are error correcting codes which allow us to choose the parameters n and \hat{n} so that the right-hand side of (37) is smaller than 1 (which requires here that $p_{\text{test}} \leq (\hat{n}/n)(1/32) < 1.5625\%$, since \hat{n} must be less than $n/2$). Equation (34) still holds if we replace \hat{n}_s by the smaller value \hat{n} . We finally get

$$SD(\rho_0^{\mathbf{E},\mathbf{w}}, \rho_1^{\mathbf{E},\mathbf{w}}) \leq 2^r [32 n p_{\text{test}} / \hat{n}]^{\hat{n}/2}, \quad (39)$$

except with a probability of $2e^{-np_{\text{test}}^2/2}$. Again, for sufficiently small error rates, there are error correcting codes which allow us to choose the parameters n , r and \hat{n} such that Eve's information is exponentially small. This completes what we want to say about the security of a single bit final key.

We shall not calculate the error-rate threshold here, but such a threshold was calculated in the proof of security against the most general attacks possible [1]. Due to various non-optimal steps in the current paper, the threshold which can be obtained here is much worse than the one obtained in [1] (which is 7.56) for more general attacks, hence calculating the threshold for the current proof is meaningless.

9 Multi-bit final key

If $m > 1$ Alice chooses for privacy amplification m linearly independent vectors $\mathbf{w}_1, \mathbf{w}_2 \dots \mathbf{w}_m$ whose span \mathbf{W} is such that $\mathbf{V} \cap \mathbf{W} = \{\mathbf{0}\}$. The final m -bit key will be $\mathbf{k} = (k_i)_{1 \leq i \leq m}$ where $k_1 = \mathbf{x} \cdot \mathbf{w}_1, k_2 = \mathbf{x} \cdot \mathbf{w}_2, \dots, k_m = \mathbf{x} \cdot \mathbf{w}_m$. Put differently, if \mathbf{W} is the matrix whose rows are $\mathbf{w}_1, \dots, \mathbf{w}_m$, then $\mathbf{k} = \mathbf{x}\mathbf{W}^\top$.

To obtain a security parameter \hat{n} , Alice and Bob must make sure that each string in $\mathbf{W} - \{\mathbf{0}\}$ has a Hamming distance of at least \hat{n} from any string in \mathbf{V} , the linear space spanned by $\{\mathbf{v}_1, \mathbf{v}_2 \dots \mathbf{v}_r\}$. Furthermore, if Alice and Bob want Eve to be ignorant of linear combinations of the different key bits, they should be more strict, and make sure that any string in the space $\mathbf{W} - \{\mathbf{0}\}$ has at least a Hamming distance \hat{n} from any other string in the space $\mathbf{V} + \mathbf{W}$.

Then, the final formula (39) is modified by replacing 2^r by 2^{r+m-1} , and the Hamming distance \hat{n} is modified to be the minimal distance between any two vectors in the space $\mathbf{V} + \mathbf{W}$ (of dimension $r + m$ over \mathbf{F}_2).

10 Acknowledgements

We would like to thank Claude Crépeau for helpful remarks. The work of M.B., G.B., J.v-d-G., and T.M. was partially supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada, the Canada Council for the Arts, and Québec's FCAR. The work of T.M. was also partially supported by Grant No. 961360 from the Jet Propulsion Lab and Grant No. 530-1415-01 from the DARPA Ultra program. The work of E.B. and T.M. was partially supported by the European Commission through the IST Programme under contract IST-1999-11234.

11 Appendix

Proof 1. — If $\mathcal{E} = (E_x)_{x \in X}$ is a POVM on \mathcal{H}_1 then $\mathcal{E} \otimes \mathbb{I}_{\mathcal{H}_2} \equiv (E_x \otimes \mathbb{I}_{\mathcal{H}_2})_{x \in X}$ is a POVM on $\mathcal{H}_1 \otimes \mathcal{H}_2$ and $Tr_1(Tr_2(\tilde{\rho}_i)E_x) = Tr(\tilde{\rho}_i(E_x \otimes \mathbb{I}_{\mathcal{H}_2}))$. Consequently $SD^{\mathcal{E}}(Tr_2(\tilde{\rho}_0), Tr_2(\tilde{\rho}_1)) = SD^{\mathcal{E} \otimes \mathbb{I}_{\mathcal{H}_2}}(\tilde{\rho}_0, \tilde{\rho}_1)$. By definition of the optimization process $\sup_{\mathcal{E}}[SD^{\mathcal{E} \otimes \mathbb{I}_{\mathcal{H}_2}}(\tilde{\rho}_0, \tilde{\rho}_1)] \leq \sup_{\tilde{\mathcal{E}}}[SD^{\tilde{\mathcal{E}}}(\tilde{\rho}_0, \tilde{\rho}_1)]$, since $\mathcal{E} \otimes \mathbb{I}_{\mathcal{H}_2}$ is not necessarily the optimal POVM on the combined space. Thus, $SD(Tr_2(\tilde{\rho}_0), Tr_2(\tilde{\rho}_1)) \leq SD(\tilde{\rho}_0, \tilde{\rho}_1)$.

Proof 2. — Note that, for emphasizing the generality of the result, the \sim sign is removed since it is not required in the proof here. In order to prove this inequality (see also [19]) let us first fix some measurement procedure $\mathcal{E} = (E_x)_{x \in X}$. Then the resulting mutual information $SD^{\mathcal{E}}(\rho_0, \rho_1) = I(B; X) = \sum_{x \in X} p(x) I_2(p_x(0))$, where (from the Bayes formula) $p_x(0) = p(B=0)p(X=x|B=0)/p(x) = (1/2)p_0^{\mathcal{E}}(x)/p(x)$. Knowing that

$$I_2(r) \leq |2r - 1|$$

for $0 \leq r \leq 1$ we conclude that

$$SD^{\mathcal{E}}(\rho_0, \rho_1) \leq \sum_{x \in X} p(x) |2p_x(0) - 1| .$$

Assigning $p_x(0) = (1/2)p_0^{\mathcal{E}}(x)/p(x)$ into the last expression [and using $p(x) = (p_0^{\mathcal{E}}(x) + p_1^{\mathcal{E}}(x))/2$ in the following equality], we obtain

$$SD^{\mathcal{E}}(\rho_0, \rho_1) \leq \sum_{x \in X} p(x) |2[p_0^{\mathcal{E}}(x)/2p(x)] - 1|$$

$$= \frac{1}{2} \sum_{x \in X} |p_0^\mathcal{E}(x) - p_1^\mathcal{E}(x)| .$$

Now, since $\rho_0 - \rho_1$ is Hermitian, it can be diagonalized and consequently written in the form $\rho_0 - \rho_1 = \sum \lambda_j |j\rangle\langle j|$ where $|j\rangle$ is an orthonormal basis and

$$\text{Tr}|\rho_0 - \rho_1| = \sum |\lambda_j| .$$

Clearly $\text{Tr}(|j\rangle\langle j|E_x) = \langle j|E_x|j\rangle$ and so

$$\begin{aligned} p_0^\mathcal{E}(x) - p_1^\mathcal{E}(x) &= \text{Tr}((\rho_0 - \rho_1)E_x) \\ &= \sum_j \lambda_j \langle j|E_x|j\rangle . \end{aligned}$$

Using the last expression for SD and using $\langle j|E_x|j\rangle \geq 0$ (since E_x is positive definite), we can now deduce

$$\begin{aligned} SD^\mathcal{E}(\rho_0, \rho_1) &\leq \frac{1}{2} \sum_j |\lambda_j| \sum_{x \in X} \langle j|E_x|j\rangle \\ &= \frac{1}{2} \text{Tr}|\rho_0 - \rho_1| . \end{aligned}$$

Since \mathcal{E} is arbitrary, we choose the one which optimizes SD and this concludes the proof.

Proof 3. — For convenience let us define $\Delta^\mathbf{w} \equiv \widetilde{\rho}_0^\mathbf{w} - \widetilde{\rho}_1^\mathbf{w}$. Using $(-1)^b = (-1)^{\mathbf{x} \cdot \mathbf{w}}$ and for the states given by Eqs(20,19), we get

$$\begin{aligned} \Delta^\mathbf{w} &= (-1)^0 \widetilde{\rho}_0^\mathbf{w} + (-1)^1 \widetilde{\rho}_1^\mathbf{w} \\ &= 2^{-n+1} \sum_{\mathbf{j}, \mathbf{k}} d_{\mathbf{j}} d_{\mathbf{k}} \sum_{\mathbf{x}} (-1)^{\mathbf{x} \cdot (\mathbf{j} \oplus \mathbf{k} \oplus \mathbf{w})} |\mathbf{j}\rangle\langle \mathbf{k}| . \end{aligned}$$

We now simplify the preceding sum using a technique similar to the one of [9]. If $\mathbf{j} \oplus \mathbf{k} \oplus \mathbf{w} \neq \mathbf{0}$, there is some string \mathbf{y} such that $(\mathbf{j} \oplus \mathbf{k} \oplus \mathbf{w}) \cdot \mathbf{y} = 1$. If we let $\mathbf{x}' = \mathbf{x} \oplus \mathbf{y}$ then $(-1)^{\mathbf{x}' \cdot (\mathbf{j} \oplus \mathbf{k} \oplus \mathbf{w})} + (-1)^{\mathbf{x} \cdot (\mathbf{j} \oplus \mathbf{k} \oplus \mathbf{w})} = 0$ and since $\mathbf{x} \neq \mathbf{x}'$ (because $\mathbf{y} \neq \mathbf{0}$) all the coefficients of $|\mathbf{j}\rangle\langle \mathbf{k}|$ cancel in pairs. If $\mathbf{j} \oplus \mathbf{k} \oplus \mathbf{w} = \mathbf{0}$, then $(-1)^{\mathbf{x} \cdot (\mathbf{j} \oplus \mathbf{k} \oplus \mathbf{w})} = 1$ for all \mathbf{x} and since there are 2^n such strings, we get

$$\begin{aligned} \Delta^\mathbf{w} &= 2 \sum_{\mathbf{i} \oplus \mathbf{j} = \mathbf{w}} d_{\mathbf{i}} d_{\mathbf{j}} |\mathbf{i}\rangle\langle \mathbf{j}| \\ &= 2 \sum_{\mathbf{j} \in \{0,1\}^n} d_{\mathbf{j}} d_{\mathbf{j} \oplus \mathbf{w}} |\mathbf{j}\rangle\langle \mathbf{j} \oplus \mathbf{w}| , \end{aligned}$$

proving the first part of the theorem.

If $\mathbf{i} \oplus \mathbf{j} = \mathbf{w}$ then clearly $\mathbf{j} \oplus \mathbf{i} = \mathbf{w}$. Therefore, $\Delta^{\mathbf{w}}$ is a sum of 2^{n-1} Hermitian ($2^n \times 2^n$ -dimensional) matrices $d_i d_j |\mathbf{i}\rangle\langle \mathbf{j}| + d_j d_i |\mathbf{j}\rangle\langle \mathbf{i}| = d_i d_j [|\mathbf{i}\rangle\langle \mathbf{j}| + |\mathbf{j}\rangle\langle \mathbf{i}|]$. For each of them the Trace-norm is $2d_i d_j = d_i d_j + d_j d_i$. Using this result and the triangle inequality (which is satisfied by any norm) we obtain

$$\text{Tr}|\Delta^{\mathbf{w}}| \leq 2 \sum_{\mathbf{i} \oplus \mathbf{j} = \mathbf{w}} d_i d_j = 2 \sum_{\mathbf{j}} d_j d_{\mathbf{j} \oplus \mathbf{w}} ,$$

proving the second part of the theorem.

Proof 4. — The first steps here follow similar steps in the previous Proof. Let us define $\Delta^{\mathbf{E}, \mathbf{w}} \equiv \widetilde{\rho}_0^{\mathbf{E}, \mathbf{w}} - \widetilde{\rho}_1^{\mathbf{E}, \mathbf{w}}$. Using $(-1)^b = (-1)^{\mathbf{x} \cdot \mathbf{w}}$ and for the states given by Eqs(27,19), we get

$$\begin{aligned} \Delta^{\mathbf{E}, \mathbf{w}} &= (-1)^0 \widetilde{\rho}_0^{\mathbf{E}, \mathbf{w}} + (-1)^1 \widetilde{\rho}_1^{\mathbf{E}, \mathbf{w}} \\ &= 2^{-n+r+1} \sum_{\mathbf{j}, \mathbf{k}} d_j d_k \sum_{\mathbf{x}: \mathbf{xV}^T = \mathbf{b}} (-1)^{\mathbf{x} \cdot (\mathbf{j} \oplus \mathbf{k} \oplus \mathbf{w})} |\mathbf{j}\rangle\langle \mathbf{k}| . \end{aligned}$$

For any $\mathbf{s} \in \{0, 1\}^r$ we let \mathbf{v}_s denote the element $\sum_{i=1}^r s_i \mathbf{v}_i$. Clearly $\mathbf{v}_s \in \mathbf{V}$. We first simplify the expression for $\Delta^{\mathbf{E}, \mathbf{w}}$. If $\mathbf{xV}^T = \mathbf{b}$, i.e. \mathbf{x} is a solution of the system \mathbf{E} , then $\mathbf{x} \cdot \mathbf{v}_s = \mathbf{s} \cdot \mathbf{b}$.

If $\mathbf{j} \oplus \mathbf{k} \oplus \mathbf{w} = \mathbf{v}_s$ and $\mathbf{xV}^T = \mathbf{b}$, then the exponent $\mathbf{x} \cdot (\mathbf{j} \oplus \mathbf{k} \oplus \mathbf{w})$ in the above expression for $\Delta^{\mathbf{E}, \mathbf{w}}$ reduces to $\mathbf{x} \cdot \mathbf{v}_s = \mathbf{s} \cdot \mathbf{b}$. This value is independent of \mathbf{x} and so the coefficient of $|\mathbf{j}\rangle\langle \mathbf{k}| = |\mathbf{j}\rangle\langle \mathbf{j} \oplus \mathbf{w} \oplus \mathbf{v}_s|$ is now $2d_j d_{\mathbf{j} \oplus \mathbf{w} \oplus \mathbf{v}_s} (-1)^{\mathbf{s} \cdot \mathbf{b}}$. If $\mathbf{j} \oplus \mathbf{k} \oplus \mathbf{w}$ is not in the span \mathbf{V} of $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ then there is a solution \mathbf{y} to the system $\{(\mathbf{j} \oplus \mathbf{k} \oplus \mathbf{w}) \cdot \mathbf{y} = 1, \mathbf{v}_1 \cdot \mathbf{y} = 0, \dots, \mathbf{v}_r \cdot \mathbf{y} = 0\}$. For any \mathbf{x} solution of \mathbf{E} , let \mathbf{x}' denote $\mathbf{x} \oplus \mathbf{y}$. Clearly \mathbf{x}' is also a solution of \mathbf{E} and $(-1)^{\mathbf{x} \cdot (\mathbf{j} \oplus \mathbf{k} \oplus \mathbf{w})} + (-1)^{\mathbf{x}' \cdot (\mathbf{j} \oplus \mathbf{k} \oplus \mathbf{w})} = 0$, and consequently the coefficient of $|\mathbf{j}\rangle\langle \mathbf{k}|$ is 0, proving

$$\Delta^{\mathbf{E}, \mathbf{w}} = 2 \sum_{\mathbf{j}, \mathbf{s}} d_j d_{\mathbf{j} \oplus \mathbf{w} \oplus \mathbf{v}_s} (-1)^{\mathbf{s} \cdot \mathbf{b}} |\mathbf{j}\rangle\langle \mathbf{j} \oplus \mathbf{w} \oplus \mathbf{v}_s| ,$$

the first part of Theorem 4.

Consequently, Eq. (21) in Theorem 3 can be used to write

$$\Delta^{\mathbf{E}, \mathbf{w}} = \sum_{\mathbf{s} \in \{0, 1\}^r} (-1)^{\mathbf{s} \cdot \mathbf{b}} (\widetilde{\rho}_0^{\mathbf{w} \oplus \mathbf{v}_s} - \widetilde{\rho}_1^{\mathbf{w} \oplus \mathbf{v}_s}) ,$$

As in the proof of Theorem 3, we define $\Delta^{\mathbf{w} \oplus \mathbf{v}_s}$ for the terms in the parenthesis. We obtain $\Delta^{\mathbf{E}, \mathbf{w}} = \sum_{\mathbf{s} \in \{0,1\}^r} (-1)^{\mathbf{s} \cdot \mathbf{b}} (\Delta^{\mathbf{w} \oplus \mathbf{v}_s})$, and due to the triangle inequality

$$\text{Tr}|\Delta^{\mathbf{E}, \mathbf{w}}| \leq \sum_{\mathbf{s} \in \{0,1\}^r} \text{Tr}|\Delta^{\mathbf{w} \oplus \mathbf{v}_s}| .$$

The Trace-Norm of these terms is given by Eq. (22) once \mathbf{w} there is replaced by $\mathbf{w} \oplus \mathbf{v}_s$. Now, using the set notation we finally get

$$\text{Tr}|\Delta^{\mathbf{E}, \mathbf{w}}| \leq 2 \sum_{\mathbf{s} \in \{0,1\}^r} \sum_{\mathbf{j}} d_{\mathbf{j}} d_{\mathbf{j} \oplus \mathbf{w} \oplus \mathbf{v}_s} , \quad (40)$$

proving the second part of Theorem 4.

References

- [1] E. Biham, M. Boyer, P. O. Boykin, T. Mor, and V. Roychowdhury, A Proof of Security of Quantum Key Distribution Against Any Attack, in *Proceedings of the 32'nd Ann. ACM Symposium on the Theory of Computing (STOC'00)*, ACM Press, 2000, pp. 715–724. The full paper was submitted to Jour. of Cryptology. See also the Los Alamos archive Quant-ph 9912053.
- [2] S. Wiesner, Conjugate coding, *Sigact News*, **15**(1983), 77–88.
- [3] C. H. Bennett and G. Brassard, Quantum cryptography: Public key distribution and coin tossing, in *Proceedings of the IEEE International Conference on Computers, Systems and Signal Processing, Bangalore, India*, IEEE Press, New York, 1984, pp. 175–179.
- [4] C. H. Bennett, G. Brassard, and A. Ekert, Quantum cryptography, *Sci. Amer.*, (Oct. 1992), 50–57.
- [5] A. Ekert, Quantum cryptography based on Bell's theorem, *Phys. Rev. Lett.*, **67**(1991), 661–663.
- [6] C. H. Bennett, Quantum cryptography using any two nonorthogonal states, *Phys. Rev. Lett.*, **68**(1992), 3121–3124.

- [7] G. Brassard, N. Lütkenhaus, T. Mor, and B. C. Sanders, Limitations of practical quantum cryptography, *Phys. Rev. Lett.*, **85**(2000), 1330–1333.
- [8] C. H. Bennett, T. Mor, and J. A. Smolin, Parity bit in quantum cryptography, *Phys. Rev. A*, **54**(1996), 2675–2684.
- [9] E. Biham and T. Mor, Security of quantum cryptography against collective attacks, *Phys. Rev. Lett.*, **78**(1997), 2256–2259.
- [10] E. Biham and T. Mor, Bounds on Information and the Security of Quantum Cryptography, *Phys. Rev. Lett.*, **79**(1997), 4034–4037.
- [11] H.-K. Lo and H. F. Chau, Unconditional security of quantum key distribution over arbitrarily long distances, *Science*, **283**(1999), 2050–2056.
- [12] D. Mayers, Unconditional security in quantum cryptography. Quant-ph/9802025 (a very preliminary version of Mayers’ proof appeared in [14]), 1998.
- [13] P. W. Shor and J. Preskill, Simple Proof of Security of the BB84 Quantum Key Distribution Protocol, *Phys. Rev. Lett.*, **85**(2000), 441–444 (also Quant-ph/0003004).
- [14] D. Mayers, Quantum key distribution and string oblivious transfer in noisy channel, in *Advances in cryptology - CRYPTO’96*, LNCS 1109, Springer-Verlag, Berlin, 1996, pp. 343–357.
- [15] A. Peres, *Quantum Theory: Concepts and Methods*, Kluwer Academic Publishers, Dordrecht, 1993.
- [16] C. A. Fuchs, *Distinguishability and Accessible Information in Quantum Theory*, Quant-ph 9601020, PhD thesis, University of New Mexico, 1995.
- [17] J. van de Graaf, *Towards a formal definition of security for quantum protocols*, PhD thesis, DIRO, Université de Montréal, 1997.
- [18] W. Hoeffding, Probability inequalities for sums of bounded random variables, *J. Amer. Stat. Assoc.*, **58**(1963), 13–20.

- [19] C. A. Fuchs and J. van de Graaf, Cryptographic distinguishability measures for quantum-mechanical states, *IEEE Trans. Inform. Theory*, **45**(1999), 1216–1227.