Security Against Collective Attacks of a Modified BB84 QKD Protocol with Information only in One Basis

Michel Boyer[1], Rotem Liss[2] and Tal Mor[2]
[1]*Département IRO, Université de Montréal, Montréal (Québec) H3C 3J7, Canada*
[2]*Computer Science Department, Technion, Haifa 3200003, Israel*

Keywords: QKD, Security, Collective Attacks, Error Rate Threshold, Information Bits.

Abstract: The Quantum Key Distribution (QKD) protocol BB84 has been proven secure against several important types of attacks: the collective attacks and the joint attacks. Here we analyze the security of a modified BB84 protocol, for which information is sent only in the z basis while testing is done in both the z and the x bases, against collective attacks. The proof follows the framework of a previous paper (Boyer et al., 2009), but it avoids the classical information-theoretical analysis that caused problems with composability. We show that this modified BB84 protocol is as secure against collective attacks as the original BB84 protocol, and that it requires more bits for testing.

1 INTRODUCTION

Quantum Key Distribution (QKD) protocols take advantage of the laws of quantum mechanics, and most of them can be proven secure even against powerful adversaries limited only by the laws of physics. The two parties (Alice and Bob) want to create a shared random key, using an insecure quantum channel and an unjammable classical channel (to which the adversary may listen, but not interfere). The adversary (eavesdropper), Eve, tries to get as much information as she can on the final shared key. The first and most important QKD protocol is BB84 (Bennett and Brassard, 1984).

Boyer, Gelles, and Mor (BGM09) (Boyer et al., 2009) discussed the security of the BB84 protocol against collective attacks. Collective attacks (Biham and Mor, 1997b; Biham and Mor, 1997a; Biham et al., 2002) are a subclass of the joint attacks; joint attacks are the most powerful theoretical attacks. BGM09 improved the security proof of Biham, Boyer, Brassard, van de Graaf, and Mor (BBBGM02) (Biham et al., 2002) against collective attacks, by using some techniques of Biham, Boyer, Boykin, Mor, and Roychowdhury (BBBMR06) (Biham et al., 2006) (that proved security against joint attacks). In this paper, too, we restrict the analysis to collective attacks, because security against collective attacks is conjectured (and, in some security notions, proved (Renner, 2008; Christandl et al., 2009)) to imply security against joint attacks. In addition, proving security against collective attacks is much simpler than proving security against joint attacks.

In many QKD protocols, including BB84, Alice and Bob exchange several types of bits (encoded as quantum systems, usually qubits): INFO bits, that are secret bits shared by Alice and Bob and are used for generating the final key (via classical processes of error correction and privacy amplification); and TEST bits, that are publicly exposed by Alice and Bob (by using the classical channel) and are used for estimating the error rate. In BB84, each bit is sent from Alice to Bob in a random basis (the z basis or the x basis).

In this paper, we extend the analysis of BB84 done in BGM09 and prove the security of a QKD protocol we shall name *BB84-INFO-z*. This protocol is almost identical to BB84, except that all its INFO bits are in the z basis. In other words, the x basis is used only for testing. The bits are thus partitioned into three disjoint sets: INFO, TEST-Z, and TEST-X. The sizes of these sets are arbitrary (n INFO bits, n_z TEST-Z bits, and n_x TEST-X bits).

We note that, while this paper follows a line of research that mainly discusses a specific approach of security proof for BB84 and similar protocols (this approach, notably, considers finite-key effects and not only the asymptotic error rate), many other approaches have also been suggested: see for example (Mayers, 2001; Shor and Preskill, 2000; Renner, 2008; Renner et al., 2005).

Boyer, M., Liss, R. and Mor, T.
Security Against Collective Attacks of a Modified BB84 QKD Protocol with Information only in One Basis.
DOI: 10.5220/0006241000230029
In *Proceedings of the 2nd International Conference on Complexity, Future Information Systems and Risk (COMPLEXIS 2017)*, pages 23-29
ISBN: 978-989-758-244-8

In contrast to the line of research adopted here (of (Biham and Mor, 1997b; Biham and Mor, 1997a; Biham et al., 2002; Biham et al., 2006; Boyer et al., 2009)), in which a classical information-theoretical analysis caused problems with composability (see definition in (Renner, 2008)), in this paper we suggest a method to avoid those problems: we calculate the trace distance between any two density matrices Eve may hold, instead of calculating the classical mutual information between Eve and the final key (as done in those previous papers). This method is implemented in this paper for the proof of BB84-INFO-z; it also directly applies to the BB84 security proof in BGM09, and it may be extended in the future to show that the BB84 security proofs of BGM09, BBBGM02, and BBBMR06 prove the composable security of BB84.

The "qubit space", $\mathcal{H}_2$, is a 2-dimensional Hilbert space. The states $|0^0\rangle, |1^0\rangle$ form an orthonormal basis of $\mathcal{H}_2$, called "the computational basis" or "the z basis". The states $|0^1\rangle \triangleq \frac{1}{\sqrt{2}}[|0^0\rangle + |1^0\rangle]$ and $|1^1\rangle \triangleq \frac{1}{\sqrt{2}}[|0^0\rangle - |1^0\rangle]$ form another orthonormal basis of $\mathcal{H}_2$, called "the x basis". Those two bases are said to be *conjugate bases*.

In this paper, bit strings of some length t are denoted by a bold letter (e.g., $\mathbf{i} = i_1 \dots i_t$ with $i_1, \dots, i_t \in \{0,1\}$) and are identified to elements of the t-dimensional $\mathbf{F}_2$-vector space $\mathbf{F}_2^t$, where $\mathbf{F}_2 = \{0,1\}$ and the addition of two vectors corresponds to a XOR operation. The number of 1-bits in a bit string $\mathbf{s}$ is denoted by $|\mathbf{s}|$, and the Hamming distance between two strings $\mathbf{s}$ and $\mathbf{s}'$ is $d_H(\mathbf{s},\mathbf{s}') = |\mathbf{s}+\mathbf{s}'|$.

2 FORMAL DESCRIPTION OF THE BB84-INFO-Z PROTOCOL

Below we describe the BB84-INFO-z protocol used in this paper.

1. Alice and Bob pre-agree on numbers n, n_z, and n_x (we denote $N \triangleq n + n_z + n_x$), on error thresholds $p_{a,z}$ and $p_{a,x}$, on a linear error-correcting code C with an $r \times n$ parity check matrix P_C, and on a linear key-generation function (privacy amplification) represented by an $m \times n$ matrix P_K. It is required that *all* the $r+m$ rows of the matrices P_C and P_K put together are linearly independent.
2. Alice randomly chooses a partition $\mathcal{P} = (\mathbf{s},\mathbf{z},\mathbf{b})$ of the N bits by randomly choosing three N-bit strings $\mathbf{s},\mathbf{z},\mathbf{b} \in \mathbf{F}_2^N$ that satisfy $|\mathbf{s}| = n, |\mathbf{z}| = n_z, |\mathbf{b}| = n_x$, and $|\mathbf{s}+\mathbf{z}+\mathbf{b}| = N$. $\mathcal{P}$ thus partitions the set of indexes $\{1,2,...,N\}$ into three disjoint sets:
   - I (INFO bits, where $s_j = 1$) of size n;
   - T_Z (TEST-Z bits, where $z_j = 1$) of size n_z; and
   - T_X (TEST-X bits, where $b_j = 1$) of size n_x.
3. Alice randomly chooses an N-bit string $\mathbf{i} \in \mathbf{F}_2^N$, and sends the N qubit states $|i_1^{b_1}\rangle, |i_2^{b_2}\rangle, \dots, |i_N^{b_N}\rangle$, one after the other, to Bob using the quantum channel. Notice that the INFO and TEST-Z bits are encoded in the z basis, while the TEST-X bits are encoded in the x basis. Bob keeps each received qubit in quantum memory, not measuring it yet[1].
4. Alice publicly sends to Bob the string $\mathbf{b} = b_1 \cdots b_N$. Bob measures each saved qubit in the correct basis (namely, if $b_i = 0$ then he measures the i-th qubit in the z basis, and if $b_i = 1$ then he measures it in the x basis).

   The bit string measured by Bob is denoted by $\mathbf{i}^B$. If there is no noise and no eavesdropping, then $\mathbf{i}^B = \mathbf{i}$.
5. Alice publicly sends to Bob the string $\mathbf{s}$. The INFO bits, used for generating the final key, are the n bits with $s_j = 1$, while the TEST-Z and TEST-X bits are the $n_z + n_x$ bits with $s_j = 0$. The substrings of $\mathbf{i}, \mathbf{b}$ that correspond to the INFO bits are denoted by $\mathbf{i_s}$ and $\mathbf{b_s}$.
6. Alice and Bob both publish their values of all the TEST-Z and TEST-X bits, and compare the bit values. If more than $n_z \cdot p_{a,z}$ of the TEST-Z bits are different between Alice and Bob *or* more than $n_x \cdot p_{a,x}$ of the TEST-X bits are different between them, they abort the protocol. We note that $p_{a,z}$ and $p_{a,x}$ (the pre-agreed error thresholds) are the maximal allowed error rates on the TEST-Z and TEST-X bits, respectively – namely, in each basis (z and x) separately.
7. Alice and Bob keep the values of the remaining n bits (the INFO bits, with $s_j = 1$) secret. The bit string of Alice is denoted $\mathbf{x} = \mathbf{i_s}$, and the bit string of Bob is denoted $\mathbf{x}^B$.
8. Alice sends to Bob the r-bit string $\xi = \mathbf{x}P_C^{\mathrm{T}}$, that is called the *syndrome* of $\mathbf{x}$ (with respect to the error-correcting code C and to its corresponding parity check matrix P_C). By using ξ, Bob corrects the errors in his $\mathbf{x}^B$ string (so that it is the same as $\mathbf{x}$).

[1] Here we assume that Bob has a quantum memory and can delay his measurement. In practical implementations, Bob usually cannot do that, but is assumed to measure in a randomly-chosen basis (z or x), so that Alice and Bob later discard the qubits measured in the wrong basis. We assume that Alice sends more than N qubits, so that N qubits are finally detected by Bob and measured in the correct basis.

9. Alice and Bob compute the m-bit final key $\mathbf{k} = \mathbf{x}P_K^{\mathrm{T}}$.

The protocol is defined similarly to BB84 (and to its description in BGM09), except that it uses the generalized bit numbers n, n_z, and n_x (numbers of INFO, TEST-Z, and TEST-X bits, respectively); that it uses the partition $\mathscr{P} = (\mathbf{s}, \mathbf{z}, \mathbf{b})$ for dividing the N-bit string $\mathbf{i}$ into three disjoint sets of indexes (I, T_Z, and T_X); and that it uses two separate thresholds ($p_{a,z}$ and $p_{a,x}$) instead of one (p_a).

3 SECURITY PROOF OF BB84-INFO-Z AGAINST COLLECTIVE ATTACKS

3.1 Results from BGM09

The security proof of BB84-INFO-z against collective attacks is very similar to the security proof of BB84 itself against collective attacks, that was detailed in BGM09. Most parts of the proof are not affected at all by the changes made to BB84 to get the BB84-INFO-z protocol (changes detailed in Section 2 of the current paper), because those parts assume fixed strings $\mathbf{s}$ and $\mathbf{b}$, and because the attack is collective (so the analysis is restricted to the INFO bits).

Therefore, the reader is referred to the proof in Section 2 and Subsections 3.1 to 3.5 of BGM09, that applies to BB84-INFO-z without any changes (except changing the total number of bits, $2n$, to N, which does not affect the proof at all), and that will not be repeated here.

We denote the rows of the error-correction parity check matrix P_C as the vectors $v_1, \ldots, v_r$ in $\mathbf{F}_2^n$, and the rows of the privacy amplification matrix P_K as the vectors $v_{r+1}, \ldots, v_{r+m}$. We also define, for every r', $V_{r'} \triangleq \mathrm{Span}\{v_1, ..., v_{r'}\}$; and we define

$$d_{r,m} \triangleq \min_{r \le r' < r+m} d_H(v_{r'+1}, V_{r'}) = \min_{r \le r' < r+m} d_{r',1}. \quad (1)$$

For a 1-bit final key $k \in \{0,1\}$, we define $\widehat{\rho}_k$ to be the state of Eve corresponding to the final key k, given that she knows ξ. Thus,

$$\widehat{\rho}_k = \frac{1}{2^{n-r-1}} \sum_{\mathbf{x} \,\big|\, \substack{\mathbf{x}P_C^T = \xi \\ \mathbf{x} \cdot v_{r+1} = k}} \rho_{\mathbf{x}}^{\mathbf{b}'}, \quad (2)$$

where $\rho_{\mathbf{x}}^{\mathbf{b}'}$ is Eve's state after the attack, given that Alice sent the INFO bits $\mathbf{x}$ encoded in the bases $\mathbf{b}' = \mathbf{b_s}$. We also defined in BGM09 the state $\widetilde{\rho}_k$, that is a lift-up of $\widehat{\rho}_k$ (which means that $\widehat{\rho}_k$ is a partial trace of $\widetilde{\rho}_k$).

In the end of Subsection 3.5 of BGM09, it was found that (in the case of a 1-bit final key, i.e., $m = 1$)

$$\frac{1}{2}\,\mathrm{tr}\,|\widetilde{\rho}_0 - \widetilde{\rho}_1| \le 2\sqrt{P\left[|\mathbf{C}_I| \ge \frac{d_{r,1}}{2} \;\mid\; \mathbf{B}_I = \overline{\mathbf{b}'}, \mathbf{s}\right]}, \quad (3)$$

where $\mathbf{C}_I$ is the random variable corresponding to the n-bit string of errors on the n INFO bits; $\mathbf{B}_I$ is the random variable corresponding to the n-bit string of bases of the n INFO bits; $\overline{\mathbf{b}'}$ is the bit-flipped string of $\mathbf{b}' = \mathbf{b_s}$; and $d_{r,1}$ (and, in general, $d_{r,m}$) was defined above.

Now, according to (Nielsen and Chuang, 2010, Theorem 9.2 and page 407), and using the fact that $\widehat{\rho}_k$ is a partial trace of $\widetilde{\rho}_k$, we find that $\frac{1}{2}\,\mathrm{tr}\,|\widehat{\rho}_0 - \widehat{\rho}_1| \le \frac{1}{2}\,\mathrm{tr}\,|\widetilde{\rho}_0 - \widetilde{\rho}_1|$. From this result and from inequality (3) we deduce that

$$\frac{1}{2}\,\mathrm{tr}\,|\widehat{\rho}_0 - \widehat{\rho}_1| \le 2\sqrt{P\left[|\mathbf{C}_I| \ge \frac{d_{r,1}}{2} \;\mid\; \mathbf{B}_I = \overline{\mathbf{b}'}, \mathbf{s}\right]}. \quad (4)$$

3.2 Bounding the Differences Between Eve's States

We define $\mathbf{c} \triangleq \mathbf{i} + \mathbf{i}^B$: namely, $\mathbf{c}$ is the XOR of the N-bit string $\mathbf{i}$ sent by Alice and of the N-bit string $\mathbf{i}^B$ measured by Bob. For each index $1 \le l \le N$, $c_l = 1$ if and only if Bob's l-th bit value is different from the l-th bit sent by Alice. The partition $\mathscr{P}$ divides the N bits into n INFO bits, n_z TEST-Z bits, and n_x TEST-X bits. The corresponding substrings of the error string $\mathbf{c}$ are $\mathbf{c_s}$ (the string of errors on the INFO bits), $\mathbf{c_z}$ (the string of errors on the TEST-Z bits), and $\mathbf{c_b}$ (the string of errors on the TEST-X bits). The random variables that correspond to $\mathbf{c_s}$, $\mathbf{c_z}$, and $\mathbf{c_b}$ are denoted by $\mathbf{C}_I$, $\mathbf{C}_{T_Z}$, and $\mathbf{C}_{T_X}$, respectively.

We define $\widetilde{\mathbf{C}_I}$ to be the random variable corresponding to the string of errors on the INFO bits *if Alice had encoded and sent the INFO bits in the x basis* (instead of the z basis dictated by the protocol). In those notations, inequality (4) reads as

$$\frac{1}{2}\,\mathrm{tr}\,|\widehat{\rho}_0 - \widehat{\rho}_1| \le 2\sqrt{P\left[|\widetilde{\mathbf{C}_I}| \ge \frac{d_{r,1}}{2} \;\mid\; \mathscr{P}\right]}$$
$$= 2\sqrt{P\left[|\widetilde{\mathbf{C}_I}| \ge \frac{d_{r,1}}{2} \;\mid\; \mathbf{c_z}, \mathbf{c_b}, \mathscr{P}\right]}, \quad (5)$$

using the fact that Eve's attack is collective, so the qubits are attacked independently, and, therefore, the errors on the INFO bits are independent of the errors on the TEST-Z and TEST-X bits (namely, of $\mathbf{c_z}$ and $\mathbf{c_b}$).

As described in BGM09, inequality (5) was not derived for the actual attack $U = U_1 \otimes \ldots \otimes U_N$ applied by Eve, but for a virtual flat attack (that depends on **b** and therefore could not have been applied by Eve). That flat attack gives the same states $\widehat{\rho}_0$ and $\widehat{\rho}_1$ as the original attack U, and gives a lower (or the same) error rate in the conjugate basis. Therefore, inequality (5) also holds for the original attack U. This means that, from now on, all our results apply to the original attack U and not the flat attack.

So far, we have discussed a 1-bit key. We will now discuss a general m-bit key $\mathbf{k}$. We define $\widehat{\rho}_{\mathbf{k}}$ to be the state of Eve corresponding to the final key $\mathbf{k}$, given that she knows ξ:

$$\widehat{\rho}_{\mathbf{k}} = \frac{1}{2^{n-r-m}} \sum_{\mathbf{x} \left| \substack{\mathbf{x}P_C^T = \xi \\ \mathbf{x}P_K^T = \mathbf{k}} \right.} \rho_{\mathbf{x}}^{\mathbf{b}'} \tag{6}$$

Proposition 1. *For any two m-bit keys* $\mathbf{k}, \mathbf{k}'$,

$$\frac{1}{2}\operatorname{tr}|\widehat{\rho}_{\mathbf{k}} - \widehat{\rho}_{\mathbf{k}'}| \leq 2m\sqrt{P\left[|\widetilde{\mathbf{C}}_I| \geq \frac{d_{r,m}}{2} \mid \mathbf{c_z}, \mathbf{c_b}, \mathscr{P}\right]}. \tag{7}$$

Proof. We define the key $\mathbf{k}_j$, for $0 \leq j \leq m$, to consist of the first j bits of $\mathbf{k}'$ and the last $m-j$ bits of $\mathbf{k}$. This means that $\mathbf{k}_0 = \mathbf{k}$, $\mathbf{k}_m = \mathbf{k}'$, and $\mathbf{k}_{j-1}$ differs from $\mathbf{k}_j$ at most on a single bit (the j-th bit).

First, we find a bound on $\frac{1}{2}\operatorname{tr}|\widehat{\rho}_{\mathbf{k}_{j-1}} - \widehat{\rho}_{\mathbf{k}_j}|$: since $\mathbf{k}_{j-1}$ differs from $\mathbf{k}_j$ at most on a single bit (the j-th bit, given by the formula $\mathbf{x} \cdot v_{r+j}$), we can use the same proof that gave us inequality (5), attaching the other (identical) key bits to ξ of the original proof; and we find that:

$$\frac{1}{2}\operatorname{tr}|\widehat{\rho}_{\mathbf{k}_{j-1}} - \widehat{\rho}_{\mathbf{k}_j}| \leq 2\sqrt{P\left[|\widetilde{\mathbf{C}}_I| \geq \frac{d_j}{2} \mid \mathbf{c_z}, \mathbf{c_b}, \mathscr{P}\right]} \tag{8}$$

where we define d_j as $d_H(v_{r+j}, V_j')$, and $V_j' \triangleq \operatorname{Span}\{v_1, v_2, \ldots, v_{r+j-1}, v_{r+j+1}, \ldots, v_{r+m}\}$.

Now we notice that d_j is the Hamming distance between v_{r+j} and some vector in V_j', which means that $d_j = |\sum_{i=1}^{r+m} a_i v_i|$ with $a_i \in \mathbf{F}_2$ and $a_{r+j} \neq 0$. The properties of Hamming distance assure us that d_j is at least $d_H(v_{r'+1}, V_{r'})$ for some $r \leq r' < r+m$. Therefore, we find that $d_{r,m} = \min_{r \leq r' < r+m} d_H(v_{r'+1}, V_{r'}) \leq d_j$.

The result $d_{r,m} \leq d_j$ implies that if $|\widetilde{\mathbf{C}}_I| \geq \frac{d_j}{2}$ then $|\widetilde{\mathbf{C}}_I| \geq \frac{d_{r,m}}{2}$. Therefore, inequality (8) implies

$$\frac{1}{2}\operatorname{tr}|\widehat{\rho}_{\mathbf{k}_{j-1}} - \widehat{\rho}_{\mathbf{k}_j}| \leq 2\sqrt{P\left[|\widetilde{\mathbf{C}}_I| \geq \frac{d_{r,m}}{2} \mid \mathbf{c_z}, \mathbf{c_b}, \mathscr{P}\right]}. \tag{9}$$

Now we use the triangle inequality for norms to find

$$\begin{aligned}\frac{1}{2}\operatorname{tr}|\widehat{\rho}_{\mathbf{k}} - \widehat{\rho}_{\mathbf{k}'}| &= \frac{1}{2}\operatorname{tr}|\widehat{\rho}_{\mathbf{k}_0} - \widehat{\rho}_{\mathbf{k}_m}| \leq \sum_{j=1}^{m} \frac{1}{2}\operatorname{tr}|\widehat{\rho}_{\mathbf{k}_{j-1}} - \widehat{\rho}_{\mathbf{k}_j}| \\ &\leq 2m\sqrt{P\left[|\widetilde{\mathbf{C}}_I| \geq \frac{d_{r,m}}{2} \mid \mathbf{c_z}, \mathbf{c_b}, \mathscr{P}\right]}.\end{aligned} \tag{10}$$

□

The value we want to bound is the expected value of difference between two states of Eve corresponding to two final keys. However, we should take into account that if the test fails, no final key is generated, and the difference between all of Eve's states becomes 0 for any purpose. We thus define the random variable $\Delta_{\text{Eve}}^{(p_{a,z}, p_{a,x})}(\mathbf{k}, \mathbf{k}')$ for any two final keys $\mathbf{k}, \mathbf{k}'$:

$$\Delta_{\text{Eve}}^{(p_{a,z}, p_{a,x})}(\mathbf{k}, \mathbf{k}' | \mathscr{P}, \xi, \mathbf{c_z}, \mathbf{c_b}) \triangleq \begin{cases} \frac{1}{2}\operatorname{tr}|\widehat{\rho}_{\mathbf{k}} - \widehat{\rho}_{\mathbf{k}'}| & \text{if } \frac{|\mathbf{c_z}|}{n_z} \leq p_{a,z} \text{ and } \frac{|\mathbf{c_b}|}{n_x} \leq p_{a,x} \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

We need to bound the expected value $\langle \Delta_{\text{Eve}}^{(p_{a,z}, p_{a,x})}(\mathbf{k}, \mathbf{k}') \rangle$, that is given by:

$$\langle \Delta_{\text{Eve}}^{(p_{a,z}, p_{a,x})}(\mathbf{k}, \mathbf{k}') \rangle = \sum_{\mathscr{P}, \xi, \mathbf{c_z}, \mathbf{c_b}} \Delta_{\text{Eve}}^{(p_{a,z}, p_{a,x})}(\mathbf{k}, \mathbf{k}' | \mathscr{P}, \xi, \mathbf{c_z}, \mathbf{c_b}) \cdot p(\mathscr{P}, \xi, \mathbf{c_z}, \mathbf{c_b}) \tag{12}$$

Theorem 2.

$$\langle \Delta_{\text{Eve}}^{(p_{a,z}, p_{a,x})}(\mathbf{k}, \mathbf{k}') \rangle \leq 2m\sqrt{P\left[\left(\frac{|\widetilde{\mathbf{C}}_I|}{n} \geq \frac{d_{r,m}}{2n}\right) \wedge \left(\frac{|\mathbf{C}_{T_Z}|}{n_z} \leq p_{a,z}\right) \wedge \left(\frac{|\mathbf{C}_{T_X}|}{n_x} \leq p_{a,x}\right)\right]} \tag{13}$$

where $\frac{|\widetilde{\mathbf{C}}_I|}{n}$ *is the random variable corresponding to the error rate on the INFO bits if they had been encoded in the x basis,* $\frac{|\mathbf{C}_{T_Z}|}{n_z}$ *is the random variable corresponding to the error rate on the TEST-Z bits, and* $\frac{|\mathbf{C}_{T_X}|}{n_x}$ *is the random variable corresponding to the error rate on the TEST-X bits.*

Proof. We use the convexity of x^2, namely, the fact that for all $\{p_i\}_i$ satisfying $p_i \geq 0$ and $\sum_i p_i = 1$, it holds that $(\sum_i p_i x_i)^2 \leq \sum_i p_i x_i^2$. We find that:

$$\langle \Delta_{\text{Eve}}^{(p_{a,z}, p_{a,x})}(\mathbf{k}, \mathbf{k}') \rangle^2$$

$$=\Big[\sum_{\mathscr{P},\xi,\mathbf{c_z},\mathbf{c_b}}\Delta_{\text{Eve}}^{(p_{a,z},p_{a,x})}(\mathbf{k},\mathbf{k}'|\mathscr{P},\xi,\mathbf{c_z},\mathbf{c_b})$$
$$\cdot p(\mathscr{P},\xi,\mathbf{c_z},\mathbf{c_b})\Big]^2 \qquad \text{(by (12))}$$
$$\le \sum_{\mathscr{P},\xi,\mathbf{c_z},\mathbf{c_b}}\left(\Delta_{\text{Eve}}^{(p_{a,z},p_{a,x})}(\mathbf{k},\mathbf{k}'|\mathscr{P},\xi,\mathbf{c_z},\mathbf{c_b})\right)^2$$
$$\cdot p(\mathscr{P},\xi,\mathbf{c_z},\mathbf{c_b}) \qquad \text{(by convexity of } x^2)$$
$$=\sum_{\mathscr{P},\xi,\frac{|\mathbf{c_z}|}{n_z}\le p_{a,z},\frac{|\mathbf{c_b}|}{n_x}\le p_{a,x}}\left(\frac{1}{2}\operatorname{tr}|\widehat{\rho}_{\mathbf{k}}-\widehat{\rho}_{\mathbf{k}'}|\right)^2$$
$$\cdot p(\mathscr{P},\xi,\mathbf{c_z},\mathbf{c_b}) \qquad \text{(by (11))}$$
$$\le 4m^2\cdot\sum_{\mathscr{P},\xi,\frac{|\mathbf{c_z}|}{n_z}\le p_{a,z},\frac{|\mathbf{c_b}|}{n_x}\le p_{a,x}}P\left[|\widetilde{\mathbf{C}_I}|\ge\frac{d_{r,m}}{2}\mid\mathbf{c_z},\mathbf{c_b},\mathscr{P}\right]$$
$$\cdot p(\mathscr{P},\xi,\mathbf{c_z},\mathbf{c_b}) \qquad \text{(by (7))}$$
$$=4m^2\cdot\sum_{\mathscr{P},\frac{|\mathbf{c_z}|}{n_z}\le p_{a,z},\frac{|\mathbf{c_b}|}{n_x}\le p_{a,x}}P\left[|\widetilde{\mathbf{C}_I}|\ge\frac{d_{r,m}}{2}\mid\mathbf{c_z},\mathbf{c_b},\mathscr{P}\right]$$
$$\cdot p(\mathscr{P},\mathbf{c_z},\mathbf{c_b})$$
$$=4m^2\cdot\sum_{\mathscr{P}}P\Big[\left(|\widetilde{\mathbf{C}_I}|\ge\frac{d_{r,m}}{2}\right)$$
$$\wedge\left(\frac{|\mathbf{C}_{T_Z}|}{n_z}\le p_{a,z}\right)\wedge\left(\frac{|\mathbf{C}_{T_X}|}{n_x}\le p_{a,x}\right)\mid\mathscr{P}\Big]\cdot p(\mathscr{P})$$
$$=4m^2\cdot P\Big[\left(|\widetilde{\mathbf{C}_I}|\ge\frac{d_{r,m}}{2}\right)$$
$$\wedge\left(\frac{|\mathbf{C}_{T_Z}|}{n_z}\le p_{a,z}\right)\wedge\left(\frac{|\mathbf{C}_{T_X}|}{n_x}\le p_{a,x}\right)\Big]$$

□

3.3 Proof of Security

Following BGM09 and BBBMR06, we choose matrices P_C and P_K such that the inequality $\frac{d_{r,m}}{2n} > p_{a,x}+\varepsilon$ is satisfied for some ε (we will explain in Subsection 3.5 why this is possible). This means that

$$P\left[\left(\frac{|\widetilde{\mathbf{C}_I}|}{n}\ge\frac{d_{r,m}}{2n}\right)\wedge\left(\frac{|\mathbf{C}_{T_Z}|}{n_z}\le p_{a,z}\right)\wedge\left(\frac{|\mathbf{C}_{T_X}|}{n_x}\le p_{a,x}\right)\right]$$
$$\le P\left[\left(\frac{|\widetilde{\mathbf{C}_I}|}{n}>p_{a,x}+\varepsilon\right)\wedge\left(\frac{|\mathbf{C}_{T_X}|}{n_x}\le p_{a,x}\right)\right]. \qquad (14)$$

We will now prove the right-hand-side of (14) to be exponentially small in n.

As said earlier, the random variable $\widetilde{\mathbf{C}_I}$ corresponds to the bit string of errors on the INFO bits if they had been encoded in the x basis. The TEST-X bits are also encoded in the x basis, and the random variable $\mathbf{C}_{T_X}$ corresponds to the bit string of errors on those bits. Therefore, we can treat the selection of the n INFO bits and of the n_x TEST-X bits as a random sampling (after the numbers n, n_z, and n_x *and* the TEST-Z bits have all already been chosen), and use Hoeffding's theorem (that is described in Appendix A of BGM09).

Therefore, for each bit string $c_1\dots c_{n+n_x}$ that consists of the errors in the $n+n_x$ INFO and TEST-X bits *if the INFO bits had been encoded in the x basis*, we apply Hoeffding's theorem: namely, we take a sample of size n without replacement from the population $c_1,\dots,c_{n+n_x}$ (this corresponds to the random selection of the INFO bits and the TEST-X bits, as defined above, given that the TEST-Z bits have already been chosen). Let $\overline{X}=\frac{|\widetilde{\mathbf{C}_I}|}{n}$ be the average of the sample (this is exactly the error rate on the INFO bits, assuming, again, the INFO bits had been encoded in the x basis); and let $\mu=\frac{|\widetilde{\mathbf{C}_I}|+|\mathbf{C}_{T_X}|}{n+n_x}$ be the expectancy of $\overline{X}$ (this is exactly the error rate on the INFO bits and TEST-X bits together). Then $\frac{|\mathbf{C}_{T_X}|}{n_x}\le p_{a,x}$ is equivalent to $(n+n_x)\mu-n\overline{X}\le n_x\cdot p_{a,x}$, and, therefore, to $n\cdot(\overline{X}-\mu)\ge n_x\cdot(\mu-p_{a,x})$. This means that the conditions $\left(\frac{|\widetilde{\mathbf{C}_I}|}{n}>p_{a,x}+\varepsilon\right)$ and $\left(\frac{|\mathbf{C}_{T_X}|}{n_x}\le p_{a,x}\right)$ rewrite to

$$(\overline{X}-\mu>\varepsilon+p_{a,x}-\mu)$$
$$\wedge\left(\frac{n}{n_x}\cdot(\overline{X}-\mu)\ge\mu-p_{a,x}\right), \qquad (15)$$

which implies $\left(1+\frac{n}{n_x}\right)(\overline{X}-\mu)>\varepsilon$, which is equivalent to $\overline{X}-\mu>\frac{n_x}{n+n_x}\varepsilon$. Using Hoeffding's theorem (from Appendix A of BGM09), we get:

$$P\left[\left(\frac{|\widetilde{\mathbf{C}_I}|}{n}>p_{a,x}+\varepsilon\right)\wedge\left(\frac{|\mathbf{C}_{T_X}|}{n_x}\le p_{a,x}\right)\right]$$
$$\le P\left[\overline{X}-\mu>\frac{n_x}{n+n_x}\varepsilon\right]\le e^{-2\left(\frac{n_x}{n+n_x}\right)^2 n\varepsilon^2} \qquad (16)$$

In the above discussion, we have actually proved the following Theorem:

Theorem 3. *Let us be given $\delta>0$, $R>0$, and, for infinitely many values of n, a family $\{v_1^n,\dots,v_{r_n+m_n}^n\}$ of linearly independent vectors in $\mathbf{F}_2^n$ such that $\delta<\frac{d_{r_n,m_n}}{n}$ and $\frac{m_n}{n}\le R$. Then for any $p_{a,z},p_{a,x}>0$ and $\varepsilon_{\text{sec}}>0$ such that $p_{a,x}+\varepsilon_{\text{sec}}\le\frac{\delta}{2}$, and for any $n,n_z,n_x>0$ and two m_n-bit final keys $\mathbf{k},\mathbf{k}'$, Eve's difference between her states corresponding to $\mathbf{k}$ and $\mathbf{k}'$ satisfies the following bound:*

$$\langle\Delta_{\text{Eve}}^{(p_{a,z},p_{a,x})}(\mathbf{k},\mathbf{k}')\rangle\le 2Rne^{-\left(\frac{n_x}{n+n_x}\right)^2 n\varepsilon_{\text{sec}}^2} \qquad (17)$$

In Subsection 3.5 we explain why this Theorem guarantees security.

We note that the quantity $\langle\Delta_{\text{Eve}}^{(p_{a,z},p_{a,x})}(\mathbf{k},\mathbf{k}')\rangle$ bounds the expected values of the Shannon Distinguishability and of the mutual information between Eve and the final key, as done in BGM09 and BBBMR06, which is sufficient for proving non-composable security; but it also avoids composability problems: Eve is not required to measure immediately after the protocol ends, but she is allowed to wait until she gets more information; and equation (17) bounds the trace distance between any two of Eve's possible states.

3.4 Reliability

Security itself is not sufficient; we also need the key to be reliable (namely, to be the same for Alice and Bob). This means that we should make sure that the number of errors on the INFO bits is less than the maximal number of errors that can be corrected by the error-correcting code. We demand that our error-correcting code can correct $n(p_{a,z}+\varepsilon_{\text{rel}})$ errors. Therefore, reliability of the final key with exponentially small probability of failure is guaranteed by the following inequality: (as said, $\mathbf{C}_I$ corresponds to the actual bit string of errors on the INFO bits in the protocol, when they are encoded in the z basis)

$$P\left[\left(\frac{|\mathbf{C}_I|}{n} > p_{a,z}+\varepsilon_{\text{rel}}\right) \wedge \left(\frac{|\mathbf{C}_{T_Z}|}{n_z} \le p_{a,z}\right)\right] \le e^{-2\left(\frac{n_z}{n+n_z}\right)^2 n\varepsilon_{\text{rel}}^2}$$

This inequality is proved by an argument similar to the one used in Subsection 3.3: the selection of the INFO bits and TEST-Z bits is a random partition of $n+n_z$ bits into two subsets of sizes n and n_z, respectively (assuming that the TEST-X bits have already been chosen), and thus it corresponds to Hoeffding's sampling.

3.5 Security, Reliability, and Error Rate Threshold

According to Theorem 3 and to the discussion in Subsection 3.4, to get both security and reliability we only need vectors $\{v_1^n,\ldots,v_{r_n+m_n}^n\}$ satisfying both the conditions of the Theorem (distance $\frac{d_{r_n,m_n}}{2n} > \frac{\delta}{2} \ge p_{a,x}+\varepsilon_{\text{sec}}$) and the reliability condition (the ability to correct $n(p_{a,z}+\varepsilon_{\text{rel}})$ errors). Such families were proven to exist in Appendix E of BBBMR06, giving the bit-rate:

$$\begin{aligned} R_{\text{secret}} &\triangleq \frac{m}{n} \\ &= 1-H_2(2p_{a,x}+2\varepsilon_{\text{sec}}) \\ &- H_2\left(p_{a,z}+\varepsilon_{\text{rel}}+\frac{1}{n}\right) \end{aligned} \quad (18)$$

where $H_2(x) \triangleq -x\log_2(x)-(1-x)\log_2(1-x)$.

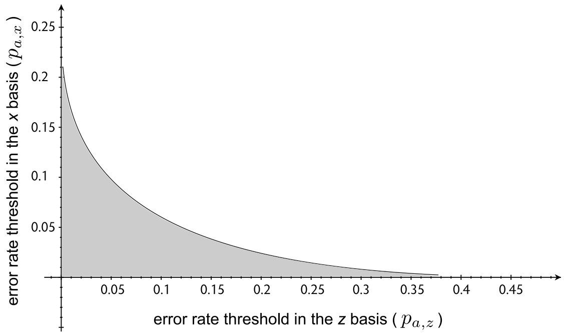

Figure 1: The secure asymptotic error rates zone (below the curve).

Note that we use here the error thresholds $p_{a,x}$ for security and $p_{a,z}$ for reliability. This is possible, because in BBBMR06 those conditions (security and reliability) on the codes are discussed separately.

To get the asymptotic error rate thresholds, we require $R_{\text{secret}} > 0$, and we get the condition:

$$H_2(2p_{a,x}+2\varepsilon_{\text{sec}})+H_2\left(p_{a,z}+\varepsilon_{\text{rel}}+\frac{1}{n}\right) < 1 \quad (19)$$

The secure asymptotic error rate thresholds zone is shown in Figure 1 (it is below the curve), assuming that $\frac{1}{n}$ is negligible. Note the trade-off between the error rates $p_{a,z}$ and $p_{a,x}$. Also note that in the case $p_{a,z}=p_{a,x}$, we get the same threshold as BB84 (BBBMR06 and BGM09), which is 7.56%.

4 CONCLUSION

In this paper, we have analyzed the security of the BB84-INFO-z protocol against any collective attack. We have discovered that the results of BB84 hold very similarly for BB84-INFO-z, with only two exceptions:

1. The error rates must be separately checked to be below the thresholds $p_{a,z}$ and $p_{a,x}$ for the TEST-Z and TEST-X bits, respectively, while in BB84 the error rate threshold p_a applies to all the TEST bits together.
2. The exponents of Eve's information (security) and of the failure probability of the error-correcting code (reliability) are different than in BGM09, because different numbers of test bits are now allowed (n_z and n_x are arbitrary). This implies that the exponents may decrease more slowly (or more quickly) as a function of n. However, if we choose $n_z=n_x=n$ (thus sending $N=3n$ qubits from Alice to Bob), then we get exactly the same exponents as in BGM09.

The asymptotic error rate thresholds found in this paper are more flexible than in BB84, because they allow us to tolerate a higher threshold for a specific basis (say, the x basis) if we demand a lower threshold for the other basis (z). If we choose the same error rate threshold for both bases, then the asymptotic bound is 7.56%, exactly the bound found for BB84 in BBBMR06 and BGM09.

We conclude that even if we change the BB84 protocol to have INFO bits only in the z basis, this does not harm its security and reliability (at least against collective attacks). This does not even change the asymptotic error rate threshold, and allows more flexibility when choosing the thresholds for both bases. The only drawbacks of this change are the need to check the error rate for the two bases separately, and the need to either send more qubits ($3n$ qubits in total, rather than $2n$) or get a slower exponential decrease of the exponents required for security and reliability.

We thus find that the feature of BB84, that both bases are used for information, is not very important for security and reliability, and that BB84-INFO-z (that lacks this feature) is almost as useful as BB84. This may have important implications on the security and reliability of other protocols that also only use one basis for information qubits, as done in some two-way protocols.

We also present a better approach for the proof, that uses a quantum distance between two states rather than the classical information. In BGM09, BBBGM02, and BBBMR06, the classical mutual information between Eve's information (after an optimal measurement) and the final key was calculated (by using the trace distance between two quantum states); although we should note that in BGM09 and BBBMR06, the trace distance was used for the proof of security of a single bit of the final key even when all other bits are given to Eve, and only the last stages of the proof discussed bounding the classical mutual information. In the current paper, on the other hand, we use the trace distance between the two quantum states until the end of the proof, which avoids composability problems that existed in the previous works.

Therefore, this proof makes a step towards making BGM09, BBBGM02, and BBBMR06 prove composable security of BB84 (namely, security even if Eve keeps her quantum states until she gets more information when Alice and Bob use the key, rather than measuring them in the end of the protocol). This approach also applies (similarly) to the BB84 security proof in BGM09.

ACKNOWLEDGEMENTS

The work of TM and RL was partly supported by the Israeli MOD Research and Technology Unit.

REFERENCES

Bennett, C. H. and Brassard, G. (1984). Quantum cryptography: Public key distribution and coin tossing. In *International Conference on Computers, Systems & Signal Processing, IEEE, 1984*, pages 175–179.

Biham, E., Boyer, M., Boykin, O. P., Mor, T., and Roychowdhury, V. (2006). A proof of the security of quantum key distribution. *J. Cryptol.*, 19(4):381–439.

Biham, E., Boyer, M., Brassard, G., van de Graaf, J., and Mor, T. (2002). Security of quantum key distribution against all collective attacks. *Algorithmica*, 34(4):372–388.

Biham, E. and Mor, T. (1997a). Bounds on information and the security of quantum cryptography. *Phys. Rev. Lett.*, 79:4034–4037.

Biham, E. and Mor, T. (1997b). Security of quantum cryptography against collective attacks. *Phys. Rev. Lett.*, 78:2256–2259.

Boyer, M., Gelles, R., and Mor, T. (2009). Security of the bennett-brassard quantum key distribution protocol against collective attacks. *Algorithms*, 2(2):790–807.

Christandl, M., König, R., and Renner, R. (2009). Postselection technique for quantum channels with applications to quantum cryptography. *Phys. Rev. Lett.*, 102:020504.

Mayers, D. (2001). Unconditional security in quantum cryptography. *J. ACM*, 48(3):351–406.

Nielsen, M. A. and Chuang, I. L. (2010). *Quantum computation and quantum information*. Cambridge university press.

Renner, R. (2008). Security of quantum key distribution. *Int. J. Quantum Inf.*, 6(01):1–127.

Renner, R., Gisin, N., and Kraus, B. (2005). Information-theoretic security proof for quantum-key-distribution protocols. *Phys. Rev. A*, 72:012332.

Shor, P. W. and Preskill, J. (2000). Simple proof of security of the bb84 quantum key distribution protocol. *Phys. Rev. Lett.*, 85:441–444.