Security of Quantum Cryptography against Collective Attacks

Eli Biham[1] and Tal Mor[2]
[1]*Computer Science Department, Technion, Haifa 32000, Israel*
[2]*Physics Department, Technion, Haifa 32000, Israel*
(Received 6 May 1996)

We present strong attacks against quantum key distribution schemes which use quantum memories and quantum gates to attack *directly* the final key. We analyze a specific attack of this type, for which we find the density matrices available to the eavesdropper and the optimal information which can be extracted from them. We prove security against this attack and discuss security against any attack allowed by the rules of quantum mechanics. [S0031-9007(97)02537-4]

PACS numbers: 89.80.+h, 03.65.Bz, 89.70.+c

Quantum cryptography [1–5] uses quantum mechanics to perform new cryptographic tasks—especially information secure key distributions—which are beyond the abilities of classical cryptography. Unfortunately, the security of such a key is still unproven: Sophisticated attacks (called *coherent* or *joint attacks*) which are directed against the final key were suggested; the analysis of such attacks is very complicated, and, by the time this work was submitted, security against them was proven only in the nonrealistic case of ideal (error-free) channels [6,7]. The security in the real case, which is crucial for making quantum cryptography practical, is commonly believed but yet unproven. A proof of security must bound the information available to the eavesdropper (traditionally called "Eve"), on the final key, to be negligible (i.e., much smaller than one bit). A protocol is considered secure if the adversary is restricted only by the rules of quantum mechanics, and a protocol is considered practical if the legitimate users are restricted to use existing technology. In this work we obtain the strongest security result for practical protocols. We suggest *collective attacks* (simpler than the joint attacks) which are simple enough to be analyzed, but are general enough to imply (or at least suggest) the security against any attack. We prove security against the simplest collective attack: We generalize methods developed in [8] in order to calculated Eve's density matrices explicitly, and to find the information which can be obtained from them; we show that it is negligible. Our result also provides better understanding of the issue of information splitting between two parties which is a fundamental problem in quantum information theory. Parts of this work were done together with Dominic Mayers.

In any quantum key distribution scheme, the sender, "Alice," sends to the receiver, "Bob," a classical string of bits by encoding them as quantum states. In the two-state scheme [2] (B92 scheme) a classical bit is represented by either of two nonorthogonal pure states, which can be written as $\psi_0 = \binom{\cos\theta}{\sin\theta}$, and $\psi_1 = \binom{\cos\theta}{-\sin\theta}$. Bob performs a test which provides him with a conclusive or inconclusive result. For instance, he can test whether a specific particle is in a state ψ_0 or a state orthogonal to it ψ_0'; a result ψ_0 is treated as inconclusive, and a result ψ_0' is identified as ψ_1. Alice and Bob use also an unjammable classical channel to inform which bits were identified conclusively, and to compare some of the common bits in order to estimate the error rate. They must accept some small error rate p_e due to imperfections in creating, transmitting, and receiving the quantum states. If the estimated error rate exceeds the allowed error rate they quit the transmission and do not use the data, thus any eavesdropping attempt is severely constrained to induce an error rate smaller than p_e. Alice and Bob are now left with similar n-bit strings which contain errors. They randomize the order of the bits and correct the errors using any error-correction code [9]. The error-correction code is usually made of r parities of substrings [where the parity bit $p(x)$ of a binary string x is zero if there is an even number of 1's in x, and one otherwise]. Alice sends these parties to Bob (using the classical channel), who uses them to obtain a (possibly shorter) string identical to Alice's, up to an exponentially small error probability. Finally, Alice and Bob can amplify the security of the final key by using privacy amplification techniques [10] by choosing some parity bits of substrings to be the final key. Their aim is to derive a final key on which Eve's average information is negligible.

Eve can measure some of the particles and gain a lot of information on them, but this induces a lot of error. Hence, she can attack only a small portion of the particles, and this reduces her information on the parity of many bits exponentially to zero. Translucent attacks [11] are much more powerful: Eve attaches a probe to *each particle* and performs some unitary transformation, after which her probe is correlated to the transmitted state. In the case where each probe is left in a pure state [11], and measured separately to obtain information on Alice's bit, it is a rather obvious conclusion (from [10]) that privacy amplification is still effective. Thus, such an *individual* translucent attack is ineffective. We deal with a much more sophisticated attack in which Eve's measurement is done *after* the processes of error correction and privacy amplification are completed. Privacy amplification techniques were not designed to stand against such attacks, hence their efficiency against them is yet unknown. Consider the following *collective* attack: (1) Eve attaches a *separate, uncorrelated* probe to each transmitted particle using a translucent attack. (2) Eve keeps the probes in a quantum memory (where nonorthogonal quantum states can be kept for a long time [5]) till receiving *all* classical data including error-correction and privacy amplification data. (3) Eve performs the optimal measurement on her probes in order to learn the maximal information on the *final* key. The case in which Eve attaches one probe (in a large-dimensional Hilbert space) to all transmitted particles is called a *joint* or *coherent* attack [4], and it is the most general possible attack. No specific joint attacks were yet suggested; the collective attack defined above is the strongest joint attack suggested so far, and there are good reasons to believe that it is the strongest possible attack.

The security of quantum cryptography is a very complicated and tricky problem. Several security claims done in the past were found later on to contain loopholes. Recently, we become aware of three new such claims [12–14]. We hope that these approaches, together with our approach, really produce the solution; yet it is important to have them all, since each of them has different advantages.

Our approach deals with error correction and privacy amplification, by calculating the density matrices which are available to the eavesdropper by the time all data transmissions (classical and quantum) are completed. We provide an example of collective attacks based on the "translucent attack without entanglement" of [11], which leave Eve with probes in a pure state, and we prove security against them. These attacks use the unitary transformation $\binom{\cos\theta}{\pm\sin\theta} \rightarrow \binom{\cos\theta'}{\pm\sin\theta'}\binom{\cos\alpha}{\pm\sin\alpha}$ with "+" for ψ_0, and "−" for ψ_1, where θ' is the angle of the states received by Bob, and α is the angle of the states in Eve's hand. The error rate, $p_e = \sin^2(\theta - \theta')$, is the probability that Alice sent ψ_0 and Bob measured ψ_0'. The connection between this induced error rate and the angle α is calculated using the unitary condition [11] $\cos 2\theta = \cos 2\theta' \cos 2\alpha$. For weak attacks which causes small error rate the angle of Eve's probe satisfies $\alpha = (p_e \tan^2 2\theta)^{1/4}$. In our case, the same translucent attack is performed on all the bits, and it leaves Eve with n probes, each in one of the two states $\binom{c}{\pm s}$, with $c = \cos\alpha$ and $s = \sin\alpha$. As result, Eve holds an n bits string x which is concatenated from its bits $(x)_1(x)_2\cdots(x)_n$. For simplicity, we choose the final key to consist of one bit, which is the parity of the n bits. Eve wants to distinguish between two density matrices corresponding to the two possible values of this parity bit. Our aim is to calculate the optimal mutual information she can extract from them.

For our analysis we need some more notations. Let $\hat{n}(x)$ be the number of 1's in x. For two strings of equal length $x \odot y$ is the bitwise "AND," so that the bit $(x \odot y)_i$ is one if both $(x)_i$ and $(y)_i$ are one. Also $x \oplus y$ is the bitwise "XOR", so that $(x \oplus y)_i$ is zero if $(x)_i$ and $(y)_i$ are the same. For k (independent) strings, $v_1 \ldots v_k$, of equal length let the set $\{v\}_k$ contain the 2^k linear combinations $(v_1), \ldots, (v_k), (v_1 \oplus v_1), (v_1 \oplus v_2), \ldots, (v_1 \oplus v_2 \ldots \oplus v_k)$. If these strings are not all different, then the original k strings are linearly dependent. The quantum state of a string is the tensor product

$$\psi_x = \begin{pmatrix} c \\ \pm s \end{pmatrix}\begin{pmatrix} c \\ \pm s \end{pmatrix}\cdots\begin{pmatrix} c \\ \pm s \end{pmatrix} = \begin{pmatrix} ccc\ldots ccc \\ \pm ccc\ldots ccs \\ \ldots \\ \pm sss\ldots sss \end{pmatrix}, \quad (1)$$

leaving in a 2^n dimensional Hilbert space. The sign of the ith bit (in the middle expression) is plus for $(x)_i = 0$ and minus for $(x)_i = 1$. The sign of the jth term ($j = 0\ldots 2^{n-1}$) in the expression at the right depends on the parity of the string $x \odot j$ and is equal to $(-1)^{p(x\odot j)}$. The density matrix $\rho_x = \psi_x \psi_x^T$ also has for any x, the same terms up to the signs. We denote the absolute values by $\rho_{jk} \equiv |(\rho_x)_{jk}|$. The sign of each term $(\rho_x)_{jk}$ is given by

$$(-1)^{p(x\odot j)}(-1)^{p(x\odot k)} = (-1)^{p[x\odot(j\oplus k)]}. \quad (2)$$

A priori, all strings are equally probable, and Eve needs to distinguish between the two density matrices describing the parities. These matrices were calculated and analyzed in Bennett, Mor, and Smolin [8] (henceforth, the BMS work), and independently in [15] for the case $\alpha = \pi/4$. In case Eve is being told what the error-correction code is, all strings consistent with the given error-correction code (the r subparities) are equally probable, and Eve needs to distinguish between the two density matrices,

$$\rho_0^{(n,r)} = \frac{1}{2^{n-r-1}} \sum_{x|\binom{p(x)=0}{x\ \mathrm{OECC}}} \rho_x;$$
$$\rho_1^{(n,r)} = \frac{1}{2^{n-r-1}} \sum_{x|\binom{p(x)=1}{x\ \mathrm{OECC}}} \rho_x, \quad (3)$$

where "OECC" is a shortcut for "obeys error-correction code." Let us look at two simple examples where $n = 5$, one with $r = 1$ and the second with $r = 2$. Suppose that the parity of the first two bits, $(x)_1$ and $(x)_2$, is $p_1 = 0$. Formally, this substring is described by the n-bit string $v_1 = 24$ which is 11000 binary; the number of 1's in the first two bits of a string x is given by $\hat{n}(x \odot v_1)$, and x obeys the error-correction code if $p(x \odot v_1) = p_1$. Let v_d be the binary string (11111 in this case) which describes the substring of the desired parity. Eve could perform the optimal attack on the three bit which are left or, in general, on $v_1 \oplus v_d$. For any such case, the optimal attack is given by the BMS work, and the optimal information depends on $\hat{n}(v_1 \oplus v_d)$, the *Hamming distance* between the two words. This information [using Eq. (53) of the BMS work] is

$$I(\hat{n}) = c\binom{2k}{k}\alpha^{2k}, \quad (4)$$

with $c = 1$ for even $\hat{n}$ (which equals $2k$) and $c = 1/\ln 2$ for odd $\hat{n}$ (that is $\hat{n} = 2k - 1$). Suppose that Eve gets another parity bit $p_2 = 1$ of the binary string 01100 ($v_2 = 12$). Now, a string x obeys the error-correction code if it also obeys $p(x \odot v_2) = p_2$. Clearly, it also satisfies $p[x \odot (v_1 \oplus v_2)] = p_1 \oplus p_2$. In the general case there are r independent parity strings, and 2^r parity strings in the set $\{v\}_r$ (including the string 00000). The BMS result cannot be directly used but still provides some intuition: For each word (i.e., each parity string) $v_l \in \{v\}_r$, let $I[\hat{n}(v_l \oplus v_d)]$ be the optimal information Eve could obtain using Eq. (4). Also let $I_{\rm sum}$ be the sum of these contributions from all such words. In reality Eve cannot obtain $I_{\rm sum}$ since each measurement changes the state of the measured bits, hence we expect that $I_{\rm sum}$ bounds her optimal information $I_{\rm total}$ from above: $I_{\rm total} < I_{\rm sum}$. On the other hand, Eve knows all these words at once, and could take advantage of it, thus we leave this as an unproven conjecture.

In the following we find an explicit way to calculate *exactly* the optimal information. However, this exact result requires cumbersome calculations, thus it is used only to verify the conjecture for short strings.

The parity of the full string is also known since the density matrix $\rho^{(n,r+1)}$ corresponds to either $\rho_0^{(n,r)}$ or $\rho_1^{(n,r)}$ depending on the desired parity p_{r+1}, thus we add the string $v_{r+1} = v_d$. There are $r + 1$ independent subparities altogether, hence 2^{r+1} parity strings in the set $\{v\}_{r+1}$. A string x is included in $\rho^{(n,r+1)}$ if $p[x \odot v_l] = p_l$ for all given substring in $\{v\}_{r+1}$. In the BMS work (where $r = 0$) the parity density matrices were put in a block diagonal form of 2^{n-1} blocks of size 2×2. This result can be generalized to the case where r parities of substrings are given. There will be 2^{n-r-1} blocks of size $2^{r+1} \times 2^{r+1}$. We shall show that the (jk)th term in a density matrix $\rho^{(n,r+1)}$ of $r + 1$ subparities is either zero, ρ_{jk} or $-\rho_{jk}$, that is, either all the relevant strings contribute exactly the same term, or half of them cancels the other half. The proof can be skipped in a first reading.

Theorem.—The element $(\rho^{(n,r+1)})_{jk}$ is zero if $j \oplus k \notin \{v\}_{r+1}$, and it is $\pm\rho_{jk}$ if $j \oplus k \in \{v\}_{r+1}$.

Proof.—In case $j \oplus k \notin \{v\}_{r+1}$ choose C such that $p[C \odot v_l] = 0$ with all (v_l)'s in $\{v\}_{r+1}$ and $p[C \odot (j \oplus k)] = 1$ (many such C's exist since C has n independent bits and it needs to fulfill only $r + 2$ constraints). For such a C and for any x which obeys the error-correction code there exists one (and only one) $y, y = x \oplus C$, which also obeys the code (due to the first demand) but has the opposite sign in the jkth element (due to the second demand), so $(\rho_y)_{jk} = -(\rho_x)_{jk}$. Since this is true for any relevant x, we obtain $(\rho^{(n,r+1)})_{jk} = 0$. In case $j \oplus k \in \{v\}_{r+1}$ such C cannot exist, and all terms must have the same sign: Suppose that there are two terms, x and y with opposite signs. Then $C = x \oplus y$ satisfies the two demands, leading to a contradiction.

This theorem tells us the place of all nonvanishing terms in the original ordering. The matrices can be reordered to a block-diagonal form by exchanges of the basis vectors. We group the vectors $s, s \oplus v_1$, etc., for all (v_l)'s in $\{v\}_{r+1}$ to be one after the other, so each such group is separated from the other groups. Now the theorem implies that all nonvanishing terms are grouped in blocks, and all vanishing terms are outside these blocks. As a result the matrix is block diagonal. This forms 2^{n-r-1} blocks of size $2^{r+1} \times 2^{r+1}$. All terms inside the blocks and their signs are given by Eqs. (1) and (2), respectively, up to reordering. The organization of the blocks depends only on the parity strings v_l and not on the parities p_l, thus $\rho_0^{(n,r)}$ and $\rho_1^{(n,r)}$ are block diagonalized in the same basis. The rank of a density matrix is the number of (independent) pure states which form it, and it is 2^{n-r-1} in case of the parity matrices [Eq. (3)]. When these matrices are put in a block-diagonal form, there are 2^{n-r-1} (all nonzero) blocks. Thus, the rank of each block is one, the corresponding state is pure, and, when fully diagonalized, the nonvanishing term a_j in the jth block is the probability that a measurement will result in this block.

In the BMS work ($r = 0$), the information , in case of small angle, was found to be exponentially small with the length of the string. When each probe is in a pure state, this result can be generalized to $r > 0$ as follows: The optimal mutual information carried by two pure states (in any dimension) in well known. The two possible pure states in the jth block of $\rho_0^{(n,r)}$ and $\rho_1^{(n,r)}$ can be written as $\binom{\cos\beta}{\pm\sin\beta}$. The optimal mutual information which can be obtained from the jth block is given by the overlap (the angle β_j) $I_j = 1 + p_j \log_2 p_j + (1 - p_j)\log_2(1 - p_j)$, where $p_j = (1 - \sin 2B_j)/2$; the overlap is calculated using Eqs. (1) and (2). Thus, for any given error-correction code, we can find the two pure states in each block, the optimal information I_j, and finally, the total information $I_{\rm total} = \sum_j a_j I_j$. We did not use the value of v_d in the proof, and thus, the final key could be the parity of any substring. Moreover, a similar method can be used to analyze keys of several bits which can be formed from parities of several substrings.

We wrote a computer program which receives any (short) error-correction code and calculates the total information as a function of the angle α between the pure states of the individual probes. We checked many short codes (up to $n = 8$) to verify whether $I_{\rm total} < I_{\rm sum}$ as we conjectured. Indeed, all our checks showed that the conjecture holds. The information for small angle α is bounded by $I_{\rm sum} = C\alpha^{2k}$ as previously explained, where C is given by summing the terms which contribute to the highest order of Eq. (4), and the Hamming distance $\hat{n}$ (which is $2k$ or $2k - 1$) can be increased by choosing longer codes to provide any desired level of security.

In addition to a desirable security level, the error-correction code must provide also a desirable reliability; a complete analysis must include also estimation of the probability p_f that Alice and Bob still has the wrong (i.e., different) final key. For enabling such analysis, one must use known error-correction codes. Random linear codes allow for such analysis but cannot be used efficiently by Alice and Bob. Hamming codes [9], H_r, which use r given parities for correcting one error in strings of length $n = 2^r - 1$, have an efficient decoding/encoding procedure and a simple way to calculate p_f. A Hamming code has 2^r words in $\{v\}_r$, all of them, except $00\ldots0$, are at the same distance $\hat{n} = 2^{r-1} - 1$ from v_d. Using our conjecture and Eq. (4) [with $k = (\hat{n} + 1)/2 = 2^{r-2}$] we obtain $I_{\rm total} < (2^r - 1)(1/\ln 2)\binom{2^{r-1}}{2^{r-2}}\alpha^{(2^{r-1})} + O(\alpha^{(2^r-1)})$. For $r = 3$ ($n = 7$) this yields $I_{\rm total} < 60.6\alpha^4$. The exact calculation done using our computer program also gives the same result, showing that the conjecture provides an extremely tight bound in this case. Using $\binom{2^{r-1}}{2^{r-2}} < 2^{(2^{r-1})}\sqrt{(\pi/2)(2^{r-1})}$ and some calculation we finally obtain

$$I_{\rm total} < \left(\frac{2}{\ln 2\sqrt{\frac{\pi}{2}}}\right)\sqrt{2^{r-1}}(2\alpha)^{(2^{r-1})}, \qquad (5)$$

bounding $I_{\rm total}$ to be exponentially small with n [which follows from $2^{r-1} = (n + 1)/2$].

The rate of errors in the string shared by Alice and Bob (after throwing inconclusive results) is the normalized error rate, $p_e^{(N)} = p_e/(p_c + p_e)$, where $p_c = \sin(\theta + \theta')$ is the probability of obtaining a correct and conclusive result. For small α it is $p_e^{(N)} = 2p_e/\sin^2 2\theta = (2\cos^2 2\theta/\sin^4 2\theta)\alpha^4$. The final error probability p_f is given by the probability to have more than one error in the initial string, since the code corrects one error. It is $p_f = [n(n - 1)/2](p_e^{(N)})^2 + O[(np_e^{(N)})^3]$, showing that we can use the Hamming codes as long as $np_e^{(N)} \ll 1$. In case it is not, better codes such as the Bose-Chaudhuri-Hocquenghem (BCH) codes [9] (which correct more than one error) are required, but their analysis is beyond the goals of this paper.

In conclusion, we presented new attacks on quantum key distribution schemes, directed against the final key, and we proved security against a specific one. This result, together with its extension to the analysis of probes in mixed state [16], suggest that the optimal information obtained by the *optimal* collective attack shall still show the same behavior as shown in our example. Let us explain the intuition that the security against collective attacks implies security against any joint attack: Most of the transmitted particles are not part of the n-bits string. The correlations between the n bits (as specified by the error correction and privacy amplification) as well as the random reordering of the bits are not known in advance. It is very reasonable that Eve can only lose by searching for such correlations when the particles are transmitted through her. Thus, the best she can do is probe the particles via the best collective attack.

We are grateful to C. H. Bennett, G. Brassard, C. Crépeau, J. Smolin, and A. Peres for many helpful discussions. We are especially grateful to D. Mayers for his great help and many suggestions, in particular, for observing that $\rho_p^{(n,r)}$ are of a block-diagonal form also for $r > 0$ [17] (he proved it independently in another context [12]). We also thank G. Brassard and the Université de Montréal for hosting a productive meeting, which had an extremely valuable contribution to this work.

[1] C. H. Bennett and G. Brassard, in *Proceedings of the IEEE International Conference on Computers, Systems and Signal Processing Bangalore, India* (IEEE, New York, 1984), p. 175.
[2] C. H. Bennett, Phys. Rev. Lett. **68**, 3121 (1992).
[3] A. K. Ekert, Phys. Rev. Lett. **67**, 661 (1991); C. H. Bennett, G. Brassard, and N. D. Mermin, Phys. Rev. Lett. **68**, 557 (1992).
[4] C. H. Bennett *et al.*, J. Cryptol. **5**, 1 (1992).
[5] E. Biham, B. Huttner, and T. Mor, Phys. Rev. A **54**, 2651 (1996).
[6] A. Yao, in *Proceedings of the 27 Symposium on the Theories of Computing* (ACM Press, Las Vegas, 1995), p. 67.
[7] D. Mayers, Proc. Crypto95 LNCS **963**, 124 (1995).
[8] C. H. Bennett, T. Mor, and J. Smolin, Phys. Rev. A **54**, 2675 (1996).
[9] F. J. MacWilliam and N. J. A. Sloane, *The Theory of Error Correction Codes* (North-Holland, Amsterdam, Amsterdam, 1977).
[10] C. H. Bennett, G. Brassard, C. Crépeau, and U. Maurer, IEEE Trans. Inf. Theory **41**, 1915 (1995).
[11] A. K. Ekert, B. Huttner, G. M. Palma, and A. Peres, Phys. Rev. A **50**, 1047 (1994).
[12] D. Mayers, Proc. Crypto 96, LNCS **1109**, 343 (1996); quant-ph 9606003.
[13] D. Deutsch, A. Ekert, R. Jozsa, C. Macchiavello, S. Popescu, and A. Sanpera, Phys. Rev. Lett. **77**, 2818 (1996). This protocol assumes nonrealistic devices with perfect accuracy (page 2819, line 18).
[14] H. K. Lo and Chau, Report No. quant-ph 9511025.
[15] D. Mayers, contribution to the ISI Quantum Computation 95 Workshop.
[16] E. Biham and T. Mor, Report No. quant-ph 9605010.
[17] D. Mayers (personal communication).