Tomographic Quantum Cryptography: Equivalence of Quantum and Classical Key Distillation

Dagmar Bruß,[1] Matthias Christandl,[2] Artur Ekert,[2,3] Berthold-Georg Englert,[3]
Dagomir Kaszlikowski,[3] and Chiara Macchiavello[4]

[1]*Institut für Theoretische Physik, Universität Hannover, 30167 Hannover, Germany*
[2]*DAMTP, University of Cambridge, Cambridge CB3 0WA, United Kingdom*
[3]*Department of Physics, National University of Singapore, Singapore 117 542, Singapore*
[4]*Dipartimento di Fisica "A. Volta," Università di Pavia, 27100 Pavia, Italy*

(Received 31 March 2003; published 25 August 2003)

The security of a cryptographic key that is generated by communication through a noisy quantum channel relies on the ability to distill a shorter secure key sequence from a longer insecure one. For an important class of protocols, which exploit tomographically complete measurements on entangled pairs of any dimension, we show that the noise threshold for classical advantage distillation is identical with the threshold for quantum entanglement distillation. As a consequence, the two distillation procedures are equivalent: neither offers a security advantage over the other.

DOI: 10.1103/PhysRevLett.91.097901 PACS numbers: 03.67.Dd, 03.65.Wj, 03.67.Hk

The ability to generate a secure cryptographic key, although the communication employs a quantum channel with a high level of noise, is crucial for all practical implementations of quantum cryptography. To be on the safe side, one must assume that all noise results from eavesdropping, that eavesdropper Eve has full knowledge of the cryptographic protocol (the "Kerckhoff principle" of cryptology), and that she acquires as much knowledge about the communication as is allowed by the laws of physics. This leads immediately to the question of where is the noise threshold below which a secure key can be generated at all. We give a definite answer for an important class of protocols, restricting, however, the discussion to incoherent attacks of the eavesdropper.

In the cryptographic protocol that we consider [1], Alice and Bob exploit entangled pairs of *qunits,* that is, n-fold quantum alternatives, the case of $n = 2$ being the elementary binary alternative of a qubit. Alice measures on her qunit, and Bob on his, an observable randomly chosen from their respective sets of $n + 1$ observables that are tomographically complete. Such sets surely exist for any dimension [2]. Adopting the notation of [3], we write $|m_k\rangle$ for the kth eigenket of Alice's mth observable and $|\overline{m}_k\rangle$ for the kth eigenket of Bob's mth observable, whereby $m = 0, 1, \ldots, n$ and $k = 0, 1, \ldots, n - 1$.

It is possible and expedient to choose these kets such that $\langle 0_j|m_k\rangle = \langle\overline{m}_k|\overline{0}_j\rangle$ for all m, j, k, and then the maximally entangled 2-qunit state $|\psi\rangle$ that Alice and Bob wish to share,

$$|\psi\rangle = \frac{1}{\sqrt{n}}\sum_{k=0}^{n-1}|0_k\overline{0}_k\rangle = \cdots = \frac{1}{\sqrt{n}}\sum_{k=0}^{n-1}|n_k\overline{n}_k\rangle, \qquad (1)$$

has the same appearance irrespective of the pair of observables that is used to define it. Therefore, their measurement results in the matched bases (same value of m for her and him) are perfectly correlated and can be used for the generation of a key in an alphabet with n letters.

On average, the measurement bases will be matched for a fraction $1/(n + 1)$ of the qunit pairs, and these data will supply the raw key sequence. Alice and Bob use part of it together with all the other measurement data, acquired for mismatched bases, to perform quantum tomography on the 2-qunit state they are actually receiving from the source. The tomographic completeness of the two sets of observables is crucial for this part of the procedure.

Alice and Bob assume that Eve distributes the qunits. They accept the raw key only if the result of their state tomography is consistent with an admixture of the chaotic state to $|\psi\rangle\langle\psi|$, thereby forcing Eve to use a symmetric strategy. In other words, they accept only a 2-qunit state ρ of the form

$$\rho = (\beta_0 - \beta_1)|\psi\rangle\langle\psi| + \frac{\beta_1}{n}I, \quad \beta_0 + (n-1)\beta_1 = 1, \qquad (2)$$

where I is the 2-qunit identity operator, β_0 is the probability that Bob gets the same value as Alice when the bases match, and β_1 is the probability that he gets a particular other one. Since $\beta_0 = \beta_1 = 1/n$ when there are no correlations whatsoever between their measurement results, we take $\beta_0 > 1/n > \beta_1$ for granted.

Although Eve fully controls the 2-qunit source, she is not free in her actions, because the state received by Alice and Bob must be of the form (2). One finds [1] that, therefore, the best Eve can do is to prepare an entangled pure state of the form

$$|\Psi\rangle = \sqrt{\frac{\beta_0}{n}}\sum_{k=0}^{n-1}|0_k\overline{0}_k\rangle|E_{kk}\rangle + \sqrt{\frac{\beta_1}{n}}\sum_{k\neq l}|0_k\overline{0}_l\rangle|E_{kl}\rangle, \qquad (3)$$

where her normalized ancilla states $|E_{kl}\rangle$ are such that those with $k \neq l$ are orthogonal to all others, whereas those with $k = l$ are not orthogonal among them selves, but obey $\langle E_{kk}|E_{ll}\rangle = 1 - (\beta_1/\beta_0)(1 - \delta_{kl})$. Thus the summations in (3) constitute two orthogonal components of $|\Psi\rangle$. The n-dimensional first component is relevant for establishing the cryptographic key, and the $n(n-1)$-dimensional second component is just noise to Alice and Bob.

We note that the invariance of $|\psi\rangle$ under base permutations is also possessed by $|\Psi\rangle$. Rather than referring to the 0th pair of observables, we could just as well use the joint eigenkets $|m_k \overline{m}_l\rangle$ of any other pair in conjunction with a suitable unitary redefinition of the ancilla states.

After Alice and Bob have given public notice of the observables they measured for each qunit pair, it is Eve's task to infer their measurement results — their nit values — whenever the bases match. To this end she must be able to identify her ancilla states. (Remember that we are only considering incoherent eavesdropping attacks.) Owing to the structure of $|\Psi\rangle$ she can distinguish unambiguously all the states belonging to the second orthogonal component, and so she can correctly infer Alice's and Bob's nit values if they are different. But if they are the same, Eve has to distinguish the $|E_{kk}\rangle$ of the first component, and then she cannot avoid errors because these states are not orthogonal to each other. In this situation, she minimizes her error probability by performing the so-called *square-root measurement* [4].

We denote by η_0 and $\eta_1 = (1-\eta_0)/(n-1)$ the probabilities that Eve infers the nit value correctly or gets a particular wrong one, respectively, provided that Bob's nit value is the same as Alice's. They are related to Bob's probabilities β_0 and β_1 of (2) by

$$\sqrt{\eta_0} - \sqrt{\eta_1} = \sqrt{\beta_1/\beta_0}. \tag{4}$$

Note that this expresses a certain complementarity between Bob's and Eve's respective knowledge about Alice's nit values. If Bob's values agree perfectly with Alice's ($\beta_0 = 1$, $\beta_1 = 0$), then Eve's values are completely random ($\eta_0 = \eta_1 = 1/n$), and conversely $\eta_0 = 1$, $\eta_1 = 0$ implies $\beta_0 = \beta_1 = 1/n$. In the more interesting intermediate situations we have $\eta_0 > 1/n > \eta_1$.

For single-particle protocols with qubits ($n = 2$) or qutrits ($n = 3$), the relation (4) is well established [5,6], and has been conjectured to hold for arbitrary dimensions [5,7]. This conjecture is proved in [1].

According to the Csiszár–Körner (CK) theorem [8], a secure key sequence can be extracted from the raw key sequence if the mutual information between Alice and Bob exceeds the mutual information between either one of them and Eve. This requires that Bob's and Eve's probabilities are such that

$$\begin{aligned}\nu \equiv\ & \beta_0 \log_n \beta_0 + (1-\beta_0)\log_n \beta_1 \\ & - \beta_0[\eta_0 \log_n \eta_0 + (1-\eta_0)\log_n \eta_1] \\ & > 0,\end{aligned} \tag{5}$$

and then ν is the yield of the CK procedure, the fraction of nit values that make it from the raw key sequence to the secure one. Since (4) implies that $\eta_0, \eta_1 \to 1/n$ as $\beta_0 \to 1$, this condition is surely met if β_0 is sufficiently large. If, however, there is too much noise in the 2-qunit state (2), the CK theorem is not immediately applicable. Rather, Alice and Bob must select a subsequence of nit values in a systematic way such that the CK theorem applies to the resulting "distilled key."

One method at their disposal for this purpose is *entanglement distillation* (ED), a quantum procedure by which they produce a smaller number of qunit pairs with stronger entanglement, by means of local operations and classical communication. (In the context of quantum cryptography, ED was originally proposed for qubits under the name of *quantum privacy amplification* [9].) By means of ED, Alice and Bob can reach a β_0 value for which the CK theorem is applicable, before they measure their respective observables. For states of the particularly simple structure (2), ED will be successful if

$$\beta_0 > 2\beta_1 \tag{6}$$

and only then [10]. If Eve can perfectly compensate for the back effect of ED on the ancillas, relation (4) also applies after ED. If she cannot, it turns into an inequality, tersely, $= \to <$.

Alternatively, Alice and Bob can produce their raw key sequences without any subensemble selection, and then perform *advantage distillation* (AD), a procedure of classical (i.e., nonquantum) cryptography [11]. As we shall see below, for β_0, η_0 values that obey (4), AD is successful whenever (6) holds, and only then, so that the thresholds for ED and AD are the same [12]. As a consequence of this coincidence, ED and AD are equivalent in the sense that neither offers a security advantage over the other.

Both ED and AD require classical two-way communication, but once the CK theorem becomes applicable, one-way communication suffices. We leave it as a moot point which method makes better use of the resources because the standard versions of both are very wasteful and hardly suited for practical implementation [13].

The AD protocol is as follows. Alice and Bob divide their raw strings of nit values into blocks of length L. For each block, Alice casts an n-sided die and then adds, modulo n, the random value thus found to the given block. Then she sends these modified blocks to Bob through a public, but authenticated, channel. Bob subtracts, modulo n, his corresponding blocks. Whenever he obtains a block consisting of L identical nit values, he enters this value into his distilled sequence. If, however, different values appear in a block, he disregards it. He tells Alice, through the public channel, which blocks contribute to the distilled key and which do not. She in turn then forms her own distilled sequence from the random values that she added to the blocks that Bob did not discard.

The two distilled sequences are identical, except at the rare positions, where Bob's whole block consisted of L

wrong nit values of the same kind. Since there are $n-1$ different wrong nit values, the relative frequency with which a particular one occurs in the distilled sequence is

$$B_L = \frac{\beta_1^L}{\beta_0^L + (n-1)\beta_1^L}, \tag{7}$$

which is, so to say, the new value of β_1 after AD. Since $\beta_1 < \beta_0$, B_L decreases exponentially with the block length L,

$$\lim_{L\to\infty} \frac{B_{L+1}}{B_L} = \frac{\beta_1}{\beta_0}. \tag{8}$$

Whenever Alice and Bob end up with a pair of different nit values after AD, Eve knows both values correctly, because she knows the values for each nit pair of the two blocks in question. But when Alice and Bob get the same value, which is the much more frequent situation for long blocks, Eve cannot be completely sure about any nit value in the blocks. Her best strategy is then to subtract Alice's block from her corresponding block, that is, to do what Bob does. Typically, Eve's block is inhomogeneous after the subtraction, and it is highly likely that the correct nit value is the value that occurs more often than any other. So, she decides by a majority vote which nit value to assign: She bets on the value that appears most frequently in the block, and if there are several most frequent values, she picks one of them at random.

To find her probability for assigning the right value, we first note that a block with m correct values and $k_1, k_2, \ldots, k_{n-1}$ ones of the $n-1$ wrong kinds, respectively, occurs with a relative frequency that is given by

$$\frac{L!\,\eta_0^m\,\eta_1^{L-m}}{m!\,k_1!\cdots k_{n-1}!}\,\delta_{L-m,k_1+k_2+\cdots+k_{n-1}}$$

$$= \binom{L}{m}\eta_0^m\left(\frac{\partial}{\partial x}\right)^{L-m} \prod_{j=1}^{n-1}\frac{(\eta_1 x)^{k_j}}{k_j!}\,\bigg|_{x=0}, \tag{9}$$

where the Kronecker delta symbol enforces the constraint $L - m = k_1 + k_2 + \cdots + k_{n-1}$.

Second, we note that Eve surely assigns a wrong value whenever $m < \max\{k_1, \ldots, k_{n-1}\}$, and in the situation of $m \geq \max\{k_1, \ldots, k_{n-1}\}$ she assigns the right value with probability $1/(l+1)$ where l is the count of k_j's that are equal to m. With the combinatorial factor

$$\binom{n-1}{l} = \binom{n}{l+1}\frac{l+1}{n} \tag{10}$$

taken into account, the summation over m and all k_j thus gives

$$1-(n-1)E_L = \sum_{m=0}^{L}\binom{L}{m}\eta_0^m\left(\frac{\partial}{\partial x}\right)^{L-m}\frac{1}{n}\sum_{l=0}^{n-1}\binom{n}{l+1}\left[\frac{(\eta_1 x)^m}{m!}\right]^l\left[\sum_{k=0}^{m-1}\frac{(\eta_1 x)^k}{k!}\right]^{n-1-l}_{|_{\mathrm{x}=0}} \tag{11}$$

for the probability that Eve assigns the right nit value to a string of length L, and E_L is then the probability that she gets a particular one of the $n-1$ wrong values. Parroting the remark after (7), we note that E_L is, so to say, the new value of η_1 after AD.

We use the generating function

$$E(t) \equiv \sum_{L=0}^{\infty}\frac{t^L}{L!}E_L \tag{12}$$

to deal with these probabilities as a set. It is given by

$$E(t) = \frac{e^t}{n-1}\sum_{m=0}^{\infty}\frac{(\eta_0 t)^m}{m!}e^{-\eta_0 t} \times \left(1 - \frac{[w_m(\eta_1 t)]^n - [w_{m-1}(\eta_1 t)]^n}{n[w_m(\eta_1 t) - w_{m-1}(\eta_1 t)]}\right), \tag{13}$$

where

$$w_m(x) = \sum_{k=0}^{m}\frac{x^k}{k!}e^{-x} \tag{14}$$

is the partially summed Poisson distribution.

For the comparison with (8), the quantity of primary interest is the limit

$$\lim_{L\to\infty}\frac{E_{L+1}}{E_L} = \lim_{t\to\infty}\frac{\partial}{\partial t}\ln E(t), \tag{15}$$

which directs our attention to the large-t behavior of $E(t)$.

Now, for large t the Poisson distribution in m, by which the parenthesized difference is weighted in (13), has its peak at $m \simeq \eta_0 t > \eta_1 t$, and the relative width of this peak shrinks with growing t. Accordingly, all relevant contributions to the sum in (13) have $m > \eta_1 t$, so that the approximation

$$w_m(\eta_1 t) = w_{m-1}(\eta_1 t) + \frac{(\eta_1 t)^m}{m!}e^{-\eta_1 t} \simeq 1 \tag{16}$$

is permissible, and

$$E(t) \simeq \tfrac{1}{2}e^{(1-\eta_0-\eta_1)t}I_0(2\sqrt{\eta_0\eta_1}t) \qquad \text{for } t \gg 1 \tag{17}$$

obtains. Since there is no difference between the modified Bessel function $I_0(z)$ and its derivative $I_1(z)$ when $z \gg 1$, this tells us that

$$\lim_{L\to\infty}\frac{E_{L+1}}{E_L} = 1 - (\sqrt{\eta_0} - \sqrt{\eta_1})^2. \tag{18}$$

In conjunction with (8), it follows that AD of this kind will be successful if

$$\frac{\beta_1}{\beta_0} < 1 - (\sqrt{\eta_0} - \sqrt{\eta_1})^2 \tag{19}$$

holds because then Bob's error probability gets exponentially smaller than Eve's with increasing block length L, and the distilled key sequence will meet the requirements of the CK theorem if L is chosen large enough. Now, if (4)

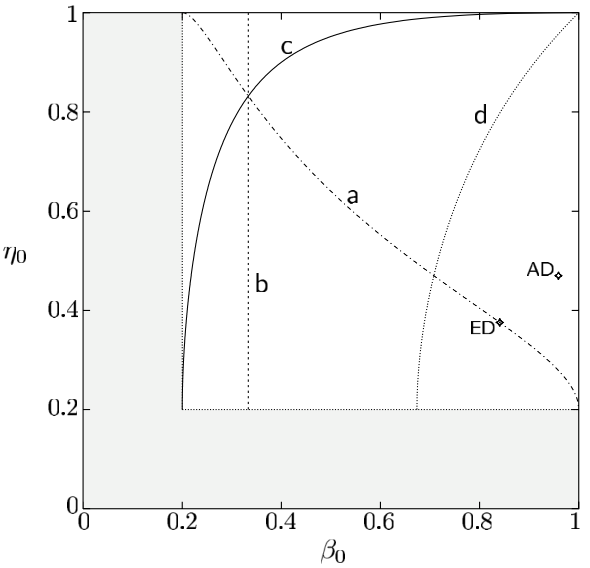

FIG. 1. The threefold coincidence, for $n = 5$, in a plot of Bob's probability β_0 vs Eve's probability η_0. The relevant values of $\beta_0 > 1/n$ and $\eta_0 > 1/n$ are outside the gray area. The dash-dotted curve a identifies the β_0, η_0 pairs for which (4) holds. To the right of the dashed vertical line b, condition (6) is obeyed and ED is possible. By contrast, AD can be successfully performed below the solid-line curve c which marks the border stated in (19). All three lines intersect at $\beta_0 = 1/3 = 0.33$, $\eta_0 = (11 + 4\sqrt{6})/25 = 0.83$, so that the part of curve a that is to the right of curve b is also the part that is below curve c.— The CK theorem is applicable to β_0, η_0 values below the dotted curve d that results from (5). It intersects curve a at $\beta_0 = 0.708$, $\eta_0 = 0.470$. From there, a single ED step takes one to the ◇ point on curve a, whereas AD with $L = 2$ moves one horizontally to the ◇ point further right (see [13]).

relates Eve's probabilities to Bob's, as it is the case for the probabilities originating in Eve's source state (3), the threshold condition (19) for classical AD is, indeed, identical with the threshold condition (6) for quantum ED, as we have asserted above. This remarkable coincidence is illustrated in Fig. 1.

It is important to note that, despite its simplicity and its lack of efficiency, the AD scheme considered correctly identifies the threshold point on curve a in Fig. 1. For, if (β_0, η_0) is to the left of the triple-coincidence point, the 2-qunit state (2) is separable. Eve can then blend it from product states and can so ensure that there is no useful mutual information between Alice and Bob, and without it they cannot generate a secure key.

For $n = 3$, a more involved argument about the same matter is given in [7]. In fact, while our paper was being written, we became aware of [7] where some of our results are conjectured and identical conclusions are reached. One wonders, of course, whether the surprising equivalence between classical and quantum distillation is more than just the coincidence as it appears here and in [7]. Perhaps it hints at a deeper connection between these fundamentally different procedures.

Finally, one might wonder if Eve has a better procedure at her disposal than the square-root measurement that is the basis of our analysis. For the following reasons we think she does not. The error-minimizing strategy takes full advantage of the built-in symmetry of the tomographic protocol. For all other fully symmetric eavesdropping attacks, the CK region is reached at smaller β_0 values, and successful AD is possible for all of them if the ED threshold (6) is crossed [14]. It seems, therefore, rather reasonable that the error-minimizing strategy is Eve's optimal choice.

We are indebted to Antonio Acín, Thomas Durt, Leong Chuan Kwek, and Norbert Lütkenhaus for sharing their insights with us. D. B., M. C., and C. M. gratefully acknowledge the splendid hospitality extended to them at the National University of Singapore. This work was supported by A*Star Grant No. 012-104-0040 and by the EC program QUPRODIS (Contract No. IST-2002-38877).

[1] Y. C. Liang, D. Kaszlikowski, B.-G. Englert, L. C. Kwek, and C. H. Oh, Phys. Rev. A (to be published).

[2] Pairwise complementary observables are ideal for this purpose, as shown by W. K. Wootters and B. D. Fields, Ann. Phys. (N.Y.) **191**, 363 (1989), who give an explicit construction if n is a power of a prime.

[3] B.-G. Englert and Y. Aharonov, Phys. Lett. A **284**, 1 (2001); P. K. Aravind, Z. Naturforsch. **58a**, 2212 (2003).

[4] See, e.g., A. Chefles, Contemp. Phys. **41**, 401 (2000), and the pertinent references therein.

[5] D. Bruß and C. Macchiavello, Phys. Rev. Lett. **88**, 127901 (2002).

[6] Put $n \rightarrow d$, $\beta_0 \rightarrow 1 - D$, and $\eta_0 \rightarrow f_d(D)$ to convert to the notation of [5].

[7] A. Acín, N. Gisin, and V. Scarani, e-print quant-ph/0303009.

[8] I. Csiszár and J. Körner, IEEE Trans. Inf. Theory **24**, 339 (1978).

[9] D. Deutsch *et al.*, Phys. Rev. Lett. **77**, 2818 (1996).

[10] M. Horodecki and P. Horodecki, Phys. Rev. A **59**, 4206 (1999).

[11] U. M. Maurer, IEEE Trans. Inf. Theory **39**, 733 (1993).

[12] For the qubit case of $n = 2$ the equivalence was shown by N. Gisin and S. Wolf, Phys. Rev. Lett. **83**, 4200 (1999).

[13] A bit of evidence—not conclusive, of course—that AD could be more efficient than ED is provided by the example of taking $n = 5$ and choosing β_0, η_0 such that $\nu = 0$ in (5). Then, accounting for both the efficiency of the distillation step and the CK efficiency (5) of the distillation product, a single ED step of the simple two-pair kind discussed in [10] gives a total yield of $\frac{1}{2} \times 0.522 \times 0.388 = 10.1\%$, whereas a single AD step with $L = 2$ has a yield of $\frac{1}{2} \times 0.522 \times 0.614 = 16.0\%$.

[14] This can be demonstrated by a suitable generalization of our analysis of AD for the error-minimizing strategy.