THE EUROPEAN
PHYSICAL JOURNAL D

On the optimality of individual entangling-probe attacks against BB84 quantum key distribution

I.M. Herbauts[1,a], S. Bettelli[2], H. Hübel[1], and M. Peev[2]

[1] Quantum Optics, Quantum Nanophysics and Quantum Information, Faculty of Physics,
University of Vienna, Boltzmanngasse 5, 1090 Vienna, Austria

[2] Austrian Research Centers GmbH - ARC, Donau-City-Straße 1, 1220 Vienna, Austria

Abstract. Some MIT researchers [Phys. Rev. A **75**, 042327 (2007)] have recently claimed that their implementation of the Slutsky-Brandt attack [Phys. Rev. A **57**, 2383 (1998); Phys. Rev. A **71**, 042312 (2005)] to the BB84 quantum-key-distribution (QKD) protocol puts the security of this protocol *"to the test"* by simulating *"the most powerful individual-photon attack"* [Phys. Rev. A **73**, 012315 (2006)]. A related unfortunate news feature by a scientific journal [G. Brumfiel, *Quantum cryptography is hacked*, News @ Nature (april 2007); Nature **447**, 372 (2007)] has spurred some concern in the QKD community and among the general public by misinterpreting the implications of this work. The present article proves the existence of a stronger individual attack on QKD protocols with encrypted error correction, for which tight bounds are shown, and clarifies why the claims of the news feature incorrectly suggest a contradiction with the established "old-style" theory of BB84 individual attacks. The full implementation of a quantum cryptographic protocol includes a reconciliation and a privacy-amplification stage, whose choice alters in general both the maximum extractable secret and the optimal eavesdropping attack. The authors of [Phys. Rev. A **75**, 042327 (2007)] are concerned only with the error-free part of the so-called sifted string, and do not consider faulty bits, which, in the version of their protocol, are discarded. When using the provably superior reconciliation approach of encrypted error correction (instead of error discard), the Slutsky-Brandt attack is no more optimal and does not "threaten" the security bound derived by Lütkenhaus [Phys. Rev. A **59**, 3301 (1999)]. It is shown that the method of Slutsky and collaborators [Phys. Rev. A **57**, 2383 (1998)] can be adapted to reconciliation with error correction, and that the optimal entangling probe can be explicitly found. Moreover, this attack fills Lütkenhaus bound, proving that it is tight (a fact which was not previously known).

PACS. 03.67.-a Quantum information – 03.67.Dd Quantum cryptography and communication security

1 Introduction

Quantum cryptography, or, more properly, quantum key distribution (QKD) is a discipline investigating techniques to grow, out of a common secret key, a larger key between two remote parties (Alice and Bob) linked by a quantum and a classical communication channel. The generated key can then be consumed to perform various classical cryptographic tasks, such as encoding messages with a one-time pad, but this is outside the scope of QKD. In the last twenty years it has been shown that it is in principle possible to grow the secret despite the channels being under the control of a non-disruptive attacker (Eve) subject only to the laws of quantum mechanics, a task deemed impossible in a completely classical setting; this ability stems ultimately from the well-known tradeoff between acquired

knowledge and state disturbance in a quantum measurement. For an introduction to the subject, the interested reader is pointed to some recent [1,2] and forthcoming [3] reviews.

Broadly speaking, QKD protocols are based on Alice transmitting quantum systems (usually photons) in randomly selected states out of an alphabet of nonorthogonal states. When Bob receives a system, he performs a measurement to infer Alice's signal; at the end of the quantum exchange, the measurement settings (but not the results) are publicly compared, and only results from compatible measurements are retained (key sifting). In the sifted key, measurement results are ideally deterministically correlated, and any eavesdropping activity, which fundamentally disturbs the exchanged systems, can be monitored. The oldest and best studied QKD procedure, described later on, is known under the name of Bennett-Brassard

[a] e-mail: isabelle.herbauts@univie.ac.at

Highlight Paper

1984 (BB84) protocol [4]; other procedures, very similar in spirit to BB84, are the entanglement-based Ekert [5] and BBM92 [6] protocols.

QKD protocols so far devised consist of (a) a quantum transmission followed by sifting over a public authenticated classical channel, establishing a highly correlated pair of keys at two remote sites; (b) a reconciliation procedure over the classical channel, allowing Alice and Bob to agree on a shared identical random key; (c) a privacy-amplification procedure over the classical channel which ensures the security of a shortened key obtained from the sifted key [7,8]. An additional necessary task for a complete secure protocol is authentication, but this is of no major consequence in the present analysis. Since the bits of the raw key are all statistically independent, no information about the sifted key can be extracted from the discarded bits of the raw key, and therefore general security analyses are concerned only with sifted keys. In both the reconciliation and the privacy amplification phases, however, information is exchanged over the classical authenticated channel, which can be perfectly spied, although not modified, by Eve. This is to be taken into account, in order that, after a sequence of appropriate procedures, both Alice and Bob possess a copy of a key, about which Eve knows only a negligible amount of information. The security of a QKD protocol, therefore, relates directly to a quantitative estimation of the amount of information potentially acquired by Eve on the sifted and reconciled key.

The conditions for the security of full QKD protocols have been extensively studied; in general, they depend on the class of allowed attacks and on the degree of non-ideality of the involved channels and cryptographic devices. In this article only individual attacks, where Eve is restricted to interact with and measure each transmitted signal independently, are considered; moreover, the channel is assumed to be noisy and potentially leaking, but the other devices are ideal and the quantum exchange is analysed only in the limit of very large keys. In this scenario, security conditions are often expressed in the form of a discarded fraction $\tau(e)$, that is the portion of the sifted and reconciled key that is to be sacrificed in order to obtain a final secret key. The discarded fraction is a function of the probability that a bit at Alice's site and the corresponding bit at Bob's site differ after sifting, i.e., the quantum-bit error rate (QBER) e; in the usual conservative approach, it must be assumed that errors in the sifted key are entirely due to Eve.

Admittedly, this is not the state of the art in QKD security proofs, since the most general class, where all signals are made to interact coherently with a very large probe which is then optimally measured by Eve (coherent attacks), has already been tackled [9,10]. Also, scenarios where Alice and Bob's devices are imperfect and potentially manipulated by Eve have been considered and partially analysed, as well as the case of finite lengths for the exchanged keys. Finally, in recent years the definition itself of what is a secure final key has changed, due to the introduction of the notion of composability. Literature on

these subjects is too large to be even cited here; the interested reader should refer to [3].

It must be remarked, however, that the case of ideal individual attacks still bears some importance because (a) proofs for realistic devices and finite key lengths are ultimately based on proofs for ideal ones; (b) security bounds for individual attacks, although conceptually very different, give results rather similar to the case of coherent attacks, which is a convincing argument about the effectiveness of eavesdropping strategies for those researchers that see coherent attacks as technologically unfeasible; and (c) individual attacks are a sufficiently simple class to be readily understood by researchers working on practical implementations, and their complete understanding helps dissipating that aura of phenomenologicality which is sometimes associated to security bounds in actual QKD protocols (as if a security bound, which is a purely mathematical statement and not an observable, could be subject to experimental investigation).

Recently, Kim et al. [11] have claimed to physically implement *"the most powerful individual-photon attack"*, therefore putting the BB84 protocol's security *"to the test"* [12]. Following their suggestion that *"the physical simulation allows investigation of the fundamental security limit of the BB84 protocol against eavesdropping in the presence of realistic physical errors, and it affords the opportunity to study the effectiveness of error correction and privacy amplification when the BB84 protocol is attacked"*, in this article this particular attack [13,14] (from now on, the *Slutsky-Brandt attack*, (SB)[1]) is analysed in the context of a complete and efficient QKD protocol.

For individual eavesdropping attacks, and using an appropriate reconciliation protocol which does not correlate signals, upper bounds on Eve's information can be estimated via the average collision probability of the sifted key. A security bound as a function of the disturbance has been derived by Lütkenhaus [15] in both scenarios when faulty bits are discarded or corrected, by modelling Eve's individual attacks by means of positive-operator-valued measurements (POVM). In Sections 2 and 3 the SB attack is analysed, and it is highlighted that this attack yields the upper value of $\tau(e)$, the discarded fraction in the privacy amplification stage, obtained by Lütkenhaus when faulty bits are rejected, therefore conferring Lütkenhaus bound the property of being sharp, as already pointed out by this author.

However, the BB84 dialect that is nowadays most commonly adopted implements the reconciliation step through error correction (instead of error discard), because this leads to a larger final secret key, as shown in Section 4. During this procedure, assumed perfect for simplicity, an amount $h(e)$ of information per sifted bit (the Shannon limit [16]) is leaked to Eve and must be discarded. In Section 5, it is shown that for such protocol the SB attack is not necessarily optimal, and in no way threatens the upper

---

[1] It is to be remarked that, despite the name of Slutsky being used for this attack in literature, the authors of [13] have never overclaimed its optimality.

bound on $\tau(e)$ as derived by Lütkenhaus for individual attacks on QKD protocols with error correction [15].

Finally, in Section 6.1, it is proven that there exist a stronger entangling-probe attack, and that this attack leads to a discarded fraction that coincides exactly with Lütkenhaus upper bound, thus abrogating the regime of hope for individual attacks against an ideal BB84 protocol with encrypted error correction subsisting thus far.

The mathematical techniques used in this article are similar to those developed by [17,18] and perfected in [13], but an important extension is introduced in Section 2.3 which allows for a significant simplification of the problem. The final result suggests an intriguing relation between the maximal collision probability achievable through an optimal measurement and the fidelity of the (mixed) states to be distinguished.

2 Modelling of individual attacks and security bounds

In general, a security proof for a given class of attacks is made out of three main ingredients. First, one needs a mathematical description (a parametrization) of all elements of the class. Then, one must estimate how dangerous each element is with respect to the final goal of establishing a secret key shared by Alice and Bob; this very much relies on the definition of security, and usually takes the form of non-tight bounds. Last, an optimisation is to be performed in the parametrized attack space in order to bound the power of the most threatening element for each value of the disturbance parameter (e.g., the QBER). The first two steps in the case of ideal individual attacks, according to the approach of Slutsky et al. [13], are reviewed in this section.

2.1 The entangling-probe model

In 1996 Fuchs and Peres, and Fuchs et al. [17,18] introduced the following individual-attack model. Eve prepares a probe and lets it interact with the signal system sent by Alice; the joint unitary evolution leaves the two systems in an entangled quantum state. The signal is then forwarded to Bob, while the probe is stored by Eve and measured after the reconciliation stage. Entanglement between the system and the probe "induces" a correlation between Eve's and Bob's measurements, allowing Eve to obtain partial information on the key. This model is known as Fuchs-Peres' entangling-probe (FPEP) attack.

The definition of individual attack does not prevent Eve from forwarding to Bob a system with a different Hilbert space from the original one, a case not covered by the FPEP model[2]. It has however been shown [15,19,20] that, if Bob's apparatus can, to some extent, reveal the presence of multiple systems in the signal, by adding a

sufficiently large penalty to the QBER in case of multiple detections it is always possible to render these attacks non-optimal for Eve[3].

That the FPEP model indeed covers the full class of individual attacks (at least among attacks where Eve is forced to measure its system at some point) is a consequence of Stinespring's dilation theorem [21], that guarantees that every completely positive and trace-preserving map can be built by embedding the input state space in the state space of a "larger" system, which is then unitarily evolved and subsequently traced down to a subsystem isomorphic to the output space. Therefore, any quantum channel can be regarded as arising from a unitary evolution on a larger (dilated) system. Embedding in a larger space can be thought of as tensoring with an auxiliary system (the probe) in a fixed initial state, because this provides an intuitive physical model. The initial state can moreover be assumed to be pure[4]. Stinespring's theorem is a generalisation of Neumark's theorem [22], that shows that every generalized measurement on a system can be implemented by letting the system interact unitarily with an ancilla, and then projectively measuring the latter[5].

The explicit FPEP parametrization for the BB84 protocol will now be introduced, following the notation of [13] as closely as possible. In BB84, Alice randomly chooses a basis from a pair $\{|u\rangle, |\bar{u}\rangle\}$ and $\{|v\rangle, |\bar{v}\rangle\}$ of mutually unbiased orthogonal bases, and a signal bit, and sends to Bob the first element of the basis if the chosen bit is 0, the second element otherwise. Bob, similarly, chooses, randomly and independently from Alice, one of the two bases, and performs a von Neumann measurement to determine the bit. The sifted key is built from those exchanges where the measurements were compatible, i.e., when both Alice and Bob chose the same basis.

If U is the unitary joint evolution of the FPEP attack, and $|w\rangle$ is the initial pure state of the probe, the overall entangled state after interaction can be decomposed as

$$U|a\rangle|w\rangle = |a\rangle|\psi_{aa}\rangle + |\bar{a}\rangle|\psi_{a\bar{a}}\rangle, \qquad (1)$$

where $a \in \{u, \bar{u}, v, \bar{v}\}$, and $|\bar{a}\rangle$ is the state corresponding to the complementary bit (the states $|\psi_{ab}\rangle$ are neither orthogonal nor normalized). When the input state $|a\rangle$ is sent by Alice, every outcome b of Bob is therefore associated to an output state of the probe proportional to $|\psi_{ab}\rangle$. It is convenient [13,17] to define an orthonormal basis $\{|e_0\rangle, |e_1\rangle\}$, oriented symmetrically with respect to

---

[2] For instance, if the signal system is a photon and the channel is an optical fibre, Eve could inject additional photons in the fibre to fool Bob's detectors.

[3] In practice, it is sufficient to insert a random bit in the sifted string for each multiple detection instead of neglecting that detection.

[4] The ability to purify each mixed state into a pure state of a larger system is again a consequence of Stinespring's theorem.

[5] In reality, the theorem asserts that a general measurements with n' possible outcomes on an n-dimensional system, with $n' > n$, can always be seen as a projective measurement on an enlarged space with n' dimensions which embeds the original state space. But the version with the measurement only on the auxiliary system is easier to visualize.

Highlight Paper

the signal states, which can then be expressed as

$$|u\rangle = + \cos\alpha \, |e_0\rangle + \sin\alpha \, |e_1\rangle, \qquad (2a)$$
$$|\bar{u}\rangle = - \sin\alpha \, |e_0\rangle + \cos\alpha \, |e_1\rangle, \qquad (2b)$$
$$|v\rangle = + \sin\alpha \, |e_0\rangle + \cos\alpha \, |e_1\rangle, \qquad (2c)$$
$$|\bar{v}\rangle = + \cos\alpha \, |e_0\rangle - \sin\alpha \, |e_1\rangle, \qquad (2d)$$

where $\alpha = \pi/8$, because the bases are unbiased. Since $|e_0\rangle$ and $|e_1\rangle$ generate the signal space, the action of a generic FPEP attack is then fully described by the action of U on them; similarly to equation (1), one defines

$$U|e_m\rangle|w\rangle = |e_0\rangle|\Phi_{m0}\rangle + |e_1\rangle|\Phi_{m1}\rangle. \qquad (3)$$

As for the $|\psi_{ab}\rangle$'s, the four states $|\Phi_{mn}\rangle$ are generally neither normalized nor orthogonal; their number shows that the probe space corresponding to a two-level signal is effectively four-dimensional.

2.2 Attack-space refinement via symmetrization

The aforementioned space of attacks is by far too complicated to be completely explored. However, standard techniques based on symmetrization are available to reduce its size without loosing potential optimal elements. The general idea is trivial: if a subset of the space is known where all attacks are *equivalent*, it is sufficient to retain only one representant of the subset during the search. What is less trivial is how to characterize and find equivalent elements. In the picture of the entangling-probe, all measurable quantities are determined by the joint state χ of the signal and probe after interaction. If $\rho_a \in \{\rho_u, \rho_{\bar{u}}, \rho_v, \rho_{\bar{v}}\}$ is a signal state and $\omega = |w\rangle\langle w|$ is the initial probe state, then

$$\chi(\rho_a, \omega, U) = U\rho_a \otimes \omega \, U^\dagger. \qquad (4)$$

The effects of an attack (U, ω), both in terms of the QBER and Eve's maximum inference power, are summarized by the statistical distribution of the χ's, which depends on the signal a priori distribution p_a, that is

$$(U, \omega) \longleftrightarrow \{ p_a; \chi(\rho_a, \omega, U) \}_{a=u,\bar{u},v,\bar{v}}. \qquad (5)$$

Since, for BB84, the a priori probabilities $p_a = 1/4$ are the same, attacks to the protocol have equivalent effects if the rays of the states are permuted (without violating the constraint that the two bases are unbiased). Readers not interested in technicalities may now just retain that the simplification of the search space implies that the vectors $|\psi\rangle$ of equation (1) can be parametrized with only two real parameters, and jump to equations (22) in Section 2.3.

All ray permutations can be generated with only two involutions, for instance (1) the basis exchange and (2) the bit exchange in the second basis; these two specific symmetries are called in the following respectively R_1 and R_2. However, the approach is more general, and can be extended to other cases, for example to the six-state variant of BB84 [23].
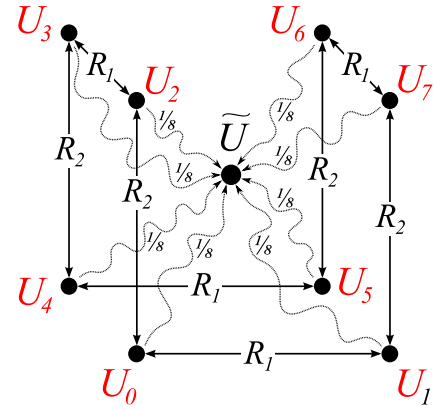

Fig. 1. (Color online) A graphical representation of the orbit $U_{g \in [1...8]}$ generated by applying the symmetry group of the BB84 protocol to a generic attack U. The whole orbit can be explored using only the involutions R_1 (basis-exchange) and R_2 (bit-exchange). The attack $\widetilde{U}$ is the average of the elements on the orbit, operates on an enlarged probe space and is symmetric under the BB84 group. The search for optimal elements can be restricted to these symmetric attacks.

Let $Q_i = R_i \otimes \mathbb{I}$ be a local operator on the joint space of the signal and the probe[6]; if Alice changes her signal convention from ρ_a into $R_i \rho_a R_i^\dagger$, and the final density matrix $\chi(R_i \rho_a R_i^\dagger, \omega, U)$ is transformed back in Bob's laboratory into $Q_i^\dagger \chi Q_i$, both the QBER and Eve's maximum inference power, which are average quantities, are statistically unchanged. It follows, very much in analogy to the passive-active picture of a reference-frame change, that the attacks (U, ω) and $(Q_i^\dagger U Q_i, \omega)$ are equivalent. In mathematical terms

$$\chi(\rho, \omega, U) \equiv Q_i^\dagger \chi(R_i \rho R_i^\dagger, \omega, U) \, Q_i$$
$$= (Q_i^\dagger U Q_i)\rho \otimes \omega \, (Q_i^\dagger U Q_i)^\dagger = \chi(\rho, \omega, Q_i^\dagger U Q_i). \qquad (6)$$

Therefore, there is a direct link between a representation of the group G of symmetries of the protocol and attack equivalence, and this remark can be exploited in a useful way. Below we consider the case of finite G, which is proper to the BB84 protocol. Since R_1 and R_2 generate the whole representation, by repeated application of equation (6) it can be shown that, for all R_g, the attack $U_g = Q_g^\dagger U Q_g$ is equivalent to $U = U_0$ (ω is omitted here, since it is always the same, and $Q_{g \in G} = R_g \otimes \mathbb{I}$). For BB84, the relevant group G is D_4 ([24], Chap. XII, Tab. 7); the action of the representation is illustrated in Figure 1. The order of the group is 8, so that the orbit of U has *at most* 8 elements.

Intuitively, a random application by Eve of attacks U_g will give another equivalent attack. The idea can be

---

[6] A local operator can be implemented without communication by Eve and Bob in their laboratories. Note that one could also define $Q_i = R_i \otimes R_i'$ with a generic unitary transformation R_i' in Eve's space, since every such transformation could be undone by Eve during her optimal measurement; but this degree of freedom does not bring additional constraints and is thus ignored here.

formalized by extending the probe space with an auxiliary space with $|G|$ dimensions. Define

$$\widetilde{U} = \sum_g U_g \otimes P_g \qquad \text{and} \qquad \widetilde{\omega} = \omega \otimes \Omega, \qquad (7)$$

where $P_g = |g\rangle\langle g|$ are orthogonal projectors in the auxiliary space, and $\Omega = |G|^{-1} \sum_{gg'} |g\rangle\langle g'|$ is the density matrix of a pure state with $\mathrm{Tr}(P_g \Omega) = 1/|G|$. The connection with the intuitive idea is that the projectors in the auxiliary space randomly select the U_g's; the construction of $(\widetilde{U}, \widetilde{\omega})$ is represented in Figure 1.

What is special about the "average" attack $(\widetilde{U}, \widetilde{\omega})$ built in this way is that it is invariant under a group, which can be built from the representation of G and some permutation operators X_g on the auxiliary space. Let

$$\widetilde{R}_g \stackrel{\text{def}}{=} Q_g^\dagger \otimes X_g = R_g^\dagger \otimes \mathbb{I} \otimes X_g \stackrel{\text{def}}{=} R_g^\dagger \otimes \widehat{R}_g. \qquad (8)$$

Operators X_g are chosen such that if $Q_g^\dagger U_\ell Q_g = U_{\pi_g(\ell)}$ then $X_g P_\ell X_g^\dagger = P_{\pi_g(\ell)}$. This is always possible due to the fundamental theorem [24, Chap. XII], that any finite group of order k is isomorphic to a subgroup of the general symmetric group of all permutations of k elements, $S(k)$, which in turn can be naturally represented by the set of all $k \times k$ permutation matrices. It is then sufficient to fix one isomorphism and chose X_g as the isomorphic image of $Q_g^\dagger$; in this way $X_g |\ell\rangle = |\pi_g(\ell)\rangle$. It is now a trivial matter to verify that

$$\widetilde{R}_g \widetilde{U} \widetilde{R}_g^\dagger = \sum_\ell Q_g^\dagger U_\ell Q_g \otimes X_g P_\ell X_g^\dagger = \widetilde{U} \qquad (9)$$

$$\text{and} \quad \widehat{R}_g \widetilde{\omega} \widehat{R}_g^\dagger = \omega \otimes X_g \Omega X_g^\dagger = \widetilde{\omega}. \qquad (10)$$

One can therefore conclude that, given a group G of protocol symmetries, for each attack (U, ω) there exists an equivalent attack $(\widetilde{U}, \widetilde{\omega})$ which is invariant under all $\widetilde{R}_g$'s as defined in equation (8). It follows that the subset containing all attacks invariant under such symmetries contains at least one optimal element; the search for optimality can thus be restricted to that subset. This finding is directly relevant to the FPEP parametrization, because it generates constraints for the $|\Phi_{mn}\rangle$'s of equation (3). In fact, for invariant attacks, replacing U with $(R_g^\dagger \otimes \widehat{R}_g) U (R_g^\dagger \otimes \widehat{R}_g)^\dagger$ and $|\omega\rangle$ with $\widehat{R}_g |\omega\rangle$ shows that

$$U R_g |e_m\rangle |\omega\rangle = \sum_n R_g |e_n\rangle \widehat{R}_g^\dagger |\Phi_{mn}\rangle, \qquad (11)$$

from which, for each symmetry R_g, the value of $\widehat{R}_g^\dagger |\Phi_{mn}\rangle$ can be calculated and used in constraints of the form

$$\langle \Phi_{mn} | \Phi_{pq} \rangle = \langle \Phi_{mn} | \widehat{R}_g \widehat{R}_g^\dagger | \Phi_{pq} \rangle. \qquad (12)$$

This formula is clearly valid for all $g \in G$, but in practice it is sufficient to restrict its application to $\widehat{R}_1$ and $\widehat{R}_2$. Also, it is more convenient to work with the symmetries of the state vectors $|a\rangle$ instead of those of the corresponding rays. This gives a representation of D_8 instead of D_4, where redundant elements are included (like $|a\rangle \to -|a\rangle$, which is physically indistinguishable from the identity); the generated constraints are however the same.

2.3 The entangling-probe parametrization

The authors of the FPEP model remarked that the BB84 protocol, as described above, is endowed with the basis-exchange symmetries R_1 (an involution corresponding to $|e_0\rangle \leftrightarrow |e_1\rangle$). Then, using essentially the same techniques described in Section 2.2, namely equation (12), they showed that an attack-dependent orthonormal basis $\{|w_i\rangle\}_{i \in 0\dots 3}$ can be found[7] such that

$$|\Phi_{00}\rangle = X_0 |w_0\rangle + X_1 |w_1\rangle + X_2 |w_2\rangle + X_3 |w_3\rangle, \qquad (13a)$$

$$|\Phi_{01}\rangle = \qquad\qquad X_5 |w_1\rangle + X_6 |w_2\rangle, \qquad (13b)$$

$$|\Phi_{10}\rangle = \qquad\qquad X_6 |w_1\rangle + X_5 |w_2\rangle, \qquad (13c)$$

$$|\Phi_{11}\rangle = X_3 |w_0\rangle + X_2 |w_1\rangle + X_1 |w_2\rangle + X_0 |w_3\rangle. \qquad (13d)$$

With analogous considerations extended to anti-unitary symmetries (complex conjugation in the probe space) they also showed that all coefficients X are real numbers. Note that this parametrization satisfies $\langle \Phi_{mn} | \Phi_{pq} \rangle = \langle \Phi_{\bar{m}\bar{n}} | \Phi_{\bar{p}\bar{q}} \rangle = \langle \Phi_{pq} | \Phi_{mn} \rangle$, given by the constraints of R_1 (as previously, the bar indicates the complementary bit). The X's are correlated by the fact that U must be a unitary operator, hence the additional constraints

$$1 = \sum_{i=0,1,2,3,5,6} X_i^2 = \|\Phi_{00}\|^2 + \|\Phi_{01}\|^2, \qquad (14a)$$

$$0 = X_1 X_6 + X_2 X_5 = \langle \Phi_{00} | \Phi_{10} \rangle = \langle \Phi_{11} | \Phi_{01} \rangle; \qquad (14b)$$

this shows that each FPEP attack, prior to Eve's measurement, can be described by only four real parameters.

However, as already said, there exists another symmetry in the BB84 protocol which has not been exploited by the authors of [13], namely R_2, the bit-exchange symmetry in one basis only. This corresponds to swapping the convention for 0 and 1 in one basis while leaving the other convention unchanged. The bit-exchange symmetry is generated by a Hadamard transformation:

$$\begin{bmatrix} |e_0\rangle \\ |e_1\rangle \end{bmatrix} \longrightarrow \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{bmatrix} |e_0\rangle \\ |e_1\rangle \end{bmatrix}. \qquad (15)$$

It is easy to check that $|v\rangle \leftrightarrow |\bar{v}\rangle$, while $|u\rangle$ and $|\bar{u}\rangle$ are invariant (actually, $|\bar{u}\rangle$ has its sign flipped, but this does not matter, since the physical state is the same). Using $R_2 |e_j\rangle = [|e_0\rangle + (-1)^j |e_1\rangle]/\sqrt{2}$, after some elementary algebraic passages, using equation (11), one obtains

$$\widehat{R}_2 |\Phi_{00}\rangle = \tfrac{1}{2} \left(|\Phi_{00}\rangle + |\Phi_{01}\rangle + |\Phi_{10}\rangle + |\Phi_{11}\rangle \right), \qquad (16a)$$

$$\widehat{R}_2 |\Phi_{01}\rangle = \tfrac{1}{2} \left(|\Phi_{00}\rangle - |\Phi_{01}\rangle + |\Phi_{10}\rangle - |\Phi_{11}\rangle \right), \qquad (16b)$$

$$\widehat{R}_2 |\Phi_{10}\rangle = \tfrac{1}{2} \left(|\Phi_{00}\rangle + |\Phi_{01}\rangle - |\Phi_{10}\rangle - |\Phi_{11}\rangle \right), \qquad (16c)$$

$$\widehat{R}_2 |\Phi_{11}\rangle = \tfrac{1}{2} \left(|\Phi_{00}\rangle - |\Phi_{01}\rangle - |\Phi_{10}\rangle + |\Phi_{11}\rangle \right). \qquad (16d)$$

Equation (12) shows how to use these relations to calculate additional constraints for $\langle \Phi_{mn} | \Phi_{pq} \rangle$ products. Of course, not all combinations of indexes are interesting, because quite a few are already fixed by other symmetries

---

[7] The absence of coefficient X_4 is due to backward compatibility.

Highlight Paper

and the unitarity of U. As already seen, there are at most four "independent" products, e.g., $\langle\Phi_{00}|\Phi_{01}\rangle$, $\langle\Phi_{01}|\Phi_{01}\rangle$, $\langle\Phi_{00}|\Phi_{11}\rangle$, and $\langle\Phi_{01}|\Phi_{10}\rangle$. The most important constraint is obtained by calculating the first one,

$$\langle\Phi_{00}|\Phi_{01}\rangle = \langle\Phi_{00}|\widehat{R}_2\widehat{R}_2^\dagger|\Phi_{01}\rangle = X_1X_5 + X_2X_6 = 0. \quad (17)$$

Together with equation (14b), this relation proves a fundamental property of the probe space for optimal attacks, i.e., this space is the direct sum of two orthogonal subspaces, one corresponding to bits received correctly by Bob and the other to errors in the sifted key,

$$\text{Span}\{|\Phi_{00}\rangle, |\Phi_{11}\rangle\} \perp \text{Span}\{|\Phi_{01}\rangle, |\Phi_{10}\rangle\}. \quad (18)$$

The symmetries analysed so far have also led to the conclusion that, within each subspace, basis vectors have the same length, $\|\Phi_{00}\| = \|\Phi_{11}\|$ and $\|\Phi_{01}\| = \|\Phi_{10}\|$, and these lengths are related by $\|\Phi_{00}\|^2 + \|\Phi_{01}\| = 1$. To determine the full geometry of the probe one therefore only needs to parametrize the intra-space products.

Applying equations (16) to the other three products, namely $\langle\Phi_{01}|\Phi_{01}\rangle$, $\langle\Phi_{00}|\Phi_{11}\rangle$, and $\langle\Phi_{01}|\Phi_{10}\rangle$ (whose calculation is greatly simplified by the previous orthogonality conditions), one obtains the desired final constraint,

$$\langle\Phi_{01}|\Phi_{10}\rangle + \langle\Phi_{00}|\Phi_{11}\rangle = 1 - 2\|\Phi_{01}\|^2. \quad (19)$$

It follows the probe space can now be parametrized with only two real parameters, the length $\|\Phi_{01}\|$ and one of the two inter-space products. In order to optimize Eve's measurement, it is handier to translate these constraints in terms of the vectors $|\psi\rangle$. Using Definitions (1, 2, 3), and solving for the $|\psi\rangle$'s, one finds

$$|\psi_{uu}\rangle = \cos^2\alpha|\Phi_{00}\rangle + \sin^2\alpha|\Phi_{11}\rangle + \sin\alpha\cos\alpha(|\Phi_{10}\rangle + |\Phi_{01}\rangle),$$
$$|\psi_{u\bar{u}}\rangle = \cos^2\alpha|\Phi_{01}\rangle - \sin^2\alpha|\Phi_{10}\rangle + \sin\alpha\cos\alpha(|\Phi_{11}\rangle - |\Phi_{00}\rangle),$$
$$|\psi_{\bar{u}u}\rangle = \cos^2\alpha|\Phi_{10}\rangle - \sin^2\alpha|\Phi_{01}\rangle + \sin\alpha\cos\alpha(|\Phi_{11}\rangle - |\Phi_{00}\rangle),$$
$$|\psi_{\bar{u}\bar{u}}\rangle = \cos^2\alpha|\Phi_{11}\rangle + \sin^2\alpha|\Phi_{00}\rangle - \sin\alpha\cos\alpha(|\Phi_{10}\rangle + |\Phi_{01}\rangle),$$

and similar relations for signals v and $\bar{v}$, which, due to the perfect symmetry of the bases, are not relevant here. Trivial but lengthy calculations show that the correspondence between the $|\psi\rangle$'s and the $|\Phi\rangle$'s is unitary (although not so easy to spot, since both vectors sets are not orthogonal and not normalized), and therefore all vector products are preserved.

Since attack optimisation is performed at constant QBER, it is better to have e as a free variable; this is easily achieved with the following reasoning. The value of the QBER cannot be changed by a local measurement by Eve after the signal-probe interaction is terminated, and, by definition, does not depend on the reconciliation procedure. From equation (1) it is immediate to understand that, if signal $|a\rangle$ is sent by Alice, an error shows up at Bob's site with probability $\langle\psi_{a\bar{a}}|\psi_{a\bar{a}}\rangle$. Considering that all signals have the same a priori probability of $1/4$, and that the parametrization, by construction, satisfies the basis-exchange symmetry, one concludes that

$$e = \frac{1}{4}\sum_{a=u,\bar{u},v,\bar{v}}\langle\psi_{a\bar{a}}|\psi_{a\bar{a}}\rangle = \frac{1}{2}\sum_{a=u,\bar{u}}\langle\psi_{a\bar{a}}|\psi_{a\bar{a}}\rangle = \|\psi_{01}\|^2. \quad (21)$$

Therefore, the vectors of the "error set", $|\psi_{01}\rangle$ and $|\psi_{10}\rangle$ have length equal to $\sqrt{e}$, and the vectors of the "good set", $|\psi_{00}\rangle$ and $|\psi_{11}\rangle$, have length equal to $\sqrt{1-e}$; moreover, the inter-space products, $\langle\psi_{00}|\psi_{11}\rangle$ and $\langle\psi_{01}|\psi_{10}\rangle$, sum up to $1 - 2e$. By introducing the inter-space imbalance δ, all these relations can be summarized as in the following table:

$$\text{Span}\{|\psi_{uu}\rangle, |\psi_{\bar{u}\bar{u}}\rangle\} \perp \text{Span}\{|\psi_{u\bar{u}}\rangle, |\psi_{\bar{u}u}\rangle\}, \quad (22a)$$
$$\|\psi_{uu}\|^2 = \|\psi_{\bar{u}\bar{u}}\|^2 = 1 - e, \quad (22b)$$
$$\|\psi_{u\bar{u}}\|^2 = \|\psi_{\bar{u}u}\|^2 = e, \quad (22c)$$
$$\langle\psi_{uu}|\psi_{\bar{u}\bar{u}}\rangle = \tfrac{1}{2} - e - \delta, \quad (22d)$$
$$\langle\psi_{u\bar{u}}|\psi_{\bar{u}u}\rangle = \tfrac{1}{2} - e + \delta. \quad (22e)$$

The imbalance is also limited by the geometrical constraint of scalar products, i.e., Schwartz inequality. The allowed values for (e, δ),

$$-\tfrac{1}{2} \leq \delta \leq +\tfrac{1}{2} - |1 - 2e|, \quad (23)$$

determined by

$$|\tfrac{1}{2} - e - \delta| \leq 1 - e, \quad (24a)$$
$$|\tfrac{1}{2} - e + \delta| \leq e, \quad (24b)$$

are represented on the left.

In the following of the article the set of equations (22) is used, still under the name of FPEP parametrization.

2.4 Estimation of Eve's inference power and the discarded fraction

As already explained in the introduction, after key reconciliation a procedure called privacy amplification is applied to reduce Eve's knowledge to negligible amount (assuming Eve is forced to measure at this point). Privacy amplification employs universal$_2$ hashing functions to compress the reconciled key, of length $\bar{n}$, to a final key, of length r. The discarded fraction τ is then defined as

$$\tau = \frac{\bar{n} - r}{\bar{n}}. \quad (25)$$

The theory of privacy amplification was developed in a seminal article by Bennett et al. [8], who found a condition for *strong security*. Lütkenhaus [25] used it to bound Eve's average[8] Shannon information on the final key: for individual attacks, the eavesdropper, on average, knows less than $1/\ln 2$ bits of the final key provided

$$\tau(e) \geq 1 + \log_2\langle P_c^1\rangle, \quad (26)$$

where $\langle P_c^1\rangle$ is the maximum *average collision probability* of Eve's knowledge of *one bit* of the reconciled key, for a

---

[8] Therefore, QKD security proofs from this period adopted *average strong security* instead of proper *strong security*, as defined in [8].
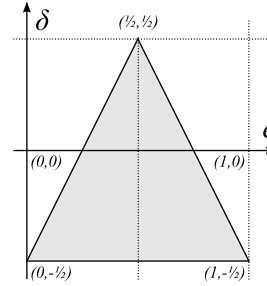
fixed value of the disturbance, the QBER e. Note that, under conservative assumptions, all noise on the quantum channel *may* be attributed to Eve, but it does not *have to*; therefore, $\tau(e)$ must be a non decreasing function. If, for instance, $\tau(e' > e) < \tau(e)$, then Eve could perform the attack causing error e, and then pass Bob's signal through a depolarazing channel with error $e' - e$. Therefore, in the following, all τ's are to be considered as monotonicized. If S is the random variable corresponding to the bit sent by Alice, with values $s = 0, 1$, and M is the random variable corresponding to all knowledge acquired by Eve, with values m, the $\langle P_c^1 \rangle$ is defined as

$$\langle P_c^1 \rangle = \sum_m P(M = m) \sum_s P^2(S = s | M = m). \quad (27)$$

However, when the approach of [13] is followed, it is not necessary to calculate the conditional probabilities $P(S = s | M = m)$ nor the marginal probabilities $P(M = m)$, because the largest possible value of $\langle P_c^1 \rangle$ can be obtained by direct inspection of the state of Eve's probe after interaction, as shown in Section 3.

3 Discarded fraction for individual attacks against a protocol using "faulty bits dumping" as reconciliation method

In Section 2.3 it was shown that the QBER e is completely determined by the signal-probe interaction during transmission. This is not the case for Eve's inference power, which depends also on the reconciliation method. Slutsky et al. [13], followed by [11,12,14], considered only the case when all errors are discarded from the sifted key. An evaluation of the cost of this procedure is postponed to Section 4; for the time being it will be assumed that it can be performed without giving Eve any piece of information other than the indexes of the retained bits.

Of course, it is very relevant to Eve that reconciliation is performed through error discard; in fact, her state of knowledge on the signal-probe system conditioned on Alice sending state $|a\rangle$ changes from that in equation (1) to a pure state, just as if Bob measurement had collapsed the signal state into $|a\rangle$,

$$U|a\rangle|w\rangle = |a\rangle|\psi_{aa}\rangle + |\bar{a}\rangle|\psi_{a\bar{a}}\rangle \xrightarrow{\text{"collapse"}} |a\rangle|\psi_{aa}\rangle. \quad (28)$$

If, for instance, the encoding basis was $\{|u\rangle, |\bar{u}\rangle\}$, Eve's probe, in Eve's view, would be in an equiprobable mixture of $|\psi_{uu}\rangle$ and $|\psi_{\bar{u}\bar{u}}\rangle$. In this case, intuitively, the largest inference power is given by a measurement that maximizes the probability to tell the first case apart from the second. It is known [26,27] that optimal ambiguous discrimination (corresponding to a minimum of the probability P_{err} of making a wrong guess) can be achieved by means of projective measurements. For two pure and normalized states, $|\phi_0\rangle$ and $|\phi_1\rangle$, assuming, without lack of generality that $\langle \phi_0 | \phi_1 \rangle \in \mathbb{R}$, and defining the pure-state fidelity as,

$$f = |\langle \phi_0 | \phi_1 \rangle|^2, \quad (29)$$

the optimal von Neumann measurement is defined by the directions $|\chi_{0/1}\rangle = \theta_{0/1}|\phi_0\rangle + \theta_{1/0}|\phi_1\rangle$, where

$$\theta_{0/1} = \frac{\sqrt{1 - \sqrt{f}} \pm \sqrt{1 + \sqrt{f}}}{2\sqrt{1 - f}}, \quad (30)$$

and the minimum error probability turns out to be $P_{\text{err}} = \frac{1}{2}[1 - \sqrt{1 - f}]$. Building on a result of Levitin [28,29], the authors of [13] showed[9] that this measurement also maximizes the average collision probability and the drop in Shannon and Rényi entropy, confirming the intuition. The maximum collision probability turns out to be

$$\langle P_c^1 \rangle = 1 - \tfrac{1}{2}f. \quad (31)$$

Therefore, in the FPEP approach, the problem of optimizing Eve's measurement is really trivial. The optimal attack is that which minimizes the value of f for a fixed value of e. Due to the intrinsic basis symmetry of the method, the value of the fidelity does not depend on the basis[10]. Using equations (22b) and (22d) one then easily finds

$$\sqrt{f} = \frac{|\langle \psi_{uu} | \psi_{\bar{u}\bar{u}} \rangle|}{\|\psi_{uu}\| \cdot \|\psi_{\bar{u}\bar{u}}\|} = \frac{|\frac{1}{2} - e - \delta|}{1 - e}, \quad (32)$$

which is minimised at fixed $e \leq 1/3$ by $\delta = 2e - 1/2$ (see the allowed range for δ in Eq. (23)), yielding:

$$\min_z|_e \sqrt{f} = \frac{1 - 3e}{1 - e} \qquad (e \leq 1/3) \quad (33)$$

(if $e > 1/3$, then, with $\delta = 1/2 - e$, the fidelity is exactly zero, i.e., the two cases are perfectly distinguishable). Substituting this result in equation (31), and then into equation (26) finally gives the maximum value of the discarded fraction (implicit in [13], and explicitly given in [12]),

$$\tau(e) = 1 + \log_2 \langle P_c^1 \rangle = \log_2(2 - f) \quad (34)$$

$$= \log_2 \frac{1 + 2e - 7e^2}{(1 - e)^2} = \log_2[1 + 4e - 4e^3 + \mathcal{O}(e^4)].$$

This formula is valid up to $e = 1/3$, where the function reaches its maximum value, $\tau(1/3) = 1$, after which Eve enjoys complete knowledge of the key established by Alice and Bob (see also the discussion of Sect. 2.4).

3.1 The Slutsky-Brandt attack

Kim et al. [11], following a proposal by Brandt [14], experimentally simulate a particular eavesdropping attack, the Slutsky-Brandt (SB) attack, that is a specific case of the general FPEP class previously described. Their practical

---

[9] Actually, the authors of [13] where interested only in Shannon and Rényi entropy; the result for the collision probability is implicit in the inequality for $\cos^2 2\zeta$ at the bottom of the first column of page 2393.

[10] Note that, since the encoding basis is known at measurement time, Eve can set up different and independent measurements for the two cases.

implementation uses a CNOT gate as entangling oper-
ation, and error-discard as reconciliation procedure. This
attack can be shown to attain the maximum collision prob-
ability, as given by equation (34), and is therefore optimal
within its class.

The SB attack is now shortly recalled. Eve employs a
probe system with the same dimensionality of the signal (a
qubit), and the entangling CNOT gate uses the signal as
control and the probe as target. The computational basis
of the CNOT is the same "symmetric" basis $\{|e_0\rangle, |e_1\rangle\}$
of equations (2); with some abuse of notation, the same
symbols $|e_0\rangle$ and $|e_1\rangle$ are used to indicate an arbitrary
basis in Eve's space. The initial probe state is

$$|w\rangle = \tfrac{1}{\sqrt{2}}\Big[(C + S)|e_0\rangle + (C - S)|e_1\rangle\Big], \qquad (35)$$

where the parameters S and C are sine and cosine of some
angle, function of the desired QBER $e \leq 1/2$:

$$S = \sqrt{2e}, \qquad C = \sqrt{1 - 2e}. \qquad (36)$$

The total system, upon Eve's action, becomes entangled,
and its state can be decomposed according to the defini-
tion of equation (1), giving

$$|\psi_{u\bar{u}}^{uu}\rangle = C\frac{|e_0\rangle + |e_1\rangle}{\sqrt{2}} \pm \frac{1}{\sqrt{2}}S\frac{|e_0\rangle - |e_1\rangle}{\sqrt{2}}, \qquad (37)$$

$$|\psi_{\bar{u}u}^{u\bar{u}}\rangle \stackrel{\text{def}}{=} |T_e\rangle \stackrel{\text{def}}{=} \frac{1}{\sqrt{2}}S\frac{|e_1\rangle - |e_0\rangle}{\sqrt{2}}. \qquad (38)$$

Similar equations hold in the other basis. The probability
of having an error is, as expected, $\langle T_e|T_e\rangle = S^2/2 = e$.
"Error states", that is the states $|\psi_{a\bar{a}}\rangle$, are characterized
by independence from the actual signal a, as they are al-
ways equal to $|T_e\rangle$. As a consequence of this, when an error
takes place, Eve has no information at all on the trans-
mitted bit – the entangling unitary is in fact optimized for
protocols which discard errors instead of correcting them.

The inference power of the SB attack can be calcu-
lated, as already seen, from the fidelity of $|\psi_{uu}\rangle$ with re-
spect to $|\psi_{\bar{u}\bar{u}}\rangle$; for $e \leq 1/3$, it is identical to that of equa-
tion (33), which proves that this attack is optimal in the
class of attacks on protocols which discard errors of the
sifted key:

$$\sqrt{f} = \frac{|\langle\psi_{uu}|\psi_{\bar{u}\bar{u}}\rangle|}{\|\psi_{uu}\| \, \|\psi_{\bar{u}\bar{u}}\|} = \frac{|2C^2 - S^2|}{2C^2 + S^2} = \frac{|1 - 3e|}{1 - e}. \qquad (39)$$

4 Reconciliation: error discard versus error correction

As emphasized earlier, a QKD protocol, like BB84, can
be implemented in many variants, by adopting different
approaches for reconciliation. Each of these dialects is a
protocol on its own, and trivially comparing the discarded
fraction for different protocols makes as much sense as

comparing apples with pears. However, a common bench-
mark can be found in the length of the final secret with
respect to the length n of the sifted key (not the length $\bar{n}$
of the reconciled key).

The problem is further complicated by the fact that
the privacy-amplification bound is based on the average
collision probability of the sifted *and* reconciled key. If
reconciliation is performed in clear, by exchanging pub-
lic messages on the classical channel, $\langle P_c^1\rangle$ of the sifted
key is modified in ways that are very difficult to account
for. For this reason, it is established practice to exchange
reconciliation information in encrypted form, with a one-
time pad. This, of course, requires a previous secret to be
shared by Alice and Bob; this secret is consumed during
the execution of the protocol, and must enter the final bal-
ance of secret key production. The alternative approach
of exchanging public messages and then reducing the final
key of an equivalent amount has never been proven to be
more efficient, but it is more difficult to justify theoreti-
cally (see, e.g., [30]).

Articles on BB84 with error discard usually do not
mention an explicit procedure for discarding faulty bits;
but it is clear that locating *all* errors in the sifted key
is exactly as difficult as correcting the string altogether
(since the output of one procedure can be directly used
to implement the other one), which implies a minimum
cost $nh(e)$, where h is the binary entropy function $h(e) = -e\log_2 e - (1-e)\log_2(1-e)$, due to the Shannon limit [16].
The secret gain is therefore at most

$$G_d = n(1 - e)(1 - \tau_d(e)) - nh(e), \qquad (40)$$

because (1) the sifted key of length n is reduced to a rec-
onciled key of length $\bar{n} = n(1 - e)$ by discarding the ne
errors, (2) the reconciled key is compressed by a factor
$1 - \tau_d$ during privacy amplification, and (3) the cost of
tight error discard, $nh(e)$, must be subtracted from the
final balance. The subscript $_d$ of τ is meant to remember
that this is the discarded fraction in case of reconciliation
through error discard. This gain can be directly compared
with that of protocols with error correction. In the latter
case, $\bar{n} = n$ (no bits are discarded), and τ becomes τ_c:

$$G_c = n(1 - \tau_c(e)) - nh(e). \qquad (41)$$

Obviously, $0 \leq \tau_c \leq \tau_d \leq 1$, because more information is
available to Eve with error discard than with error correc-
tion (i.e., the location of all bits received as errors, and the
fact that all retained bits were received without errors).
One can consider also a case in which errors are corrected,
but the positions of the corrected spots is leaked to Eve[11];
the previous considerations are not invalidated. It is im-
mediate to see that error correction is always better than
error discard, because

$$\frac{G_c - G_d}{n} = (1 - e)(\tau_d - \tau_c) + e(1 - \tau_c) \geq 0. \qquad (42)$$

---

[11] The case of "leaked errors" is considered because it simpli-
fies a lot of calculations, and is anyway an upper bound to the
case of perfectly encrypted error correction.

Therefore, it makes sense to see what happens to the "optimal BB84 attack" when reconciliation is done through error correction, a case analysed in Section 5. One may legitimately think that other reconciliation procedures could lead to an even larger gain; for instance, an algorithm could select an error-free part of the sifted string of length $\bar{n}$ by exchanging a message shorter than $\bar{n}h(e)$, as long as $\bar{n} < n$. The overall secret gain is most probably not larger than G_c, but this statement has never been formally proved. Other variants might be explored, like reconciling Alice's key to the sifted key of Bob, instead of the opposite, or changing both to a third common string, or merging reconciliation and privacy amplification into a single step, or even replacing standard privacy amplification with some other procedure in order to get closer to the $I(A : B) - I(A : E)$ bound. However, one should also remember that QKD proofs are not after finding the "optimal" protocol, but after proving that a given, probably sub-optimal but reasonably efficient protocol is secure under some conditions.

Changing the focus from one protocol to another is moreover often not a good idea because QKD proofs are a lengthy and expensive collective effort, which must be to some extent restarted when the protocol is changed. And all this, not to speak of the apparent impossibility to parametrize the space of "all possible QKD protocols". For QKD protocols, standardization is more important than optimisation.

5 The Slutsky-Brandt attack with an error-correction procedure

The SB attack will now be analysed in the context of a BB84 protocol using encrypted error correction, in order to investigate its claimed optimality. Because of this choice for reconciliation, the amount of information leaked to Eve during the raw exchange plus the knowledge of the encoding basis is all what concerns the calculation of the average collision probability. Since only individual attacks are allowed, one can consider Eve's activity as being performed on two separate strings of $n(1-e)$ correct bits and ne faulty bits respectively. The discarded fraction can thus be written as

$$\tau(e) = (1-e)\tau_= + e\tau_{\neq}, \qquad (43)$$

where the first term is related to correct bits and the second one to faulty bits; this expression is equivalent to

$$\tau(e) = 1 + \log_2\left(\langle P_{c=}^1\rangle^{1-e}\langle P_{c\neq}^1\rangle^e\right), \qquad (44)$$

where $\langle P_{c=}^1\rangle$ and $\langle P_{c\neq}^1\rangle$ are the individual average collision probabilities for error-free and faulty bits respectively. $\tau_=$ is obviously the same quantity determined in Section 2 for the SB attack, see equation (34). To calculate the amount of information leaked to Eve from erroneous bits, note that when the bit measured by Bob is wrong, the state of the probe collapses to $|T_e\rangle$, equation (38), independently
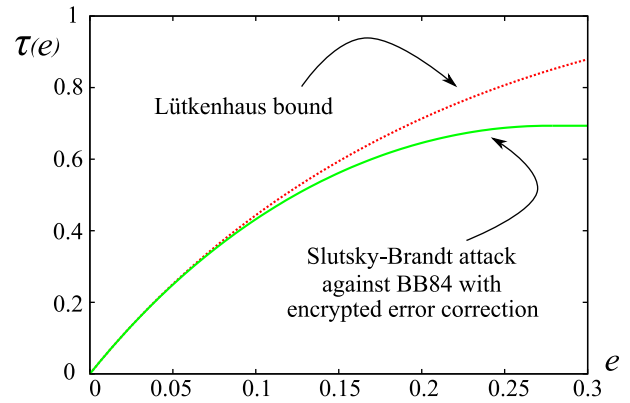

Fig. 2. (Color online) The fraction of the sifted key that must be discarded during privacy amplification in order to counter a SB attack against a protocol with encrypted error correction, equation (45), versus the QBER e compared with Lütkenhaus bound ([15], Eq. (46)). The first curve reaches its maximum at $e \sim 0.277$, the bound at $e = 0.5$, where its value is 1. The curves are non-decreasing, see the discussion in Section 2.4.

from the bit sent by Alice and the encoding basis. Therefore, Eve has no mean to distinguish between Alice's two equiprobable bits, and consequently $\tau_{\neq} = 0$. Using equations (43) and (34) one finds

$$\tau(e) = (1-e)\log_2(1 + 4e - 4e^3 + \mathcal{O}(e^4))$$
$$= \log_2\left(1 + 4e - 4e^2 - 12e^3 + \mathcal{O}(e^4)\right). \qquad (45)$$

This discarded fraction can now be compared to the general scenario of individual attacks considered by Lütkenhaus in the momentous paper [15], where the author concludes that $\tau(e)$ is bounded by

$$\tau(e) \leq \log_2(1 + 4e - 4e^2). \qquad (46)$$

In Figure 2, the discarded fraction necessary to counter a SB attack is compared with Lütkenhaus bound (which was not claimed to be tight). The latter is always higher, hence stronger, than the security curve derived from the SB attack, the two curves merging only at $e = 0$. For small error rates, most bits are exchanged correctly and, as the SB attack on correct bits is optimal, the curves converge. When more errors are introduced, Eve's lack of information on faulty bits weakens her attack.

This shows that in a QKD protocol with encrypted error correction, the SB attack does not fill the known upper bound, leaving potential room for stronger individual attacks. The SB curve is however a lower bound, since the eavesdropping strategy is given explicitly. In the next section, the question will be investigated whether a stronger FPEP attack can be found, by appropriately balancing the amount of information Eve can gain from error-free bits and from bits received incorrectly by Bob.

Highlight Paper

6 An optimal attack against BB84 with error correction

6.1 With leakage of error positions

This section revisits the FPEP class of attacks against a BB84 QKD protocol where errors of the sifted key are corrected; however, it is assumed that the positions of these errors become known to the eavesdropper. This latter apparently peculiar hypothesis is investigated also in [15], where the author shows that, due to spoiling information, this case can be used to draw an upper bound also for more secure protocols where Eve has no information about which bits were received incorrectly by Bob.

The approach to the security proof is very similar to that presented in Section 3, the difference being, that here, for a given, known encoding basis, Eve must distinguish between two pure states for bits received correctly, and two different pure states for bits received incorrectly, since there are two possibilities for the "collapse" of equation (28). For instance, if the basis is $\{|u\rangle, |\bar{u}\rangle\}$ and the bit was received incorrectly (which happened with probability e), Eve must distinguish between $|\psi_{u\bar{u}}\rangle$ and $|\psi_{\bar{u}u}\rangle$; if the bit was instead received correctly, the two states are, as before, $|\psi_{uu}\rangle$ and $|\psi_{\bar{u}\bar{u}}\rangle$. The results for the second encoding basis are identical, due to the intrinsic symmetry of the FPEP method. Equation (31) is thus changed into

$$\langle P_c^1 \rangle = \left(1 - \tfrac{1}{2}f_{[=]}\right)^{1-e} \left(1 - \tfrac{1}{2}f_{[\neq]}\right)^e, \qquad (47)$$

with $f_{[=]}$ and $f_{[\neq]}$ defined by the following expressions (which are then simplified with Eqs. (22b, c, d, e)), where the imbalance δ is constrained by equation (23):

$$\sqrt{f_{[=]}} = \frac{|\langle \psi_{uu}|\psi_{\bar{u}\bar{u}}\rangle|}{\|\psi_{uu}\| \, \|\psi_{\bar{u}\bar{u}}\|} = \frac{|\tfrac{1}{2} - e - \delta|}{1 - e}, \qquad (48a)$$

$$\sqrt{f_{[\neq]}} = \frac{|\langle \psi_{u\bar{u}}|\psi_{\bar{u}u}\rangle|}{\|\psi_{u\bar{u}}\| \, \|\psi_{\bar{u}u}\|} = \frac{|\tfrac{1}{2} - e + \delta|}{e}. \qquad (48b)$$

In order to find the optimal attack, it is now sufficient to maximize the collision probability in equation (47) over δ. It is easier to visualize the optimisation problem through the discarded fraction. In fact, note that

$$\tau = (1-e)\log_2(2 - f_{[=]}) + e\log_2(2 - f_{[\neq]})$$

$$\leq \log_2\left[(1-e)(2 - f_{[=]}) + e(2 - f_{[\neq]})\right]. \qquad (49)$$

Finding the maximum $2\delta = -(1-2e)^2$ of the upper bound is trivial since the argument is a second-degree polynomial in δ. But, for this value of δ, the two fidelities are equal, and therefore inequality (49) is filled, and the optimisation problem is solved. One obtains

$$f_{[=]} = f_{[\neq]} = f_{\min}(e) = (1-2e)^2, \quad \text{and} \qquad (50)$$

$$\tau(e) = \log_2\left[2 - f_{\min}(e)\right] = \log_2(1 + 4e - 4e^2), \qquad (51)$$

which is exactly Lütkenhaus bound of equation (46). Whereas previously this upper bound allowed some margin for lower security bounds to be found, the present optimisation proves it to be tight when error positions

are leaked[12]. Note that, due to the symmetry $e \leftrightarrow 1 - e$, the discarded fraction cannot be a monotonous curve in this case. Above $e = 50\%$, Eve's tactic for total knowledge cannot be modelled by the unitary matrix of the FPEP parametrization; an additional dissipative evolution on Bob's bit is necessary.

6.2 Without leakage of error positions

The previous section considered the implementation of a QKD protocol with error correction and leakage of the positions of the errors, because that assumption makes the mathematical derivation particularly simple. However, more secure error-correcting protocols can be devised, in which Eve has no access to this piece of information. This section investigates whether a different bound is proper to this instance.

With Eve's assumed lack of knowledge on the error positions, the final state of the probe after the entangling evolution and the "collapse" at Bob's site is the density matrix $\sigma = \mathrm{Tr}_{\mathrm{Bob}}(\chi)$, with χ being the joint state of the probe and the signal. The state σ will be a statistical mixture, over Bob's possible outcomes; namely

$$\sigma_a = |\psi_{aa}\rangle\langle\psi_{aa}| + |\psi_{a\bar{a}}\rangle\langle\psi_{a\bar{a}}|, \qquad (52)$$

when the input state $|a\rangle$ is sent by Alice; note that $|\psi_{aa}\rangle$ and $|\psi_{a\bar{a}}\rangle$ are not normalized; if the normalized vectors were used instead, the two addends would have a factor $1 - e$ and e respectively in front. Eve must distinguish between the two density matrices ensuing from the two equiprobable states $|a\rangle$ of Alice, with $a \in \{u, \bar{u}\}$.

Suppose that Eve implements the following measurement strategy, on which there is, a priori, no claim of optimality. First she performs a projective measurement to separate the $\{|\psi_{uu}\rangle, |\psi_{\bar{u}\bar{u}}\rangle\}$ subspace from the

---

[12] From the derivations of Section 2.3, explicit optimal attacks can be devised using the expressions for the unitary matrix U given by the authors of the FPEP model [17]. Indeed, these authors express the X's in terms of only four real angles $\{\lambda, \mu, \phi, \theta\}$ in the range $[0, 2\pi]$. The problem of finding an optimal attack is solved by finding matrix elements X's, satisfying the conditions equations (17) and (19), and for which

$$2\delta = -(1-2e)^2 = -2X_0X_3 - X_1X_2 + 2X_5X_6.$$

Using the parameters used in [13], it follows from equations (14b) and (17) that the parameters a, b, c, d are further constrained, so that $c = 0$, and $a = d = 1 - 2e$. The optimisation problem is therefore reduced to the simple task of finding values of some angles $\{\lambda, \mu, \phi, \theta\}$ for which

$$b = \sin^2\lambda\sin 2\mu + \cos^2\lambda\sin 2\phi = (1-2e)^2,$$

given the conditions $a = d$ and $c = 0$ on

$$a = \sin^2\lambda\sin 2\mu + \cos^2\lambda\cos 2\theta\sin 2\phi,$$

$$d = \sin^2\lambda + \cos^2\lambda\cos 2\theta,$$

$$\text{and} \quad c = \cos^2\lambda\sin 2\theta\cos 2\phi.$$

$\{|\psi_{u\bar{u}}\rangle, |\psi_{\bar{u}u}\rangle\}$ subspace (finding the first case with probability $1-e$, and the second one with probability e, but this is irrelevant); the separation is possible because the two subspaces are orthogonal, as shown in Section 2.3. Then, if the first outcome was found, she proceeds with the same measurement of Section 6.1 for this case, achieving a minimum fidelity equal to $f_{[=]}$; similarly, for the second case, she achieves $f_{[\neq]}$. Given that, for these two measurements, both $f_{[=]}$ and $f_{[\neq]}$ have the same value $f_{\min}(e) = (1-2e)^2$, the average $\langle P_c^1 \rangle$ turns out to be the same as for the case of error correction with leakage of error positions.

Therefore, there exists a measurement strategy which is ignorant of the positions of the errors and fills the bound of equation (51). It is however obvious that all attack strategies that can be implemented without this piece of knowledge can be implemented also if it is available: in other words, the set of allowed attacks without leakage is strictly included in the set with leakage, and therefore, the security bound for the current case cannot exceed the security bound of Section 6.1. Thus, the explicit attack just shown implies that the two bounds are the same, and that the attack itself is optimal.

It is remarkable that, similarly to equation (31), also in this case the maximum collision probability is linked to the fidelity [31] of the conditional density matrices σ_u and $\sigma_{\bar{u}}$. The calculation is greatly simplified by the subspaces $\{|\psi_{uu}\rangle, |\psi_{\bar{u}\bar{u}}\rangle\}$ and $\{|\psi_{u\bar{u}}\rangle, |\psi_{\bar{u}u}\rangle\}$ being orthogonal; using equations (48) one obtains

$$f(\sigma_u, \sigma_{\bar{u}}) = \mathrm{Tr}^2 \sqrt{\sqrt{\sigma_u}\, \sigma_{\bar{u}}\, \sqrt{\sigma_u}} = \left[(1-e)f_{[=]}^{\frac{1}{2}} + e f_{[\neq]}^{\frac{1}{2}} \right]^2$$
$$= \left(|\frac{1}{2} - e - \delta| + |\frac{1}{2} - e + \delta| \right)^2 = (1-2e)^2,$$

and therefore $\langle P_c^1 \rangle = 1 - \frac{1}{2}f$. This identity may be true here only due to the large number of constraints dictated by the symmetries of the BB84 protocol. However, it would be interesting to know whether the result holds more generally. This problem is somehow similar to that of minimum error probability or accessible information. Despite intuition, it is known [29,32] that these two are not equivalent for mixed states. It is likely that the maximization of the collision probability is still a different problem. Formally, the problem would read like this: provided a flat bit S is transmitted through a quantum channel, encoded in non-orthogonal density matrices ρ_0 and ρ_1, what is the maximum collision probability of the distribution of S that can be reconstructed by the receiver by means of quantum measurements?

7 Conclusions

It has been shown that no real "threat" to the security of BB84 QKD protocols stems from recent developments in implementing an entangling probe attack. Not only is this attack (claimed to be the "most powerful individual attack" [11,12]) not threatening the security bound derived previously by Lütkenhaus [15], but it is also shown to be sub-optimal in an efficient and complete QKD implementation. The SB attack is only an optimal attack for those specific types of QKD protocols in which the reconciliation procedure is to somehow discard all faulty bits, which is a less desirable scheme as it leads to a shorter final shared key.

It should also be pointed out that experiments cannot allow for the investigation of fundamental security limits, as "security" is not an observable; they can only shed light on the technological feasibility of specific eavesdropping attacks.

In view of the previous considerations, the recent headline in Nature purporting that "quantum cryptography is hacked" [33,34] as a result of the successful implementation of an SB attack is an unfortunate misunderstanding. In fact, the researchers whose work is highlighted in the news feature do not themselves make any such sensationalistic claim, even though they fail to mention existing security proofs and do not comment on the consequences their attack has on existing security bounds.

In this paper it has been shown that improved analysis of FPEP attacks leads to finding explicit optimal attacks for the case considered in [15], filling the bound introduced there, which therefore turns out to be sharp. This holds independently of whether error positions are leaked to Eve. The analysis gives a simple recipe for devising optimal individual attacks, the most powerful eavesdropping attacks that could be implemented with nowadays technology. The complete statement is the following. An ideal BB84 QKD exchange where the dimensionality of the signal space is not changed and the imperfection of the experimental apparatus consists at most in a noisy and lossy channel, and for which reconciliation through error correction is performed, followed by privacy amplification, is strongly secure on average against individual attacks if and only if the discarded fraction $\tau(e)$ satisfies

$$\tau(e) \geq \log_2(1 + 4e - 4e^2)$$

(where e is the QBER of the sifted key) both in the case that the positions of errors come to be known to the eavesdropper, and in the case that they do not. A byproduct of this analysis is the question whether the maximum collision probability in distinguishing two mixed density matrices is always one minus one half of the fidelity of the carrier states.

References

1. N. Gisin, G. Ribordy, W. Tittel, H. Zbinden, Rev. Mod. Phys. **74**, 145 (2002)
2. M. Dušek, N. Lütkenhaus, M. Hendrych, in *Quantum Cryptography, Progress in Optics*, edited by E. Wolf (Elsevier, 2006), Vol. 49, Chap. 5

3. V. Scarani, H. Bechmann-Pasquinucci, N.J. Cerf, M. Dušek, N. Lütkenhaus, M. Peev, *A framework for practical quantum cryptography*, in preparation
4. C.H. Bennett, G. Brassard, in *Proc. of IEEE International Conference on Computers, Systems, and Signal Processing, Bangalore, India*, pp. 175–179 (1984)
5. A.K. Ekert, Phys. Rev. Lett. **67**, 661 (1991)
6. C.H. Bennett, G. Brassard, D.N. Mermin, Phys. Rev. Lett. **68**, 557 (1992)
7. U.M. Maurer, IEEE Trans. Inf. Theory **39** 733 (1993)
8. C.H. Bennett, G. Brassard, C. Crépeau, U.M. Maurer, IEEE Trans. Inf. Theory **41**, 1915 (1995)
9. H. Inamori, N. Lütkenhaus, D. Mayers, Eur. Phys. J. D **41**, 599 (2007)
10. D. Gottesman, H.-K. Lo, N. Lütkenhaus, J. Preskill, Quantum Inform. Comput. **4**, 325 (2004)
11. T. Kim, I.S. genannt Wersborg, F.N.C. Wong, J.H. Shapiro, Phys. Rev. A **75**, 042327 (2007)
12. J.H. Shapiro, F.N.C. Wong, Phys. Rev. A **73**, 012315 (2006)
13. B.A. Slutsky, R. Rao, P.-C. Sun, Y. Fainman, Phys. Rev. A **57**, 2383 (1998)
14. H.E. Brandt, Phys. Rev. A **71**, 042312 (2005)
15. N. Lütkenhaus, Phys. Rev. A **59**, 3301 (1999)
16. C.E. Shannon, Bell Syst. Tech. J. **27**, 379 and 623 (1948).
17. C.A. Fuchs, A. Peres, Phys. Rev. A **53**, 2038 (1996)
18. C.A. Fuchs, N. Gisin, R.B. Griffiths, C.-S. Niu, A. Peres, Phys. Rev. A **56**, 1163 (1997)
19. N. Lütkenhaus, *Generalised Measurements and Quantum Cryptography*, Ph.D. thesis, University of Strathclyde, Glasgow (1996)
20. E. Waks, A. Zeevi, Y. Yamamoto, Phys. Rev. A **65**, 052310 (2002)
21. W.F. Stinespring, Proc. Amer. Math. Soc. **6**, 211 (1955)
22. M.A. Neumark, Izv. Akad. Nauk. SSSR, Ser. Mat. **4**, 277 (1940)
23. D. Bruss, Phys. Rev. Lett. **81**, 3018 (1998)
24. L.D. Landau, E.M. Lifshitz, *Quantum Mechanics. Non-relativistic Theory, Course of Theoretical Physics*, 3rd edn. (Butterworth Heinemann, Oxford, 1981), Vol. 3
25. N. Lütkenhaus, Phys. Rev. A **54**, 97 (1996)
26. C.W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, 1976)
27. C.A. Fuchs, *Distinguishability and Accessible Information in Quantum Theory*, Ph.D. thesis, University of New Mexico (1996)
28. L.B. Levitin, in *IEEE Intern. Symp. on Information Theory*, Santa Monica, CA, USA (1981)
29. L.B. Levitin, in *Quantum Communication and Measurement*, edited by V.P. Belavkin, O. Hirota, R.L. Hudson (Plenum, New York, 1995) pp. 439–448. Proceedings of QCM94
30. C. Cachin, U.M. Maurer, J. Cryptol. **10**, 97 (1997)
31. R. Jozsa, J. Mod. Opt. **41**, 2315 (1994)
32. C.A. Fuchs, C.M. Caves, Phys. Rev. Lett. **73**, 3047 (1994)
33. G. Brumfiel, *Quantum cryptography is hacked*, News @ Nature (april 2007). Online feature whose summary reads: "*Simulation proves it's possible to eavesdrop on super-secure encrypted messages*"
34. G. Brumfiel, Nature **447**, 372 (2007). The editor's summary starts with: "*Quantum cryptography is 100% hack-proof. Or at least it was, until the hackers got cracking. Recent simulations suggest that it is only a matter of time before a quantum-mechanical method of eavesdropping on super-secure encrypted messages is developed...*"

Highlight Paper