Quantum Cryptography Protocols Robust against Photon Number Splitting Attacks for Weak Laser Pulse Implementations

Scarani, Valerio; Acin, Antonio; Ribordy, Grégoire; Gisin, Nicolas

Quantum Cryptography Protocols Robust against Photon Number Splitting Attacks for Weak Laser Pulse Implementations

Valerio Scarani,[1] Antonio Acín,[1] Grégoire Ribordy,[2] and Nicolas Gisin[1]

[1]*Group of Applied Physics, University of Geneva, 20, rue de l'Ecole-de-Médecine, CH-1211 Geneva 4, Switzerland*

[2]*Id-Quantique, rue Cingria 10, CH-1205 Geneva, Switzerland*

(Received 22 November 2002; published 6 February 2004)

We introduce a new class of quantum key distribution protocols, tailored to be robust against photon number splitting (PNS) attacks. We study one of these protocols, which differs from the original protocol by Bennett and Brassard (BB84) only in the classical sifting procedure. This protocol is provably better than BB84 against PNS attacks at zero error.

DOI: 10.1103/PhysRevLett.92.057901 PACS numbers: 03.67.Dd, 03.67.Hk

Quantum cryptography, or more precisely quantum key distribution (QKD) is the only physically secure method for the distribution of a secret key between two distant partners, Alice and Bob [1]. Its security comes from the well-known fact that the measurement of an unknown quantum state modifies the state itself: thus an eavesdopper on the quantum channel, Eve, cannot get information on the key without introducing errors in the correlations between Alice and Bob. In equivalent terms, QKD is secure because of the no-cloning theorem of quantum mechanics: Eve cannot duplicate the signal and forward a perfect copy to Bob.

In recent years, several long-distance implementations of QKD have been developed, that use photons as information carriers and optical fibers as quantum channels [1]. Most often, although not always [2], Alice sends to Bob a *weak laser pulse* in which she has encoded the bit. Each pulse is *a priori* in a coherent state $|\sqrt{\mu}e^{i\theta}\rangle$ of weak intensity, typically $\mu \approx 0.1$ photons. However, since no reference phase is available outside Alice's office, Bob and Eve have no information on θ. Consequently, they see the mixed state $\rho = \int \frac{d\theta}{2\pi} |\sqrt{\mu}e^{i\theta}\rangle\langle\sqrt{\mu}e^{i\theta}|$. This state can be rewritten as a mixture of Fock states, $\sum_n p_n |n\rangle\langle n|$, with the number n of photons distributed according to the Poissonian statistics of mean μ, $p_n = p_n(\mu) = e^{-\mu}\mu^n/n!$. Because two realizations of the same density matrix are indistinguishable, QKD with weak pulses can be reinterpreted as follows: Alice encodes her bit in one photon with frequency p_1, in two photons with frequency p_2, and so on, and does nothing with frequency p_0. Thus, in weak pulses QKD, a rather important fraction of the nonempty pulses actually contains more than one photon. For these pulses, Eve is then no longer limited by the no-cloning theorem: she can simply keep some of the photons while letting the others go to Bob. Such an attack is called *photon number splitting* (PNS) attack. Although PNS attacks are far beyond today's technology [3], if one includes them in the security analysis, the consequences are dramatic [4,5].

In this Letter, we present new QKD protocols that are secure against PNS attack up to significantly longer distances, and that can thus lead to a secure implementation of QKD with weak pulses. These protocols are better tailored than the ones studied before to exploit the correlations that can be established using ρ. The basic idea is that Alice should encode each bit into a pair of *nonorthogonal* states belonging to *two or more* suitable sets.

The structure of the paper is as follows. First, we review the PNS attack on the first and best-known QKD protocol, proposed by Bennett and Brassard in 1984 and thence called BB84 [6], in order to understand why this attack is really devastating when the bit is encoded into pairs of orthogonal states. Then we present the benefits of using nonorthogonal states, mostly by focusing on a specific new protocol which is a simple modification of the BB84.

PNS attacks on the BB84 protocol.—Alice encodes each bit in a qubit, either as an eigenstate of σ_x ($|+x\rangle$ coding 0 or $|-x\rangle$ coding 1) or as an eigenstate of σ_z ($|+z\rangle$ coding 0 or $|-z\rangle$ coding 1). The qubit is sent to Bob, who measures either σ_x or σ_z. Then comes a classical procedure known as "sifting" or "basis reconciliation": Alice communicates to Bob through a public classical channel the basis, x or z, in which she prepared each qubit. When Bob has used the same basis for his measurement, he knows that (in the absence of perturbations, and, in particular, in the absence of Eve) he has the correct result. When Bob has used the wrong basis, the partners simply discard that item.

Consider now the implementation of the BB84 protocol with weak pulses. Bob's raw detection rate is the probability that he detects a photon per pulse sent by Alice. In the absence of Eve, this is given by

$$R_{\mathrm{raw}}(\delta) = \sum_{n\geq 1} p_n [1 - (1 - \eta_{\mathrm{det}}\eta_\delta)^n] \simeq \eta_{\mathrm{det}}\eta_\delta\,\mu, \quad (1)$$

where η_{det} is the quantum efficiency of the detector (typically 10% at telecom wavelengths), and η_δ is attenuation due to the losses in the fiber of length ℓ:

$$\eta_\delta = 10^{-\delta/10}, \qquad \delta = \alpha\ell\,[\mathrm{dB}]. \quad (2)$$

Below, when we give a distance, we assume the typical

value $\alpha = 0.25$ dB/km. The approximate equality in (1) is valid if $\eta_{\rm det}\eta_\delta\, p_n\, n \ll 1$ for all n, which is always the case in weak pulses QKD.

If we endow Eve with unlimited technological power within the laws of physics, the following PNS attack (*storage attack*) is in principle possible [4,5]: (i) Eve counts the number of photons, using a photon number quantum nondemolition (QND) measurement; (ii) she blocks the single-photon pulses, and for the multiphoton pulses she stores one photon in a quantum memory; she forwards the remaining photons to Bob using a perfectly transparent quantum channel, $\eta_\delta = 1$ [7]; (iii) she waits until Alice and Bob publicly reveal the used bases and correspondingly measures the photons stored in her quantum memory: she has to discriminate between two orthogonal states, and this can be done deterministically. In this way, Eve has obtained full information about Alice's bits, thence no processing can distill secret keys for the legitimate users; moreover, Eve has not introduced any error on Bob's side.

The unique constraint on PNS attack is that Eve's presence should not be noticed; in particular, Eve must ensure that the rate of photons received by Bob (1) is not modified [9]. Thus, the PNS attack can be performed on all pulses only when the losses that Bob expects because of the fiber are equal to those introduced by Eve's storing and blocking photons, that is, when the attenuation in the fiber is larger than a critical value δ_c^{BB84} defined by

$$R_{\rm raw}(\delta_c^{BB84}) = \sum_{n\geq 2} p_n\,[1-(1-\eta_{\rm det})^{n-1}] \simeq \eta_{\rm det}\, p_2. \quad (3)$$

For $\mu = 0.1$, we find $\delta_c^{BB84} = 13$ dB, that is $\ell_c^{BB84} \approx 50$ km. For shorter distances, Eve can optimize her attack, but will not be able to obtain full information; Alice and Bob can therefore use a *privacy amplification* scheme to retrieve a shorter secret key from their data. In conclusion, for $\delta \geq \delta_c^{BB84}$, the weak-pulse implementation of the BB84 protocol becomes in principle insecure, even for zero quantum-bit error rate (QBER).

Encoding in nonorthogonal states.—The extreme weakness of the BB84 protocol against PNS attacks is due to the fact that whenever Eve can keep one photon, she gets all the information, because after the sifting phase she has to discriminate between two eigenstates of a known Hermitian operator. Intuition suggests then that the robustness against PNS attacks can be increased by using protocols that encode the classical bit into pairs of nonorthogonal states, that cannot be discriminated deterministically. We prove that this intuition is correct.

To fix the ideas, consider the following protocol using four states: Alice encodes each bit in the state of a qubit, belonging either to the set $\mathcal{A} = \{|0_a\rangle, |1_a\rangle\}$ or to the set $\mathcal{B} = \{|0_b\rangle, |1_b\rangle\}$, with $|\langle 0_a|1_a\rangle| = |\langle 0_b|1_b\rangle| = \chi \neq 0$ (Fig. 1, left). In the absence of an eavesdropper, Bob can be perfectly correlated with Alice: in fact, although the two states are not orthogonal, one can construct a generalized measurement that unambiguously discriminates between the two. The price to pay is that sometimes one gets an inconclusive result [10]. Such a measurement can be realized by a selective filtering, that is a filter whose effect is not the same on all states, followed by a von Neumann measurement on the photons that pass the filter [11]. In the example of Fig. 1, the filter that discriminates between the elements of $\mathcal{A}$ is given by $F_{\mathcal{A}} = (1/\sqrt{1+\chi})(|+x\rangle\langle 1_a^\perp| + |-x\rangle\langle 0_a^\perp|)$, where $|\psi^\perp\rangle$ is the state orthogonal to $|\psi\rangle$. When the photons are prepared in a state of the pair $\mathcal{A}$, a fraction $1-\chi$ of them pass this filter, and in this case the von Neumann measurement of σ_x achieves the discrimination. It is then clear how the cryptography protocol generalizes BB84: Bob randomly applies on each qubit one of the two filters $F_{\mathcal{A}}$ or $F_{\mathcal{B}}$, and measures σ_x on the outcome. Later, Alice discloses for each bit the set $\mathcal{A}$ or $\mathcal{B}$: Alice and Bob discard all the items in which Bob has chosen the wrong filter and all the inconclusive results.

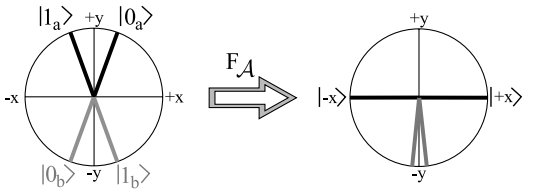

FIG. 1. Two pairs of nonorthogonal states on the equator of the Poincaré sphere, and the effect of the filter $F_{\mathcal{A}}$.

Of course, since not all the qubits will pass the filter even when it was correctly chosen, there is a small nuisance on Bob's side because the net key rate is decreased. This is compensated by increasing μ by a factor $1/(1-\chi)$. However, the nuisance is by far bigger on Eve's side, even when the increased mean number of photons μ is taken into account. We shall give a detailed analysis of the PNS attacks below for a specific protocol, but a simple estimate shows the origin of the improved robustness. Eve can obtain *full information* only when (i) she can block all the pulses containing one and two photons, and (ii) on the pulses containing three or more photons, she performs a suitable unambiguous discrimination measurement (see below) and obtains a conclusive outcome, which happens only with probability $p_{ok} < 1$. Consequently, the critical attenuation is defined by $R_{\rm raw}(\delta_c) \simeq \eta_{\rm det}\, p_3[\mu/(1-\chi)]\, p_{ok}$, and is determined by p_3 instead of p_2 as in the BB84; see (3). For typical values, $\delta_c - \delta_c^{BB84} \approx 10$ dB, which means an improvement of some 40 km in the distance [8].

A specific protocol.— Here is an astonishingly simple protocol using four nonorthogonal states. Alice sends randomly one of the four states $|\pm x\rangle$ or $|\pm z\rangle$; Bob measures either σ_x or σ_z. Thus, at the "quantum" level, the protocol is identical to BB84, and can be immediately implemented with the existing devices. However, we *modify the classical sifting procedure*: instead of revealing the basis, Alice announces publicly one of the four pairs of nonorthogonal states $\mathcal{A}_{\omega,\omega'} = \{|\omega x\rangle, |\omega' z\rangle\}$,

with $\omega, \omega' \in \{+, -\}$, and with the convention that $|\pm x\rangle$ code for 0 and $|\pm z\rangle$ code for 1. Within each set, the overlap of the two states is $\chi = (1/\sqrt{2})$. Because of the peculiar choice of states, the usual procedure of choosing randomly between σ_x or σ_z turns out to implement the most effective unambiguous discrimination. For definiteness, suppose that for a given qubit Alice has sent $|+x\rangle$, and that she has announced the set $\mathcal{A}_{+,+}$. If Bob has measured σ_x, which happens with probability $\frac{1}{2}$, he has certainly got the result $+1$; but since this result is possible for both states in the set $\mathcal{A}_{+,+}$, he has to discard it. If Bob has measured σ_z and got $+1$, again he cannot discriminate. But if he has measured σ_z and got -1, then he knows that Alice has sent $|+x\rangle$ and adds a 0 to his key. By symmetry, we see that after this sifting procedure Bob is left with $\frac{1}{4}$ of the raw list of bits, compared to the $\frac{1}{2}$ of the original BB84 protocol. Thus, for a fair comparison with BB84 using $\mu = 0.1$, we shall take here $\mu = 0.2$, so that the net key rates without eavesdropper at a given distance are the same for both protocols. In spite of the fact that a larger μ is used (that is, multiphoton pulses are more frequent), this new protocol is provably better than BB84 against PNS attacks at QBER $= 0$. This is our main claim, and is demonstrated in the following.

PNS attacks at QBER=0.—First, let us prove something that we mentioned above, namely: for protocols using four states such as the one under study, Eve can obtain full information from three-photon pulses by using strategies based on unambiguous state discrimination. Such strategies have also been considered for BB84, because (although worse than the storage attack for an all-powerful Eve) they do not require a quantum memory [12], and in their simplest implementation the photon number QND measurement is not required either [13]. The most powerful of these attacks, against which any protocol using four states becomes completely insecure for the three-photon pulses, goes as follows [14]. A pulse containing three photons is necessarily in one of the four states $|\Psi_1\rangle = |+z\rangle^{\otimes 3}$, $|\Psi_2\rangle = |+x\rangle^{\otimes 3}$, $|\Psi_3\rangle = |-z\rangle^{\otimes 3}$, $|\Psi_4\rangle = |-x\rangle^{\otimes 3}$; that is, in the *symmetric* subspace of 3 qubits. The dimension of this subspace is 4, and it can be shown that all the $|\Psi_k\rangle^{\otimes 3}$ are linearly independent [15]. Therefore, there exist a measurement $\mathcal{M}$ that distinguishes unambiguously among them, with some probability of success. In the present case, there exist even four orthogonal states of three qubits, $|\Phi_k\rangle$, $k = 1, ..., 4$, such that $|\langle\Phi_i|\Psi_j\rangle| = (1/\sqrt{2})\,\delta_{ij}$ [16]. The measurement $\mathcal{M}$ is then any von Neumann measurement discriminating the $|\Phi_k\rangle$; it will give a conclusive outcome with probability $p_{ok} = \frac{1}{2}$, which is optimal [15,17].

It is then clear that Eve can obtain full information if she can block all the one- and two-photon pulses and half of the three-photon pulses, by applying the following PNS attack: (i) she measures the number of photons; (ii) she discards all pulses containing less than 3 photons; (iii) on the pulses containing at least 3 photons, she performs $\mathcal{M}$, and if the result is conclusive (which happens with probability $p_{ok} \gtrsim \frac{1}{2}$) she sends a new photon prepared in the good state to Bob. We refer to this attack as to *intercept-resend with unambiguous discrimination (IRUD) attack*. Neither the quantum memory nor the lossless channel are needed, since the new state can be prepared by a friend of Eve located close to Bob.

The critical attenuation δ_c at which the IRUD attack becomes always possible is defined by $\eta_{\delta_c}\mu = p_{ok}p_3(\mu)$; for $\mu = 0.2$, this gives $\delta_c = 25.6$ dB $\approx 2\delta_c^{BB84}$. Thus, the ultimate limit of robustness (in the case of zero errors) is shifted from $\sim$50 km up to $\sim$100 km by using our simple modification of the BB84 protocol. To further increase the limit of 100 km, one can move to protocols using six or more nonorthogonal states [17].

Figure 2 plots Eve's information for the best PNS attack at QBER $= 0$, as a function of the attenuation. Note that the new protocol is better than BB84 at any distance. For almost all $\delta < \delta_c$, the best PNS attack is not the IRUD but a storage attack, in which Eve keeps one or two photons in a quantum memory and waits for the announcements of the sifting phase. Recall that in BB84, this kind of attack provides Eve with full information. In our protocol Alice announces sets of two nonorthogonal states, so storage attacks give Eve only a limited amount of information. If Eve keeps n photons and the overlap is χ (here, $1/\sqrt{2}$), the largest information she can obtain is $I(n, \chi) = 1 - H(P, 1 - P)$ with $P = \frac{1}{2}(1 + \sqrt{1 - \chi^{2n}})$ [10]. In particular, Eve obtains $I[1, (1/\sqrt{2})] \approx 0.4$ bits/pulse for the attenuation δ_1 at which she can always keep one photon ($\delta_1 \simeq 11$ dB for $\mu = 0.2$).

In conclusion, in the limiting case of QBER $= 0$, our protocol is always more secure than BB84 against PNS attacks, and can be made provably secure against such attacks in regions where BB84 is already provably insecure. Recall that the comparison is made by fixing the net key rates without eavesdropper at a given distance.

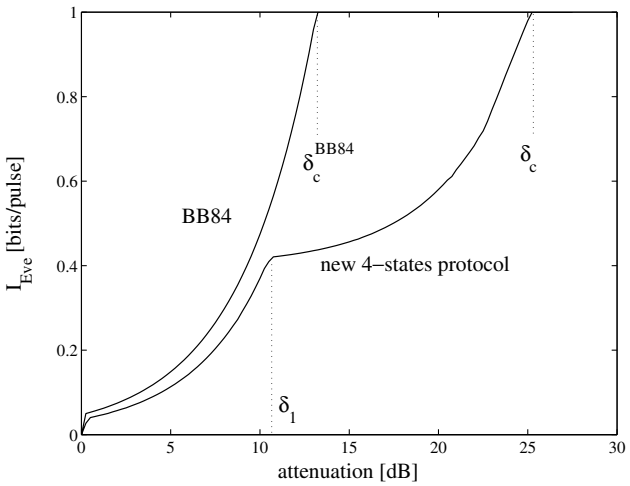

FIG. 2. PNS attacks with QBER $= 0$ on the BB84 protocol for $\mu = 0.1$ and on the new protocol for $\mu = 0.2$: Eve's information as a function of the attenuation $\delta = \alpha\ell$.

Attacks at QBER > 0 on the new protocol.—In real experiments, dark counts in the detectors and misalignement of optical elements always introduce some errors. It is then important to show that the specific protocol we presented does not break down if a small amount of error on Bob's side is allowed. Several attacks at nonzero QBER are described in detail in Ref. [17]. Here, we sketch the analysis of two individual attacks.

First, let us suppose that Eve uses the *phase-covariant cloning machine* that is the optimal individual attack against BB84 [18]. In the case of the present protocol, Eve can extract less information from her clones, again because Alice does not disclose a basis but a set of nonorthogonal states. As a consequence, the condition $I_{\mathrm{Bob}} = I_{\mathrm{Eve}}$ is fulfilled up to QBER $= 15\%$ [17], a value which is slightly higher than the 14.67% obtained for BB84. So our new protocol, designed to avoid PNS attacks in a weak-pulses implementation, seems to be robust also against individual eavesdropping in a single-photon implementation. Incidentally note that, in the case of a single-photon implementation, our protocol is at least as secure as the protocol using two nonorthogonal states proposed by Bennett in 1992 (B92), in the sense of "unconditional security" proofs [19]. This is because our protocol can be seen as a modified B92, where Alice chooses randomly between four sets of nonorthogonal states [20].

The second kind of individual attacks that we would like to discuss, and that we call *PNS + cloning attacks*, are specific to imperfect sources. Focus on the range $\delta \simeq$ 10–20 dB (see Fig. 2), where one-photon pulses can be blocked and the occurrence of three or more photons is still comparatively rare. Because for the BB84 Eve has already full information in this range, such attacks have never been considered before. Eve could take the two photons, apply an asymmetric $2 \rightarrow 3$ cloning machine and send one of the clones to Bob; she keeps two clones and some information in the machine. By a suitable choice of the cloning machine, the QBER at which $I_{\mathrm{Bob}} = I_{\mathrm{Eve}}$ is lowered down to $\sim 9\%$ [17]. In Ref. [21], a successful qubit distribution over 67 km with $\mu = 0.2$ and QBER $= 5\%$ has been reported. Under the considered PNS attacks, such distribution is provably insecure using the sifting procedure of BB84, while it can yield a secret key if our sifting procedure is used.

In summary, we have shown that by encoding a classical bit in sets of nonorthogonal qubit states, quantum cryptography can be made significantly more robust against photon number splitting attacks. We have presented a specific protocol, which is identical to the BB84 protocol for all the manipulations at the quantum level and differs only in the classical sifting procedure. Under the studied attacks, our protocol is secure in a region where BB84 is provably insecure. Preliminary studies of more complex attacks suggest that it is at least as robust as BB84 in any situation, and could then replace it. Moreover, our encoding can easily be combined with more complex procedures on the quantum level, e.g., [22].

We thank Norbert Lütkenhaus and Daniel Collins for insightful comments. We acknowledge financial support by the Swiss OFES and NSF within the European IST project EQUIP and the NCCR "Quantum Photonics."

[1] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, Rev. Mod. Phys. **74**, 145 (2002).
[2] G. Ribordy *et al.*, Phys. Rev. A **63**, 012309 (2001); F. Grosshans and P. Grangier, Phys. Rev. Lett. **88**, 057902 (2002); E. Waks *et al.*, Nature (London) **420**, 762 (2002).
[3] S. Félix *et al.*, J. Mod. Opt. **48**, 2009 (2001).
[4] G. Brassard, N. Lütkenhaus, T. Mor, and B. C. Sanders, Phys. Rev. Lett. **85**, 1330 (2000).
[5] N. Lütkenhaus, Phys. Rev. A **61**, 052304 (2000).
[6] Ch.H. Bennett and G. Brassard, in *Proceedings of the IEEE International Conference on Computers, Systems and Signal Processing, Bangalore, India* (IEEE, New York, 1984), pp. 175–179.
[7] We suppose that Eve does not have access to Bob's detectors. Though dropped in some theoretical work [e.g., D. Gottesman *et al.*, quant-ph/0212066], this hypothesis is generally held as most natural: the detectors cannot be modified from outside, and if Eve has access to Bob's laboratory, she had better put an antenna in his computer and read all the data. Moreover, this assumption does not modify our result, see [8] below. See also M. Curty and N. Lütkenhaus, quant-ph/0311066.
[8] If we allow Eve access to Bob's detectors, both critical attenuations become smaller, but the improvement $\delta_c - \delta_c^{BB84} \approx 10$ dB holds unchanged, because it is independent of η_{det}.
[9] In a realistic setup, Bob can monitor only the detection rate and the fraction of double counts in the wrong basis.
[10] A. Peres, *Quantum Theory: Concepts and Methods* (Kluwer, Dordrecht, 1998), Sec. 9-5.
[11] B. Huttner *et al.*, Phys. Rev. A **54**, 3783 (1996)
[12] M. Dušek, M. Jahma, and N. Lütkenhaus, Phys. Rev. A **62**, 022306 (2000)
[13] C. H. Bennett *et al.*, J. Cryptology **5**, 3 (1992); H. P. Yuen, Quantum Semiclass. Opt. **8**, 939 (1996)
[14] N. Lütkenhaus (private communication).
[15] A. Chefles, Phys. Lett. A **239**, 339 (1998); quant-ph/0105016.
[16] With $|+z\rangle = |0\rangle$, $|-z\rangle = |1\rangle$: $|\Phi_1\rangle = (1/\sqrt{2})(|000\rangle - |011\rangle)$, $|\Phi_2\rangle = \frac{1}{2}(|101\rangle + |010\rangle + |100\rangle + |110\rangle)$, $|\Phi_3\rangle = (1/\sqrt{2})(|111\rangle - |001\rangle)$, $|\Phi_4\rangle = \frac{1}{2}(|101\rangle - |010\rangle - |100\rangle + |110\rangle)$.
[17] A. Acín, N. Gisin, and V. Scarani, quant-ph/0302037 [Phys. Rev. A (to be published)].
[18] C. A. Fuchs *et al.*,Phys. Rev. A **56**, 1163 (1997); N. J. Cerf *et al.*, Phys. Rev. Lett. **88**, 127902 (2002).
[19] K. Tamaki *et al.*, Phys. Rev. Lett. **90**, 167904 (2003)
[20] S. Iblisdir (private communication).
[21] D. Stucki *et al.*, New J. Phys. **4**, 41 (2002).
[22] W.-Y. Hwang, Phys. Rev. Lett. **91**, 057901 (2003).