Quantum key distribution with realistic states: photon-number statistics in the photon-number splitting attack

Norbert Lütkenhaus¹ and Mika Jahma²

¹ Quantum Information Theory Group, Centre for Modern Optics (ZEMO), Universität Erlangen-Nürnberg, Staudtstr. 7/B2, D-91058 Erlangen, Germany
² Helsinki Institute of Physics, PL 64, FIN-00014 Helsingin yliopisto, Finland
E-mail: luetkenhaus@kerr.physik.uni-erlangen.de

New Journal of Physics 4 (2002) 44.1–44.9 (http://www.njp.org/)
Received 2 January 2002, in final form 20 March 2002
Published 12 July 2002

Abstract. Quantum key distribution can be performed with practical signal sources such as weak coherent pulses. One example of such a scheme is the Bennett–Brassard protocol that can be implemented via polarization of the signals, or equivalent signals. It turns out that the most powerful tool at the disposition of an eavesdropper is the photon-number splitting attack. We show that this attack can be extended in the relevant parameter regime so as to preserve the Poissonian photon number distribution of the combination of the signal source and the lossy channel.

1. Introduction

Quantum key distribution (QKD) allows us to generate a long secret, shared key between two parties, conventionally named Alice and Bob, from a short initial secret key. Part of this newly generated key can then be used up by sending an unconditionally secure secret message via the one-time pad, also called the Vernam cipher [1]. The remaining part is retained to repeat the QKD protocol to generate a new key. The first complete protocol is that by Bennett and Brassard, BB84 [2], although Wiesner formulated the basic ideas earlier [3].

In ideal QKD protocols we are required to use particular states for which the preparation is beyond our present experimental capability, such as single-photon states on which we can imprint signals in the form of specific polarizations. For example, for the BB84 protocol we would use two pairs of orthogonal polarizations, e.g. horizontal/vertical linear polarization and right/left circular polarization.
Recently it has been proven that one can use realistic signal sources such as weak laser pulses polarized in the four signal polarizations to perform QKD even in the presence of loss and noise in the quantum channel. Indeed, in most experiments demonstrating the technique required for QKD this signal source has been used [4]–[9]. For eavesdropping attacks on such signals, the security of the BB84 protocol in a realistic setting has been explored regarding attacks on individual signals in [10]. The proof of unconditional security of QKD with the BB84 protocol in this framework has been presented in [11]. It turns out that the combination of multi-photon signals of the source, such as weak laser pulses, together with loss in the quantum channel in the presence of errors, leads to limitations of rate and distance that can be covered by these techniques. These restrictions are due not only to the proving techniques but are of fundamental nature [12], at least in a conservative approach to security where all errors and losses are assumed to be due to eavesdropping activity.

The limitation comes from the fact that the combination of multi-photon signals of the source and loss in the transmission line opens the door for a powerful eavesdropping attack, the photon number splitting (PNS) attack that was first mentioned in [13]. In this attack an eavesdropper, usually referred to as Eve, can replace the noisy and lossy transmission line by a superior one. Therefore, she can cut out part of the losses, and we shall refer to this as accessible loss. The non-accessible loss may contain detector inefficiencies or minimum transmission losses†. The improved transmission gives her room to launch a special attack. The basic step is that a signal consisting of two or more photons (multi-photon signal) can be split via a physical interaction [10] by an eavesdropper such that Eve retains one photon and Bob receives the other photons such that the polarization of both parts remains undisturbed. The photon in Eve’s hand will reveal its signal polarization to Eve if she waits until she learns the polarization basis during the public discussion part of the BB84 protocol. In the presence of accessible loss, this attack can put Eve in a position where she knows the complete information about all signals received by Bob and no secure key can be generated. For this let us assume the model where any non-accessible loss of the quantum channel is considered to be part of Bob’s detection apparatus, which can be done without loss of generality. The transmission becomes totally insecure if the accessible loss is strong enough that fewer non-vacuum signals are expected at the entrance to Bob’s apparatus than multi-photon signals are prepared by the signal source. Then Eve can replace the lossy quantum channel by an ideal one, block all single-photon signals and use only multi-photon signal to match Bob’s expectation of non-vacuum pulses. If the loss is not high enough for this, then Eve can block only a fraction $b$ of the single-photon signals, but she can perform some optimal eavesdropping attack on the remaining single-photon pulses. This constitutes her optimal attack [10]. Despite this powerful attack, in this situation, if the error rate is not too high, Alice and Bob can establish a secure key, as has been shown in [10, 11], in the standard BB84 protocol where no photon-number statistics is monitored. Note that the PNS attack can be well approximated with linear optics, a rudimentary QND measurement and a short-time quantum memory. Moreover, even with linear optics alone powerful attacks in this spirit can be launched [14].

One remaining open question is whether Alice and Bob might be able to detect that Eve has performed the PNS attack. After all, the photon number statistics changes under the PNS attack as described above. This manifests itself in a change of the coincidence rate of the standard detection setup that uses two detectors. Moreover, one can think of other set-ups specifically

† In a conservative approach all loss is accessible loss.
designed to allow to measure higher moments of the photon number distribution. In this paper, we shall show that it is possible to extend the PNS attack such that the complete photon number statistics, as seen by Bob, is indistinguishable from that resulting from weak laser pulses and a lossy channel. This result holds in the relevant parameter regime of mean photon number $\mu$ of the Poissonian photon number distribution of the weak laser pulse and of the transmission factor $\eta$ of the quantum channel. Therefore, in the relevant parameter regime, the extended PNS attack allows the eavesdropper to remain undetected even if Alice and Bob measure the complete photon number distribution. One important consequence of our result is that the yield of secure bits as given by the unconditional security proof in [11] cannot be increased by monitoring the received photon number distribution.

2. Extended PNS attack

We consider a photon source emitting signals with a Poissonian photon number distribution with mean value $\mu$. Weak laser pulses are well described by Fock states with this photon number distribution, but our analysis can be extended to other distributions. The quantum channel is described by a single-photon transmission efficiency $\eta$. As a reminder, this efficiency refers to the accessible loss, meaning that detection inefficiency or minimum loss in the channel is not included here. Then we again find at Bob’s end of the quantum channel a Poissonian photon number distribution with mean photon number $\mu\eta$, that is

$$P_{\text{loss}}[n] = \frac{(\eta\mu)^n}{n!} \exp[-\eta\mu].$$

(1)

On the other hand, the PNS attack described above will give another photon number statistics. Let Eve cut out the accessible loss and perform the PNS attack in which she blocks a fraction $b$ of the single-photon signals. Then we find a resulting photon number distribution that is not Poissonian, namely

$$P_{\text{PNS}}[n] = \begin{cases} 
(1 + b\mu) \exp[-\mu] & n = 0 \\
((1 - b)\mu + \mu^2/2) \exp[-\mu] & n = 1 \\
\mu^{n+1} \frac{\exp[-\mu]}{(n+1)!} & n > 1. 
\end{cases}$$

(2)

As a first step to remain undetected, Eve adjusts $b$ to match the number of vacuum signals of the PNS attack to that of the lossy channel, $P_{\text{loss}}[0] = P_{\text{PNS}}[0]$. This leads to the expression

$$b_{\text{match}} = \frac{1}{\mu} (\exp[\mu(1 - \eta)] - 1).$$

(3)

Clearly, it is only possible to fulfil this matching condition if $b_{\text{match}} \in [0, 1]$. We find $b_{\text{match}} = 0$ for $\eta = 1$, which expresses the fact that for a lossless channel the PNS attack cannot (and need not) be accompanied by the blocking of single-photon signals. On the other hand we find that all single-photon signals can be blocked ($b_{\text{match}} = 1$) if there are exactly as many multi-photon signals leaving the source as non-vacuum signals are arriving at the receiver, $1 - (1 + \mu) \exp[-\mu] = 1 - P_{\text{loss}}[0]$. In this case the complete information falls into Eve’s hands. This situation occurs at a channel transmission efficiency of $\eta_{\text{lim}} := 1 - (1/\mu) \ln[1 + \mu]$. For values of $\eta$ in the interval $[\eta_{\text{lim}}, 1]$, $b$ takes on values in the interval $[1, 0]$. If the channel transmission efficiency falls below $\eta_{\text{lim}}$, leading to $b_{\text{match}} > 1$, Eve needs to suppress not only...
Table 1. Example for the photon number distribution of a lossy channel and of the matched PNS attack. The parameters are $\mu = 0.4$ and $\eta = 0.2$, resulting in $b_{\text{match}} = 0.94$. One finds that within the non-vacuum signals the fraction of single-photon signals is too low initially by 6.6% in $P_{\text{match}}$. In this case Eve can make the distribution Poissonian by extracting additional photons.

| $n$ | $P_{\text{loss}}$ | $P_{\text{match}}$ | $d_n$ | $D_n$ |
|-----|-------------------|-------------------|------|------|
| 0   | 0.9231163         | 0.9231163         | 0    | 0    |
| 1   | 0.0738493         | 0.0689573         | -0.0048920 | 0.0048920 |
| 2   | 0.0029540         | 0.0071501         | +0.0041961 | 0.0006959 |
| 3   | 0.0000788         | 0.0007150         | +0.0006362 | 0.0000596 |
| 4   | 0.0000016         | 0.0000572         | +0.0000556 | 0.0000040 |

Table 2. Example for the photon number distribution of a lossy channel and of the matched PNS attack. The parameters are $\mu = 0.4$ and $\eta = 0.4$, resulting in $b_{\text{match}} = 0.68$. Within the non-vacuum signals the fraction of single-photon signals is too high initially by 2.6% in $P_{\text{match}}$. Therefore we cannot match $P_{\text{match}}$ to $P_{\text{loss}}$ by extracting more photons from multi-photon signals.

| $n$ | $P_{\text{loss}}$ | $P_{\text{match}}$ | $d_n$ | $D_n$ |
|-----|-------------------|-------------------|------|------|
| 0   | 0.852144          | 0.852144          | 0    | 0    |
| 1   | 0.136343          | 0.139930          | +0.0035869 | -0.0035869 |
| 2   | 0.0109074         | 0.0071501         | +0.0037574 | +0.0001705 |
| 3   | 0.0005817         | 0.0007150         | +0.0001333 | +0.0000372 |
| 4   | 0.0000233         | 0.0000572         | +0.0000339 | +0.0000033 |

single-photon signals, but also multi-photon signals to match $P_{\text{loss}}[0] = P_{\text{PNS}}[0]$. This is only to Eve’s advantage, as we shall see later on. For now we concentrate on the case where it is sufficient for Eve to block only single-photon signals.

In this situation, with the appropriate choice for $b$, the photon number distribution after Eve’s attack takes the form

$$
P_{\text{match}}[n] = \begin{cases} 
\exp[-\eta \mu] & n = 0 \\
(1 + \mu + \mu^2/2) \exp[-\mu] - \exp[-\eta \mu] & n = 1 \\
\frac{\mu^{n+1}}{(n+1)!} \exp[-\mu] & n > 1.
\end{cases}
$$

This photon number distribution is not Poissonian (see the two examples in tables 1 and 2).

The question is whether Eve can make it Poissonian without losing any advantage of the PNS attack, so her goal is not only to match the probability of non-vacuum signals, but to match the whole photon number distribution. It is easy to come up with a possible solution: Eve can extract not only one, but two or more photons from pulses depending on the photon number in each pulse. If she does not empty a pulse completely, then the initial match of the non-vacuum signals via the parameter $b$ will not be disturbed and Eve’s information about the signals remains unchanged. With this method it is possible to implement all redistributions of...
probabilities from higher to lower photon numbers, but no redistributions the other way round. Therefore the necessary and sufficient condition for the redistribution to be possible between an input distribution $P_{in}[n]$ and an output distribution $P_{out}$ is that

$$\sum_{i=0}^{n} P_{in}[i] \leq \sum_{i=0}^{n} P_{out}[i]$$

(5)

is satisfied for all $n$. To see that this is indeed the relevant condition let us define $d_n$ as the difference of the two probability distributions for given $n$,

$$d_n = P_{in}[n] - P_{out}[n].$$

(6)

The meaning of $d_n$ is that of an excess probability for the photon number $n$. Clearly, we require $d_0 \leq 0$ since we cannot transfer vacuum states to non-vacuum states. Now we can interpret $D_n := -\sum_{i=0}^{n} d_i$ as the required flow of probability into the photon number interval $[0, n]$ from higher photon numbers. It is necessary that $D_n \geq 0$ for all $n$, otherwise we would need to insert photons into signals rather than extracting them. To see that equation (5) is indeed sufficient, note that $D_n$ can be viewed as a prescription to implement the redistribution. For each value of $n$ we find

$$P_{out}[n] = P_{in}[n] + D[n] - D[n-1].$$

(7)

For positive values of $D[n]$ and $D[n-1]$ this can be understood as a prescription to have an influx of probability $D[n]$ from higher photon numbers and an out-flux of $D[n-1]$ into lower photon numbers so as to reach the (positive) target probability $P_{out}[n]$. This establishes that equation (5) is also sufficient. (See tables 1 and 2 for two examples, one where redistribution is possible, and another where it is not possible.)

In order to work as an eavesdropping attack, we need to make sure that the number of non-vacuum signals remains unchanged. For this reason we made the match of the non-vacuum signals before redistributing the photon number statistics. This guarantees that the information gain by Eve on the signals remains unchanged in this second step. The only change is that of the photon number statistics of the signals arriving at Bob’s end of the quantum channel.

3. Evaluation of extended PNS attack

In this section we shall show that the conditions (5), as applied to our tasks, are satisfied in a parameter regime that we shall show in the following section to be relevant to practical QKD. Thus we show that Eve can mimic a Poissonian photon number distribution even while performing the PNS attack via our extension. Note that $d_0 = 0$ due to the matching via the blocking parameter $b$. We shall show that in a relevant parameter regime we find that $d_n \leq 0$ for $n \in [1, n_l]$ and $d_n \geq 0$ for $n \in [n_l + 1, \infty]$. This is sufficient (though not necessary) to fulfil the conditions (5). In other words, with increasing value of $n$, the difference $d_n$ vanishes for $n = 0$, then takes negative values until for $n \geq n_l + 1$ it becomes positive.

Let us first show by induction that once the function has become positive it will not turn back negative for $n \geq 2$ and the parameter regime $\eta \leq 3/4$. Assume that $d_n \geq 0$ for $n \geq 2$. This means

$$\frac{\mu^{n+1}}{(n+1)!} \exp[-\mu] \geq \frac{(\eta \mu)^n}{n! \exp[-\eta \mu]},$$

(8)
Figure 1. As a function of the mean photon number $\mu$ and the transmission efficiency $\eta$, we see the area (grey shade) where the original PNS attack yields fewer single-photon signals than the corresponding lossy channel. In this region the extended PNS attack is successful in that it mimics the lossy channel not only in the fraction of non-vacuum signals but also in the whole photon number statistics.

Then it follows that

$$d_{n+1} = \frac{\mu^{n+1}}{(n+1)!} \exp[-\mu] \frac{\mu}{n+2} \left( \frac{(\eta\mu)^n}{n!} \exp[-\eta\mu] \frac{\mu\eta}{n+1} \right) \geq \frac{(\eta\mu)^n}{n!} \exp[-\eta\mu] \left( \frac{\mu}{n+2} - \frac{\mu\eta}{n+1} \right) \geq 0 \quad \text{for } \eta \leq \frac{3}{4}. \quad (9)$$

We do not need to prove directly that there is some $d_{n_i} \geq 0$ with $n_i \geq 2$. Instead, we shall show that in a suitable parameter regime we find $d_1 \leq 0$. This proves, together with the normalization $\sum_{n=0}^{\infty} d_n = 0$ and $d_0 = 0$, that there must be some positive $d_n$ for $n \geq 2$. In other words, for $\eta < 3/4$ we find that $d_1 \leq 0$ is a sufficient condition to allow a redistribution of the photon number distribution after the PNS attack to make it Poissonian without changing the flow of information between the parties.

The required condition $d_1 \leq 0$ can only be analysed numerically. It is given after a regrouping as

$$(1 + \mu + \mu^2/2) \exp[-\mu] - (1 + \eta\mu) \exp[-\eta\mu] \leq 0. \quad (10)$$

The first term describes the fraction of signals containing zero or one photon after the original PNS attack while the second term describes the target value for this fraction. Note that we fixed the fraction of vacuum signals so that this is, indeed, a statement about the fraction of single-photon signals. The region where $d_1$ is negative is plotted in figure 1.

The border curve shown in figure 1 can be evaluated more closely in the typical regime where $\mu, \eta \ll 1$. If we expand $d_1$ in $\mu, \eta$ and neglect terms $\eta^k \mu^l$ with $k + l > 4$, then we obtain

$$d_1 \approx \mu^2/2(-\mu/3 + \mu^2/4 + \eta^2) \quad (11)$$

so that we find that the value $\eta_0$ for which $d_1$ vanishes can be approximated by

$$\eta_0 \approx \sqrt{\mu/3 - \mu^2/4}. \quad (12)$$
Figure 2. We compare the approximation (dashed curve) for the critical value \( \eta_0 \) for which \( d_1 \) vanishes with a numerical solution of \( d_1(\eta_0) = 0 \) (solid curve).

As we see from figure 2, this is a good approximation for low photon numbers. Note that the lowest order of the approximation (11) is given by \( d_1 \approx -\mu^3/6 \), which guarantees that the extended PNS attack is always successful as long as \( \mu \) and \( \eta \) are small. This is mirrored in the infinitely steep rise of the limiting line shown in figure 2.

What happens if the losses in the quantum channel are so high that it is not sufficient for Eve to block only single-photon signals to match the probability of non-vacuum signals? It turns out that it even simpler to analyze this situation. In this case we extend the PNS attack in the following way: as a first step Eve uses the above-described extension to mimic a lossy quantum channel with the transmission efficiency \( \eta_{\text{lim}} \), which is just the limiting case. This leads to a Poissonian distribution with a mean photon number too high. Then Eve inserts a beam-splitter into the path to Bob that mimics additional loss to bring the total transmission efficiency down the desired value. This means no loss of information on Eve’s side, since the first step already assures Eve that she extracts the information about all non-vacuum signals that are subjected to the second step. Therefore, if for parameters \((\mu, \eta_{\text{lim}})\) we can extend the PNS attack so as to mimic a Poissonian distribution, then we can also do so for any parameter pair \((\mu, \eta)\) with \( \eta < \eta_{\text{lim}} \).

The situation \((\mu, \eta = \eta_{\text{lim}})\) can be analysed without the help of numerics. Indeed, in this case we find that the anchor of the induction in equation (8) already takes the given form for \( n = 1 \) and consequently we can conclude the induction of equation (9) in a similar fashion. In this case, due to the changed range for \( n \), we find in summary that, for \( \eta = \eta_{\text{lim}} \), the occurrence of \( d_n \geq 0 \) for \( n \geq 1 \) implies that \( d_{n+1} \geq 0 \) as long as \( \eta \leq 2/3 \). This simplifies our argumentation, since in this case we have \( d_0 = 0 \) and \( d_1 \) cannot be positive, otherwise the induction would show that all other \( d_n \) are non-negative, which would violate \( \sum_n d_n = 4 \). Therefore, a redistribution to Poissonian statistics without information loss is possible for high losses \( (\eta \leq 2/3) \) in a regime where the PNS attack already gives complete information to an eavesdropper.

4. Application to security proofs and discussion

From the security proofs [11, 10] we know the optimal choice of the mean photon number \( \mu \) from the point of view of Alice and Bob. This optimization can be understood starting from the idea that Alice and Bob would like to optimize the gain rate of the QKD process. This gain rate \( G \) is
bounded in a conservative scenario from above as $G \leq \frac{1}{2}(S_m - p_{exp})$ with $S_m$ as the multi-photon probability of the source and $p_{exp}$ as the fraction of non-vacuum signals detected by Bob. The factor $1/2$ arises from a sifting state of the QKD protocol and is specific for the BB84 protocol. For other polarization-based protocols such as the six-state protocol [15] we find other factors; however, the reasoning is independent of this factor. For our Poissonian-distributed signals, we find

$$\frac{1}{2}(S_m - p_{exp}) = \frac{1}{2}(\exp[-\mu \eta] - (1 + \mu) \exp[-\mu])$$

(13)

and this expression is optimized for small values of $\eta$ by $\mu_{opt} \approx \eta$. As it turns out, this value remains approximately optimal in a detailed analysis taking other error sources into account [10, 11]. From figure 1 we find that for these optimal choices and for $\eta$ below 0.25 Eve can achieve a Poissonian photon number distribution within the PNS attack. In typical experiments we find that photon numbers even higher than the optimal ones are used. This pushes the working point even further away from the critical line in figure 1.

Still, one might wonder whether it is of advantage to choose the mean photon number $\mu$ such that the PNS attack cannot be amended to yield Poissonian statistics. This is not clear a priori, but one has to be warned that the goal of optimization is to maximize the gain rate. Therefore, if the change of $\mu$ together with the monitoring of the photon statistics can exclude the extended PNS attack as presented in this paper, this might be a Phyrric victory. The gain rate might be lower than for a value of $\mu$ that allows the PNS attack.

The statements above are made for a conservative scenario where all loss and all errors are attributed to Eve. It will be desirable to extend this analysis by making assumptions stating that Eve cannot change the dark count rate of Bob’s detectors and cannot increase the detection efficiency. In a recent paper, Félix et al [16] approach this problem by studying specific attacks restricting Eve’s toolbox. However, it would be necessary to study whole classes of eavesdropping attacks that are defined only by a consistent defined restricted toolbox for Eve. Though we can use studies of specific attacks to develop a feeling for the situation and to find lower bounds on Eve’s information (and therefore upper bounds on achievable gain rates), one has to be careful not to interpret the results in the sense of obtainable secure gain rates. It turns out that the rigorous analysis for restricted eavesdroppers is not trivial at all. The analysis in our paper, however, puts some bounds on the results of such an analysis. Once Eve can perform an extended PNS attack such that her action mimics the accessible loss of a lossy quantum channel in both the transmission efficiency and the Poissonian photon number statistics, and she can block all single-photon signals in the process, then the transmission cannot be secure.

At the heart of our paper is the statement that stochastic processes such as loss in a quantum channel can be explained by a rather cunning strategy of a third party, for example an eavesdropper. This means that the eavesdropper can access some preferred signals while suppressing others such that the resulting action is indistinguishable even in principle from a normal lossy channel for given input signals. It is unclear so far what input signals have to be chosen to make such a situation impossible.

Acknowledgments

We gladly acknowledge stimulating discussion with Miloslav Dušek and John Calsamiglia. NL thanks Marcos Curty and Peter van Loock for their critical discussion of this article. This work has been supported under project no 43336 of the Academy of Finland and by the German Research Council (DFG) via the Emmy–Noether Programme.
References

[1] Vernam G S 1926 J. Am. Inst. Electr. Eng. 45 109
[2] Bennett C H and Brassard G 1984 Proc. IEEE Int. Conf. on Computers, Systems, and Signal Processing (Bangalore) (New York: IEEE) pp 175–9
[3] Wiesner S 1983 Sigact News 15 78
[4] Chiangga S, Zarda P, Jennewein T and Weinfurter H 1999 Appl. Phys. B 69 389
[5] Bourennane M, Gibson F, Karlsson A, Hening A, Jonsson P, Tsegaye T, Ljunggren D and Sundberg E 1999 Opt. Express 4 383
[6] Marand C and Townsend P T 1995 Opt. Lett. 20 1695
[7] Townsend P D 1998 IEEE Photon. Technol. Lett. 10 1048
[8] Hughes R J, Morgan G L and Peterson C G 2000 J. Mod. Opt. 47 533
[9] Ribordy G, Gautier J-D, Gisin N, Guinnard O and Zbinden H 2000 J. Mod. Opt. 47 517
[10] Lütkenhaus N 2000 Phys. Rev. A 61 052304
[11] Inamori H, Lütkenhaus N and Mayers D 2001 Preprint quant-ph/0107017
[12] Brassard G, Lütkenhaus N, Mor T and Sanders B 2000 Phys. Rev. Lett. 85 1330
[13] Huttner B, Imoto N, Gisin N and Mor T 1995 Phys. Rev. A 51 1863
[14] Calsamiglia J, Barnett S M and Lütkenhaus N 2002 Phys. Rev. A 65 012312
[15] Bruß D 1998 Phys. Rev. Lett. 81 3018
[16] Félix S, Gisin N, Stefanov A and Zbinden H 2001 J. Mod. Opt. 48 2009 (quant-ph/0102062)