Proof-of-principle experiment of a modified photon-number-splitting attack against quantum key distribution

Wei-Tao Liu,* Shi-Hai Sun, Lin-Mei Liang, and Jian-Min Yuan
Department of Physics, College of Science, National University of Defense Technology, Changsha, 410073, P. R. China
(Received 14 July 2010; published 21 April 2011)

Any imperfections in a practical quantum key distribution (QKD) system may be exploited by an eavesdropper to collect information about the key without being discovered. We propose a modified photon-number-splitting attack scheme against QKD systems based on weak laser pulses taking advantage of possible multiphoton pulses. Proof-of-principle experiments are demonstrated. The results show that the eavesdropper can get information about the key generated between the legitimate parties without being detected. Since the equivalent attenuation introduced by the eavesdropper for pulses of different average photon numbers are different, the decoy-state method is effective in fighting against this kind of attack. This has also been proven in our experiments.

DOI: 10.1103/PhysRevA.83.042326 PACS number(s): 03.67.Dd, 03.67.Hk

I. INTRODUCTION

Quantum key distribution (QKD) [1–4] allows the establishment of a secret key between two legitimate parties, called Alice (the sender) and Bob (the receiver). The security of the key is guaranteed by the laws of quantum mechanics, such that the presence of any eavesdropping in the quantum communication channel can be detected by the legitimate users, in principle. However, in the real world, every QKD system is not perfect. Imperfections in the sources and/or detectors may be exploited by an eavesdropper (called Eve) to collect information about the key without being discovered. Research on attacking practical QKD systems, especially experimental research taking advantage of drawbacks in those systems, is of great importance, as it will help us to understand possible practical threats for practical QKD systems, keep us aware of the power of a real eavesdropper, and help us to find methods for identifying and patching technological loopholes in real systems, thus improving the security. The impact of several imperfections has been discussed previously [5–12], and several experimental attempts on hacking real QKD systems against imperfect detectors have been reported [8,11,12].

Since a single photon source is not available for QKD nowadays, weak laser pulses are widely used in practical systems. This kind of real photon source has a finite probability of emitting more than one photon in a single pulse, all of which are usually encoded as the same qubit. So long as the quantum channel between Alice and Bob is not lossless, Eve can take advantage of those multiphoton pulses. She can pick out one or more photons from these pulses, while replacing the quantum channel with a channel of lower loss. If the probability that a nonempty pulse which has more than one photon (at Alice's output) is greater than the probability that a nonempty pulse is detected by Bob, Eve can get full information without introducing any perturbation. These kinds of eavesdropping are called quantum nondemolition attacks, among which the photon-number-splitting (PNS) attack [13–15] is typical. Although PNS attack has not been experimentally performed due to technical challenges, several protocols have been proposed against it. The decoy-state method [16–20], theoretically proven to be the most effective one, has been demonstrated in many QKD systems [21–25].

In this paper, a modified PNS attack scheme against QKD systems based on weak laser pulses is proposed and proof-of-principle experiments are demonstrated. The results show that Eve can get information without being discovered by the legitimate parties. The validity of the decoy-state method against this kind of attack is also verified in our experiments. In the next section, we will present our modified PNS attacking scheme theoretically. Then, in Sec. III, proof-of-principle experiments will be explained, followed by the results and some discussions.

II. THE SCHEME

The brief idea is shown in Fig. 1. In a common practical QKD system, Alice sends weak laser pulses to Bob with an average photon number of μ, via a quantum channel with a transmitting efficiency of η. Eve takes advantage of those multiphoton pulses to steal information about the key. First, she inserts a beamsplitter (BS) with a transmitting efficiency of α, which splits each pulse into two parts. Then Eve performs quantum nondemolition detection on one of the two parts to determine whether there is (are) photon(s) in it. If Eve gets a nonempty pulse, she will keep the photon(s) with a quantum memory while sending the other part of the pulse to Bob via a channel of lower loss or a lossless channel (dashed line in Fig. 1). Otherwise, she will block the quantum channel. Blocking the channel or not is controlled via an optical switch. After Alice and Bob announce the bases they used for each pulse, Eve withdraws the photon(s) from the memory and performs proper measurements to find out the key. Under this attack, all the single-photon pulses from Alice will not cause valid counts at Bob's side. Those multiphoton pulses are randomly split into two parts and only when Eve obtains nonempty pulses can Bob get nonempty pulses [26], therefore Eve can get full information about the key.

In order not to be discovered, Eve should keep the qubit error rate between Alice and Bob as well as the raw key rate unchanged. Obviously this attack will not change any qubit encoded in each pulse, thus introducing no error rate changes.

*mugualaw@hotmail.com

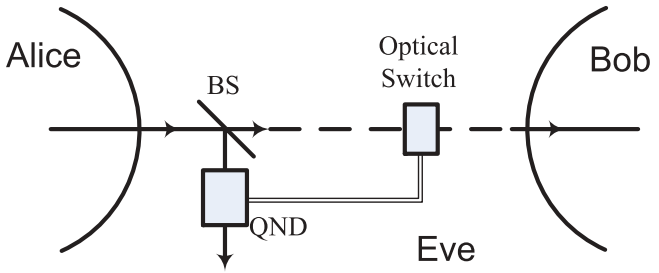

FIG. 1. (Color online) A schematic diagram of the modified PNS attack. Eve splits the pulse from Alice into two parts with a beamsplitter (BS) and keeps one of them, on which she performs quantum nondemolition detection (QND). In the case of obtaining an empty pulse, she blocks the quantum channel with an optical switch. Otherwise, she sends the other part of the pulse to Bob via a channel of lower loss or a lossless channel. After the legitimate parties announce the bases, Eve performs measurements on the pulses she kept and thus gets the information. Beamsplitting will introduce no error between Alice and Bob, and Eve can keep the count rate at Bob's side unchanged by adjusting the transmitting efficiency of BS; therefore, Eve will not be discovered.

To maintain the raw key rate, Eve has to find out a proper value of α. Without Eve's intervention, the count rate or the probability of getting one click for each pulse at Bob's side is

$$\begin{aligned} Q_B &= P(0,\mu\eta)Y_0 + \sum_{n=1}^{\infty} P(n,\mu\eta)[Y_0 + 1 - (1-\eta_d)^n] \\ &= 1 + Y_0 - e^{-\mu\eta\eta_d}, \end{aligned} \tag{1}$$

where $P(n,\gamma) = \gamma^n e^{-\gamma}/n!$ shows Poissonian distribution of photon number in laser pulses with an average photon number of γ, while η_d and Y_0 are the efficiency and dark count rate of Bob's detectors, respectively. With intervened from Eve, the photon-number distribution of laser pulses transmitted to Bob changes into

$$\begin{aligned} P'(0) &= P[0,\mu(1-\alpha)] + \{1 - P[0,\mu(1-\alpha)]\}P(0,\mu\alpha), \\ P'(n) &= \{1 - P[0,\mu(1-\alpha)]\}P(n,\mu\alpha), \quad n \geqslant 1 \end{aligned} \tag{2}$$

with loss of new channel being zero. Therefore the count rate at Bob's side becomes

$$\begin{aligned} Q_B^E &= P'(0)Y_0 + \sum_{n=1}^{\infty} P'(n)[Y_0 + 1 - (1-\eta_d)^n] \\ &= Y_0 + [1 - e^{-\mu(1-\alpha)}][1 - e^{-\mu\alpha\eta_d}]. \end{aligned} \tag{3}$$

Letting $Q_B^E = Q_B$, Eve can find out the value of α. For example, if the loss of the original quantum channel is 20 dB, and $\mu = 0.1, \eta_d = 0.1$, Eve can choose $\alpha = 0.886$. In this way, as long as the loss of the original quantum channel employed by Alice and Bob is not very low, Eve can get full information without changing the raw key rate and qubit error rate, and therefore will not be discovered.

This modified PNS attack is closer to practice than the original one. For quantum nondemolition detection, what we need is only to determine whether the pulse is empty or not. Picking out one photon from the pulse is also no longer necessary, so a usual beam splitter can take the place. As for quantum memory, experiments are being pursued using several techniques, such as atomic ensembles [27,28], NV centers [29], and doped crystals [30,31]. If a heralded quantum memory is possible, even quantum nondemolition detection is no longer necessary. This modified attack can also be used for hacking QKD systems based on other photon sources maintaining multiphoton pulses, as the original one.

III. EXPERIMENTAL RESULTS AND DISCUSSIONS

Based on the eavesdropping scheme proposed above, we demonstrated proof-of-principle experiments. The experiments are performed on a polarization-encoded system, the sources of which are several laser diodes, and the loss of the quantum channel is simulated with a variable attenuator. The experimental arrangements are shown in Fig. 2. One of the laser diodes is turned on by Alice's computer for each pulse, operated at a wavelength of 850 nm and a repetition frequency of 1 MHz with a pulse width of 50 ns. During the experiments, electronic signals are used for synchronization. Eve first splits each laser pulse from Alice into two parts with a 50:50 beamsplitter and performs measurement on one of them. Due to the lack of nondemolition detection technique in our laboratory, Eve makes threshold detection. The polarization of each pulse is analyzed via a polarization beamsplitter (PBS) and the following two single-photon detectors. Since Eve cannot know the basis Alice employed for each pulse in advance, the experiments are carried out on the same basis among three parties in each run of quantum communication. When one and only one of Eve's two detectors clicks, she will switch on the channel and send the other part of the pulse to Bob. Otherwise, she blocks the channel. An acousto-optic modulator (AOM) serves as an optical switch at Eve's output side. When the AOM is switched on, the first-order diffraction of the laser will be sent to Bob, with a diffraction efficiency of 70%. Since there are time delays in the chips as well as between the electrical signal and the optical switching on AOM, a single-mode optical fiber with a length of 400 m is employed as an optical delay line to match the optical and electrical signals. The birefringence of the fiber is passively compensated with a polarization controller (PC in Fig. 2). For

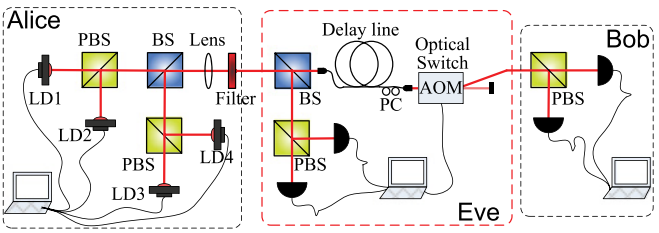

FIG. 2. (Color online) Experimental arrangements. The system is based on polarization encoding and only one basis is used for each run of quantum communication due to the lack of quantum memory in our laboratory. Only when one of Eve's detectors clicks will Eve turn on the optical switch, thus the pulse passes on to Bob. An optical fiber of 400 m is used as optical delay line to match the electrical signal. An acousto-optic modulator (AOM) serves as an optical switch. For experiments without attack, a variable attenuator is inserted into the channel to simulate the loss. LD1 and LD2 are the sources for signal states, while LD3 and LD4 are sources for decoy states.

each run, Alice sends a randomly chosen sequence with a size of 2M bit in one basis. Then the results of 12 such different sequences are tested.

First, an attack on the system without decoy state is performed. Two laser diodes (LD1 and LD2 in Fig. 2) are used as photon sources. The experiments are carried out under different intensities of photon sources. To validate this kind of attack, the system is operated with and without eavesdropping for comparison. For convenience, the same experimental arrangements are used for both cases. For experiments without attack, the optical switch is kept on and the variable attenuator is set such that the count rate at Bob's side is just the same as that under attack, for the same intensity of photon sources [32]. For experiments under attack, the variable attenuator is removed, thus the quantum channel itself is almost lossless. For both cases, the qubit error rate between Alice and Bob, as well as the information between Eve and Alice, is investigated. The information between Eve and Alice is measured with mutual information,

$$I^{\mathrm{EA}} = 1 + e^{\mathrm{EA}} \log_2 e^{\mathrm{EA}} + (1 - e^{\mathrm{EA}}) \log_2 (1 - e^{\mathrm{EA}}), \quad (4)$$

where e^{EA} is the error rate between Eve and Alice. As is shown in Fig. 3, when the average photon number sent by Alice is not very low, the mutual information between Eve and Alice can be as high as 0.92 while the qubit error rate between Alice and Bob maintains around 6.5%, being the same as that of without attack. That is, Eve can obtain almost the full information about the key generated between Alice and Bob without introducing obvious errors. Since the count rate at Bob's side can also be kept unchanged, the existence of eavesdropping cannot be discovered by the legitimate parties.

When the average photon number sent by Alice is rather low, Eve cannot get so much information, as shown in Fig. 3. It is caused by dark counts of Eve's detectors. When Eve gets a dark count, there is some probability for Bob to obtain a nonempty pulse without leaking information to Eve. These cases will introduce a decrease in the information Eve can steal. For the cases of not so low average photon number, the percentage of dark counts among all the counts is rather low and Eve can obtain almost all the information. While the average photon number is very low, dark counts become prominent or even dominant; thus Eve's information becomes very low. However, under the same level of loss, a QKD system itself without attack is no longer available since the error rate between Alice and Bob becomes high due to the loss and the dark counts of Bob's detectors.

It should be noticed that only one basis is used for each run of quantum communication here. In practice, this kind of attack can also work. However, Eve's information will be cut by half due to random choosing of bases between the legitimate parties in practice.

Although the legitimate parties cannot discover the existence of this kind of eavesdropping, Eve does leave behind footprints in the pulses Bob obtains. The equivalent attenuation introduced by Eve differs according to the average photon number of laser pulses Alice sends. To illuminate this, we measured the count rates at Bob's side for different pulse intensities with and without attack, as is shown in Fig. 4. The diamonds show the count rate under attack, which appears as a quadratic function of average photon number. However, the loss of a real quantum channel will be a constant for any pulse intensity, thus the count rate at Bob's side will change linearly.

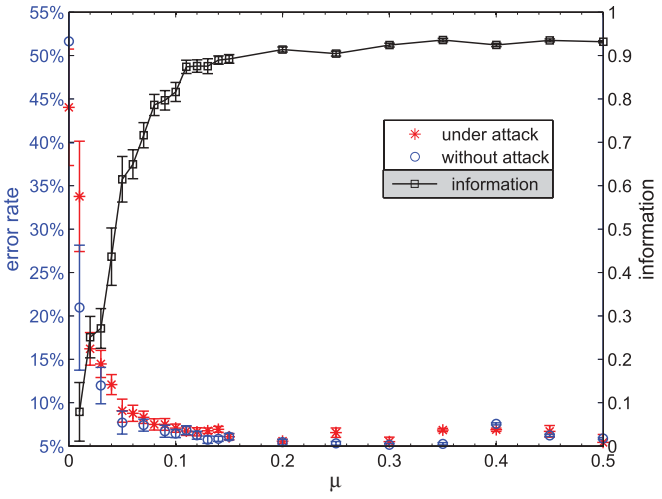

FIG. 3. (Color online) The qubit error rate between legitimate parities and the information Eve obtains. The experiments are carried out at different intensities of photon sources. Stars and circles show the qubit error rate between Alice and Bob, with and without attack, respectively. Squares represent the information Eve obtains, measured with mutual information between Eve and Alice. Since only one basis is employed for each run of quantum communication, Eve's information will be cut by half due to random choosing of bases between the legitimate parties in practice with no quantum memory. Error bars in the figure show the statistical errors and systematic errors are not included.

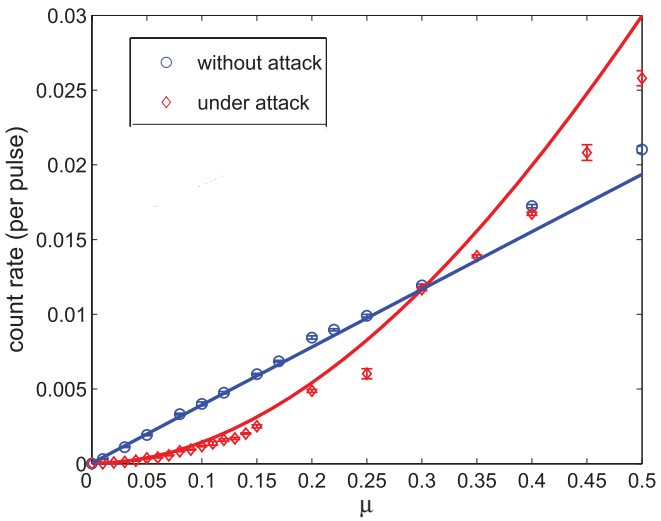

FIG. 4. (Color online) The count rates at Bob's side. The lines show the theoretical results while the dots show experimental results. The diamonds represent the count rates under attack, which appear approximately as a quadratic function of average photon number of photon source. The circles show the count rates without attack, which display linear variation with average photon number. For the data of without attack, the variable attenuator is fixed such that the count rate of with and without attack will be the same when the average photon number is 0.3. Error bars in the figure show the statistical errors and systematic errors are not included.

The circles in Fig. 4 show Bob's count rate for different pulse intensities without attack. For the data of without attack, the variable attenuator is fixed such that the count rate at Bob's side will be the same for under and without attack when the average photon number is 0.3, which are 0.0117 ± 0.0001 and 0.0119 ± 0.0001 per pulse in our experiments (only statistical errors are shown here; the systematic errors discussed below were not quantitatively analyzed), respectively. For the data of under attack, the attenuator is removed, therefore the loss is caused by Eve's intervention. When the average photon number emitted from Alice is lower than 0.3, the equivalent loss caused by Eve is higher than the fixed attenuator, thus the count rate is lower than that of without attack. While the average photon number is higher than 0.3, the equivalent loss caused by Eve is lower than the fixed attenuator since she obtains a higher count rate herself. Therefore Bob can obtain a higher count rate than that of without attack. From the different trends of data for two cases, we can figure out the footprints left in the pulses sent to Bob. Upon that, the legitimate parties can find out Eve's intervention by switching between two or more intensities of laser pulses, which will be influenced in different ways thus marking Eve's intervention. That is, this kind of attack can be discovered using the decoy-state method, which is also verified in our experiments. LD1 and LD2 are used as photon sources for signal states and the average photon number is set as $\mu = 0.3$. Another two laser diodes (LD3 and LD4 in Fig. 2) are introduced as photon sources for decoy states. Two decoy states are employed in our experiments, namely, vacuum and a state of an average photon number ν lower than signal state. Under different values of ν, the count rate as well as the qubit error rate for signal states and decoy states are measured then the secure key rate is calculated with the Gottesman-Lo-Lütkenhaus-Preskill (GLLP) formula [17–20,33], and the results are shown in Fig. 5. For experiments without attack, the loss of quantum channel is set such that the count rate of signal states at Bob's side is the same as that of under attack. It can be seen in Fig. 5 that the secure key rate between Alice and Bob turns to be negative under Eve's intervention. Therefore Eve's intervention can be discovered with the decoy-state method.

Compared with the original PNS attacking scheme, our modified attack also blocks all the single-photon pulses. The difference is that for some multiphoton pulses Eve obtains more than one photon instead of one in the original PNS attack. It leaves the distribution of intervened pulses closer to the Poissonian function while introducing higher equivalent attenuation than the original one. The theoretical secure key rates between Alice and Bob are calculated under either attack, which is also shown in Fig. 5. The necessary parameters are obtained from our experiments and the count rates of signal state at Bob's side are kept the same for both attacks by adjusting the parameter of attenuation. The results show that the theoretical secure key rate displays not much difference between two kinds of attack. Therefore our experiments can also be regarded as experimental validating of the decoy-state method against PNS attack.

There are three reasons that bring up errors of the count rates. The first one is the fluctuations of photon sources. The second one comes from the timing jitter of the signals. Only when all the signals are perfectly synchronized will Bob get the highest count rate. That is, any timing jitter of the signals will decrease the count rates. The third reason is due to statistical errors among finite samples. Those errors lead to the fluctuations of the calculated key rate shown in Fig. 5. For example, we happened to obtain a rather low count rate at the point of $\nu = 0.25$ (it can be seen in Fig. 4) and the calculated key rate becomes quite different from the theory. Comparing Fig. 4 and Fig. 5, we can also find that the final key rate is rather sensitive to the count rate. To reduce these errors, efforts should be made to cut down the fluctuations of photon sources by precisely controlling the temperature of laser diodes and the driven current, and to enhance the performance of synchronization. As for the qubit error, it is mainly caused by the birefringence of the optical fiber, which was passively compensated for in our experiments. Employing real-time active polarization compensation will be helpful to reduce such errors.

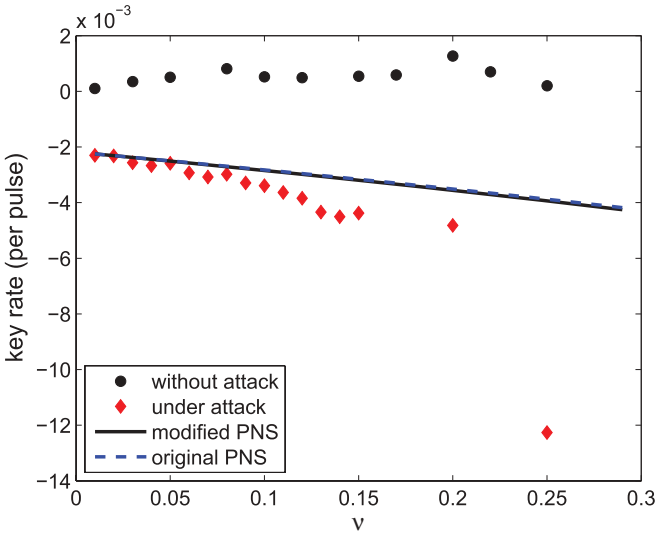

FIG. 5. (Color online) The secure key rate calculated with GLLP formula. The dots show the experimental results. Two decoy states are used in the experiments, namely, vacuum and another decoy state with an average photon number of ν. The average photon number of signal state is fixed at $\mu = 0.3$. The lines show the theoretical results. All those necessary parameters are obtained from the experiments. Error bars of the experimental data are not shown here because systematic errors were not quantitatively analyzed.

In addition, the AOM used in our experiments will lead to a frequency shift of the laser pulses sent to Bob, which could potentially be used to discover Eve, although it might be difficult in practice to do this. The frequency of acoustic waves employed for AOM is 78.5 MHz and only the first order of diffraction is sent to Bob, therefore there will be a frequency shift of 78.5 MHz for the pulses sent to Bob. In almost all the practical systems, the bandwidths of the laser pulses are much larger than 78.5 MHz. To discover such an attack by detecting the frequency shift means that Alice should possess a laser source which can work in pulse mode and the frequency of itself should be smaller than a typical frequency employed for AOM, and Bob should be capable for detecting a rather small frequency shift for a broadband signal of very low intensity.

On the other hand, Eve can make use of another kind of optical switch for this attack.

IV. CONCLUSION

In conclusion, we proposed a modified PNS attack scheme against practical QKD systems based on weak laser pulses and performed proof-of-principle experiments. This attack scheme can also work for other photon sources obtaining multiphoton pulses. In our experiments carried out in one-basis quantum communication, results show that Eve can obtain almost full information about the key generated between Alice and Bob without being discovered. In practice, information Eve can get will be cut by half since she cannot know which basis Alice used for each pulse in the case of having no quantum memory. The influences introduced by Eve are also explored by inspecting the count rates at Bob's side. For pulses of different average photon numbers, the equivalent attenuations introduced by Eve are different, therefore the decoy-state method can work well against this kind of attack, which has also been verified in our experiments. Using two decoy states, the final key rate between legitimate parties under attack becomes negative, thus they can successfully discover Eve's intervention.

ACKNOWLEDGMENTS

This work is supported by National Natural Science Foundation of China under Grants No. 61072071 and No. 11004248.

[1] C. H. Bennett and G. Brassard, in *Proceedings of the IEEE International Conference on Computers, Systems, and Signal Processing* (IEEE, New York, 1984), pp. 175-179.
[2] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, Rev. Mod. Phys. **74**, 145 (2002).
[3] M. Dušek, N. Lütkenhaus, and M. Hendrych, Prog. Opt. **49**, 381 (2006).
[4] W. Mauerer, W. Helwig, and C. Silberhorn, Ann. Phys. **17**, 158 (2008).
[5] V. Scarani, H. Bechmann-Pasquinucci, N. J. Cerf, M. Dušek, N. Lütkenhaus, and M. Peev, Rev. Mod. Phys. **81**, 1301 (2009).
[6] V. Makarov and D. Hjelme, J. Mod. Opt. **52**, 691 (2005).
[7] N. Gisin, S. Fasel, B. Kraus, H. Zbinden, and G. Ribordy, Phys. Rev. A **73**, 022320 (2006).
[8] V. Makarov *et al.*, New J. Phys. **11**, 065003 (2009).
[9] V. Makarov, A. Anisimov, and J. Skaar, Phys. Rev. A **74**, 022313 (2006).
[10] B. Qi *et al.*, Quantum Inf. Comput. **7**, 73 (2007).
[11] Y. Zhao, Chi-Hang Fred Fung, B. Qi, C. Chen, and H. K. Lo, Phys. Rev. A **78**, 042333 (2008).
[12] V. Makarov *et al.*, e-print arXiv:0809.3408.
[13] N. Lütkenhaus, Phys. Rev. A **61**, 052304 (2000).
[14] G. Brassard, N. Lütkenhaus, T. Mor, and B. C. Sanders, Phys. Rev. Lett. **85**, 1330 (2000).
[15] N. Lütkenhaus and M. Jahma, New J. Phys. **4**, 44 (2002).
[16] W.-Y. Hwang, Phys. Rev. Lett. **91**, 057901 (2003).
[17] H.-K. Lo, X. Ma, and K. Chen, Phys. Rev. Lett. **94**, 230504 (2005).
[18] X.-B. Wang, Phys. Rev. Lett. **94**, 230503 (2005).
[19] X. Ma, B. Qi, Y. Zhao, and H.-K. Lo, Phys. Rev. A **72**, 012326 (2005).
[20] X.-B. Wang, Phys. Rev. A **72**, 012322 (2005).
[21] X.-B. Wang, e-print arXiv:quant-ph/0509084.
[22] Y. Zhao, B. Qi, X. Ma, H. K. Lo, and L. Qian, Phys. Rev. Lett. **96**, 070502 (2006).
[23] D. Rosenberg *et al.*, Phys. Rev. Lett. **98**, 010503 (2007).
[24] T. Schmitt-Manderbach *et al.*, Phys. Rev. Lett. **98**, 010504 (2007).
[25] C. Z. Peng *et al.*, Phys. Rev. Lett. **98**, 010505 (2007).
[26] This is not valid in the cases that there are dark counts in Eve's detectors, especially when the average photon number sent by Alice is rather low. It will be discussed below based on the experimental results.
[27] B. Julsgaard *et al.*, Nature (London) **432**, 482 (2004).
[28] C.-W. Chou *et al.*, Science **316**, 1316 (2007).
[29] L. Childress, J. M. Taylor, A. S. Sorensen, and M. D. Lukin, Phys. Rev. Lett. **96**, 070504 (2006).
[30] A. L. Alexander, J. J. Longdell, M. J. Sellars, and N. B. Manson, Phys. Rev. Lett. **96**, 043602 (2006).
[31] M. U. Staudt *et al.*, Phys. Rev. Lett. **98**, 113601 (2007).
[32] The count rate is matched in a way opposite to the reality in our experiments for convenience. It does not influence the proof of the attack scheme. There is no technical challenge to keep the count rate unchanged in practice.
[33] D. Gottesman, H. K. Lo, N. Lütkenhaus, and J. Preskill, Quantum Inf. Comput. **5**, 325 (2004).