Effect of source tampering in the security of quantum cryptography

Effect of source tampering in the security of quantum cryptography

Shi-Hai Sun,[1,*] Feihu Xu,[2,3,†] Mu-Sheng Jiang,[1] Xiang-Chun Ma,[1] Hoi-Kwong Lo,[2,‡] and Lin-Mei Liang[1,4,§]

[1]*College of Science, National University of Defense Technology, Changsha 410073, China*

[2]*Center for Quantum Information and Quantum Control, Department of Electrical and Computer Engineering and Department of Physics, University of Toronto, Toronto, Ontario, M5S 3G4, Canada*

[3]*Research Laboratory of Electronics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA*

[4]*State Key Laboratory of High Performance Computing, National University of Defense Technology, Changsha 410073, China*

(Received 24 January 2015; published 4 August 2015)

The security of source has become an increasingly important issue in quantum cryptography. Based on the framework of measurement-device-independent quantum key distribution (MDI-QKD), the source becomes the only region exploitable by a potential eavesdropper (Eve). Phase randomization is a cornerstone assumption in most discrete-variable (DV) quantum communication protocols (e.g., QKD, quantum coin tossing, weak-coherent-state blind quantum computing, and so on), and the violation of such an assumption is thus fatal to the security of those protocols. In this paper, we show a simple quantum hacking strategy, with commercial and homemade pulsed lasers, by Eve that allows her to actively tamper with the source and violate such an assumption, without leaving a trace afterwards. Furthermore, our attack may also be valid for continuous-variable (CV) QKD, which is another main class of QKD protocol, since, excepting the phase random assumption, other parameters (e.g., intensity) could also be changed, which directly determine the security of CV-QKD.

DOI: 10.1103/PhysRevA.92.022304 PACS number(s): 03.67.Dd, 03.67.Hk, 42.50.Ex

I. INTRODUCTION

Quantum key distribution (QKD) [1] allows two remote parties to share an unconditional secret key, which has been proven in theory [2–4] and demonstrated in experiment [5]. However, the imperfections of practical devices will compromise the security of QKD systems [6–14]. So far, three main approaches have been proposed to bridge the gap between theory and practice. The first one is to close specific loopholes of devices with security patches [15], but it could not close potential and unnoticed loopholes. The second one is device-independent (DI-) QKD [16–18]. By testing Bells inequality in a loophole-free setting, security could be obtained without detailed information about the implementation devices. But DI-QKD is impractical because an almost perfect single-photon detector (SPD) is required, and even so the secret key rate is limited [19,20]. The third approach is to remove as many device loopholes and assumptions as possible by either modifying the QKD protocol or refining the security proof. One of the best results with this approach is measurement-device-independent (MDI) QKD [21], which can remove all detector loopholes. Since the detection system is widely regarded as the Achilles' heel of QKD [6,8,9,13], MDI-QKD is of great importance. Indeed, recently, MDI-QKD has been demonstrated both in the laboratory and in the field [22].

Based on the framework of MDI-QKD, the source becomes the final battlefield for the legitimate parties and Eve. And the major flaw of the source is that a semiconductor laser diode (S-LD), which generates a weak coherent state, is normally used as a single-photon source in most commercial and research QKD systems [5,22]. The security of MDI-QKD as well as BB84 based on S-LD has been proven with decoy state method [23]. Hence, it has been convinced that if the source can be well characterized (for example the source flaws could be taken care of with the loss-tolerant QKD protocol [24]), perfect security can still be obtained.

Generally speaking, there are two main classes of QKD protocols: one is discrete-variable (DV) QKD (including BB84, decoy state BB84, MDI-QKD, Scarani-Acin-Ribordy-Gisin (SARG04) [25], differential phase shift (DPS) [26], and so on), and the other one is continuous-variable (CV) QKD [27]. In most DV-quantum communication protocols (e.g., DV-QKD, quantum coin tossing (QCT) [28], weak-coherent-state blind quantum computing (BQC) [29]), the phase randomization is a cornerstone assumption. By assuming that the overall phase is uniformly distributed from 0 to 2π (in fact, discrete randomization with finite points, e.g., 10, is sufficient to guarantee QKD security [30]), a coherent state with intensity $|\alpha|^2$ is reduced into a classical mixture state, that is, $\rho = \int_0^{2\pi} \frac{d\theta}{2\pi} |\alpha e^{i\theta}\rangle\langle\alpha e^{i\theta}| = \sum_{n=0}^{\infty} \frac{e^{-|\alpha|^2}|\alpha|^{2n}}{n!} |n\rangle\langle n|$. Then it allows one to apply classical statistics theory to analyze quantum mechanics. Note that although the security of QKD with nonrandom phase had been proven [31], the performance is very limited in distance and key rate.

In this paper, however, we demonstrate a simple quantum hacking strategy, with both a commercial and homemade pulsed laser based on S-LD, that allows Eve to actively violate the phase randomization assumption, without leaving a trace afterwards. Thus it is effective for most DV quantum communication protocols. Our attack may also be effective for CV-QKD, since other parameters of the source (e.g., intensity) could also be changed. For example, it had been proven that the local oscillator fluctuation will compromise the security of CV-QKD [14]. Since S-LDs are widely used in most

*shsun@nudt.edu.cn
†tigerfeihuxu@gmail.com
‡hklo@ece.utoronto.ca
§nmliang@nudt.edu.cn

quantum information protocols (e.g., DV-QKD, CV-QKD, QCT, BQC, and so on) and the security of these protocols is closely related to S-LD parameters [4], our work constitutes an important step towards secure quantum information processing.

Our attack differs from previous attacks [6–14]. First, in our attack, Eve actively violate some basic assumptions required in the security proof by tampering with an initial perfect source. Second, unlike the laser damage attack [13] in which Eve also actively creates loopholes for a perfect SPD, the loopholes created by our attack are temporary; this makes our attack impossible for Alice and Bob to detect during the off-time of the QKD system. Third, our attack also differs from the Trojan horse attack [32,33]. In our attack, Eve directly breaks some basic assumptions of QKD protocols, whereas in the Trojan horse attack, backreflected light is measured to analyze Alice's information. And as the best we know, the Trojan horse attack is invalid for Alice with multilasers [34], but our attack remains applicable to such systems. Fourth and most importantly, our attack targets the source instead of SPD. This makes our attack a serious threat for most quantum information protocols (not only QKD, but also QCT and BQC).

Here we emphasize that the phase randomization is a cornerstone assumption in the security of many quantum communication protocols including QKD, QCT, and BQC. It is important for not only weak coherent pulse–based protocols, but also, for instance, parametric down-conversion–based protocols [35]. And continuous or discrete phase randomization is also crucial for the loss-tolerant protocol [24]. In fact, without the phase randomization, the performance of a quantum communication protocol will be dramatically reduced in distance and key rate [31]. However, we demonstrate experimentally in a clear manner how easy it is for Eve to violate such a fundamental assumption in a practical setting. Thus our work is very general for most quantum information processing protocols. It works for most DV-QKD, with various encoding schemes (polarization, phase, and time bin) and various kinds of lasers (pulsed laser and cw laser). It is also possibly a serious threat for CV-QKD and other quantum information processing protocols (such as QCT and BQC).

The basic principle of our attack is as follows. In the interdriven mode, the semiconductor medium of the S-LD is excited from loss to gain by each driving current pulse. A laser pulse is generated from *seed* photons originating from spontaneous emission. The phase of the laser pulse is determined by the seed photons. Since the phase of the seed photons is random, the phase of each laser pulse is random inherently [36–39]. However, if a certain number of photons are injected from an external source into the semiconductor medium, these photons will also be amplified to generate laser pulses. Consequently, the seed photons consist of two parts: one from spontaneous emission and the other part from the external source. Both parts will affect the phase of the resulting laser pulse. If the injected photons greatly outnumber the photons from spontaneous emission, the phase of the output laser pulse is largely determined by the phase of the injected photons. Therefore, Eve can control the phase of Alice's signal laser by illuminating the S-LD from an external "control source" and successfully violate the phase randomization assumption.

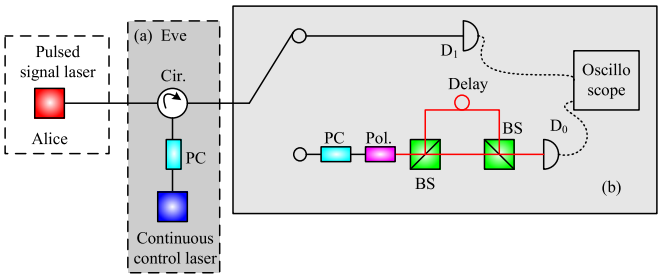

FIG. 1. (Color online) Schematic setup of our experiment. Part (a) shows Eve's control devices, in which Eve uses a continuous wave (cw) laser to tamper with the parameters of Alice's pulsed signal laser. Part(b) shows the experimental setups to measure the parameters of Alice's signal pulses. The phase of adjacent pulse is measured by an unbalanced Mach-Zehnder interferometer [lower arm of part(b)], and the waveform of Alice's signal pulse is directly measured with a photodiode [upper arm of part(b)]. The output of photodiodes (D_0 and D_1) are recorded with an oscilloscope. Cir.: circulator; PC: polarization controller; Pol.: polarizer; BS: beam splitter. Solid lines are optical fibers (single-mode fiber for black color and polarization-maintaining fiber for red color), and dashed lines are electrical lines. Here we consider Eve's control laser working at continuous wave (cw) mode. However, in later parts of this paper, we will consider the possibility that Eve modulates her control laser into short photon pulses. This can make it harder for Alice to detect Eve's attack.

II. EXPERIMENT AND MAIN RESULTS

Figure 1 shows the schematic setup of our experiment. We test four sample S-LDs operating in interdriven mode, two ID300 pulsed lasers from IdQuantique [40] (numbers ID300-1 and ID300-2) and two homemade pulsed lasers with S-LDs from Sunstar Communication Technology Co., Ltd. (model SDLP55HMBIFPN, numbers HM-1 and HM-2). To measure the phase relationship between adjacent pulses, an unbalanced Mach-Zehnder interferometer is used [see Fig. 1(b)]. The repetition rate of the signal laser is set to be 206.34 MHz to match the delay of the interferometer. The output light is detected by a photodiode (D_0) with a bandwidth of 1 GHz, and the voltage of each pulse is recorded using an oscilloscope with bandwidth 33 GHz and sample rate 80 GHz (Agilent, model DSOX93304Q).

Because the central frequency (with a finite linewidth) and polarization of the signal laser are unstable in experiment, Eve needs to carefully modulate the frequency and polarization of her control laser to match her control laser with Alice's signal laser. In our experiment, a tuning laser module (model: 81600B-201, Agilent) is used as Eve's control laser. Furthermore, in Fig. 1 of the main text, we consider Eve's control laser working at cw mode. However, at the end of this paper we consider the possibility that Eve modulates her control laser into short photon pulses. This can reduce Alice's ability to detect Eve's attack.

In theory, the output voltage after D_0 is $V_P \propto [1 + \cos(\Delta\phi + \theta_0)]/2$, where $\Delta\phi$ is the phase difference between adjacent pulses, and θ_0 is the inherent phase difference between the two paths of the interferometer. By passively controlling the interferometer with a temperature controller and vibration

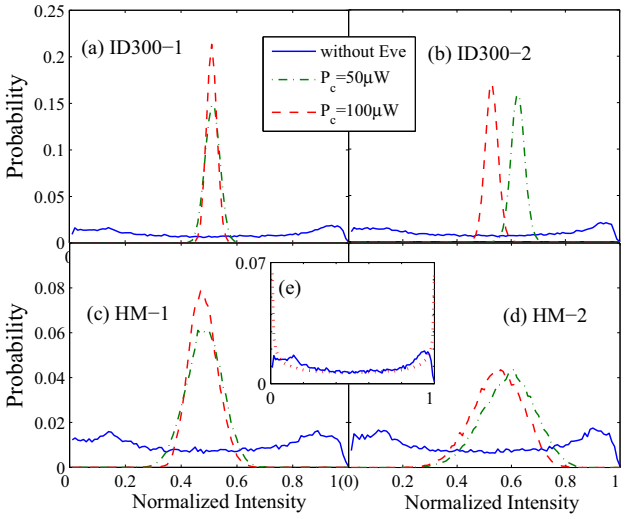

FIG. 2. (Color online) Experimental results for normalized intensity distribution of V_P^s. P_c is the power of Eve's control laser. Parts (a)–(d) show the intensity distribution of four S-LDs with Eve's different control intensities. Part (e) shows the theoretical simulation (dashed line) of the probability distribution when the phase of each pulse follows a uniform distribution from 0 to 2π, and the experimental results of ID300-1 (solid line) when Eve is absent. These results clearly show that when photons are injected into Alice's signal laser, the phase of the signal laser becomes correlated. Here P_c is not minimized for Eve [41], and a further experiment about the minimal power is discussed in the following text (see Fig. 4).

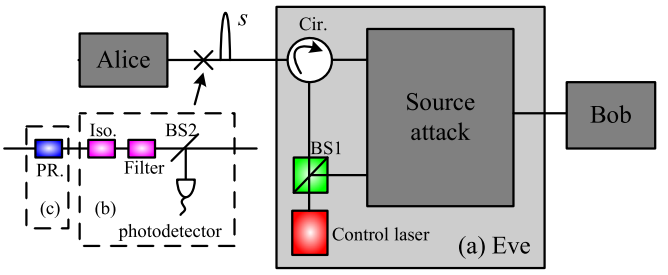

FIG. 3. (Color online) (a) Principle scheme to attack a complete QKD system by combining our attack with those of [42]. s is Alice's quantum signal pulse. Eve splits her bright control pulse into two parts with a beam splitter (BS1); one part serves as control laser to tamper with the parameters of Alice's signal pulse, while the other part serves as phase reference for Eve to perform the source attack [42]. (b) A possible countermeasure for Alice to monitor our attack. Alice splits parts of the light with BS2 and monitors the power with a photodetector. The optical frequency filter is used to remove all wavelength-dependent flaws of Alice's source. The isolator (Iso.) is used to prevent light from entering Alice's laboratory from the quantum channel. (c) Active phase randomization scheme (PR.), which can guarantee the phase randomization assumption and partially reduce the risk of our attack, but it cannot entirely remove our attack (see text for details).

isolator, we can stabilize the interferometer within about 2 min. In the test we set the number of pulses to be 25 791 in each experimental point of Fig. 2. (In each experimental point of Fig. 2, we collect and store 10 M data.) Note that the repetition rate of the laser is 206.34 MHz and the sample rate of the oscilloscope is 80 GHz. The number of data is about (1/206.34 MHz)/(1/80 GHz) $\approx$ 388 in each pulse cycle. Thus the number of pulses is about 10 M/388 $\approx$ 25 791), and the time interval is about 0.125 ms (25 791/206.34 MHz), which is much lower than the time scale of the interferometer. Thus we could set $V_P^s \propto [1 + \sin(\Delta\phi)]/2$ for $\theta_0 = \pi/2$.

A uniform distribution of $\Delta\phi$ from 0 to 2π will produce a U-type intensity distribution, due to the fact that the mapping from phase to intensity is nonlinear, $V_P \propto \sin(\Delta\phi)$. Indeed, when Eve is absent, the same distributions (solid lines of Fig. 2) are obtained in experiments with both ID-300 and the homemade pulsed laser. However, a bright light from Eve could correlate the phase of each pulse and violate the phase randomization assumption (dashed lines of Fig. 2). In fact, when photons are injected into Alice's signal laser, the intensity distribution of V_P^s for both ID300 and the homemade signal laser becomes Gaussian. Consequently, various quantum hacking strategies can be applied to spy on the final key [42]. Figure 3(a) shows a schematic setup to attack a complete QKD system.

Theoretically speaking, Eve can perfectly control the phase of Alice's source, and then the intensity distribution should be a sharp line. However, owing to the following two main reasons, the measured intensity distribution in Fig. 2 of the main text follows Gaussian distribution: (1) There exists phase noise in Eve's controlling laser, which follows Gaussian distribution. The measured intensity is the interference of adjacent pulses (the interval of adjacent pulses is about 5 ns); thus the experimental results depend on the phase noise of Eve's control laser at different times. (2) The interference is imperfect, including the loss of two paths of the interferometer, the time jitter of the optical pulse, and so on. Therefore, a practical Eve cannot perfectly control the phase of Alice's source, and the phase noise decides how much information will be leaked to Eve. Furthermore, although the security of the BB84 protocol had been proven based on a uniformed random phase from 0 to 2π [4] and nonrandom phase [31], the key rate (or mutual information between Alice and Eve) is still unknown, if the phase of source follows Gaussian distribution or a general probability distribution, which will be studied in future.

Furthermore, we note that when the LD is operated in interdriven mode, the emitted pulses have random phase, and such phase noise had been used as a quantum random number generator by many groups [36–39]. However, Fig. 2(e) of the main text does not prove that the phase of each pulse follows uniform distribution from 0 to 2π. In fact, if the phase is uniformly distributed from 0 to π, the same probability distribution could also be obtained. Thus the phase randomization assumption must be carefully evaluated, particularly for a high-speed QKD system [39]. Active phase randomization [43] is a good countermeasure to guarantee the phase randomization assumption.

III. COUNTERMEASURE

Figure 3(b) shows a possible countermeasure for Alice to monitor our attack. It includes three main devices: an isolator

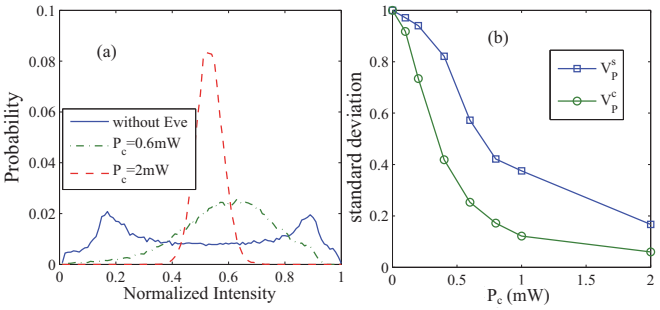

FIG. 4. (Color online) (a) Experimental results for V_P^s, when a 25-dB isolator is placed after the signal laser ID300-1. (b) The standard deviation of $V_P^s \propto \sin(\Delta\phi)$ and $V_P^c \propto \cos(\Delta\phi)$ with different powers of control light. The standard deviation has been normalized by that of $P_c = 0$. The experimental results clearly show that, even if a 25-dB isolator is used by Alice, the intensity distribution is still Gaussian type but not U type when Eve uses a cw laser with a power of 0.6 mW, which means that Eve could still introduce a nonrandom phase in Alice's quantum signal. In the test, only a 25-dB isolator is put after the output of Alice (the photodetector and the filter will be discussed later). Other setups used here are the same as those for Fig. 2.

(Iso.), a filter, and a photodetector. But these devices could not defeat our attack completely, if they are not carefully configured (see Appendix A for details). (1) The isolator could not entirely stop Eve's photons due to its finite isolation (see Fig. 4), and other imperfections of practical isolators have been found in a recent paper [33]. (2) Since the wavelength of Eve's control laser is the same as that of Alice's signal laser in our attack, an optical frequency filter is also ineffective. (3) Both an optical power meter and classical photodetector could be foiled by Eve so that they could not accurately show the power of light from the channel. For example, a short pulse light might reduce the average power of Eve's light, and the finite bandwidth of these monitor devices might worsen the monitoring results. Furthermore, a recent paper also shows other imperfections of a practical monitoring photodetector [44].

An active phase randomization [Fig. 3(c)] [43], or the cw laser followed by an external intensity modulator and an active phase randomization scheme, is another important choice for practical QKD systems, especially when the QKD system works in a high repetition rate [39]. Then phase randomization assumption is automatically guaranteed. But such a countermeasure may not remove our attack entirely, since Eve can tamper with other parameters (e.g., intensity and shape, see Fig. 5) to compromise the security of such systems. For example, the key rate of both CV-QKD and DV-QKD depends on the intensity of the signal pules [14,45,46]. But the stability of S-LD (no matter whether it works on pulsed mode or cw mode) could be damaged by bright light so that the intensity of Alice's laser is unstable. Therefore, in this sense, our attack is also effective for the QKD system with a cw laser and an active phase randomization scheme. Another countermeasure is to use a protocol (or security proof) with an unrandom phase, but the performance of such a protocol is dramatically reduced in distance and key rate [31].

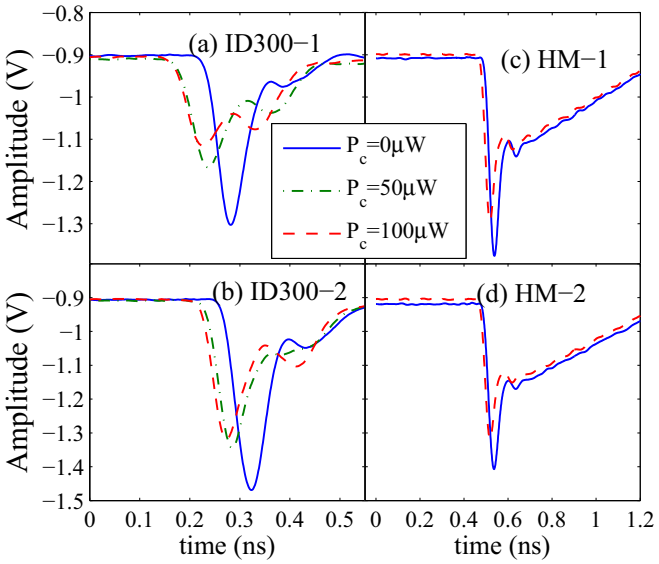

FIG. 5. (Color online) Measured signal pulse waveforms when Alice's signal laser is illuminated by a bright light. Eve sends a bright cw light to Alice's signal lasers (including both commercial and homemade pulsed lasers), then the signal pulse of Alice is directly measured using a photodiode (D_1) with bandwidth 40 GHz, an oscilloscope with bandwidth 33 GHz, and sample rate 80 GHz (model DSOX93303Q, Agilent). The repetition rate of the signal laser is 10 MHz. It is clearly seen that when Alice's signal laser is illuminated, the pulse amplitude and width will be changed.

IV. DISCUSSION

Figure 5 shows that the pulse shape would also be changed by Eve's bright light. These changed parameters are also helpful. For example, the signal pulse is emitted earlier than that without Eve [47], and the time shift is different for each S-LD. Furthermore, in the absence of an external field, the first oscillation is much stronger than the following oscillation, and a few oscillations appear [48]. But when Eve is present, more oscillations are observed, and different laser diodes have different oscillation waveforms. Thus it is possible for Eve to compromise the security of QKD systems with multilasers [34] by measuring the characters of signal pules (e.g., time shift, pulse width, optical frequency).

Here we remark that, generally speaking, the changes of pulse shape are helpful for both Eve and Alice. Although more imperfection could be exploited by Eve, more parameters could be monitored by Alice to discover the existence of Eve. In fact, both Eve and Alice must be very careful in the cat-and-mouse game (see Appendix B for details). First, if Alice wants to completely monitor the changes of pulse shape, some advanced devices with high speed and bandwidth are required which may dramatically increase the technology challenge and cost of a practical Alice. Second, Eve could carefully configure her attack to ensure that her attack could not increase the error rate and the changes of pulse shape could not be discovered by Alice. Third, generally speaking, the changed shape may actually benefit Eve more than Alice and Bob. This is because Eve could well be a spy or national security agency such as the NSA, and so Eve has a much larger power and budget than Alice and Bob. Thus Eve is probably in a better position

to exploit the imperfections that she has introduced in the quantum signal. Furthermore, note that even a tiny violation of the phase randomization assumption or other parameters of the source will undermine the very foundation of security proofs in QKD and it will no longer be fair for Alice and Bob to claim unconditional security.

Finally, in addition to using a laser, Eve can also attack the QKD system by using temperature, microwave radiation, and so on. At the same time, although most quantum hackers focus on the optical devices of the legitimate parties, Eve can also exploit imperfections in the electrical devices of the QKD system. For example, if the electromagnetic shielding of devices of Alice and Bob is imperfect, Eve could use microwave radiation from outside to control the parameters of these devices. These are the subjects of future investigations.

V. CONCLUSION

In summary, phase randomization is a cornerstone assumption for many quantum communication protocols, and a tiny violation of such an assumption is fatal to the security of such protocols. However, here we demonstrate experimentally, with both commercial and homemade pulsed lasers, how easy it is for Eve to violate such a fundamental assumption in a practical setting. Additionally, besides the random phase, other parameters (e.g., intensity) of the source could also be changed. Our attack works for most DV-QKD protocols and possibly for CV-QKD and other quantum information processing protocols (e.g., QCT and BQC). Thus our work constitutes an important step towards secure quantum information processing.

ACKNOWLEDGMENTS

We thank Z. Yuan and V. Makarov for helpful discussions. This work is supported by the National Natural Science Foundation of China, Grant No. 11304391. L.M.L. is supported by the NCET program. H.-K.L. is supported by NSERC.

APPENDIX A: THE SCHEME FOR EVE TO FOIL ALICE'S MONITOR DEVICES

Now we show that Alice's countermeasure of the main text (including an isolator, an optical filter, and a photodetector), shown in Fig. 3(b), cannot remove our attack entirely.

(i) *Isolator.* In general, an optical isolator serves to prevent backreflected photons from returning to Alice's laboratory. However, owing to the finite isolation of practical isolators, this approach only reduces the probability that photons infuse into Alice's zone but cannot eliminate this probability entirely. We perform a proof-of-principle experiment by inserting a 25-dB isolator after the output port of the signal laser ID300-1. The experimental results of Fig. 4 of the main text show that the intensity distribution is still Gaussian type but not U type when Eve uses a cw laser with a power of 0.6 mW. Thus the phase of adjacent pulses can be still correlated. Although isolation of some commercial isolators reaches 50 dB (or Alice can use two or more isolators in series to increase the isolation), it cannot totally foil our attack, because Eve can always increase the power of her control laser. Furthermore, other imperfections of the practical isolator have been found in a recent paper [33].

(ii) *Filter.* An optical frequency filter is often used by Alice to remove any wavelength-dependent flaws. By doing so, only the light within a narrow band of frequencies can enter Alice's laboratory. However, the wavelength of Eve's control laser is the same as that of Alice's signal laser in our attack. Thus an optical frequency filter is not an effective countermeasure against our attack.

(iii) *Photodetector.* Alice can use both an optical power meter and photodetector to monitor the intensity of light from a quantum channel, but the optical power meter measures the average power of light. Thus it could be foiled by Eve who uses a pulsed laser. For example, Fig. 4 of the main text shows that a cw laser with an optical power of 0.6 mW is sufficient to correlate the phase of Alice's signal pulse. Now, suppose that the repetition rate of the QKD system is 10 MHz and Eve uses a pulsed control laser with width of 100 ps. Then the duty circle of Eve's pulse is $100\ \text{ps}/10\ \text{ns} = 0.001$. Thus the average optical power is reduced to $0.6\ \text{mW} \times 0.001 = 0.6\ \mu\text{W}$.

A classical photodetector with a discrimination voltage can be used to monitor the intensity of pulsed light. However, the classical photodetector could also be cheated due to the following two reasons:

First, the classical photodetector can be damaged by bright light so that it may not work as expected. There are two kinds of classical photodetectors: one based on the PIN, and the other one based on the APD. Both can be damaged by bright light [13]. For example, the detector based on InGaAs-APD from Thorlabs has a maximal input power of 10 mW (model APD310) and 1 mW (model APD110C). The maximal input power for the detector based on InGaAs-PIN from Thorlabs (model PDA8GS) is about 1 mW for cw and 20 mW for 60 ms [49].

Second, the finite bandwidth of the classical photodetector may worsen the monitoring results. We experimentally measure the amplitude of an electrical signal using an oscilloscope with various bandwidths [Fig. 6(a)]. Furthermore, the theoretical amplitudes of an ideal Gaussian pulse which passes a linear time-invariant ideal low-pass filter are also shown in Figs. 6(b)–6(c). Generally speaking, when a signal pulse $f(t)$ passes a linear time-invariant device, its amplitude function becomes

$$g(t) = \int_{-\infty}^{\infty} G(\omega) F[f(t)] e^{i\omega t} d\omega, \tag{A1}$$

where $F[\cdot]$ is the Fourier transformation and $G(\omega)$ is the frequency response function of device. It clearly shows that devices with finite bandwidth will filter high-frequency signals and reduce the amplitude of a signal pulse. For simply, we assume that the signal is a Gaussian pulse and the device is an ideal low-pass filter, that is,

$$\begin{aligned} f(t) &= \exp\left[-\frac{t^2}{2\sigma^2}\right], \\ G(\omega) &= \begin{cases} 1 & |\omega| \leqslant \omega_0 \\ 0 & |\omega| > \omega_0 \end{cases}. \end{aligned} \tag{A2}$$

Here σ is the standard deviation of a signal pulse $f(x)$. If the 3-dB width of $f(x)$ is noted as Δx, it is easy to check that $\Delta x = \sqrt{8\ln(2)}\sigma$. ω_0 is the maximal bandwidth of the ideal low-pass filter.

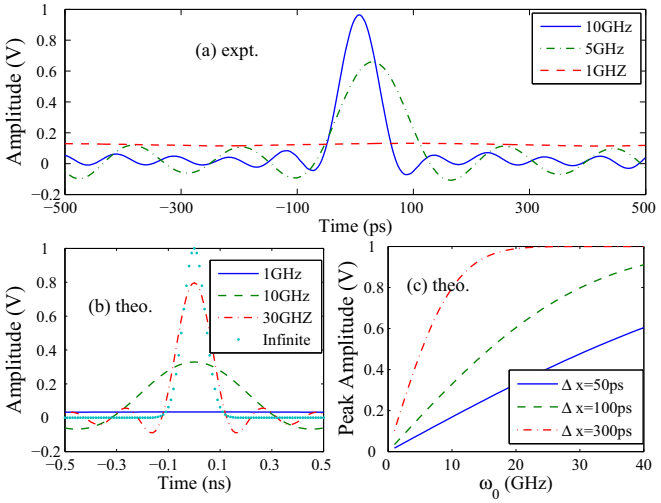

FIG. 6. (Color online) Part (a) shows the experimentally measured pulse amplitudes by directly inputting an electrical signal with amplitude 1 V and 3-dB width 100 ps into an oscilloscope (model DSOX93304Q, Agilent) with various bandwidths: 1, 5, and 10 GHz. The electrical signal is generated from a pattern generator (model 12050, Picosecond Pulse Labs). Parts (b) and (c) show the theoretical amplitude of an ideal Gaussian pulse which passes an ideal low-pass filter. Δx is the 3-dB width of the Gaussian signal. ω_0 is the maximal bandwidth of the low-pass filter. In part (b), we set $\Delta x = 100$ ps. The mismatch between experiment [part (a)] and theory [part (b)] is mainly due to the simplified version of our model. All the results show that the monitor devices with finite bandwidth cannot faithfully characterize the input signal.

The theoretical amplitude of $g(t)$ is shown in Figs. 5(b)–5(c) of the main text. The results clearly show that monitoring devices with finite bandwidth could not faithfully characterize the factual amplitude of the input signal, and Eve could foil the monitoring devices with a sharp pulsed signal. Although the test of Fig. 6 is performed for an electrical signal, the results can be directly applied to the photodetector with finite bandwidth. For example, suppose that the gain and discrimination voltage of the photodetector are 10^4 V/W and 0.2 V, and that Eve uses a pulsed control light with a 3-dB width of 100 ps and a peak power of 100 μW. Then the expected output voltage of the photodetector should be 1 V, which is much larger than the discrimination voltage, 0.2 V.

Figure 5 of the main text also shows that if the bandwidth of Alice's photodetector is high enough (e.g., >5 GHz), Eve can be discovered. (Note that generally speaking, the gain of the photodetector will be decreased when the bandwidth is increased, but here we simply assume the gain is independent of the bandwidth.) However, if the bandwidth of the photodetector is limited (e.g., 1 GHz), the factual output voltage is lower than the discrimination voltage, 0.2 V. Alice cannot discover the existence of Eve. Note that Fig. 2 of the main text has shown that 100 μW is sufficient for Eve to break the phase randomization assumption. Furthermore, a recent paper also shows other imperfections in a practical monitoring photodetector [44].

Therefore, the possible countermeasure of Fig. 3(b) of the main text could be cheated by Eve if the devices are not carefully configured. Furthermore, illumination by a bright light changes not only the phase but also the pulse waveform, including its width, amplitude, and shape. Although we still do not know how Eve can obtain more information by exploiting such a modified waveform, it remains possible for Eve to attack the QKD system.

APPENDIX B: A SIMPLE DISCUSSION ABOUT FIG. 5

Figure 5 of the main text clearly shows that when the signal laser is illuminated by bright light, the pulse shape would also be changed. Generally speaking, the additional changes are helpful for both Eve and the legitimate parties. More imperfections can be exploited by Eve to spy the final key, and more parameters can be monitored by Alice to discover the existence of Eve. But it is still possible for Eve to perform our attack.

Theoretically speaking, Eve could perform a suitable attack to ensure that the modification of the pulse shape would not increase the error rate between Alice and Bob. In fact, Eve can perform the intercept-and-resend attack and ensure that the error rate is lower than a reasonable value. For example, in the system with multilaser diodes, she first measures the time shift of each laser diode to determine Alice's state. Then she can resend a faked state to Bob according to her measurement results. In this case, if the time shift is distinguishable for each laser diode (it is possible according to Fig. 5), Eve could know the state sent by Alice. Then she can resend a perfectly faked state to Bob according to her measurement. Thus no additional error will be introduced, and the legitimate parties could not discover the existence of Eve by monitoring the error rate.

Therefore, the main battlefield for Alice and Eve is the monitor devices, and both of them must be very careful in the cat-and-mouse game.

For Alice, she may discover the existence of Eve by carefully monitoring the parameters of the signal laser. But since the change is tiny in some parameters, some advanced devices with high speed and bandwidth (e.g., photodetectors, analog-digital convertors, or time-amplitude convertors, and so on) are required for Alice, which may dramatically increase the technology challenge and cost of a practical Alice. For example, the time shift for ID300 lasers is about 100 ps; thus if Alice wanted to characterize the time shift of her pulses, the bandwidth and sample rate of Alice's analog-digital convertor should be larger than 40 GHz (generally speaking, at least four points are needed to recover a pulse). Furthermore, the bandwidth and sample rate should be increased for homemade lasers (see Fig. 5 of the main text for HM-1 and HM-2), since much smaller changes are introduced.

For Eve's part, she should carefully configure her attack to foil Alice's monitor devices. (1) Eve may carefully stable her controlling laser and match the optical frequency of her controlling laser with that of Alice's signal laser, so that, excepting the random phase, many tiny changes will be introduced on the pulse shape. Taking the homemade lasers (HM-1 and HM-2) as an example, Eve's light will correlate the phase of each of the pulses [see Figs. 2(c) and 2(d) of the main text], but Figs. 5(c) and 5(d) of the main text show that the changes of pulse shape are very tiny. (At least, compared with ID300-1 and ID300-2, we do not find any obvious changes in

the pulse shape using a photodetector with 40-GHz bandwidth, an oscilloscope with 33-GHz bandwidth, and a sample rate of 80 GHz); thus if Alice wants to discover the changed shape of HM-1 and HM-2, advanced devices with higher bandwidth and sample rate are required. (2) Eve may reduce the risk of being discovered by spying parts (not all) of the final key. For example, it has been proven that a small fluctuation of intensity will dramatically reduce the secret key rate of decoy state BB84 protocol [50]. Thus she still could obtain parts of the final key by trivially changing the intensity of Alice's signal laser. In fact, it has been shown that if the intensity of Alice's signal pulses fluctuates 1%, 2%, and 3%, the final key rate will be reduced by 11.86%, 23.91%, and 36.17%, respectively [50]. (The simulation was performed based on the experimental parameters of Ref. [34b].)

Furthermore, generally speaking, the ability for Eve to change other parameters in an optical signal may actually benefit Eve more than Alice and Bob. This is because Eve could well be a spy or work for a national security agency such as the NSA, and so Eve has a much larger budget than Alice and Bob and thus is probably in a better position to exploit the imperfections that she has introduced in the quantum signal.

[1] C. H. Bennett and G. Brassard, in *International Conference on Computers, Systems and Signal Processing, Bangalore, India* (IEEE, New York, 1984), p. 175–179.

[2] H. K. Lo and H. F. Chau, Science **283**, 2050 (1999).

[3] P. W. Shor and J. Preskill, Phys. Rev. Lett. **85**, 441 (2000).

[4] D. Gottesman, H. K. Lo, N. Lütkenhaus, and J. Preskill, Quantum Inf. Comput. **4**, 325 (2004).

[5] S. Wang, W. Chen, J. F. Guo, Z. Q. Yin, H. W. Li, Z. Zhou, G. C. Guo, and Z. F. Han, Opt. Lett. **37**, 1008 (2012); Y. Liu, T.-Y. Chen, J. Wang, W.-Q. Cai, X. Wang, L.-K. Chen, J.-H. Wang, S.-B. Liu, H. Liang, L. Yang *et al.*, Opt. Express **18**, 8587 (2010); Z. L. Yuan, A. R. Dixon, J. F. Dynes, A. W. Sharpe, and A. J. Shields, Appl. Phys. Lett. **92**, 201104 (2008).

[6] Y. Zhao, Chi-Hang Fred Fung, B. Qi, C. Chen, and H.-K. Lo, Phys. Rev. A **78**, 042333 (2008).

[7] F. H. Xu, B. Qi, and H. K. Lo, New J. Phys. **12**, 113026 (2010).

[8] L. Lydersen, C. Wiechers, C. Wittmann, D. Elser, J. Skaar, and V. Makarov, Nat. Photonics **4**, 686 (2010).

[9] I. Gerhardt, Q. Liu, A. Lamas-Linares, J. Skaar, C. Kurtsiefer, and V. Makarov, Nat. Commun. **2**, 349 (2011).

[10] S. H. Sun, M. S. Jiang, and L. M. Liang, Phys. Rev. A **83**, 062331 (2011).

[11] N. Jain, C. Wittmann, L. Lydersen, C. Wiechers, D. Elser, C. Marquardt, V. Makarov, and G. Leuchs, Phys. Rev. Lett. **107**, 110501 (2011).

[12] H. Weier, H. Krauss, M. Rau, M. Fürst, S. Nauerth, and H. Weinfurter, New J. Phys. **13**, 073024 (2011).

[13] A. N. Bugge, S. Sauge, A. M. M. Ghazali, J. Skaar, L. Lydersen, and V. Makarov, Phys. Rev. Lett. **112**, 070503 (2014).

[14] X. C. Ma, S. H. Sun, M. S. Jiang, and L. M. Liang, Phys. Rev. A **88**, 022339 (2013).

[15] Z. L. Yuan, J. F. Dynes, and A. J. Shields, Nat. Photonics **4**, 800 (2010); T. F. da Silva, G. B. Xavier, G. P. Termorao, and J. P. von der Weid, Opt. Express **20**, 18911 (2012); T. F. da Silva, G. C. do Amaral, G. B. Xavier, G. P. Temporao, and J. P. von der Weid, IEEE J. Sel. Top. Quantum Electron. **21**, 6600309 (2015); C. C. W. Lim, N. Walenta, M. Legré, N. Gisin, and H. Zbinden, *ibid.* **21**, 6601305 (2015).

[16] D. Mayers and A. Yao, in *Proceedings of the 39th Annual Symposium on Foundations of Computer Science FOCS'98* (IEEE, Washington, US, 1998), p. 503.

[17] A. Acín, N. Brunner, N. Gisin, S. Massar, S. Pironio, and V. Scarani, Phys. Rev. Lett. **98**, 230501 (2007).

[18] S. Pironio, A. Acín, N. Brunner, N. Gisin, S. Massar, and V. Scarani, New J. Phys. **11**, 045021 (2009).

[19] N. Gisin, S. Pironio, and N. Sangouard, Phys. Rev. Lett. **105**, 070501 (2010).

[20] M. Curty and T. Moroder, Phys. Rev. A **84**, 010304(R) (2011).

[21] H. K. Lo, M. Curty, and B. Qi, Phys. Rev. Lett. **108**, 130503 (2012).

[22] A. Rubenok, J. A. Slater, P. Chan, I. Lucio-Martinez, and W. Tittel, Phys. Rev. Lett. **111**, 130501 (2013); Y. Liu, T.-Y. Chen, L.-J. Wang, H. Liang, G.-L. Shentu, J. Wang, K. Cui, H.-L. Yin, N.-L. Liu, L. Li *et al.*, *ibid.* **111**, 130502 (2013); Z. Y. Tang, Z. F. Liao, F. H. Xu, B. Qi, L. Qian, and H. K. Lo, *ibid.* **112**, 190503 (2014); Y.-L. Tang, H.-L. Yin, S.-J. Chen, Y. Liu, W.-J. Zhang, X. Jiang, L. Zhang, J. Wang, L.-X. You, J.-Y. Guan *et al.*, *ibid.* **113**, 190501 (2014); T. Ferreira da Silva, D. Vitoreti, G. B. Xavier, G. C. do Amaral, G. P. Temporao, and J. P. von der Weid, Phys. Rev. A **88**, 052303 (2013).

[23] W. Y. Hwang, Phys. Rev. Lett. **91**, 057901 (2003); H. K. Lo, X. F. Ma, and K. Chen, *ibid.* **94**, 230504 (2005); X.-B. Wang, *ibid.* **94**, 230503 (2015).

[24] K. Tamaki, M. Curty, G. Kato, H.-K. Lo, and K. Azuma, Phys. Rev. A **90**, 052314 (2014).

[25] V. Scarani, A. Acin, G. Ribordy, and N. Gisin, Phys. Rev. Lett. **92**, 057901 (2004).

[26] H. Takesue, E. Diamanti, T. Honjo, C. Langrock, M. M. Fejer, K. Inoue, and Y. Yamamoto, New J. Phys. **7**, 232 (2005).

[27] A. Leverrier and P. Grangier, Phys. Rev. Lett. **102**, 180504 (2009).

[28] A. Pappa, P. Jouguet, T. Lawson, A. Chailloux, M. Legré, P. Trinkler, I. Kerenidis, and E. Diamanti, Nat. Commun. **5**, 3717 (2014).

[29] V. Dunjko, E. Kashefi, and A. Leverrier, Phys. Rev. Lett. **108**, 200502 (2012).

[30] Z. Cao, Z. Zhang, H. K. Lo, and X. F. Ma, New J. Phys. **17**, 053014 (2015).

[31] H. K. Lo and J. Preskill, Quantum Inf. Comput. **7**, 431 (2007).

[32] N. Gisin, S. Fasel, B. Kraus, H. Zbinden, and G. Ribordy, Phys. Rev. A **73**, 022320 (2006).

[33] N. Jain, B. Stiller, I. Khan, V. Makarov, C. Marquardt, and G. Leuchs, IEEE J. Sel. Top. Quantum Electron. **21**, 6600710 (2015).

[34] (a) T. Schmitt-Manderbach, H. Weier, M. Fürst, R. Ursin, F. Tiefenbacher, T. Scheidl, J. Perdigues, Z. Sodnik, C. Kurtsiefer, J. G. Rarity *et al.*, Phys. Rev. Lett. **98**, 010504 (2007); (b) C. Z. Peng, J. Zhang, D. Yang, W.-B. Gao, H.-X. Ma, H. Yin, H.-P. Zeng, T. Yang, X.-B. Wang, and J.-W. Pan, *ibid.* **98**, 010505 (2007).

[35] Q. Wang, W. Chen, G. Xavier, M. Swillo, T. Zhang, S. Sauge, M. Tengner, Z.-F. Han, G.-C. Guo, and A. Karlsson, Phys. Rev. Lett. **100**, 090501 (2008).
[36] C. R. S. Williams, J. C. Salevan, X. Li, R. Roy, and T. E. Murphy, Opt. Express **18**, 23584 (2010).
[37] F. H. Xu, B. Qi, H. Xu, H. X. Zheng, and H. K. Lo, Opt. Express **20**, 12366 (2012).
[38] Z. L. Yuan, M. Lucamarini, J. F. Dvnes *et al.*, Appl. Phys. Lett. **104**, 261112 (2014).
[39] T. Kobayashi, A. Tomita, and A. Okamoto, Phys. Rev. A **90**, 032320 (2014).
[40] http://www.idquantique.com.
[41] The minimal power of Eve's control laser (P_c) depends on the parameters of both the signal and control lasers, such as the polarization, linewidth, isolation of the S-LDs, and so on.
[42] S. H. Sun, M. Gao, M. S. Jiang, C. Y. Li, and L. M. Liang, Phys. Rev. A **85**, 032304 (2012); Y.-L. Tang, H.-L. Yin, X. Ma, C.-H. Fred Fung, Y. Liu, H.-L. Yong, T.-Y. Chen, C.-Z. Peng, Z.-B. Chen, and J.-W. Pan, *ibid*. **88**, 022308 (2013).
[43] Y. Zhao, B. Qi, and H. K. Lo, Appl. Phys. Lett. **90**, 044106 (2007); S. H. Sun and L. M. Liang, *ibid.* **101**, 071107 (2012).
[44] S. Sajeed, I. Radchenko, S. Kaiser, J.-P. Bourgoin, A. Pappa, L. Monat, M. Legré, and V. Makarov, Phys. Rev. A **91**, 032326 (2015).
[45] X. B. Wang, C. Z. Peng, J. Zhang, L. Yang, and J. W. Pan, Phys. Rev. A **77**, 042311 (2008).
[46] A. Mizutani, M. Curty, C. C. Wen Lim, N. Imoto, and K. Tamaki, arXiv:1504.08151 [quant-ph].
[47] The time shift appears because when the laser cavity is seeded with an external field, the relaxation oscillations are dampened and the first oscillation occurs earlier. And the amount of the time shift is much larger for the two ID300 lasers than those of the homemade lasers, because the linewidth of two ID300 lasers is larger than that of the homemade lasers. Thus the probability that a photon is infused into the ID300 lasers is larger than that of the homemade lasers.
[48] The reason is that it takes time for the initial field to appear (owing to the finite spontaneous emission rate and geometry of the laser chip). While the field takes time to build up, there is little or no stimulated emission, yet the electrical pumping continues at the full rate. The population inversion then rises above the steady-state value and far overshoots it. Then the field far overshoots the steady state, because for a short time the laser has much higher gain; then the stronger field depletes it by stimulated emission below the steady-state value, and so on, there are a few oscillations before the emission settles to a stable value. But in the presence of the seed field in the cavity, the population inversion does not initially overshoot as high because the emission stimulated by the seed field begins earlier.
[49] http://www.thorlabschina.cn.
[50] X. B. Wang, L. Yang, C. Z. Peng, and J. W. Pan, New J. Phys. **11**, 075006 (2009).